Knotting and Linking in Macromolecules

K. C. Millett[a]

[a] *University of California, Santa Barbara, Department of Mathematics, UCSB, Santa Barbara, CA 93106, USA*

Abstract

In the 1980's, knotting in DNA became a fundamental research dimension in the study of the mechanisms by which enzymes act on it. Later, the first compelling identification of knotting in proteins, in 2000, launched the study of knotting in protein structures and linear macromolecules more generally following on theoretical efforts of the 1960's. While the linking occurring in structures such as DNA, with the articulation of the relationship between linking, twisting, and writhe, and, more directly, linking in Olympic gels has been of interest to geometers, molecular biologists and, polymer physicists since the 1960's, a new mathematical analysis of both global and local facets of knotting and linking is again providing new promising discoveries. Following a discussion of the two topological structures of knotting and linking, we will consider some of their applications, and close with a consideration of new questions that suggest attractive directions for future research.

Keywords: DNA, protein, macromolecule, polymer, knots, slipknots, links, lassos, entanglement, knotting fingerprint, local linking fingerprint, Gauss linking number

1. Introduction

In the 1960's, Edwards [24] undertook a theoretical study of the effects on knotting on the properties of polymer gels, but is was not until 1981 that Liu and David [38] demonstrated the presence of a knot in DNA in the lab. This discovery launched an ongoing theoretical and experimental research effort to understand the occurance and character of these DNA knots and to employ them to discover the mechanisms through which enzymes act on DNA in vivo [34, 74, 75, 25, 66]. Buck [7] gives a very helpful introduction to this direction of research. It was not until 2000 that Taylor [69] identified the first deep knots in protein structures thereby launching another thread in the application of topology to biology. Assessing the presence of knotting and linking in proteins has provided an ongoing stream of theoretical and experimental research into the functional role such structures might play in living organisms. The objective

Email address: millett@math.ucsb.edu (K. C. Millett)

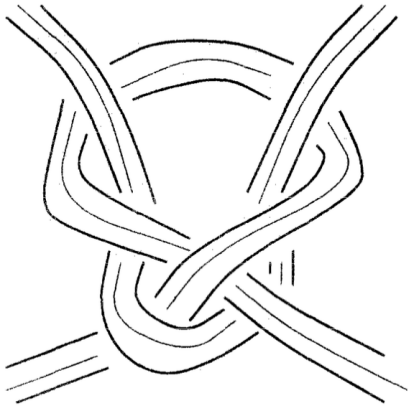

Figure 1: One of Gauss' 1797 knot drawings [61].

of this report is to describe key facets of knotting and linking in polymeric systems and in protein structures. As these are usually linear macromolecules, we describe one approach to identifying knotting in such structures [50] as well as applications of refinements of this method [51, 65, 67, 64, 30].

Historically, the first published consideration of knots appears to be due to Vandermonde [61] but Gauss is recognized for initiating the mathematical study of knots in his notes, figure 1. In the 1870's [26], he gave an integral, equation 1, that captures the integer linking between two oriented rings. This integral can be employed to define a real number quantifying the extent of linking between two oriented spatial arcs, such as those in figure 1, or the self-linking of a single arc, for example the open chains in models of proteins or of filamental systems that employ periodic boundary conditions (PBC) such as polymer melts [5, 62, 53, 58, 56]. These enable one to define periodic linking and periodic self-linking numbers that quantify the linking between pairs of filaments, either open or closed, and, thereby, define the periodic linking matrix. Here, we will describe an application to Olympic gels [16, 4, 29], see Figure 2, an application to the local linking in mathematical knots, and its extension to proteins. Finally, we will also briefly discuss the analysis of linking in proteins when cystene bridges are added to the structure.

In the next section we describe give an introduction to knots [1] and the study of knots and slipknots in open polymers such as proteins by employing the knotting fingerprint [67, 65, 64, 30, 28]. Next, we discuss the Gauss linking [26] and self-linking numbers, one-dimensional periodic boundary condition models, the extension to periodic linking and self linking, and the definition of the periodic linking matrix whose eigenvalues quantify the extent of entanglement in the systems to which they are applied [58, 56]. These are applied to the analysis of Olympic gels [16, 4, 29]. In another direction, the local linking number and linking fingerprint is described and applied to classical knots, to

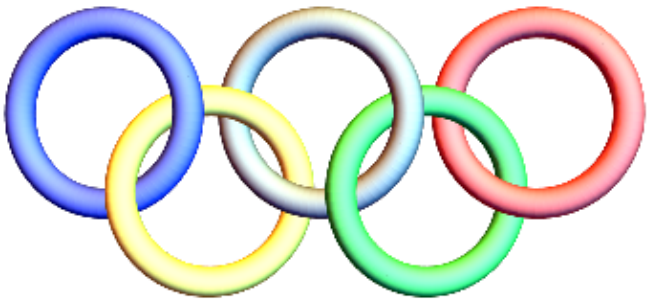

Figure 2: Olympic gels are collections of simply linked ring polymers [16]

illustrate its implications, and to proteins. In the third section, we will briefly describe the application of linking to lassos and their extensions occuring in proteins when cystene bridges enrich the structural analysis.

2. Knots, Slipknots, and Knotting Fingerprints

2.1. Mathematical Knots

While, since thousands of years, humans have employed knotting and entanglement of materials and, later, in their artistic representations, it is not until the time of Gauss and, later, Kelvin that the mathematical study of knots was systematically undertaken in the context of electromagnetism and as proposed models of atoms. Knots are closed rings in space with two knots being equivalent if one can be deformed to the other without breaks or singularities in the evolution. Although Gauss provided a method to symbolically codify knots, i.e. the Gauss code, it was the later purpose that gave rise to the first efforts by William Thomson (Lord Kelvin) classify knots and links by formulating a study of indivisible, or prime, knots and links [70], see figure 3. The mathematical study of knots developed steadily as a subfield of topology with connections to geometry, the two areas most relevant to their later application to polymer gels, DNA, and other macromolecules. Important features of this knot theory, for our purposes, are the decomposition of a knot into indecomposable subknots via the "connected sum," as illustrated in figure 3 where the two upper left knots are prime, a right trefoil and a seven crossing knot, while the upper right knot is the composition of two three crossing knots, a left and a right trefoil. Note that, below them, are a two component link and a three component link, the later being known as the Borromean rings. The indecomposible, i.e. prime, knots through sixteen crossings have been classified [27], there are $1,701,936$, only the simpler ones are most often encountered. They are currently identified using "knot polynomials" such as Alexander, Jones, or HOMFLYPT or their convenient evaluations [2, 31, 35]. We note that there are more powerful and more complex methods now available.

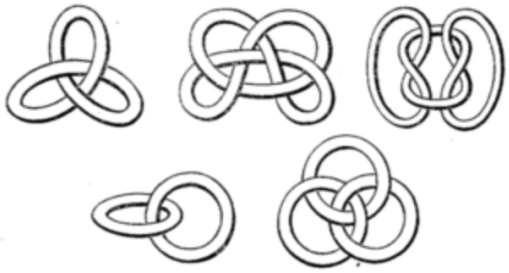

Figure 3: Knots and links from Kelvin's "On vortex montion" [70].

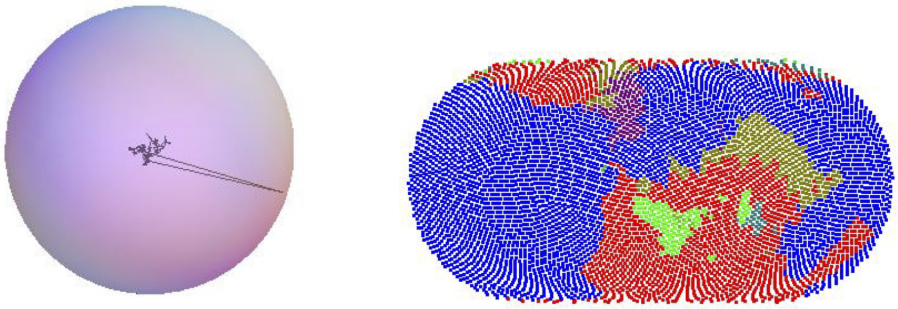

Figure 4: Knotting in open arcs: one closure of an open arc and distribution of knot types [23].

