Topological decoding of biomolecular fold complexity

Scalvini, B.

CHAPTER 1:

TOPOLOGY OF FOLDED MOLECULAR CHAINS: FROM SINGLE BIOMOLECULES TO ENGINEERED ORIGAMI

The topology of biological polymers such as proteins and nucleic acids is an important aspect of their three-dimensional structure. Recently, two applications of topology to molecular chains have emerged as major theoretical developments which are beginning to find utility in heteropolymer characterization and design, namely circuit topology and knot theory. Here, we review the application of these two theories to protein, RNA, and DNA/genome structure, focusing on connections to conventional 3D structural information, relevance to function, and highlighting recent experimental findings. We conclude with a discussion of recent applications to molecular origami and engineering.

Publications associated to this chapter:
Barbara Scalvini, Vahid Sheikhhassani, Jaie Woodard, Jana Aupič, Remus T. Dame, Roman Jerala, Alireza Mashaghi, *Topology of Folded Molecular Chains: From Single Biomolecules to Engineered Origami*, Trends in Chemistry, 2, 7, P609-622, July 01, 2020, https://doi.org/10.1016/j.trechm.2020.04.009

1. TOPOLOGY: A KEY PROPERTY TO DISENTANGLE FOLDING COMPLEXITY

Despite their apparent simplicity, linear heteropolymer chains may fold into distinct topologically diverse structures. In polymer chemistry, the diverse collection of linear polymers is supplemented by branched and cyclical structures, while in biological chemistry linear protein and nucleic acid chains adopt various topologies via chain folding. Folding involves rearrangements of the chain and formation of contacts. In biology, we encounter a vast multiplicity of folded polymer identities, with chemical and structural, as well as functional, relations. Topology can not only help us make sense of the complex network of structural relationships but can provide insights into folding mechanisms, conformational dynamics, and folding stability, ultimately aiding protein and drug design[1][2][3][4]. A particular challenge, and opportunity for innovation, has been the application of topology to folded linear chains where three-dimensional structures are stabilized by non-covalent intra-chain contacts[5].

In this introduction, we highlight two recent applications of topology within the biomolecular sciences and molecular engineering. These topological approaches promise to categorize linear polymer structures. Knot theory categorizes molecular structures based on whether and how they are knotted. Circuit topology, in the context of polymer structure, categorizes folded linear chains based on their contact arrangement (Figure 1A), allowing structural summaries and comparisons in terms of topological building blocks and sets of permutation operations.

2. KNOT THEORY & CIRCUIT TOPOLOGY: BASIC DEFINITIONS

Formally, a **knot** is an embedding of the circle in three-dimensional space. A knot may be equivalent (through stretching and bending operations, without allowing the knot to pass through itself) to the trivial knot, or circle, or to other knots with greater minimal numbers of crossings in their projections onto the plane. In contrast to proteins, RNA, and linear DNA, such knots lack a start and end point. However, linear molecules, upon connecting the endpoints across an external arc traversing the 3D surface, may be said to be knotted or unknotted (trivial knot), according to the topology of the backbone[6].

A folded chain is formed when a polymer establishes intra-chain contacts: two contact-sites along the chain come in close proximity, creating either pair-wise or higher order connections. The **circuit topology** (CT) of a folded polymer chain defines the arrangement of intramolecular contacts with respect to the

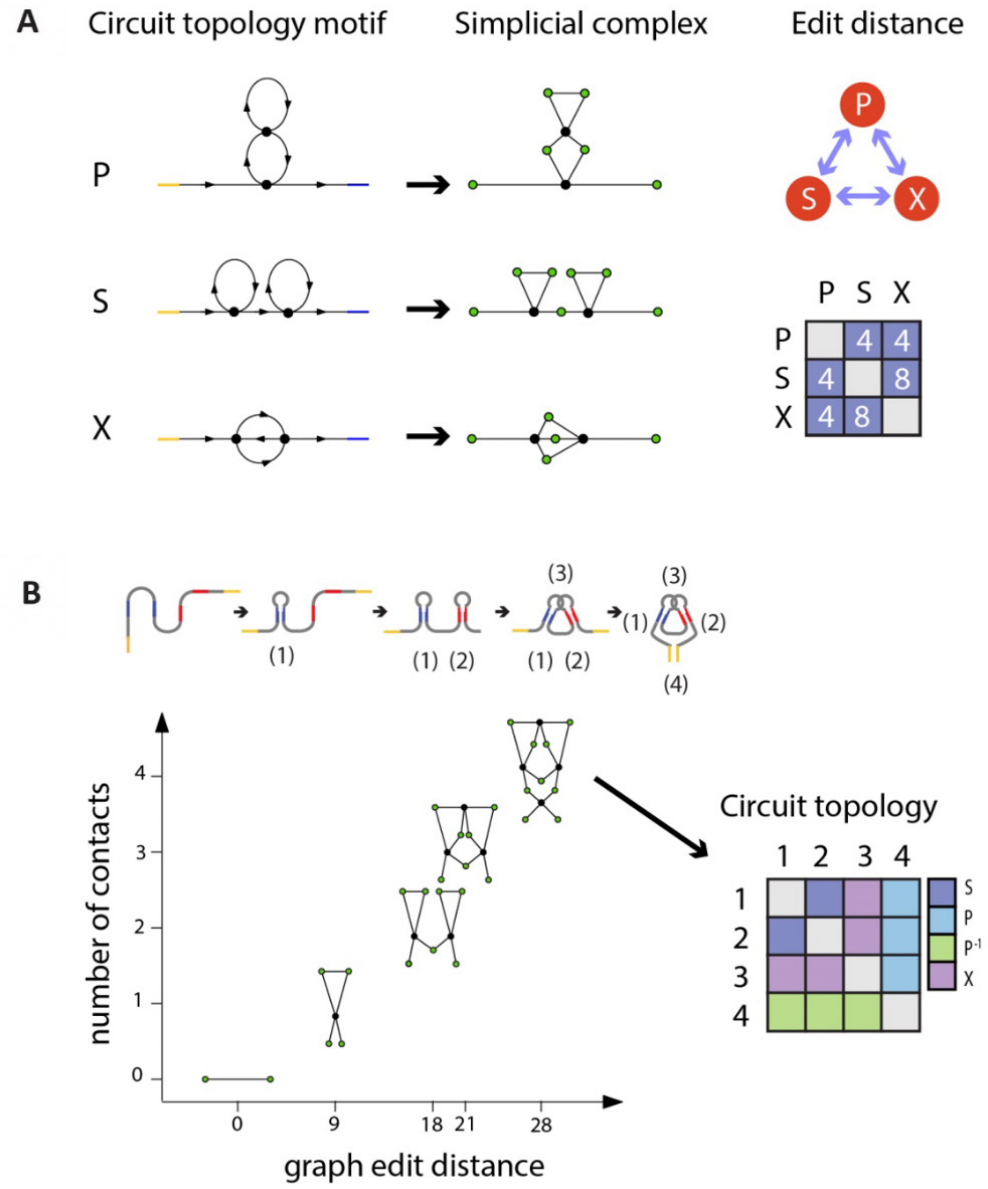


Figure 1. Simplicial complex representation of the circuit topology formalism.
A Simplicial complex representation of two contacts in parallel (P), series (S) and cross (X) relation. In order to transition from one configuration to another, vertices and nodes need to be edited (added or removed). Such analysis provides a framework for calculating distances between structures via a graph editing approach. In ***B*** a further example of this concept is shown, applied to the folding process: the folding of E adenine riboswitch is represented in a graph where the x-axis represents the graph edit distance, and the y axis the number of contacts (see [14] for further information).

path between polymer ends. The approach is simple and generic and can be represented according to an algebraic formalism, providing quantitative measures for comparative analysis and experimental studies. For a given pair of binary contacts, three arrangements can be identified: parallel (P), series (S), and cross (X) (Figure 1A)[5]. In addition, two contacts may be in concerted series (CS) or parallel (CP) relation if they share a site[7]. Here, a "site" or node may have one of several definitions: for instance, it may be a protein residue, a single nucleotide, or an element of secondary structure. Furthermore, the definition of a contact may incorporate a cutoff distance (or atom type-specific distances) and number of atom-atom contacts or may focus on a particular type of contact such as a disulfide bond[8]. When needed, one can simplify the representation, for example by treating the asymmetric parallel relation as a symmetric one, or by extending the definitions of P and S to merge them with CP and CS relations respectively. Given suitable definitions, a matrix of relations between pairs of contacts may be constructed (Figure 1B).

Here, we focus on circuit topology and geometric topology approaches, or more specifically, on knot theory. There are a wide variety of other topological methods that have been developed for molecular sciences, including algebraic topology (e.g., persistent homology), differential topology (e.g., de Rham Hodge theory, quantum topology, topological order) [9]. Among these methods, persistent homology appears to be more promising for biomolecules [10][11][12][13]. Persistent homology approaches are reviewed elsewhere [9]. However, we note that CT analysis can be readily combined with existing persistent homology tools. CT motifs introduced above can be readily represented in the form of simplicial complexes and subjected to algebraic topology analysis (Figure 1)[14].

3. TOPOLOGICAL ANALYSIS OF PROTEINS

Proteins, known as the primary machinery of life [15], often need to fold transiently or permanently into one or more specific spatial conformations, mostly driven by non-covalent interactions [16][17]. Among the unlimited possibilities of arrangements, a limited number of motifs and domains is exhibited by nature, evidencing some general rules which govern the complexities of protein structure [18]. Different theoretical methods including knot theory [19], knotoids [20] and, recently, circuit topology [5] have been developed to formalize the structural relationships among diverse proteins. In this brief introduction to biomolecular topology, we make use of the Alexander-Briggs notation[21] to characterize knots (as, e.g., in Figure 2A). In this notation, a knot is represented by two numbers: the main one indicates the crossing number and its subscript provides the identification number of the knots with the same crossing number.