2.2. Knots and slipknots in arcs

The fundamental challenge in defining the knot type associated to an open arc in 3-space is to translate the question into the case of an associated 3-space closed curve. One approach, see [23], is to define a closure using points on a very large sphere containing the finite chain and add segments from the two endpoints of the arc to the chosen point on the sphere. This defines a knot type for almost all points, the exceptions being a set of measure zero on the sphere. The designation is locally constant allowing to estimate the area of the regions on the sphere associated to each knot type using, for example, the HOMFLYPT polynomial. Since there are only finitely many possible types, one is able to determine the proportion of the sphere associated to each knot type, the knotting spectrum of the arc, and to associate the dominate knot type to arc. This provides a powerful method in that it almost always successfully identifies a specific knot type, even for random walks, [52]. One may modify this method to improve the computational speed, for example [71].

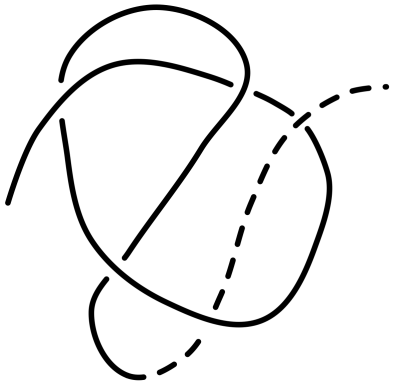

Figure 5: A trefoil slipknot: solid arc forms an ephemeral trefoil which becomes unknotted when the dashed segment is added [51].

Empowered with the ability to determine the knot type of an open arc, one is now in position to identify slipknots and their associated ephemeral knots as follows: an unknotted segment in an arc containing a knotted segment, the associated *ephemeral knot*, is called a *slipknot*, figure 5. Mathematicians have proved that the probability that a random arc or ring is knotted or contains a slipknot goes to one as the length of the arc or ring goes to infinity, [68, 59, 22, 21, 51].

3. Gauss Linking, Periodic Boundary Condition (PBC) Models, Periodic Linking, and the Periodic Linking Matrix

3.1. Gauss Linking and Self-linking

The linking number between two oriented chains, l_1 and l_2, is defined using parameterizations of the chains, $\gamma_1(t)$ *and* $\gamma_2(s)$, via the Gauss linking integral:

Definition 3.1. The Gauss *linking number* of two disjoint (closed or open) oriented curves l_1 and l_2, whose arc-length parametrizations are $\gamma_1(t), \gamma_2(s)$ respectively, is defined as a double integral over l_1 and l_2 [26]:

$$L(l_1, l_2) = \frac{1}{4\pi} \int_{[0,1]} \int_{[0,1]} \frac{(\dot{\gamma}_1(t), \dot{\gamma}_2(s), \gamma_1(t) - \gamma_2(s))}{||\gamma_1(t) - \gamma_2(s)||^3} dtds, \tag{1}$$

where $(\dot{\gamma}_1(t), \dot{\gamma}_2(s), \gamma_1(t) - \gamma_2(s))$ is the triple product of the derivatives, $\dot{\gamma}_1(t)$ and $\dot{\gamma}_2(s)$, and of the difference $\gamma_1(t) - \gamma_2(s)$.

Note that the Gauss linking number definition applies, with a choice of orientation of the two strands, to the two arc in figure 1 to give a real number quantifying the extent to whch they are linked. The rings in the Olympic rings, figure 2, have linking ± 1 is they are adjacent and 0 if not. The two links at the

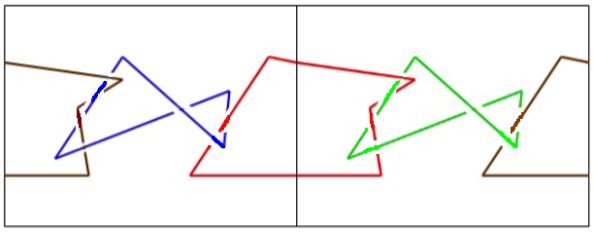

Figure 6: A 1-dimensional PBC model [56].

bottom in figure 3 have linking ± 1 for the Hopf link, on the bottom left, while the Borromean rings have pairwise linking 0 for any choice of orientations.

Definition 3.2 (Self-linking number)**.** Let l denote a chain, parameterized by $\gamma(t)$, then the *self-linking number* of l is defined as:

$$\begin{aligned} Sl\,(l) \quad &= \tfrac{1}{4\pi}\int_{[0,1]^*}\int_{[0,1]^*}\tfrac{(\dot\gamma(t),\dot\gamma(s),\gamma(t)-\gamma(s))}{||\gamma(t)-\gamma(s)||^3}dtds \\ &\quad +\tfrac{1}{2\pi}\int_{[0,1]}\tfrac{(\dot\gamma(t),\ddot\gamma(t),\dddot\gamma(t))}{||\dot\gamma(t)\times\ddot\gamma(t)||^2}dt. \end{aligned} \tag{2}$$

where $\dot\gamma(t)$, $\ddot\gamma(t)$, and $\dddot\gamma(t)$ are the first, second, and third derivatives of $\gamma(t)$, respectively, and $(\dot\gamma(t),\ddot\gamma(t),\dddot\gamma(t))$ is their triple product.

The self-linking number consists of two terms, the first being the Gauss integral and the second being the total torsion of the curve. As above, this can be applied to indivitual arcs as well as rings such as those at the top of figure 3. Note that the standard round planar circle has zero selflinking.

3.2. Periodic Boundary Condition (PBC) models

The underlying structure of the *Periodic Boundary Condition*, PBC, model consists of a unit cube in 3-space. The three-dimensional cube contains a collection of arcs whose endpoints either lie in the interior or intersect the $x = 0$ or $x = 1$ faces under the constraint that the pattern on both faces is identical, see figure 6. The latter condition allows one to create and infinite structure by taking the union of integer translates of the cells and taking the unions of the resulting one-chains to define a collection of one dimensional chains. In general, these chains may be open or closed or even non-compact. One can also require that $y = 0$ and $y = 1$ give analogous additional boundary conditions to create a structure using the translation in the two dimensional lattice, Z^2, to fill out a thick planar structure. Similarly, add the requirement that there are $z = 0$ and $z = 1$ give a final boundary condition so that one fills 3-space using the translation in the three dimensional lattice, Z^3, to fill out a 3-dimensional structure. While a knot theory for such structures is quite exotic, one can extend the Gauss linking to these structures.

3.3. Periodic Linking and Self-linking

In a PBC model, each chain is translated to give an infinite collection copies of itself. As a consequence, one is faced with quantifying the linking of one chain, l_0 with infinitely many translation copies of itself, $l_v = l_0 + \vec{v}$, or with infinitely many copies of another chain, $J_v = J_0 + \vec{v}$. This is achieved by employing Panagiotou's periodic linking and self-linkings numbers described next. In the periodic system we define linking at the level of free chains (i.e. the collection of translation copies of a chain, l_0; see [56] for a discussion of the motivation for this definition). The underlying idea is to calculate the linking between a chain in one free chain with all the chains in the other free chain.

Definition 3.3 (Periodic linking number)**.** Let I and J denote two (closed, open or infinite) free chains in a periodic system. Suppose that I_0 is an image of the free chain I in the periodic system. The *periodic linking number*, LK_P, between two free chains I and J is defined as:

$$LK_P(I, J) = \sum_{\vec{v} \neq \vec{0}} L(I_0, J_0 + \vec{v}), \tag{3}$$

where the sum is taken over all the images of the free chain J in the periodic system.