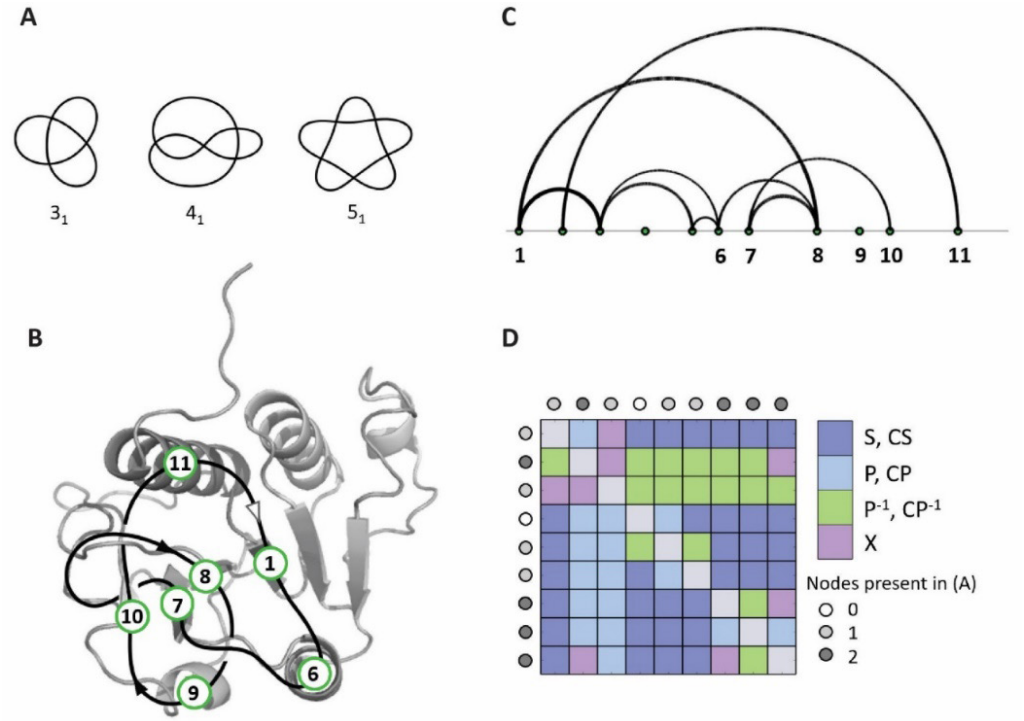


Figure 2. Knot and circuit topology representation: a comparison. Topology representations of the YibK methyltranferase, which exhibits a 3_1 knot forming the co-factor binding site. ***A*** Projections of three knots. YibK is an example of a 31 or trefoil knot. ***B*** Protein structure, with knot diagram overlayed. Secondary structural elements along the knot diagram are numbered according to their position along the backbone. ***C*** Circuit topology diagram with numbered elements in (B) numbered underneath the diagram. Cutoff: 3.7 Angstroms, 4 contacts. ***D*** Circuit topology matrix of YibK methyltranferase, retrieved from the diagram in (C). The grayscale dots represent the number of nodes for each interacting loop pair which are part of the knot.

3.1. Knot and knotoids in proteins

Structural analysis of 400 knotted proteins, diverse in sequence and family, showed that knotting pattern in proteins is strictly evolutionarily conserved [22]. There are several families of proteins that reproducibly form simple knots, complex knots, and slipknots; in these proteins, the disadvantage of less efficient folding may be balanced by a functional advantage connected with the presence of these knots [23]. Knot theory appears to be a powerful approach to explore structural, mechanical, and functional roles of such entangled topological features in proteins [24]. Most of the knots are located in functionally important positions in protein structure. Recent research has established that knotted cores, and especially their borders, show strong enrichment in the number of contacts

with surrounding structural elements. Buried inside the protein structure, these regions showed increased thermal stability, providing a favorable environment for the protein active site [25]. Despite advances, knot theory has limitations. The fraction of knotted proteins is only 0.77% of all proteins [26]. Also, to adhere with formal mathematical definitions, knots must be closed rings, which are rare in protein structures. In 2012, as a generalization of knot theory, knotoids were introduced as diagrams representing projections of open curves in 3D space [27]. Due to the open and dynamic nature of protein structure, knotoids have attracted interest for studies of global and local entanglements of proteins [28]. Results are now accessible through online databases [29][30].

In recent years, several knot theory inspired topological descriptors have been suggested, which proved to be quite useful for the characterization of folding processes [31]–[34]. Among these, the *Gaussian entanglement* indicator proved to correlate moderately with folding rate of a protein [33], while not requiring computational procedures for the artificial closure of the curve (connecting the two ends of the protein), thus overcoming one of the main limitations of knot theory.

3.2. Circuit topology of proteins

Despite many applications, both knot and knotoids theories ignore intra-chain interactions. Circuit topology is well suited to address this challenge. Within this framework, a wide range of biologically important interactions could be considered: e.g., contacts inclusive of the covalent S-S bond, interactions between secondary structural elements (e.g. β-β and α-α), and connections between coevolving groups known as sectors [35]. Knot theory and CT focus on two complementary aspects of biomolecular chains, that is, backbone entanglement and contact formation. However, there have been substantial attempts in recent years to bridge these two aspects in one unifying theory. The concept of quandles [36], algebraic objects used to distinguish between knots, was adapted for the classification of chains within the circuit topology framework [37]. Moreover, a generalization of circuit topology relations to characterize backbone crossings in the form of soft contacts has been formalized[38]. These extensions show promise to promote a unification of knot-based and contact-based topology frameworks (such as CT), providing a uniform language for a working topological description. An example of how these two approaches can provide complementary descriptions of a protein is portrayed in Figure 2. Figure 2B shows the knotted protein YibK, with the overlaid knot diagram of the knotted portion of the structure, alongside the protein topology diagram and matrix (Figure 2C, 2D. Strands 1, 8, 7, and 10 form part of the protein's beta sheet, including two contacts in cross relation. However

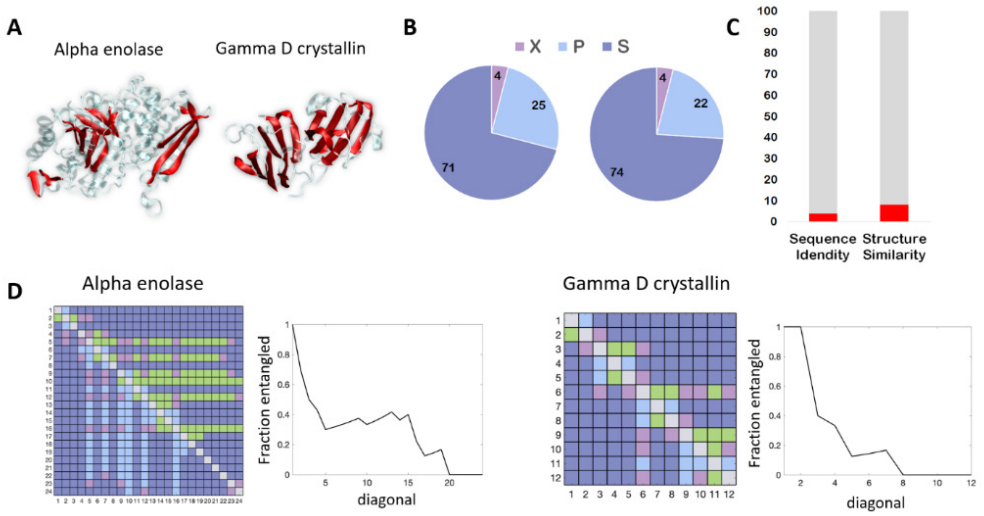


Figure 3. Topological comparison between two moonlighting proteins. Preliminary data showing two proteins that are different in sequence and structure, yet similar in topology and function. ***A*** Crystal structure of α-enolase and gamma crystallin D: red parts indicate extended beta strands. ***B*** Striking similarity in the frequencies of topology motifs extracted from circuit topology matrix for atom-atom topologies. P: parallel, S: series, X: crossing arrangements; ***C*** The two proteins have negligible sequence overlap and structural similarity as estimated by jFATCAT_rigid comparison method. ***D*** Circuit topology matrix of the two proteins: the 'entangled' relations are clustered along the diagonal and decay with the distance from the diagonal in a similar way for both proteins. Our analysis raises the question as to whether these topological similarities can be generically related to protein function.

intriguing, such developments have yet to be applied to real protein structures and will most likely be the object of future research. In this thesis we base our analysis on CT in its original definition [5], therefore focusing specifically on *hard contacts* – fixed contacts between strands defined by the folding process – and ignoring backbone entanglement. However intriguing, such developments have yet to be applied to real protein structures, and will most likely be the object of future research.