The periodic linking number has the following properties with respect to the structure of the cell, see [56], which follow directly by its definition:
(i) The infinite sum defining LK_P converges, i.e. LK_P makes sense mathematically.
(ii) LK_P captures all the linking that all the images of a free chain impose to an image of the other.
(iii) LK_P is independent of the choice of the image I_0 of the free chain I in the periodic system.
(iv) LK_P is independent of the choice, the size and the shape of the generating cell.
(v) LK_P is symmetric.

The quantification of the linking of a free chain with itself is a bit special and requires a bit more care as there are two contributing cases, the linking of a chain with itself and the linking of a chain with translations of itself. As a consequence, one is led to the following definition [56]:

Definition 3.4 (Periodic self-linking number)**.** Let I denote a free chain in a periodic system and let I_0 be an image of I, then the *periodic self-linking number* of I is defined as:

$$SL_P(I) = Sl(I_0) + \sum_{v \neq u} L(I_0, I_v), \tag{4}$$

where the index v runs over all the images of I, except I_0, in the periodic system.

As with the periodic linking number, the mathematical proof of the existence of this quantity and its properties are proved in [56].

3.4. Periodic Linking Matrix

In order to analyze the linking entanglement present in a PBC system, L, consisting of a finite number of free chains, $l_1, l_2, ..., l_n$, one employs an n x n real symmetric matrix, LM_C, whose i, jth entry is defined by equation

$$\begin{aligned} LM_{C_{i,i}} &= SL_P(l_i) \\ LM_{C_{i,j}} &= LK_P(l_i, l_j) \end{aligned} \tag{5}$$

In the case of a single generating chain, l, the periodic linking matrix consists of a single entry, the periodic self-linkng number, $Sl(l)$. From the definitiion, there are two contributing factors, the self-linking given by the equation 2 and the linking between distinct copies, reflecting distinct features of periodic self-linking.

For systems with two independent chain types, see figure 6, the periodic linking matrix adds entanglement information due to the linking between the two distinct chains. Associated to the periodic linking matrix are two real eigenvalues, $e_1(L)$ and $e_2(L)$, given in decreasing order. The larger of these, $e_1(L)$ is proposed as the dominant characterization of the linking entanglement of the PBC system. The set of eigenvalues is the *periodic linking spectrum* of the system.

Similarly, for systems with n independent chain types, one defines the periodic linking matrix, LM_C. The associated ordered collection of eigenvalues, $e_1(L), ..., e_n(L)$ define the spectrum of the PBC system.

4. Some applications of knotting and linking to macromolecules

We first discuss the appliction of the ability to identify knotting in linear chains to polymers in θ conditions, in mathematical models of spatial rings and diagrams, and in proteins employing the knotting fingerprint.

4.1. Knotting in polymeric systems

Under θ conditions, individual linear polymer chains can be modeled as random walks providing the basis for many numerical studies estimating the average spatial properties such as the growth of the radius of gyration as a function of the length of the chain. With theorems proving that the probability that a random walk is knotted is asymptotically one, random walks have been studied to characterize the proportion of unknots and other specific knot types as length increases. Using the methods described in section 2.2, one can estimate the size of the knots, ephemeral knots, and slipknots in random walks [47], see figure 7 in which comparison is made with another subknot scaling [41, 42]. Here, one finds that the average size of subknots, ephemeral knots, and slipknots appear to be bounded as the length of the random walk increases.

Empowered by the ability to characterize the presence of knotting in open arcs, it is possible to study the presence and position of knots in rings to assess the complexity and character of the knotting that may be present in ideal

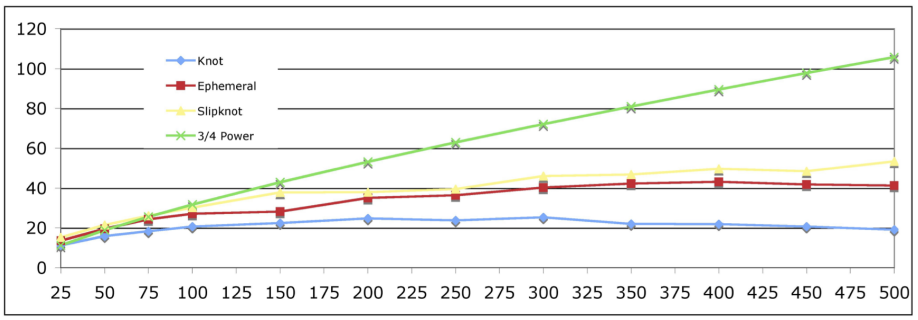

Figure 7: The average size of knots, ephemeral knots, and slipknots in random walks [47].

knots and random knots [64]. The key assessment vehicle is the knotting fingerprint, see figure 8, in which the colors of the cells in a disk represent the knotting present in the subsegment indicated by coordinates of the cell. The planar graph defined by this knotting fingerprint can be analyzed by translating the relationships into the knotting fingerprint graph to which graph theoretic methods can then be applied to quantify the characteristics of the complexity of the knot [28]. For example, one can calculate the associated Cheeger constant, a measure the constraints placed on the indipendent unknotting pathways or, by inversion, the knotting pathways.

These considerations naturally lead to the analysis of subknots of minimal crossing knot diagrams, i.e. the presentations employed in knot and link tables, [1], and often employed to test theories. They are essentially unique for knots with minimal crossing alternating diagrams but the story is rather more complicated for non-alternating knots, i.e. those without alternating crossing diagrams [1]. Careful analysis of subknots in the classical minimal crossing prime knot diagrams through 15 crossings [48, 49] supports the conjecture that *Every minimal crossing prime knot diagram contains either a trefoil or a figure-eight knot.* Indeed, 99.83% of the $313,258$ minimal prime knot diagrams contain a trefoil subknot. This fact shows that the simplest knot, the trefoil, is a critical component in the formation of most knots. While there are infinite families of minimal crossing prime knot presentations, with rapidly increasing number of crossings, it may still be the case that, as the number of crossings increases, the probability of containing a trefoil knot asymptotically goes to 1.

4.2. Knots and slipknots in proteins

The first instance of a knot present deep inside a protein was identified by Taylor, [69], in 2000 and the first slipknot was identified in a protein by King et. al., [32] in 2007. Since then, new knotting in proteins has been identified provoking speculation as to the evolutionary purpose of the systematic presense of these knots in certain families of protiens [67, 12]. A form of the knotting fingerprint, adapted to the circumstance of the open macromolecule structure

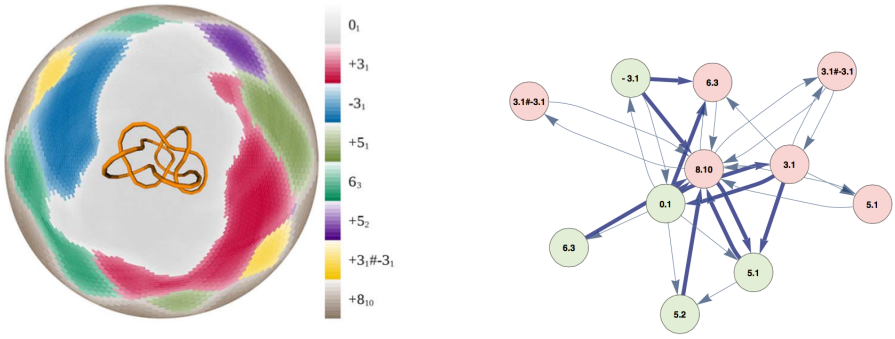

Figure 8: The knotting fingerprint of an ideal 8_{10} knot from [64]. The distance from the center corresponds to the length of the segment and the angular coordunate corresponds to the position of the center of the segment in the circular parameterization of the knot. The information in the fingerprint can be expressed in the form of a graph to which powerful graph methods can be applied to extract new quantitative expressions of the complexity of the knotting, for example a Cheeger constant [28].

of the protein, provides the foundation of the identification and localization of knots and slipknots in protein structures, see figure 9. A website, KnotProt: http://knotprot.cent.uw.edu.pl [30], has been created that provides a regularly updated analysis of protein structures appearing in the Protein Data Bank including the knotting fingerprint and links to associated structures:

The knotting fingerprint provides a highly effective method to study knot localization in proteins [65] and can be applied to any linear macromolecule spatial structure.