The native circuit topology of a protein may inform on its function. For instance, crystallin is a moonlighting protein [39] mainly known as a structural protein but also, in some cases, exhibiting enolase activity. Comparison of the CT map of α-crystallin and human enolase showed negligible sequence and geometric similarities (Figure 3A, 3B, 3C) but striking similarities in the frequencies of topology motifs extracted from the CT matrix (Figure 3D). In this thesis, we make a fundamental step towards disentangling the role that native topology plays in folding mechanisms (**Chapter 2**). It has been previously shown how various structural descriptors such as absolute and relative contact order[40]–[42], as

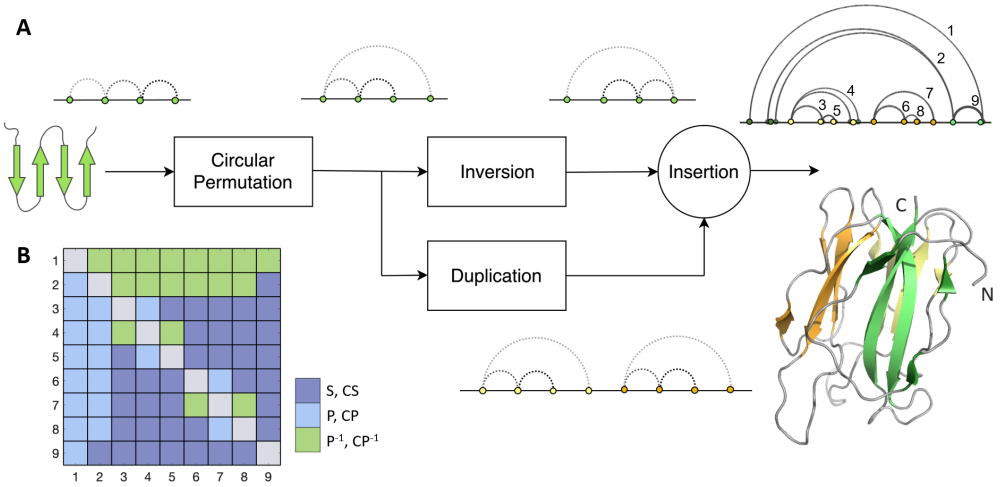


Figure 4. Molecular operations as topological permutations. A more complex protein topology is built from permutations of a simple circuit topology motif. ***A*** Diagram showing the construction of the circuit topology of membrane protein VMO-I (PDB ID 1VMO) based on permutations of the concerted series arrangement belonging to the up-down-up-down four-strand motif. ***B*** Relations between contacts in the circuit topology of VMO-I, with sites numbered as in (C). P-1 indicates the inverse of the parallel relation (loop i includes loop j), while CP and CS are parallel and series relations in which one of the contact sites is shared between the two loops (See [6]). ***C*** Circuit topology structure of VMO-I, with nodes corresponding to beta strand segments and edges weighted according to the number of contacts. An atom-atom distance cutoff of 3.5 Angstroms and number-of-contacts threshold of 5 contacts were employed.

well as size [43], can be effective folding rate predictors. Being size invariant and flexible in defining the contacts, CT complements geometrical predictors such as contact order in estimating the folding rates and number of unfolding paths of a macromolecule [3]. We show in **Chapter 2** how the addition of CT parameters to size and contact order increases the accuracy of folding rate predictions [44]. However, the applications of circuit topology to folding mechanisms provide a deeper insight than folding rate alone. The folding process in proteins is by nature a dynamic and collaborative process. Molecular machines such as chaperones transiently interact with proteins during this process [45]. CT analysis of model systems mimicking protein-chaperon interaction revealed how the latter guides the conformational search towards certain topologies and away from others [46] [47]. Moreover, there is growing interest in the topology of protein-protein interaction in the context of macromolecular complexes and condensates[48], [49], in virtue of the fundamental role they play in many cellular processes[50]–[52]. In order to provide a complete topological characterization of protein complexes,

CT was expanded to consider both intra- and inter-chain interactions [53].

Biomolecular systems are often dynamic in nature. Therefore, an effective topological framework needs to provide tools for the representation and analysis of topological evolution. Recently, Schullian et al. developed a mathematical framework to describe the circuit topology of a biomolecule and topological changes such as standard and circular permutation, duplication, and addition/elimination of contacts [7]. It was found that topology permutations underlie aspects of protein evolution and dynamics such as domain swapping upon mutation and hairpin flipping within a beta barrel. Figure 4 shows how a relatively complex protein can be built from permutations and combinations of a simple topology. Figure 4A and 3C show the progression from a simple topology to the final product, for the protein membrane protein VMO-I. Figure 4C shows the topology matrix of the protein.

Folding pathways can also be mapped onto a topology landscape, allowing for identification of topological transitions and topological traps (misfolds) [54]. This technical advantage appears to be crucial especially for those molecular systems that are highly dynamic in nature and lack a stable native structure, such as intrinsically disordered proteins (IDPs). In **Chapter 3** we present dynamic circuit topology (dCT), an extension of the CT framework for the characterization and comparison of disordered chains [55]. This framework allows us to uncouple the temporal evolution of short-lived contacts from those which, however transient, live longer, shaping its trajectory on the conformational space. We identify transient states in the trajectory, enabling comparison between the topological and dynamic fingerprints of different proteins. Understanding the mechanistic principles that govern such systems is crucial, considering the fundamental role that IDPs play as drug targets in many diseases [56]–[59].

The recent development of single molecule techniques to study protein folding has led to an increased demand for theoretical tools to interpret experimental data [60][61]. Optical tweezers experiments on model and some human proteins demonstrated different steps of conformational change at distinct lengths [62] [63]. Since force jumps between different lengths are related to the breaking of connections [39], contact-based frameworks such as CT provide attractive opportunities for gaining molecular insight from force spectroscopy data. In **Chapter 4** we take a first step in this direction, by showing how to obtain a topology matrix from experimentally retrieved force-extension diagrams.

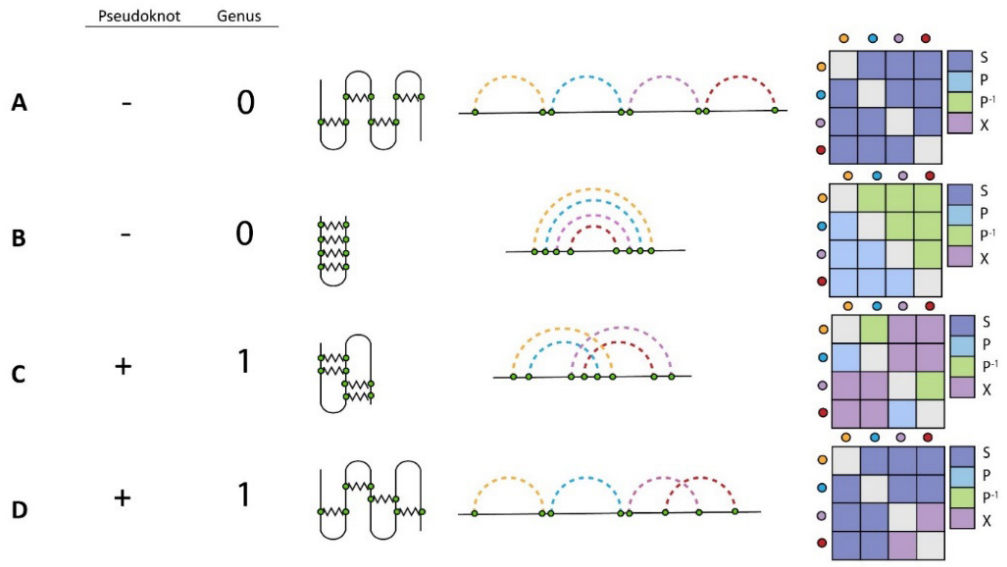


Figure 5. Pseudoknots, genus and circuit topology: a comparison. Four examples of RNA structures and their diagram representation. The first two structures (A,B) are not pseudoknotted, while structure C and D contain pseudoknots. Pseudoknots correspond to cross relations in the CT matrix. We show that structures with the same genus can have dramatically different topologies. Structure A has genus 0, but only contains Series relations. Structure B on the other hand, while still having genus 0, only contain the so called "entangled" relations: Parallel and Cross. Similarly, figure C and D have genus 1, even though structure C only contains Cross and Parallel relations, structure D is dominated by Series relations. The only common trait between C and D is the presence of Cross relations, which indicate the pseudoknot.

4. TOPOLOGICAL ANALYSIS OF NUCLEIC ACIDS

Cellular nucleic acids often fold into globular structures to achieve function. Folding happens at various scales, from small RNA molecules to large eukaryotic genomes. Various topological concepts, including supercoiling, knot theory, and contact arrangement, have been developed to describe folded nucleic acids. In what follows, we summarize these developments and discuss how circuit topology can be used as a universal topology framework.

4.1. Topology of RNA

RNA molecules may fold back on themselves to form complex 3D shapes capable of ligand/target recognition and catalysis. These structures can be achieved by means of several mechanisms, including hydrogen bonding and stacking

interactions as in tRNA tertiary structure; ions bound to specific sites (as found in a ribosomal RNA fragments); and pseudoknot folds, as seen in mRNA fragments with extensive non-canonical pairing structure (in contrast to canonical Watson-Crick pairing) [64]. **Pseudoknots**, which are segments in the secondary structure where half of one stem is intercalated between the two halves of another stem, are abundant in RNA molecules and can have important functional implications [65]; thus, topological classifications of RNA have been mainly focused on pseudoknots and on the concept of topological genus. RNA secondary structure can be schematically represented by a planar diagram, with straight lines representing the backbone of the molecule and arches representing the bonds that give the molecule its characteristic folded shape. The RNA secondary structure is said to be pseudo-knotted if the diagram indicates crossing among base pairs. These crossings, in diagram representation, are equivalent to cross relations in the CT framework [66]. Figure 5A and 5B display RNA structures with no pseudoknots (no cross relations). Figure 5C and 5D show examples of pseudo-knotted RNA diagrams. A given RNA molecule thus represented can then be characterized by the **genus** of the auxiliary two-dimensional surface associated to the diagram, that is, a sphere with handles. The genus g of a diagram is the minimum number of handles a sphere must have in order to be able to draw the diagram on it without any crossing [66][67][68][69].