4.3. Linking in polymeric systems

In [57], Panagiotou et. al. show that the average squared linking number, the averages squared writhe, and the average squared selflinking number of uniform random walks or polygons of length n is a convex confined space grow on the order of n^2. These theoretical results are confirmed by number simulations and provide rigorous confirmation of the average values one encounters for polymer models in θ conditions. If one wishes to remove possible boundary effects due to confinement, a periodic boundary condition model provides a method to acquire simulation data. In [58], Panagiotou et. al. employ the periodic linking number and the associated periodic linking matrix to formulate a new measure the entanglement of a collection of open or closed chains and demonstrate, using numerical simulations, that these provide effective tools to measure the homogeniety of entanglement as a function of the chain length. These methods have been shown to give information consistent that that achieved by more traditional methods defining entanglement in polymer melts [55], see figure 11.

Another important application of linking in PBC models is provided by the study of Olympic systems. *Olympic systems* are collections of small ring

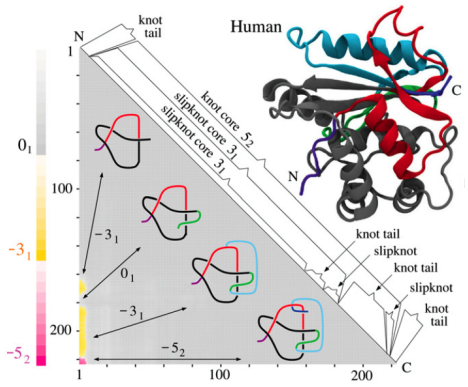

Figure 9: The knotting fingerprint in ubiquitin C-terminal hydrolase in H. sapiens from [67]. The fingerprint is annoted to show the location of specific knot and slipknots within the structure as well as a cartoon representation from the Protein Data Bank [6]

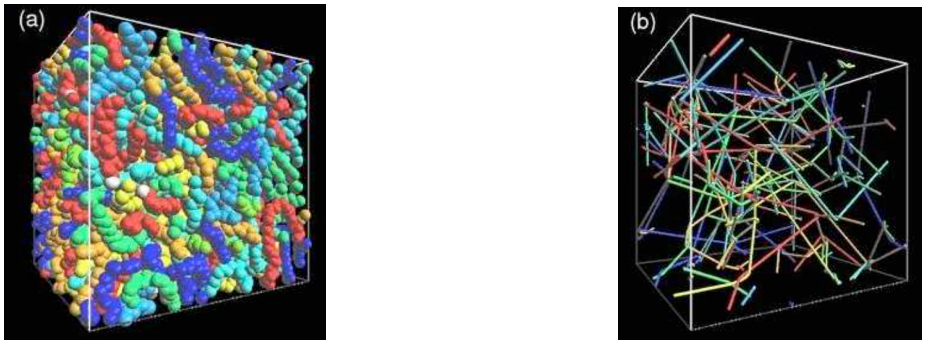

Figure 10: A representative atomistic polyethylene sample, (a) and its correspoinding reduced network (b) in which bends in the piecewise linear paths denote the presence of local linking [55].

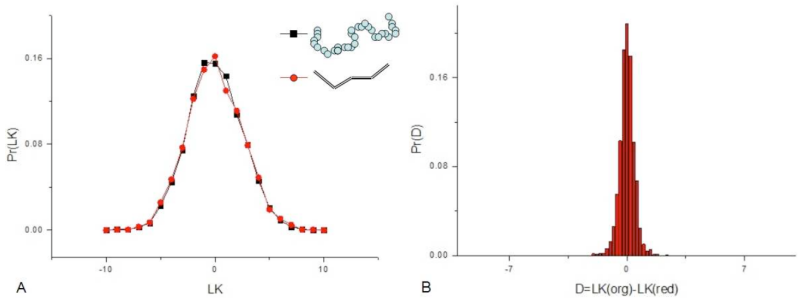

Figure 11: Graph (a) shows the normalized probability distribution of the periodic linking number for the original and the corresponding reduced chains in a polyethylene model. In (b) the distribution of the differences between these two shows that the periodic linking number captures information that is statistically equivalent to that provided by the traditional chain shortening method, [55]

polymers whose aggregate properties are largely characterized by the extent (or absence) of topological linking in contrast with the topological entanglement arising from physical movement constraints associated with excluded volume contacts or arising from chemical bonds. These were first discussed by de Gennes [17] and have been of interest ever since due to their particular properties and their occurrence in natural organisms, for example as intermediates in the replication of circular DNA in the mitochondria of malignant cells or in the kinetoplast DNA networks of trypanosomes [11, 39, 45, 18, 19, 4, 20]. In Igram et. al. [29], simulations of Olympic systems with one, two, and three-dimensional boundary conditions are studied using the periodic linking number and associated linking matrix. In these, there are observed critical densities at which the dimensional character of the entangled subsystems change. For example, in the three dimensional case, one first observes a one dimensional linked subsystem, then a two dimensional one and, finally, the entangling of the entire three dimensional system, see figure 12.

4.4. Linking in proteins

When disulfide bridges are added to the structure of a protein a loop is formed, a lasso [54], through which both the C and N termini of the backbone can link, see figure 13. The extent of linking can be measured, for example, using the Gauss linking integral, equation 3.1, or by counting the oriented intersection of these tails with an auxiliary oriented surface whose boundary is the covalent loop as illustrated in the figure. Lassoprot [14] has been created to enable research on the structure of lassos created by these disulfide bridges in proteins. We note that none of these structures contain knots, as described in our earlier discussion of proteins, so that the lasso structures provide a new dimension through which the spatial structure of the protein may be connected to its

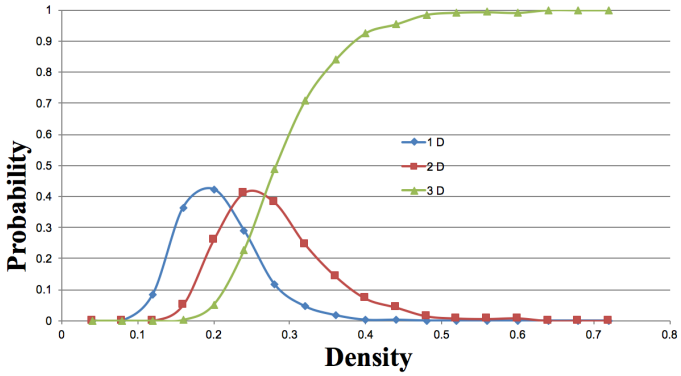

Figure 12: In 3-dimensional PBC models of Olympic systems, one observes the onset of one dimensional entangled subsystems, then two dimensional entangled subsystems, before finally achieving full percolation, [29].

function.

When two or more disulfide bridges are present in a protein, even more complex structures, modeled by three valent mathematical graphs, are encountered. In [15] the simplest of such structures, those forming topological links, are analyzed, see figure 14.

4.5. Thick polymers

Except under θ conditions, the effective thickness of a polymer gives rise to excluded volume constraints that play an important in characterizing the spatial structure of the physical polymer. Randomly sampling such physical polymers over over a physically important range of effective thicknesses has been an important problem. Recently, however, fundamental progress has occurred. First, for open chains, using Rawdon's determination of thickness of open polygonal chains [63], Plunkett and Chapman [60] define a new sampling algorithm for the space of random walks with any specifiec thickness and prove that this algorithm is ergodic, i.e. faithfully samples the continuous space of conformations. The sampling algorithm is inspired by the reflections employed in lattice structures but is applied to off lattice, or 3-space, conformations and requires a new continuous structure argument to prove that it is ergodic. One is not able to study the consequence of thickness on such spatial meansures as the squared radius of gyration, see figure 15.

Chapman [10] has formulated a new ergodic algorithm, also employing 3-space reflections, that faithfully samples the space of equilateral rings with any specified thickness. Employing these new algorithms, one is now able to study the relationship between polymer length and thickness and, for example, the scaling of the radius of gyration, the presence of knots and slipknots and their average radii of gyration or other spatial properties.

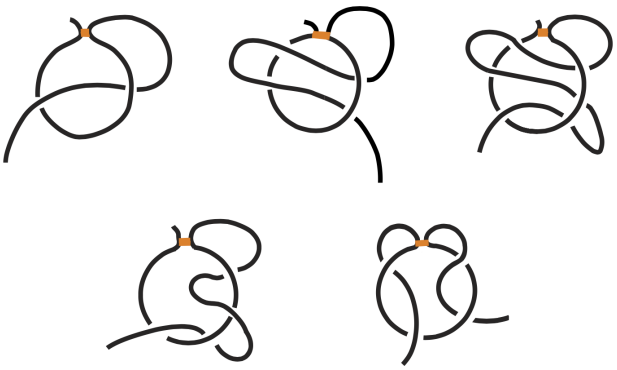

Figure 13: When a disulfide bridge is added to the protein structure, a lasso can be formed with which the C and N termini can link, [54]. The orange segments represent the disulfide bridges.

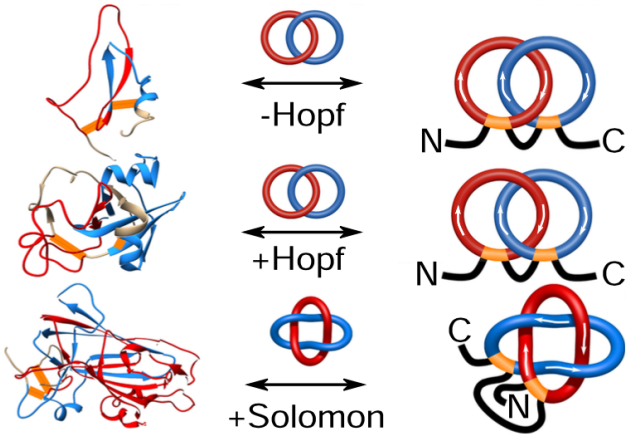

Figure 14: When two or more disulfide bridges are added to the protein structure, mathematical links can be formed, [15]. The orange segments represent the disulfide bridges.

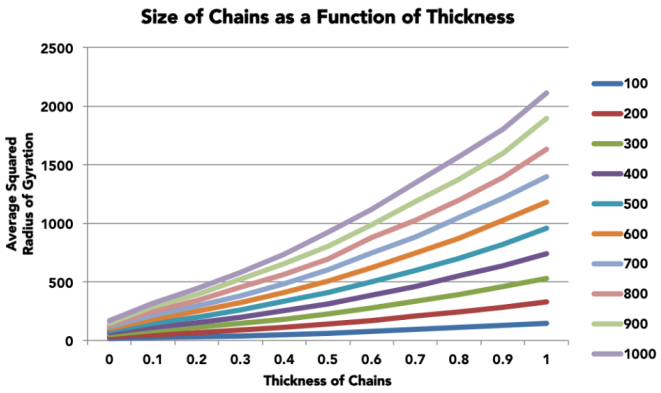

Figure 15: The growth of average squared radius of gyration as a function of thickness, [60].

Deguchi and Uehara [72, 73], using an off-lattice algorithm employing reflections and folds, have also studied the knotting under excluded volume constraints. While the ergodicity of this algorithm is unknown, one expects that the results will be consistent with those achieved using an ergodic algorithm. Among many other interesting matters, they consider the equilibrium length of an equilateral polygonal knot type, defined as the number of segments at which the average squared radius of gyration of the knot type equals the average squared radius of gyration of all configurations of that length. From their data, it appears that the equilibrium length may be additive, at least for trefoil knots, as this relationship seems to hold for the trefoil, 3_1, the sum of two trefoils, $3_1\#3_1$ and, even, for three trefoils, $3_1\#3_1\#3_1$. Perhaps this is a reflection of knot localization and the average character of the connected sum.

5. A few attractive research directions

5.1. Knot localization

One question that arises is "How to define the locus of a knot?" in an open or closed polygon. For open or closed polygons containing a prime knot, one can determine the knotting fingerprint and select the shortest subsegment representing the given knot type. This is method employed, for example, in the study of knotted proteins [67]. To render this question accessible to an initial exploration, one can begin by a study of minimal crossing planar presentations of the knot type. For the sum of two trefoils in a closed polygon, this method provides the "correct" answer even though there are three distinct presentations of the connected sum, see figure 16. One observes that they are distinct by virtue of the number of edges bounding the exterior unbounded region of the presentation. In the first case, one has a hexagon, in the second case one has a tetragon, while in the third case one has a pentagon region.

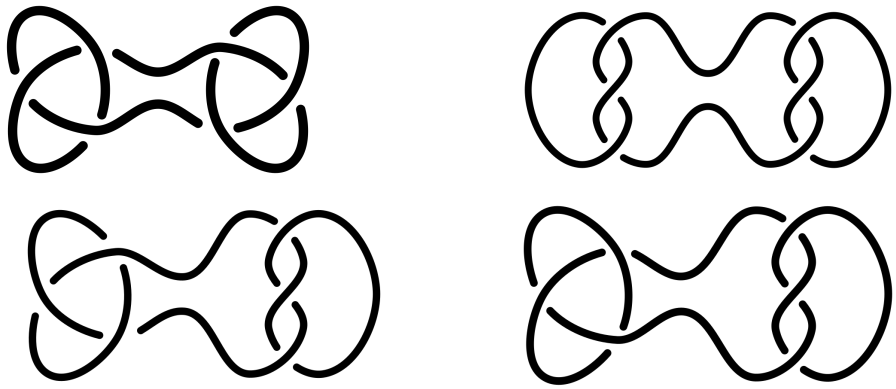

Figure 16: The closed connected sum of two trefoils has three different presentations in $\mathbf{R}^2$. The upper two have hexagonal and tetragonal unbounded complementary regions whereas the middle one has a pentagonal unbounded complementary region. This last sum is an alternating crossing connected sum of a left (negative crossing) and a right (positive crossing) trefoil. The corresponding minimal crossing presentation of two right trefoils, bottom right, is not alternating but is minimal.

The classification of $\mathbf{S}^2$- equivalent connected sums of three trefoils is much more complex with a much larger number of options for configuring a minimal crossing planar presentation of the connected sum. A critical component in the analysis is Menasco's theorem [43] stating that a reduced alternating knot or link projection is irreducible if and only if there is no planar circle meeting the presentation in exactly two transverse points with crossings in each of the two complementary regions. In fact, for composite knots or links, such circles are very easily recognized and allow one to factor the presentation into a non-trivial presentations within each of the complementary regions. One can then continue the analysis with simpler non-trivial presentations until achieving an irreducible factorization of the presentation. It is the combinatorics of these separating circles and the eventual prime two string tangles that is the principal source of the complexity of the classification of the presentations. Observe that there arises an additional complication demonstrated by the fact that the connected sum of the two different presentations of the right (positive) trefoil, see figure 16, might be not an alternating presentation. The analogous alternating presentation must be the connected sum of a right and a left trefoil. In as much as matters of chirality are of, at least, biological importance, this is an important complication to be addressed.