By comparing the planar diagram representation with the CT diagram (Figure 5) it is possible to draw a parallel between planar diagrams and CT: cross relations, where the number of arches n is equal to the number of loops in the chain. Therefore, if we were to calculate the topology matrix of a pseudo-knotted RNA molecule, we would obtain a n×n matrix, such as the one represented in Figure 5. Given the similarity between these two representations, the genus of a CT diagram can be readily calculated, and consequently a topology matrix. Recently, pseudoknot classification and comparison in RNA molecules was given a new algebraic formalism [70]. Here, RNA structures are represented as expressions of an algebraic language with three operators (concatenation, nesting and crossing) and simple hairpin loops as operands. In the language of CT, concatenation, nesting and crossing correspond to series, parallel and cross relations. These relations were also given an operator representation [7]. Other creative frameworks exploit graph theory. The RNA-As-Graph approach involves the translation of an RNA 2D structure into tree and dual graph objects. In tree graphs, stems are the edges, while junctions, bulges, and loops are the vertices [71]. Once again, we draw a parallel with CT, in its simplicial complex representation [14], where transitions from one topology relation to another are described and quantified in terms of graph editing, e.g. addition or removal of vertices and edges from a graph (Figure 1B). CT as such provides a unified language for the description of RNA folds.

4.2. Topology of cellular DNA

The genome topology plays a fundamental role in the regulation of gene expression. The genomic spatial arrangement is shaped by chromatin long-range interactions, which are mediated by architectural proteins such as CTCF and Cohesin [72][73]. These interactions causes the formation of chromatin loops and eventually organize in topologically associating domains (TADs) in mammals. Similar types of domain arrangements are found in lower eukaryotes [72] and prokaryotes [74][75]. Alterations in chromatin topology are key to cell differentiation [76] and have been implied as drivers of oncogenic programs [77]. Genetic mutations that affect chromatin topology potentially lead to changes in gene expression, therewith facilitating disease susceptibility and evolutionary adaptation [78]. Providing a rigorous topological framework is therefore a fundamental step to shed light on the link between genome topology and function.

4.2.1. Supercoiled DNA

Early efforts to characterize topology in DNA were focused on supercoiling. DNA supercoiling is the consequence of twisting DNA: it describes the coiling of the axis of the double helix. Supercoiling occurs in the DNA of organisms at all levels of evolutionary complexity. Human interphase chromosomes are divided into domains with different levels of supercoiling, where under-wound domains are transcriptionally active, cytologically decondensed, and topologically constrained [79]. These domains were shown to frequently correspond to TADs detected by 3C (Chromosome Conformation Capture) methods [79]. A topological constant commonly used to characterize supercoiling is the **linking number Lk**, which represents the number of times the two strands of the DNA double helix are intertwined [80]. This parameter can be expressed as the sum of two geometric parameters: writhe (Wr) and twist (Tw) [80]. Wr measures the coiling of the DNA axis, and Tw the helical winding of the DNA strands around each other. Although this formalism was developed for circular DNA, supercoiling has also been observed and studied experimentally in linear segments of double-stranded DNA (dsDNA) [81][82]. The twisted linear DNA forms intertwined loops called plectonemes, the dynamics of which could be studied as they diffuse or hop along the DNA strand. From formal point of view, plectonemic loops and their dynamics can be readily represented with CT terminology.

4.2.2. Knotted DNA

DNA at short length scales (<50 nm) is a stiff polymer, but its considerable

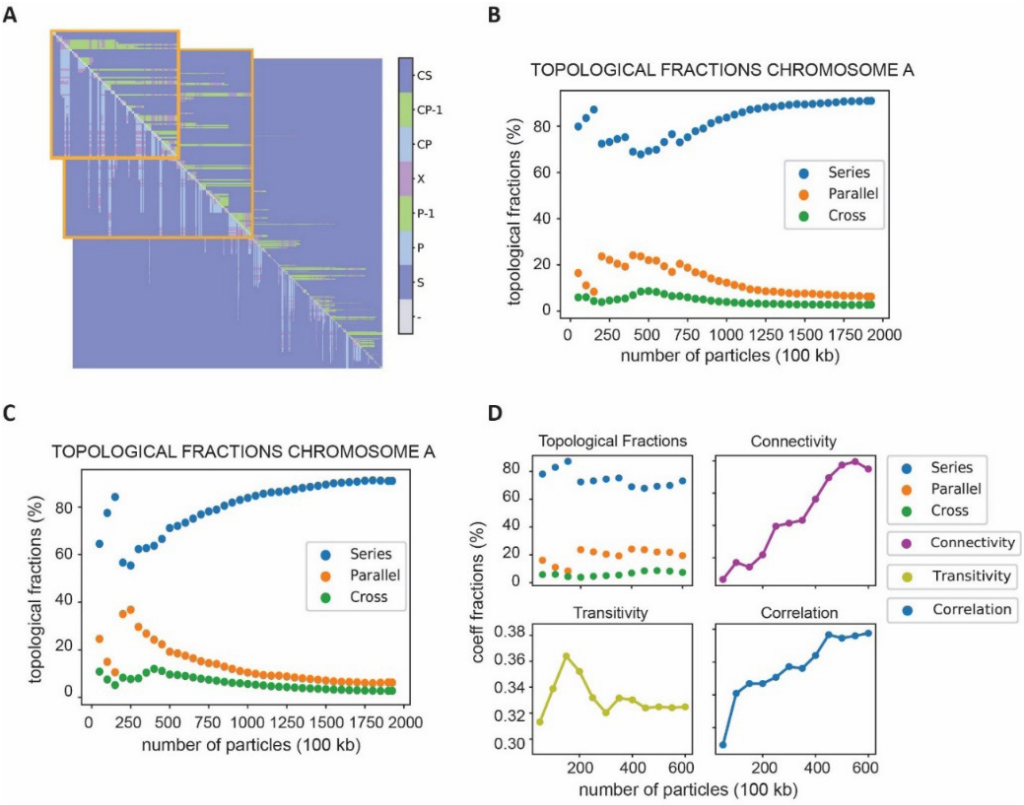


Figure 6. Circuit topology analysis of a chromosome. Circuit topology analysis of the first chromosome of a single murine ES cell [68]. ***A*** Circuit topology matrix, calculated by choosing a cutoff radius (1 particle radius) to define contacts between 100-kb particles. The three images show progressive close-ups on smaller areas of the matrix. We can see how parallel and cross relations cluster around the diagonal, suggesting a domain-like structure. In ***B*** and ***C*** we can see the calculated topological fractions for a progressively higher number of particles. This cumulative analysis shows how we can have an indication of the domain- like structures when we have a high resolution (few particles). When the resolution lowers (right hand side of the graph), the fractions remain constant, indicating that the relative proportion of series, cross and parallel does not depend on the number of kb included in the analysis, after a certain threshold (which, for this chromosome, is about 750 particles). In B the particles were computed from the top of the PDB file to the bottom, while in C from the bottom up.
D Comparison between CT analysis and network analysis for the first 600 particles. The network was build using contact sites as nodes and contacts as edges. Two network parameters, average connectivity and Pearson correlation, show a step-wise behavior similar to that displayed by the topological fractions. Transitivity presents a rise at small scales.

length, of the order of millimeters in bacteria and meters in humans, makes it very liable to self-entanglement and knotting [83]. Knots in packaged viral DNA have been widely documented in the literature [84]. The microns-long viral DNA molecules are tightly packed and condensed inside a capsid, which has a size

in the range of tens of nanometers [85]. This strong confinement facilitates the occurrence of knots, with a distribution of knot types which is biased towards complex knots: gel electrophoresis characterization revealed a predominance of the torus knot 5_1 and scarcity of the achiral knot 4_1 [86]. More recently, small steady state fractions of DNA knots were also found in chromatin inside cells [87]. There is debate about the extent and scale to which knots are present at the chromosome scale. The 100 kb resolution analysis of individual chromosomes in the nuclei of single haploid mouse embryonic stem (ES) cells obtained by Hi–C contact data [88] revealed that chromosomes do contain knots, with the fraction of unknotted chromosomes being less than 20% [89]. Moreover, knots with more than five crossings or even multiple knots appeared to be the most popular kind, representing more than 50% of the knot population [89].

Various single molecule techniques have also been used to characterize DNA knots. For instance, nanopore sensors have been used to map the equilibrium configurations of DNA knots, revealing a wide distribution in tightness. The persistence of very loose knots might have implications for understanding the efficiency of the biological mechanisms accountable for unknotting the molecules [90], like, for example, the action of type II DNA topoisomerases [91]. Considering the new wealth of information we have on contacts in genomic structures (see below) and the high likelihood these structures have of producing complex knots, a generalized knot theory approach such as the one presented by Adams et al. [36] may be a useful direction for future research.

4.2.3. Contact arrangement in cellular DNA

The development of innovative technologies such as fluorescence in situ hybridization, in vivo tagging of genomic loci, and 3C and Hi-C based technologies have led to a rapid increase in available structural information. Hi-C technology is particularly suited as a source of topological data for chromosomes, since it allows for the identification of long-range interactions in a genome-wide fashion [92]. This process results in large libraries of pairwise chromatin interactions, which reveal highly reproducible features such as TADs [93]. CT represents a natural framework for this type of data, since it only relies on contact indexes to encode topological information (Figure 6). In this thesis we provide the first description in terms of circuit topology of chromosome data derived from single cell genome structures (**Chapter 5**). We show how the topology matrix of a chromosome encapsulates the fingerprint of conformational motifs (Figure 6A), which can be extrapolated to build a model for chromatin looping, the L-pattern [94]. Figure 6A shows progressive close-ups into the CT matrix of a chromosome; in a conceptually similar fashion, Figure 6B and 6C show the topological fractions

for chromosome sections of different sizes, while presenting a comparison with parameters derived from network topology (Figure 6D). We exploited such a cumulative analysis to highlight periodic features in the structural make-up of chromosomes, and calculate their characteristic scale length (**Chapter 5**).

5. TOPOLOGY OF ORGANIC AND BIOINSPIRED POLYMERS

Advances in molecular engineering enabled synthesis of molecular knots and topological polymers, and have led the way towards applications in several fields, including chemical biology, medicine, and materials science.