Inspired by this example, one approach to the identification problem might be via *tangle theory* [13, 36, 37], see figure 17. As a first test case, one should analyze the alternating crossing diagrams of connected sums of, say, the positive trefoil. Unlike the classical case of alternating diagrams of prime knots and the flype-conjecture, proved by Menasco and Thistlethwaite [44], there does not

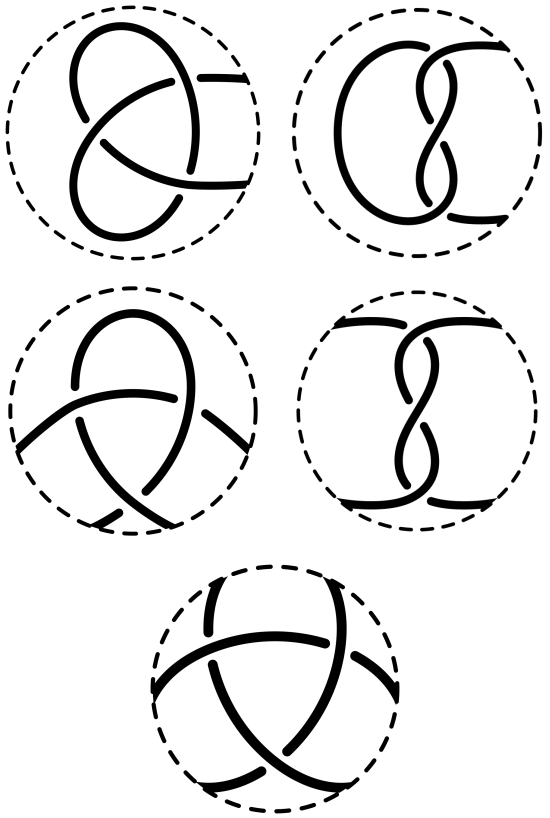

Figure 17: At the top are two distinct one string tangles representing the trefoil knot. In the middle are two two string tangles representing the trefoil. And at the bottom, a three string tangle also representing the trefoil knot.

appear to be a classification of alternating composite knots or links, a rather more complex problem even for closed connected sums of trefoils, see figures 18 and 19.

With respect to the question of classification of the spatial configurations, the specific spatial geometries occur with different probabilities and have distinct spatial properties, such as the radius of gyration, that depend not only upon the number of summands but also the way in which they are organized in space, as illustrated in figures 18 and 19.

In addition to the above sources of complexity, for the open polygonal connected sum, the character and classification is rendered even more complex as it depends on the location of ends of the arc. For a hint as to what one faces, see figure 20, in which we consider two of the simplest cases and the problem that arises in identifying the size of the trefoils that form the connected sum.

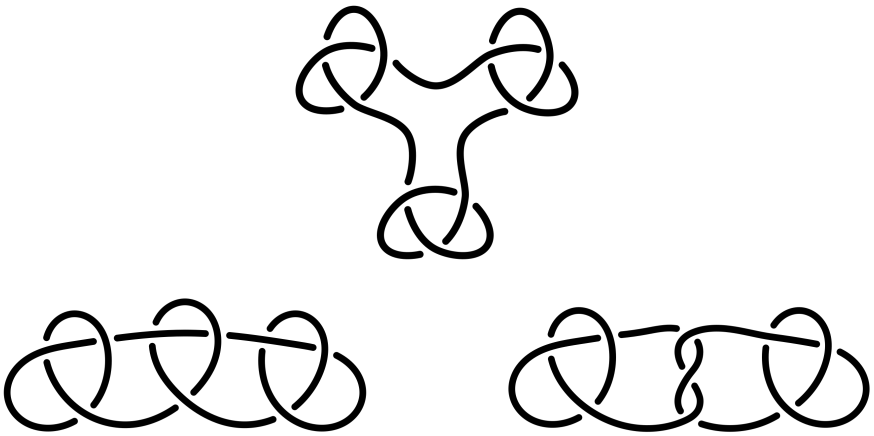

Figure 18: Connected sums of three trefoils. On the top is a representation of the "standard" connected sum. Below are two different connected sum representations having a two string tangle in the center, analogous to the two string structures for tangles in figure 17. Note that the second of these is minimal but not alternating.

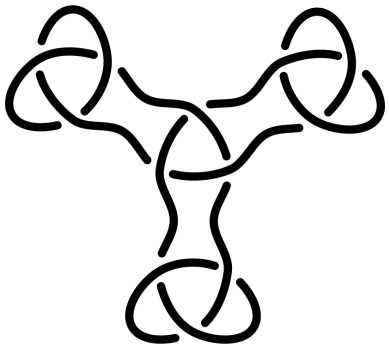

Figure 19: Connected sum of four trefoils having a three string tangle in the center, analogous to the third structure for tangles in figure 17.

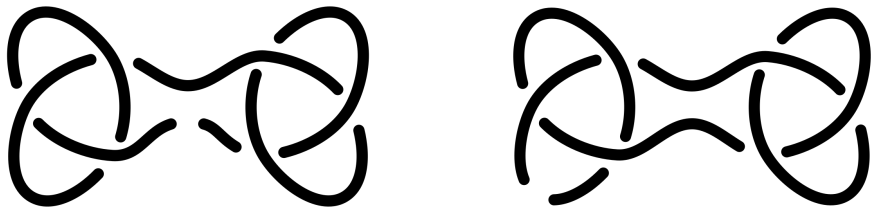

Figure 20: Two distinct open connected sums of two trefoils. The standard open sum, on left, gives the usual size supports for the trefoils while the current algorithm, applied to the sum on the right would not recognize the presence, of the second trefoil subknot, only the entire connected sum.

5.2. Physical knots and links

One of the enduring challenges in the simulation of three-space, off lattice, random walks and polygons has been the absence of a proof that, until now, any of the popular sampling algorithms is ergodic, i.e. faithfully samples the continuous space of configurations. While such proofs for lattice sampling have been available for some time, [40], rigorous off lattice methods are only now available. Unfortunately, neither the very useful crankshaft [33, 3] or the polygonal fold [46, 3] are known to be ergodic. Employing the symplectic geometry of the configuration space, Cantarella et. al. [9, 8] now provide one powerful ergodic method. In the study of physical knots and in an entirely independent approach inspired by the lattice methods, Chapman and Plunkett [60, 10], have described algorithms that they can rigourously prove ergodically sample the spaces of thick walks and polygons using Rawdon's definition of thickness for polygonal structures.

For applications to physical polymeric structures, the question as to whether, as popularly believed, the chrankshaft and the polygonal fold algorithms are ergodic in the continuous space of closed polygons continues to be an important theoretical problem. Furthermore, in the study of physical polymers, the bending contribution to Rawdon's thickness is not employed but, rather, a tube or cylinder constraint is used to insure an excluded volume. As a consequence, one is lead to ask whether the methods of Chapman and Plunkett ergodically sample the spaces of "thick" structures where one uses only the cylinder definition of thickness. If so, their methods would be more widely applicable. Deguchi and Uehara [72, 73] employ a sampling method that uses both reflections, similar to Chapman and Plunkett, and polygonal folds in the context of the cylinder thickness but, here too, one lacks a rigorous proof that this method is ergodic. It seems, from a mathematical physics or polymer science perspective, there remain some fundamental foundational matters to be rigorously resolved.

The increasing progress in the synthesis of complex polymeric structures leads one to revist the implications of the presence of knotting and linking. In the context of this discussion, it is apparent that the spatial structure and consequences of spatially distinct connected sum structures of the simplest formations

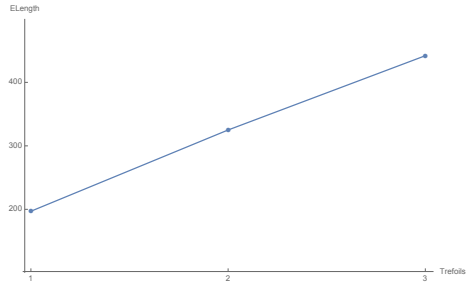

Figure 21: The average equilibrium length as a function of the number of trefoil summands

such as trefoil knots merits careful analysis, especially as these structures will depend upon the thickness of the macromolecules. Deguchi and Uehara have considered the average properties of trefoil knots and have noted that the average equlibruim length grows roughly linearly with the number of summands, see figure 21. This is consistent with the localization of the trefoil summands but, as we have seen above, the precise spatial structure of the connected sum could be complex. As a consequence, one may wonder about the proportion of the structures occurring and how one might go about selecting those whose physical properties are desired.