5.1. Engineered folded DNA structures

DNA has been demonstrated as a versatile building block for objects such as two-dimensional crystals [96], nanotubes [97], and three-dimensional nanopolyhedra [98]. Many DNA-based materials involve branched molecules (DNA bricks), in which branch points represent the vertices of various types of polyhedra [98][99]. These building blocks are created with techniques that combine hybridization (sticky-ended cohesion) and synthetic stable branched DNA (as, e.g., Holliday junctions). Others rely on the design of scaffolded DNA origami, where one long, single-stranded DNA molecule is folded into arbitrary two-dimensional shapes, which are then the building blocks for larger assemblies [100]. Here, we will focus on those cases where one single molecule is folded.

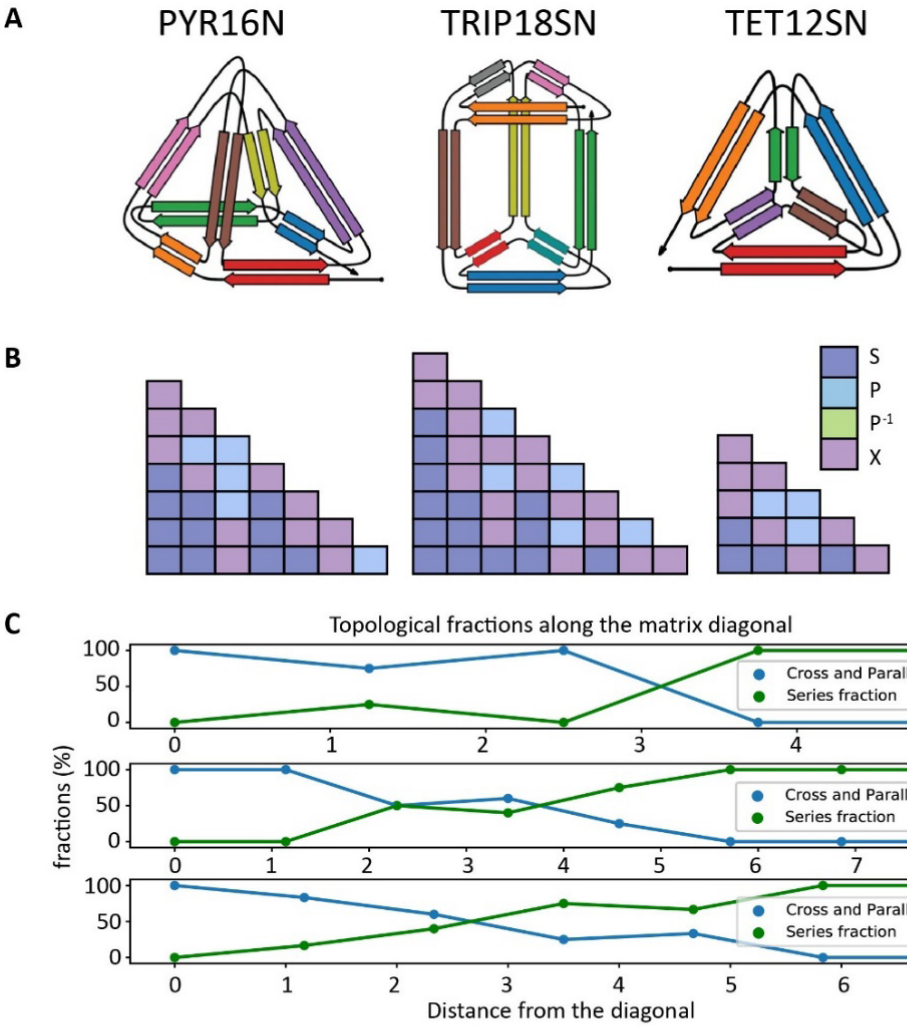
There is a growing interest in designing knotted nucleic acids [101]. Kočar et al. presented the design principles to fold highly knotted single-chain DNA nanostructures. One of the key principles hereby demonstrated is the identification of favorable and unfavorable folding steps, from a topological and kinetic point of view. These steps are identified by considering the pairwise connections that are created during folding, and by classifying these connections by using circuit topology [5]. This strategy demonstrated that highly knotted structures can be formed based on the stepwise formation of connections defined by their decreasing stability as the alternative folding pathways that results in structures of the same stability could not form the knotted structures. This is an example of how CT and knot theory combined can be used to engineer the topological features of a chain.

For what concerns contact-based topology, Han et al. presented a new strategy to design and synthesize a single DNA or RNA strand to self-fold into a complex (user-prescribed) structure [102]. In their approach single molecules of ssDNA and RNA with synthetic sequences ranging in length from ~1000 to ~10,000 nt were folded into origamis. Knotting in these structures is prevented, in order to

avoid kinetic traps and assure smooth folding. While these origami structures are unknotted, their contact arrangement topology is quite elaborate. The design principle is based on parallel crossover, with layers which are covalently linked in a raster filling pattern. From a knot perspective, these structure all belong to the same class (the unknot). Therefore, a contact base topology such as CT is necessary to detect their distinguishing features.

5.2. Synthetic proteins

Advances in nucleic acid engineering have inspired analogous designs for proteins. Proteins are programmable polymers, which can fold into elaborate three-dimensional structures, and are therefore particularly versatile for the engineering of materials with tailor-made structure and function. Design principles correlating loop geometries and secondary structure packing orientation allows for accurate protein size and length control, as investigated by Baker et al. [103] [104]; this loop-based characterization is highly compatible with CT. Ljubetic et al. designed self-assembling coiled-coil protein-origami (CCPO) cages of various geometries (tetrahedra, a four-sided pyramid, and a triangular prism), and provided a computational platform for the design of arbitrary complex CCPO polyhedra [105]. These structures combine the modular building strategy of DNA (DNA bricks) nanotechnology with the programmable functionality of amino-acids. They have interesting physical properties, which can be studied from a circuit topological point of view. In Figure 7A we show the CCPO cage structures from [105] and the corresponding CT matrices (Figure 7B). We can see in Figure 7C and 7D that these three cages are strikingly similar for what concerns topological fractions (percentages of series, cross and parallel relations) and show relatively low contact order. This proof-of-concept study demonstrates that the CCPO cages can be constructed with desired contact order and topology. We note that topology determines folding pathway; furthermore, the topology and contact order may independently affect the folding rate. However, a systematic analysis of CCPO cages based on topological traces with different CT but shared contact order has not been performed, and whether some combinations of the CT topological fractions might be more helpful than others in promoting stability and other kinetic properties remains to be seen. While knots could potentially form between linker regions located at the vertexes of CCPO polyhedra, they cannot be programmed into the designs at the current stage. Extending the length of CC building modules to encompass a full turn has the potential to design knotted protein structures with the possibility to design the folding pathway and make highly knotted proteins or polypeptide-based materials.



D

Protein	Cross	Parallel	Series
PYR16N	43%	18%	39%
TRIP18SN	46%	15%	39%
TET12SN	56%	17%	27%

Figure 7. Circuit topology analysis of origami proteins. A Three examples of CCPO cages [82], namely a tetrahedron (TET12SN) a pyramid (PYR16N) and a trigonal prism (TRIP18SN). ***B*** The CT matrices (of which half of it is shown, since they are symmetrical), show remarkable similarities both in topological fractions and in how they are located in the matrix. First of all, most the dominant topological fractions appear to be Cross in all three cases. Secondly, most of Parallel and Cross relations are clustered along the diagonal of the matrix, indicating that most short-range contacts have this type of arrangement. Series contacts are only present

in the corner of the matrix, indicating that Series dominates long range distances along the chain. If we calculate the percentage of the fractions along diagonal lines in the matrix, from the matrix diagonal (i=j), towards the periphery of the matrix, and we plot the percentages (***C***), we see that Cross and Parallel start from a maximum and decrease to zero, while Series has a specular behavior, starting from zero and reaching a maximum for maximum distance from the diagonal. ***D*** Percentage of topological relations corresponding to the CCPO structures displayed in (A).

5.3. Topology and organic chemistry

New topological features at the molecular level can introduce new material properties. Many efforts in this direction have been focused on molecules created by interlocked chains (as opposed to single folded chains), such as catenanes and rotaxanes [106] and on networks of interconnected molecules called polymer networks [107]. On the other hand, the field is also starting to obtain a better understanding of the strategies needed for the synthesis of a single entangled molecular strand, as in the case of molecular knots. The steric restrictions imparted on the molecule by knotting hinder the range of movement of the molecular components, significantly influencing physicochemical properties [108]. So far, four types of knots have been successfully synthesized using small-molecule building blocks: the trefoil [109][110] figure-eight [111], pentafoil [112], and 8_{19} knot (a knot with eight crossings) [113]. A comprehensive theoretical framework would not only allow characterization but would also be beneficial for the practical purpose of purifying polymers with different topologies, as exemplified in [114]. In that study, it was shown using simulations how nanopores can be used for sensing and enriching certain circuit topologies.

6. CONCLUDING REMARKS AND FUTURE PERSPECTIVES

Contact-based circuit topology and knot theory form two complementary frameworks for describing, understanding, and engineering linear biopolymers such as proteins and nucleic acids. An important future development will be further integration(s) of these two applied theories and establishment of how they can be more generally utilized in prediction and design. Towards this goal, it is likely that machine learning and artificial intelligence (AI), including recent advances protein structure prediction (such as the AlphaFold 2 algorithm [115]) will play key roles. We believe a smart topological encoding such as the one provided by Circuit Topology will prove fundamental in both increasing efficiency of such techniques and providing a new perspective on the role of topology in the function of biomolecular polymers. The potential of Circuit Topology is not limi-

ted to the field of biopolymers. The flexibility of the framework makes it suitable for application to any chain-like system where it is possible to define a direction and intra (or inter) chain contacts. Possible applications to abstract chains encompass for example the fields of chemical reactions [116] and natural language processing (**Chapter 6**).

7. REFERENCES

[1] X.-W. Wang and W.-B. Zhang, "Chemical Topology and Complexity of Protein Architectures," Trends Biochem Sci, vol. 43, no. 10, pp. 806–817, Oct. 2018, doi: 10.1016/j.tibs.2018.07.001.

[2] P. Dabrowski-Tumanski and J. I. Sulkowska, "Topological knots and links in proteins," Proceedings of the National Academy of Sciences, vol. 114, no. 13, pp. 3415–3420, Mar. 2017, doi: 10.1073/pnas.1615862114.

[3] A. Mugler, S. J. Tans, and A. Mashaghi, "Circuit topology of self-interacting chains: Implications for folding and unfolding dynamics," Physical Chemistry Chemical Physics, vol. 16, no. 41, pp. 22537–22544, 2014, doi: 10.1039/c4cp03402c.

[4] E. Flapan, A. He, and H. Wong, "Topological descriptions of protein folding," Proc Natl Acad Sci U S A, vol. 116, no. 19, pp. 9360–9369, 2019, doi: 10.1073/pnas.1808312116.

[5] A. Mashaghi, R. J. Van Wijk, and S. J. Tans, "Circuit topology of proteins and nucleic acids," Structure, vol. 22, no. 9, pp. 1227–1237, 2014, doi: 10.1016/j.str.2014.06.015.

[6] P. F. N. Faísca, "Knotted proteins: A tangled tale of Structural Biology," Comput Struct Biotechnol J, vol. 13, pp. 459–468, 2015, doi: 10.1016/j.csbj.2015.08.003.

[7] O. Schullian, J. Woodard, A. Tirandaz, and A. Mashaghi, "A Circuit Topology Approach to Categorizing Changes in Biomolecular Structure," Front Phys, vol. 8, no. 5, Jan. 2020, doi: 10.3389/fphy.2020.00005.

[8] D. Moes, E. Banijamali, V. Sheikhhassani, B. Scalvini, J. Woodard, and A. Mashaghi, "ProteinCT: An implementation of the protein Circuit Topology framework," MethodsX, p. 101861, Sep. 2022, doi: 10.1016/j.mex.2022.101861.

[9] D. D. Nguyen, Z. Cang, and G.-W. Wei, "A review of mathematical representations of biomolecular data," Physical Chemistry Chemical Physics, vol. 22, no. 8, pp. 4343–4367, 2020, doi: 10.1039/C9CP06554G.

[10] P. G. Cámara, "Topological methods for genomics: Present and future directions," Curr Opin Syst Biol, vol. 1, pp. 95–101, Feb. 2017, doi: 10.1016/j.coisb.2016.12.007.

[11] K. Xia and G.-W. Wei, "Persistent homology analysis of protein structure, flexibility, and folding," Int J Numer Method Biomed Eng, vol. 30, no. 8, pp. 814–844, Aug. 2014, doi: 10.1002/cnm.2655.

[12] Z. Cang and G.-W. Wei, "TopologyNet: Topology based deep convolutional and multi-task neural networks for biomolecular property predictions," PLoS Comput Biol, vol. 13, no. 7, p. e1005690, Jul. 2017, doi: 10.1371/journal.pcbi.1005690.

[13] D. D. Nguyen, Z. Cang, K. Wu, M. Wang, Y. Cao, and G.-W. Wei, "Mathematical deep

learning for pose and binding affinity prediction and ranking in D3R Grand Challenges," J Comput Aided Mol Des, vol. 33, no. 1, pp. 71–82, Jan. 2019, doi: 10.1007/s10822-018-0146-6.

[14] S. K. Verovšek and A. Mashaghi, "Extended Topological Persistence and Contact Arrangements in Folded Linear Molecules," Front Appl Math Stat, vol. 2, May 2016, doi: 10.3389/fams.2016.00006.

[15] L. Brocchieri, "Protein length in eukaryotic and prokaryotic proteomes," Nucleic Acids Res, vol. 33, no. 10, pp. 3390–3400, 2005, doi: 10.1093/nar/gki615.

[16] L. Mollica, L. M. Bessa, X. Hanoulle, M. R. Jensen, M. Blackledge, and R. Schneider, "Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment," Front Mol Biosci, vol. 3, no. SEP, pp. 1–18, 2016, doi: 10.3389/fmolb.2016.00052.

[17] D.-H. Kim and K.-H. Han, "Transient Secondary Structures as General Target-Binding Motifs in Intrinsically Disordered Proteins," Int J Mol Sci, vol. 19, no. 11, p. 3614, Nov. 2018, doi: 10.3390/ijms19113614.

[18] J. Tubiana, S. Cocco, and R. Monasson, "Learning protein constitutive motifs from sequence data," Elife, vol. 8, 2019, doi: 10.7554/eLife.39397.

[19] R. Mishra and S. Bhushan, "Knot theory in understanding proteins," J Math Biol, vol. 65, no. 6–7, pp. 1187–1213, 2012, doi: 10.1007/s00285-011-0488-3.

[20] D. Goundaroulis, J. Dorier, F. Benedetti, and A. Stasiak, "Studies of global and local entanglements of individual protein chains using the concept of knotoids," Sci Rep, vol. 7, no. 1, p. 6309, 2017, doi: 10.1038/s41598-017-06649-3.

[21] C. C. Adams, The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots. American Mathematical Society, Providence, R.I., 2004.

[22] J. I. Sułkowska, E. J. Rawdon, K. C. Millett, J. N. Onuchic, and A. Stasiak, "Conservation of complex knotting and slipknotting patterns in proteins," Proc Natl Acad Sci U S A, vol. 109, no. 26, 2012, doi: 10.1073/pnas.1205918109.

[23] J. I. Sulkowska, E. J. Rawdon, K. C. Millett, J. N. Onuchic, and A. Stasiak, "Conservation of complex knotting and slipknotting patterns in proteins," Proceedings of the National Academy of Sciences, vol. 109, no. 26, pp. E1715–E1723, Jun. 2012, doi: 10.1073/pnas.1205918109.

[24] V. P. Patil, J. D. Sandt, M. Kolle, and J. Dunkel, "Topological mechanics of knots and tangles," Science (1979), vol. 367, no. 6473, pp. 71–75, 2020, doi: 10.1126/science.aaz0135.

[25] P. Dabrowski-Tumanski, A. Stasiak, and J. I. Sulkowska, "In Search of Functional Advantages of Knots in Proteins," PLoS One, vol. 11, no. 11, p. e0165986, Nov. 2016, doi: 10.1371/journal.pone.0165986.

[26] K. Alexander, A. J. Taylor, and M. R. Dennis, "Proteins analysed as virtual knots," Sci Rep, vol. 7, no. 1, p. 42300, Sep. 2017, doi: 10.1038/srep42300.

[27] T. Vladimir, "KNOTOIDS," Osaka Journal of Mathematics, vol. 49, pp. 195–223, 2012.

[28] D. Goundaroulis, J. Dorier, F. Benedetti, and A. Stasiak, "Studies of global and local entanglements of individual protein chains using the concept of knotoids," Sci Rep, vol. 7, no. 1, p. 6309, Dec. 2017, doi: 10.1038/s41598-017-06649-3.

[29] P. Dabrowski-Tumanski et al., "KnotProt 2.0: A database of proteins with knots and other entangled structures," Nucleic Acids Res, vol. 47, no. D1, pp. D367--D375, 2019, doi: 10.1093/nar/gky1140.

[30] J. Dorier, D. Goundaroulis, F. Benedetti, and A. Stasiak, "Knoto-ID: a tool to study the entanglement of open protein chains using the concept of knotoids," Bioinformatics, vol. 34, no. 19, pp. 3402–3404, 2018, doi: 10.1093/bioinformatics/bty365.

[31] M. Baiesi, E. Orlandini, A. Trovato, and F. Seno, "Linking in domain-swapped protein dimers," Sci Rep, vol. 6, pp. 1–11, 2016, doi: 10.1038/srep33872.

[32] M. Baiesi, E. Orlandini, F. Seno, and A. Trovato, "Sequence and structural patterns detected in entangled proteins reveal the importance of co-translational folding," Sci Rep, vol. 9, no. 1, pp. 1–12, 2019, doi: 10.1038/s41598-019-44928-3.

[33] M. Baiesi, E. Orlandini, F. Seno, and A. Trovato, "Exploring the correlation between the folding rates of proteins and the entanglement of their native states," ArXiv, 2017.

[34] E. Panagiotou and K. W. Plaxco, "A topological study of protein folding kinetics," ArXiv, pp. 1–13, 2018, doi: 10.1090/conm/746/15010.

[35] N. Halabi, O. Rivoire, S. Leibler, and R. Ranganathan, "Protein Sectors: Evolutionary Units of Three-Dimensional Structure," Cell, vol. 138, no. 4, pp. 774–786, 2009, doi: 10.1016/j.cell.2009.07.038.

[36] C. Adams, J. Devadoss, M. Elhamdadi, and A. Mashaghi, "Knot Theory for Proteins: Gauss Codes, Quandles and Bondles," Dec. 2019.

[37] L. Chains, "SS symmetry Coloring Invariant for Topological Circuits in Folded," pp. 1–12, 2021.

[38] A. Golovnev and A. Mashaghi, "Generalized Circuit Topology of Folded Linear Chains," iScience, vol. 23, no. 9, p. 101492, 2020, doi: 10.1016/j.isci.2020.101492.

[39] D. H. E. W. Huberts and I. J. van der Klei, "Moonlighting proteins: An intriguing mode of multitasking," Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, vol. 1803, no. 4, pp. 520–525, Apr. 2010, doi: 10.1016/j.bbamcr.2010.01.022.