6. Conclusion

Although somewhat selective, we have seen that there is a rather broad spectrum of ways in which knotting and linking are fundamental to charactizing the spatial structure of macromolecules. From experimental methods to discern the mechanisms by which enzymes act upon DNA to knotting and linking existing with in protein structures and other polymeric materials, one sees a variety of physical manifestations that include consequences for their radius of gyration, topological swelling, and the variety of distinct structures one may encounter in connected sums of even the simplest of knots. While there has been important progress recently, there remain some fundamental problems and important new research opportunities across polymer chemistry, biology, and physics as well as within pure and applied mathematics. We have tried to provide an introduction and some insight into all of these in this short reflection of the state of the art.

7. Acknowledgements

This report was begun while the author was in residence at Ochanomizu University, Tokyo in conjunction with the international workshop: Knots and Polymers: Aspects of topological entanglement in DNA, proteins and graph-shaped polymers. Thanks to Tetsuo Deguchi for his hospitality and conversations regarding many of these topics. Thanks also to Pat Plunkett for the rendering of a number of the figures.

References

[1] C. C. Adams. *The knot book*. Amer. Math. Soc., Providence, RI, 2004.

[2] J. W. Alexander. Topological invariants of knots and links. *Trans. Amer. Math. Soc.*, 30(2):275–306, 1928.

[3] S. Alvarado, J. A. Calvo, and K. C. Millett. The generation of random equilateral polygons. *J. Stat. Phys.*, 143(1):102–138, 2011.

[4] J Arsuaga, Y Diao, and K Hinson. The effect of angle restriction on the topological characteristics of minicircle networks. *Journal of Statistical Physics*, 146(2):434–445, 2012.

[5] C. Benham and D. Miller. Writhing and linking densities for closed, circular dna. *J. Knot Theory Ramifications*, 9(5):577–585, 2000.

[6] H. M. Berman, Z. Westbrook, Z. Feng, g. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Research*, 28:23500242, 2000.

[7] D. Buck. *Applications of knot theory*, volume 66, chapter DNA topology. Amer. Math. Soc., Providence, RI, 2009.

[8] J. Cantarella, B. Duplantier, C. Shonkwiler, and E. Uehara. A fast direct sampling algorithm for equilateral closed polygons. *J. Phys. A.*, 49(27):275202, 9, 2016.

[9] J. Cantarella and C. Shonkwiler. The symplectic geometry of closed equilateral random walks in 3-space. *Ann. Appl. Probab.*, 26(1):549–596, 2016.

[10] K. Chapman. An ergodic algorithm for generating knots with a prescribed injectivity radius.

[11] J Chen, CA Rauch, JH White, PT Englund, and NR Cozarelli. The topology of kinetoplast dna network. *Cell*, 80:61 – 69, 1995.

[12] Thomas Christian, Reiko Sakaguchi, Agata P Perlinska, Georges Lahoud, Takuhiro Ito, Erika A Taylor, Shigeyuki Yokoyama, Joanna I Sulkowska, and Ya-Ming Hou. Methyl transfer by substrate signaling from a knotted protein fold. *Nature structural & molecular biology*, 23(10):941, 2016.

[13] J. H. Conway. An enumeration of knots and links, and some of their algebraic properties. In *Computational Problems in Abstract Algebra (Proc. Conf. Oxford, 1967)*, pages 329–358. Pergamon, Oxford, 1970.

[14] P. Dabrowski-Tumanski, W. Niemyska, P. Pasznik, and J. I. Sulkowska. Lassoprot: server to analyze biopolymers with lassos. *Nucleic Acids Research*, 308, 2016.

[15] P. Dabrowski-Tumanski and J. I. Sulkowska. Topological knots and links in proteins. *Proc. Natl. Acad. Sci. USA*, 114(13):3415–3420, 2016.

[16] Pierre-Gilles De Gennes. *Scaling concepts in polymer physics*. Cornell University Press, 1979.

[17] Pierre-Gilles De Gennes. *Scaling concepts in polymer physics*. Cornell University Press, 1979.

[18] Y Diao, K Hinson, and J Arsuaga. The growth of minicircle networks on regular lattices. *Journal of Physics A: Mathematical and Theoretical*, 45(3):035004, 2012.

[19] Y Diao, K Hinson, R Kaplan, M Vazquez, and J Arsuaga. The effects of density on the topological structure of the mitochondrial dna from trypanosomes. *Journal of mathematical biology*, 64(6):1087–1108, 2012.

[20] Y Diao, K Hinson, Y Sun, and J Arsuaga. The effect of volume exclusion on the formation of dna mini circle networks: implications to kinetoplast dna. *J. Phys. A: Math. Theor.*, 48:1 – 11, 2015.

[21] Yuanan Diao. The knotting of equilateral polygons in $\mathbf{R}^3$. *J. Knot Theory Ramifications*, 4(2):189–196, 1995.

[22] Yuanan Diao, Nicholas Pippenger, and De Witt Sumners. On random knots. *J. Knot Theory Ramifications*, 3(3):419–429, 1994. Random knotting and linking (Vancouver, BC, 1993).

[23] Akos Dobay, Jacques Dubochet, Kenneth Millett, Pierre-Edouard Sottas, and Andrzej Stasiak. Scaling behavior of random knots. *Proc. Natl. Acad. Sci. USA*, 100(10):5611–5615, 2003.

[24] S F Edwards. Statistical mechanics with topological constraints: I. *Proceedings of the Physical Society*, 91(3):513, 1967.

[25] C. Ernst and D. W. Sumners. A calculus for rational tangles: applications to dna recombination. *Math. Proc. Cambridge Philos. Soc.*, 108:489 – 515, 1990.

[26] Gauss K F. Zur mathematischen theorie der electrodynamischen wirkungen. *Werke vol 5 Konigl. Ges. Wiss. Gottingen*, page 605, 1877.

[27] J. Hoste, M. B. Thistlethwaite, and J. Weeks. The first 1,701,936 knots. *The Math. Intelligencer*, 20(4):33–48, 1998.

[28] D. A. B. Hyde, J. Henrich, E. J. Rawdon, and K. C. Millett. Knotting fingerprints resolve complexity and knotting pathways in ideal knots. *J. Phys.: Condens. Matter*, 27(354112):13, 2015.

[29] Spencer Igram, Kenneth C Millett, and Eleni Panagiotou. Resolving critical degrees of entanglement in olympic ring systems. *Journal of Knot Theory and its Ramifications*, 25(14):1650081, 2016.

[30] M. Jamroz, W. Niemyska, E.J. Rawdon, A. Stasiak, K.C. Millett, P. Sulkowski, and J.I. Sulkowska. Knotprot: a database of proteins with knots and slipknots. *Nucleic Acids Research*, 43(1):D306–D314, 2015.

[31] V. Jones. Hecke algebra representations of braid groups and link polynomials. *Ann. of Math.*, 126(2):335–388, 1987.

[32] N.P. King, E.O. Yeates, and T.O. Yeates. Identification of rare slipknots in proteins and their implications for stability and folding. *J. Mol. Biol.*, 373:153–166, 2007.

[33] K. V. Klenin, A. V. Vologodskii, V. V. Anshelevich, A. M. Dykhne, and M. D. Frank-Kamenetskii. Effect of excluded volume on topological properties of circular dna. *J. Biomol. Struct. Dyn.*, 5:1175–1185, 1988.

[34] M. A. Krasnow, A. Stasiak, S. J. Spengler, F. Dean, T. Koller, and N. R. Cozzarelli. Determination of the absolute handedness of knots and catenanes of dna. *Nature*, 304:559–60., 1983.