[40] K. W. Plaxco, K. T. Simons, and D. Baker, "Contact order, transition state placement and the refolding rates of single domain proteins," J Mol Biol, vol. 277, no. 4, pp. 985–994, 1998, doi: 10.1006/jmbi.1998.1645.

[41] D. N. Ivankov, S. O. Garbuzynskiy, E. Alm, K. W. Plaxco, D. Baker, and A. V. Finkelstein, "Contact order revisited: Influence of protein size on the folding rate," Protein Science, vol. 12, no. 9, pp. 2057–2062, 2003, doi: 10.1110/ps.0302503.

[42] D. Baker, "A surprising simplicity to protein folding," Nature, vol. 405, no. 6782, pp. 39–42, May 2000, doi: 10.1038/35011000.

[43] O. V. Galzitskaya, S. O. Garbuzynskiy, D. N. Ivankov, and A. V. Finkelstein, "Chain length is the main determinant of the folding rate for proteins with three-state folding kinetics," Proteins: Structure, Function, and Genetics, vol. 51, no. 2, pp. 162–166, May 2003, doi: 10.1002/prot.10343.

[44] B. Scalvini, V. Sheikhhassani, and A. Mashaghi, "Topological principles of protein folding," Physical Chemistry Chemical Physics, vol. 23, no. 37, pp. 21316–21328, 2021, doi: 10.1039/

d1cp03390e.

[45] A. Mashaghi, G. Kramer, D. C. Lamb, M. P. Mayer, and S. J. Tans, "Chaperone action at the single-molecule level," Chemical Reviews. 2014. doi: 10.1021/cr400326k.

[46] V. Satarifard, M. Heidari, S. Mashaghi, S. J. Tans, M. R. Ejtehadi, and A. Mashaghi, "Topology of polymer chains under nanoscale confinement," Nanoscale, vol. 9, no. 33, pp. 12170–12177, 2017, doi: 10.1039/C7NR04220E.

[47] M. Heidari, V. Satarifard, S. J. Tans, M. R. Ejtehadi, S. Mashaghi, and A. Mashaghi, "Topology of internally constrained polymer chains," Physical Chemistry Chemical Physics, vol. 19, no. 28, pp. 18389–18393, 2017, doi: 10.1039/C7CP02145C.

[48] D. Scholl and A. A. Deniz, "Conformational Freedom and Topological Confinement of Proteins in Biomolecular Condensates," J Mol Biol, vol. 434, no. 1, p. 167348, 2022, doi: 10.1016/j.jmb.2021.167348.

[49] L. Jawerth et al., "Protein condensates as aging Maxwell fluids," Science (1979), vol. 370, no. 6522, pp. 1317–1323, 2020, doi: 10.1126/science.aaw4951.

[50] I. A. Klein et al., "Partitioning of cancer therapeutics in nuclear condensates," Science (1979), vol. 368, no. 6497, pp. 1386–1392, 2020, doi: 10.1126/science.aaz4427.

[51] A. Boija, I. A. Klein, and R. A. Young, "Biomolecular Condensates and Cancer," Cancer Cell, vol. 39, no. 2, pp. 174–192, 2021, doi: 10.1016/j.ccell.2020.12.003.

[52] B. R. Sabari, "Biomolecular condensates in the nucleus," 2020.

[53] M. Heidari, D. Moes, O. Schullian, B. Scalvini, and A. Mashaghi, "A topology framework for macromolecular complexes and condensates," Nano Res, 2022, doi: 10.1007/s12274-022-4355-x.

[54] A. Mashaghi and A. Ramezanpour, "Distance measures and evolution of polymer chains in their topological space," Soft Matter, vol. 11, no. 33, pp. 6576–6585, 2015, doi: 10.1039/C5SM01482D.

[55] V. Sheikhhassani et al., "Topological dynamics of an intrinsically disordered N-terminal domain of the human androgen receptor," Protein Science, vol. 31, no. 6, pp. 1–17, 2022, doi: 10.1002/pro.4334.

[56] V. N. Uversky, C. J. Oldfield, and A. K. Dunker, "Intrinsically Disordered Proteins in Human Diseases: Introducing the D 2 Concept," Annu Rev Biophys, vol. 37, no. 1, pp. 215–246, Jun. 2008, doi: 10.1146/annurev.biophys.37.032807.125924.

[57] S. J. Metallo, "Intrinsically disordered proteins are potential drug targets," Curr Opin Chem Biol, vol. 14, no. 4, pp. 481–488, 2010, doi: 10.1016/j.cbpa.2010.06.169.

[58] P. Santofimia-Castaño et al., "Targeting intrinsically disordered proteins involved in cancer," Cellular and Molecular Life Sciences, vol. 77, no. 9, pp. 1695–1707, May 2020, doi: 10.1007/s00018-019-03347-3.

[59] B. Mészáros, B. Hajdu-Soltész, A. Zeke, and Z. Dosztányi, "Mutations of Intrinsically Disordered Protein Regions Can Drive Cancer but Lack Therapeutic Strategies," Biomolecules, vol. 11, no. 3, p. 381, Mar. 2021, doi: 10.3390/biom11030381.

[60] A. Mashaghi et al., "Reshaping of the conformational search of a protein by the chaperone trigger factor," Nature, vol. 500, no. 7460, pp. 98–101, Aug. 2013, doi: 10.1038/nature12293.

[61] A. Mashaghi et al., "Alternative modes of client binding enable functional plasticity of Hsp70," Nature, vol. 539, no. 7629, pp. 448–451, Nov. 2016, doi: 10.1038/nature20137.

[62] T. Suren, D. Rutz, P. Mößmer, U. Merkel, J. Buchner, and M. Rief, "Single-molecule force spectroscopy reveals folding steps associated with hormone binding and activation of the glucocorticoid receptor," Proc Natl Acad Sci U S A, vol. 115, no. 46, pp. 11688–11693, 2018, doi: 10.1073/pnas.1807618115.

[63] A. Mashaghi, S. Mashaghi, and S. J. Tans, "Misfolding of Luciferase at the Single-Molecule Level," Angewandte Chemie International Edition, vol. 53, no. 39, pp. 10390–10393, Sep. 2014, doi: 10.1002/anie.201405566.

[64] D. E. Draper, "Strategies for RNA folding," Trends Biochem Sci, vol. 21, no. 4, pp. 145–149, Apr. 1996, doi: 10.1016/S0968-0004(96)80169-1.

[65] D. P. Giedroc and P. V. Cornish, "Frameshifting RNA pseudoknots: Structure and mechanism," Virus Res, vol. 139, no. 2, pp. 193–208, Feb. 2009, doi: 10.1016/j.virusres.2008.06.008.

[66] M. Bon, G. Vernizzi, H. Orland, and A. Zee, "Topological Classification of RNA Structures," J Mol Biol, vol. 379, no. 4, pp. 900–911, Jun. 2008, doi: 10.1016/j.jmb.2008.04.033.

[67] R. D. Ringeisen, "Survey of results on the maximum genus of a graph," J Graph Theory, vol. 3, no. 1, pp. 1–13, 1979, doi: 10.1002/jgt.3190030102.

[68] S. Zając, C. Geary, E. S. Andersen, P. Dabrowski-Tumanski, J. I. Sulkowska, and P. Sułkowski, "Genus trace reveals the topological complexity and domain structure of biomolecules," Sci Rep, vol. 8, no. 1, p. 17537, Dec. 2018, doi: 10.1038/s41598-018-35557-3.

[69] P. Rubach, S. Zajac, B. Jastrzebski, J. I. Sulkowska, and P. Sułkowski, "Genus for biomolecules," Nucleic Acids Res, vol. 48, no. D1, pp. D1129–D1135, Jan. 2020, doi: 10.1093/nar/gkz845.

[70] M. Quadrini, L. Tesei, and E. Merelli, "An algebraic language for RNA pseudoknots comparison," BMC Bioinformatics, vol. 20, no. S4, p. 161, Apr. 2019, doi: 10.1186/s12859-019-2689-5.

[71] T. Schlick, "Adventures with RNA graphs," Methods, vol. 143, pp. 16–33, Jul. 2018, doi: 10.1016/j.ymeth.2018.03.009.

[72] C. Cubeñas-Potts and V. G. Corces, "Architectural proteins, transcription, and the three-dimensional organization of the genome," FEBS Lett, vol. 589, no. 20PartA, pp. 2923–2930, Oct. 2015, doi: 10.1016/j.febslet.2015.05.025.

[73] J. Dekker and T. Misteli, "Long-Range Chromatin Interactions," Cold Spring Harb Perspect Biol, vol. 7, no. 10, p. a019356, Oct. 2015, doi: 10.1101/cshperspect.a019356.

[74] R. T. Dame and M. Tark-Dame, "Bacterial chromatin: converging views at different scales," Curr Opin Cell Biol, vol. 40, pp. 60–65, Jun. 2016, doi: 10.1016/j.ceb.2016.02.015.

[75] R. T. Dame, F.-Z. M. Rashid, and D. C. Grainger, "Chromosome organization in bacteria: mechanistic insights into genome structure and function," Nat Rev Genet, Nov. 2019, doi: 10.1038/s41576-019-0185-4.

[76] R. Stadhouders, G. J. Filion, and T. Graf, "Transcription factors and 3D genome conformation in cell-fate decisions," Nature, vol. 569, no. 7756, pp. 345–354, May 2019, doi: 10.1038/s41586-019-1182-7.

[77] W. A. Flavahan et al., "Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs," Nature, vol. 575, no. 7781, pp. 229–233, Nov. 2019, doi: 10.1038/s41586-019-1668-3.

[78] Z. Tang et al., "CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription," Cell, vol. 163, no. 7, pp. 1611–1627, Dec. 2015, doi: 10.1016/j.cell.2015.11.024.