[35] W. B. R. Lickorish and Kenneth C. Millett. A polynomial invariant of oriented links. *Topology*, 26(1):107–141, 1987.

[36] WB Raymond Lickorish and Kenneth C Millett. A polynomial invariant of oriented links. *Topology*, 26(1):107–141, 1987.

[37] WBR Lickorish and KC Millett. The new polynomial invariants of knots and links. *Mathematics Magazine*, pages 3–23, 1988.

[38] L. F. Liu and J. L. Davis. Novel topologically knotted dna from bacteriophage p4 capsids: studies with topoisomerases. *Nucleic Acids Research*, 9:3979–3989, 1981.

[39] J. Lukes, D.L. Guilbride, J. Votypka, A. Zikova, R. Benne, and P. T. Englund. Kinetoplast dna network: Evolution of an improbable structure. *Eukaryotic Cell*, 1(4):495–502, 2002.

[40] N. Madras and A. D. Sokal. The pivot algorithm: a highly efficient monte carlo method for the self-avoiding walk. *J. Stat. Phys.*, 50(1-2):109–186, 1998.

[41] B Marcone, E Orlandini, A L Stella, and F Zonta. What is the length of a knot in a polymer? *Journal of Physics A: Mathematical and General*, 38(1):L15, 2005.

[42] B. Marcone, E. Orlandini, A. L. Stella, and F. Zonta. Size of knots in ring polymers. *Phys. Rev. E*, 75:041105, 2007.

[43] W. Menasco. Closed incompressible surfaces in alternating knot and link complements. *Topology*, 23(1):37 – 44, 1984.

[44] W. Menasco and M. B. Thistlethwaite. The classification of alternating links. *Ann. of Math.*, 138(1):113–171, 1993.

[45] C Micheletti, D Marenduzzo, and E Orlandini. Polymers with spatial or topological constraints: Theoretical and computational results. *Physics Reports*, 504:1 – 73, 2011.

[46] K. C. Millett. Knotting of regular polygons in 3-space. *J. Knot Theory Ramifications*, 3(3):263–278, 1994.

[47] K. C. Millett. The length scale of 3-space knots, ephemeral knots, and slipknots in random walks. *Prog. of Theo. Physics Sup.*, 191:182–191, 2011.

[48] K. C. Millett. Knots in knots: A study of classical knot diagrams. *J. Knot Theory Ramifications*, 25(9):1641013, 10, 2016.

[49] K. C. Millett and A. Rich. More knots in knots: a study of classical knot diagrams. *J. Knot Theory Ramifications*, 26(8):1750046, 18, 2017.

[50] Kenneth Millett, Akos Dobay, and Andrzej Stasiak. Linear random knots and their scaling behavior. *Macromolecules*, 38(2):601–606, 2005.

[51] Kenneth C. Millett. Knots, slipknots, and ephemeral knots in random walks and equilateral polygons. *J. Knot Theory Ramifications*, 19(5):601–615, 2010.

[52] Kenneth C. Millett and Benjamin M. Sheldon. Tying down open knots: A statistical method of identifying open knots with applications to proteins. In *Physical and numerical models in knot theory*, volume 36 of *Ser. Knots Everything*, pages 203–217. World Sci. Publishing, Singapore, 2005.

[53] B. G. Nielsen, P. Rõgen, and H. G. Bohr. Gauss-integral based representation of protein structure for predicting the fold class from the sequence. *Math. Comput. Modelling*, 43(3 - 4):401 – 412, 2006.

[54] Wanda Niemyska, Pawel Dabrowski-Tumanski, Michal Kadlof, Ellinor Haglund, Piotr Sułkowski, and Joanna I. Sulkowska. Complex lasso: new entangled motifs in proteins. *Scientific Reports*, 6:36895, 2016.

[55] E. Panagiotou, K. C. Millett, S. Lambropoulou, C. Tzoumanekas, and D. N. Theodorou. A study of the entanglement in systems with periodic boundary conditions. *Prog. of Theo. Physics Sup.*, 191:10, 2011.

[56] Eleni Panagiotou. The linking number in systems with periodic boundary conditions. *Journal of Computational Physics*, 300:533 – 573, 2015.

[57] Eleni Panagiotou, Kenneth C Millett, and Sofia Lambropoulou. The linking number and the writhe of uniform random walks and polygons in confined spaces. *Journal of Physics A: Mathematical and Theoretical*, 43(4):045208, 2010.

[58] Eleni Panagiotou, Kenneth C Millett, and Sofia Lambropoulou. Quantifying entanglement for collections of chains in models with periodic boundary conditions. *Procedia IUTAM*, 7:251–260, 2013.

[59] Nicholas Pippenger. Knots in random walks. *Discrete Appl. Math.*, 25(3):273–278, 1989.

[60] L. Plunkett and K. Chapman. Off-lattice random walks with excluded volume: a new method of generation, proof of ergodicity and numerical results. *J Phys, A: Math. Theor.*, 49(13):135203, 28, 2016.

[61] J. Przytycki. History of the knot theory from vandermnde to jones. In *XIVth National Congress of the Mexican Mathematical Society*, volume 11, pages 173–185. So. Mat. Mexicana, Mexico, 1992.

[62] P. Rõgen and H. Bohr. A new family of global shape descriptors. *Math. Biosci.*, 182(2):167–187, 2003.

[63] E. J. Rawdon. Approximating the thickness of a knot. In *Ideal Knots*, volume 19, pages 143–150. World Sci. Publ., River Edge, NJ, 1998.

[64] Eric J. Rawdon, Kenneth C. Millett, and Andrzej Stasiak. Subknots in ideal knots, random knots, and knotted proteins. *Sci. Rep.*, 5:8928, 2015.

[65] Eric J. Rawdon, Kenneth C. Millett, Joanna I. Sulkowska, and Andrzej Stasiak. Knot localization in proteins. *Biochem. Soc. Trans.*, 41(2):538–541, 2013.

[66] R. Stoltz, M. Yoshida, R. Brasher, M. Flanner, M. Ishihara, D. J. Sherratt, K. Shimokawa, and M. Vazquez. Pathways of dna unlinking: A story of stepwise simplification. *Scientific Reports*, 7:12420, 2017.

[67] Joanna I. Sułkowska, Eric J. Rawdon, Kenneth C. Millett, Jose N. Onuchic, and Andrzej Stasiak. Conservation of complex knotting and slipknotting patterns in proteins. *Proc. Natl. Acad. Sci. USA*, 109(26):E1715–E1723, 2012.

[68] D. W. Sumners and S. G. Whittington. Knots in self-avoiding walks. *J. Phys. A*, 21(7):1689–1694, 1988.

[69] William R. Taylor. A deeply knotted protein and how it might fold. *Nature*, 406:916–919, 2000.

[70] W. H. Thomson. On vortex motion. *Trans. R. Soc. Edin.*, 25:217–260, 1867.

[71] L. Tubiana, Orlandini E., and C. Micheletti. Probing the entanglement and locating knots in ring polymers: a comparative study of different arc closure schemes. *Prog. of Theo. Physics Sup.*, 191:192 – 204, 2011.

[72] Erica Uehara and Tetsuo Deguchi. Knotting probability and the scaling behavior of self-avoiding polygons under a topological constraint. 147, 04 2017.

[73] Erica Uehara and Tetsuo Deguchi. Scaling behavior of knotted random polygons and self-avoiding polygons: Topological swelling with enhanced exponent. 07 2017.

[74] S. A. Wasserman and N. R. Cozzarelli. Biochemical topology: applications to DNA recombination and replication. *Science*, 232(4753):951–960, May 1986.

[75] J. H. White, K. C. Millett, and N. R. Cozzarelli. Description of the topological entanglement of dna catenanes and knots by a powerful method involving strand passage and recombination. *Journal of molecular biology*, 197(3):585–603, 1987.