[79] C. Naughton et al., "Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures," Nat Struct Mol Biol, vol. 20, no. 3, pp. 387–395, Mar. 2013, doi: 10.1038/nsmb.2509.

[80] J. H. White, N. R. Cozzarelli, W. R. Bauer, J. H. White, N. R. Cozzarelli, and W. R. Bauer, "Helical Repeat and Linking Number of Surface-Wrapped DNA," vol. 241, no. 4863, pp. 323–327, 2019.

[81] M. T. J. van Loenhout, M. V. de Grunt, and C. Dekker, "Dynamics of DNA Supercoils," Science (1979), vol. 338, no. 6103, pp. 94–97, Oct. 2012, doi: 10.1126/science.1225810.

[82] R. Vlijm, A. Mashaghi, S. Bernard, M. Modesti, and C. Dekker, "Experimental phase diagram of negatively supercoiled DNA measured by magnetic tweezers and fluorescence," Nanoscale, vol. 7, no. 7, pp. 3205–3216, 2015, doi: 10.1039/C4NR04332D.

[83] S. Shaw and J. Wang, "Knotting of a DNA chain during ring closure," Science (1979), vol. 260, no. 5107, pp. 533–536, Apr. 1993, doi: 10.1126/science.8475384.

[84] L. F. Liu, L. Perkocha, R. Calendar, and J. C. Wang, "Knotted DNA from bacteriophage capsids.," Proceedings of the National Academy of Sciences, vol. 78, no. 9, pp. 5498–5502, Sep. 1981, doi: 10.1073/pnas.78.9.5498.

[85] Y. Tao, N. H. Olson, W. Xu, D. L. Anderson, M. G. Rossmann, and T. S. Baker, "Assembly of a Tailed Bacterial Virus and Its Genome Release Studied in Three Dimensions," Cell, vol. 95, no. 3, pp. 431–437, Oct. 1998, doi: 10.1016/S0092-8674(00)81773-0.

[86] J. Arsuaga, M. Vazquez, P. McGuirk, S. Trigueros, D. W. Sumners, and J. Roca, "DNA knots reveal a chiral organization of DNA in phage capsids," Proceedings of the National Academy of Sciences, vol. 102, no. 26, pp. 9165–9169, Jun. 2005, doi: 10.1073/pnas.0409323102.

[87] A. Valdés, J. Segura, S. Dyson, B. Martínez-García, and J. Roca, "DNA knots occur in intracellular chromatin," Nucleic Acids Res, vol. 46, no. 2, pp. 650–660, Jan. 2018, doi: 10.1093/nar/gkx1137.

[88] T. J. Stevens et al., "3D structures of individual mammalian genomes studied by single-cell Hi-C," Nature, vol. 544, no. 7648, pp. 59–64, Apr. 2017, doi: 10.1038/nature21429.

[89] J. Siebert, A. Kivel, L. Atkinson, T. Stevens, E. Laue, and P. Virnau, "Are There Knots in Chromosomes?," Polymers (Basel), vol. 9, no. 12, p. 317, Aug. 2017, doi: 10.3390/polym9080317.

[90] R. Kumar Sharma, I. Agrawal, L. Dai, P. S. Doyle, and S. Garaj, "Complex DNA knots detected with a nanopore sensor," Nat Commun, vol. 10, no. 1, 2019, doi: 10.1038/s41467-019-12358-4.

[91] G. Witz, G. Dietler, and A. Stasiak, "Tightening of DNA knots by supercoiling facilitates their unknotting by type II DNA topoisomerases," Proceedings of the National Academy of Sciences, vol. 108, no. 9, pp. 3608–3611, Mar. 2011, doi: 10.1073/pnas.1016150108.

[92] N. L. van Berkum et al., "Hi-C: A Method to Study the Three-dimensional Architecture of Genomes.," Journal of Visualized Experiments, no. 39, May 2010, doi: 10.3791/1869.

[93] J. R. Dixon, D. U. Gorkin, and B. Ren, "Chromatin Domains: The Unit of Chromosome Organization," Mol Cell, vol. 62, no. 5, pp. 668–680, Jun. 2016, doi: 10.1016/j.molcel.2016.05.018.

[94] B. Scalvini, H. Schiessel, A. Golovnev, and A. Mashaghi, "Circuit topology analysis of cellular genome reveals signature motifs , conformational heterogeneity , and scaling," iScience, vol. 25, no. 3, p. 103866, 2022, doi: 10.1016/j.isci.2022.103866.

[95] O. L. Kantidze et al., "The anti-cancer drugs curaxins target spatial genome organization," Nat Commun, vol. 10, no. 1, pp. 1–11, 2019, doi: 10.1038/s41467-019-09500-7.

[96] E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman, "Design and self-assembly of two-dimensional DNA crystals," Nature, vol. 394, no. 6693, pp. 539–544, Aug. 1998, doi: 10.1038/28998.

[97] P. W. K. Rothemund, A. Ekani-Nkodo, N. Papadakis, A. Kumar, D. K. Fygenson, and E. Winfree, "Design and Characterization of Programmable DNA Nanotubes," J Am Chem Soc, vol. 126, no. 50, pp. 16344–16352, Dec. 2004, doi: 10.1021/ja044319l.

[98] S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, and W. M. Shih, "Self-assembly of DNA into nanoscale three-dimensional shapes," Nature, vol. 459, no. 7245, pp. 414–418, May 2009, doi: 10.1038/nature08016.

[99] N. C. Seeman, "Nanomaterials Based on DNA," Annu Rev Biochem, vol. 79, no. 1, pp. 65–87, Jun. 2010, doi: 10.1146/annurev-biochem-060308-102244.

[100] P. W. K. Rothemund, "Folding DNA to create nanoscale shapes and patterns," Nature, vol. 440, no. 7082, pp. 297–302, Mar. 2006, doi: 10.1038/nature04586.

[101] V. Kočar et al., "Design principles for rapid folding of knotted DNA nanostructures," Nat Commun, vol. 7, no. 1, p. 10803, Apr. 2016, doi: 10.1038/ncomms10803.

[102] D. Han et al., "Single-stranded DNA and RNA origami," Science (1979), vol. 358, no. 6369, p. eaao2648, Dec. 2017, doi: 10.1126/science.aao2648.

[103] Y.-R. Lin et al., "Control over overall shape and size in de novo designed proteins," Proceedings of the National Academy of Sciences, vol. 112, no. 40, pp. E5478–E5485, Oct. 2015, doi: 10.1073/pnas.1509508112.

[104] D. Baker, "Prediction and design of macromolecular structures and interactions," Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 361, no. 1467, pp. 459–463, Mar. 2006, doi: 10.1098/rstb.2005.1803.

[105] A. Ljubetič et al., "Design of coiled-coil protein-origami cages that self-assemble in vitro and in vivo," Nat Biotechnol, vol. 35, no. 11, pp. 1094–1101, Nov. 2017, doi: 10.1038/nbt.3994.

[106] M. Denis and S. M. Goldup, "The active template approach to interlocked molecules," Nat Rev Chem, vol. 1, no. 8, p. 0061, Aug. 2017, doi: 10.1038/s41570-017-0061.

[107] L. Voorhaar and R. Hoogenboom, "Supramolecular polymer networks: hydrogels and bulk materials," Chem Soc Rev, vol. 45, no. 14, pp. 4013–4031, 2016, doi: 10.1039/C6CS00130K.

[108] L. Zhang, J.-F. Lemonnier, A. Acocella, M. Calvaresi, F. Zerbetto, and D. A. Leigh, "Effects

of knot tightness at the molecular level," Proceedings of the National Academy of Sciences, vol. 116, no. 7, pp. 2452–2457, Feb. 2019, doi: 10.1073/pnas.1815570116.

[109] Y. Segawa et al., "Topological molecular nanocarbons: All-benzene catenane and trefoil knot," Science (1979), vol. 365, no. 6450, pp. 272–276, Jul. 2019, doi: 10.1126/science.aav5021.

[110] P. E. Barran et al., "Active-Metal Template Synthesis of a Molecular Trefoil Knot," Angewandte Chemie, vol. 123, no. 51, pp. 12488–12492, Dec. 2011, doi: 10.1002/ange.201105012.

[111] N. Ponnuswamy, F. B. L. Cougnon, G. D. Pantoş, and J. K. M. Sanders, "Homochiral and meso Figure Eight Knots and a Solomon Link," J Am Chem Soc, vol. 136, no. 23, pp. 8243–8251, Jun. 2014, doi: 10.1021/ja4125884.

[112] J.-F. Ayme, J. E. Beves, D. A. Leigh, R. T. McBurney, K. Rissanen, and D. Schultz, "A synthetic molecular pentafoil knot," Nat Chem, vol. 4, no. 1, pp. 15–20, Jan. 2012, doi: 10.1038/nchem.1193.

[113] J. J. Danon et al., "Braiding a molecular knot with eight crossings," Science (1979), vol. 355, no. 6321, pp. 159–162, Jan. 2017, doi: 10.1126/science.aal1619.

[114] N. Nikoofard and A. Mashaghi, "Topology sorting and characterization of folded polymers using nano-pores," Nanoscale, vol. 8, no. 8, pp. 4643–4649, 2016, doi: 10.1039/C5NR08828C.

[115] J. Skolnick, M. Gao, H. Zhou, and S. Singh, "AlphaFold 2: Why It Works and Its Implications for Understanding the Relationships of Protein Sequence, Structure, and Function," J Chem Inf Model, vol. 61, no. 10, pp. 4827–4831, 2021, doi: 10.1021/acs.jcim.1c01114.

[116] M. Ghafari and A. Mashaghi, "On the role of topology in regulating transcriptional cascades," Physical Chemistry Chemical Physics, vol. 19, no. 36, pp. 25168–25179, 2017, doi: 10.1039/C7CP02671D.