

Circuit Topology Analysis of Polymer Folding Reactions

Maziar Heidari, Helmut Schiessel, and Alireza Mashaghi*



Cite This: *ACS Cent. Sci.* 2020, 6, 839–847



Read Online

ACCESS |

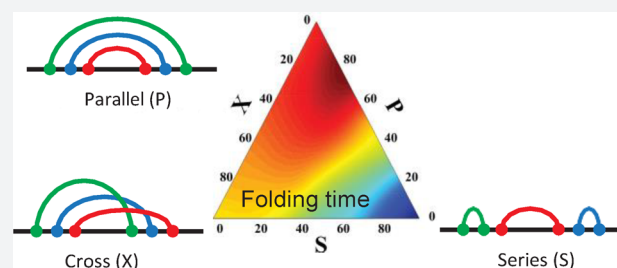


Metrics & More



Article Recommendations

ABSTRACT: Circuit topology is emerging as a versatile measure to classify the internal structures of folded linear polymers such as proteins and nucleic acids. The topology framework can be applied to a wide range of problems, most notably molecular folding reactions that are central to biology and molecular engineering. In this Outlook, we discuss the state-of-the-art of the technology and elaborate on the opportunities and challenges that lie ahead.



INTRODUCTION

Linear polymer chains are building blocks of life and an important class of synthetic macromolecules in chemistry. The chemical and physical properties of these molecules are determined partly by the chemical nature of their monomers and partly by their arrangements along the chain and in 3D. Natural proteins and synthetic protein origami are good examples of such linear polymers, in which topological arrangement of interacting non-neighboring monomers (contact sites) determine their structure, stability, surface properties, and folding kinetics and pathways. Chain segments can come into physical contact via electrostatic, van der Waals, hydrogen bond, or covalent interactions, but they can also cross each other and form knots. Circuit topology categorizes the arrangement of contacts, while knot theory categorizes the arrangement of chain crossings. Thus, a combination of both conceptual frameworks provides a complete theory for folded linear polymer chains.

Arrangement of contacts and knot crossings are functionally important properties of folded linear macromolecules. For proteins, topology has been used to describe folding pathways;^{1–4} supercoiling and knotting are linked to protein stability;^{4,5} intriguing observations revealed possible connections between topology, function, and evolution of some proteins.⁶ Similar links are being explored for RNA molecules. For example, the statistics of loops has been linked to the thermodynamics of RNA folding.⁷ RNA molecules involving base pairs between loops are likely to become topologically trapped in persistent frustrated states.⁸ The folding structure of the genomic DNA has also been a subject of intense studies and discussions.^{9,10} A topology analysis has been used to extract conformational reaction pathways that lead to the formation of the chromosome. A combination of modeling and experimental analysis of contacts suggest that the organization of the genome is determined by an interplay of loop extrusion

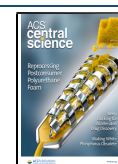
and compartmental segregation.^{11,12} Condensin I and II proteins together form nested DNA loops around a helical DNA scaffold leading to the formation of the mitotic chromosome.¹³ In addition to protein and genome studies, much progress has been made in the field of polymer science by studying synthetic and model polymers. For example, a recent study revealed the interplay between topology and mechanics in elastic knots. By combining an optomechanical analysis of knotted fibers with modeling, Patil et al. identified simple topological counting rules to predict the relative mechanical stability of knots and tangles.¹⁴ The synthesis of molecules with different topologies is a challenging task. Some good progress has been made,^{15,16} but there is still much to be learned on this emerging research frontier.

Circuit topology is emerging as a versatile measure to classify the internal structures of folded linear polymers such as proteins and nucleic acids.

Studies on the topology of linear molecules are naturally linked to studies on folding processes. As mentioned earlier, folding may happen through knotting or contact formation or both. Until now, most studies focused on the knot formation pathway and kinetics. Contacts are ignored in knot theory, and most proteins and RNA molecules do not form knots.¹⁷ Thus,

Received: March 17, 2020

Published: May 12, 2020



understanding the contact arrangement (circuit topology) of folded molecules and their implications for folding is critically important for biochemistry and bioinspired molecular engineering and beyond. Research on contact arrangements and their implication for folding has gained momentum only recently thanks to theoretical and technological advancements. Here, we focus on circuit topology and discuss its applications to folding studies.

We address the question of how folding and unfolding processes are mapped on the topological space.

The organization of the paper is as follows. We start with defining the statistics of loop formation and circuit topology. Then we will provide a topological description of single-molecule folding and unfolding dynamics. Next, we will discuss how one can assist folding toward preferred topologies by introducing various forms of confinement. Afterward, we will show how one can sort and enrich certain topologies. Finally, we will discuss possible extensions of current topology frameworks and particularly, how knot theory and circuit topology can merge into a complete topology theory for folded chains.

FORMATION AND ARRANGEMENT OF CONTACTS

A given linear polymer, biological or synthetic, can be considered as a fluctuating chain (path), whose equilibrium statistics are defined by the Boltzmann weight, which is the exponential of the chain energy in units of the thermal energy. As the result of the thermal fluctuations, different segments of the chains can meet and form contacts. The contact probability of two points separated by contour length s along an ideal chain in a dilute solution decays with $s^{-3/2}$.¹⁸ However, a polymer chain in a poor solvent folds into a globule, whose structure is either close to equilibrium or out-of-equilibrium. In the former case, the contact probability decays still as $s^{-3/2}$ while in the latter, which is also known as fractal or collapsed globule, it decays as s^{-1} .¹⁹ The stability of these contacts is determined by the free energy changes between bound (after contact) and unbound (free) states; the free energy change includes enthalpic and entropic contributions. The formation

of a contact is associated with the formation of a closed loop along the contour length; the statistics of such loops carry critical information with functional relevance (e.g., see ref 7). The loops can be characterized by their geometry (e.g., contact order defined as the ratio of the loop size to the total contour length of the chain) and by their topology, that is, the way in which the loops are mutually closed. Loop geometry has been discussed in several studies; for example, statistical physics frameworks have been introduced to study loop geometry in nucleic acids and proteins.^{7,20} However, the topological arrangements of the loops and their importance in determining folding kinetics have only recently received attention since the introduction of “circuit topology” by Mashaghi et al.¹⁷

The topological arrangements of the loops in the circuit topology framework are defined by considering a linear polymer chain with N contact sites and M loops. Here, the contact sites are labeled as $C = C_1, C_2, \dots, C_N$ and loops as $L = L_1, L_2, \dots, L_M$. If loop L_1 connects sites C_i and C_j with contour segment $[C_i, C_j]$ and loop L_2 connects sites C_k and C_l with contour segment $[C_k, C_l]$, the circuit topology relation between L_1 and L_2 is defined as follows:

$$\text{series: } L_1SL_2 \Leftrightarrow [C_i, C_j] \cap [C_k, C_l] = \{\emptyset\}$$

$$\text{parallel: } L_1PL_2 \Leftrightarrow [C_i, C_j] \subset [C_k, C_l]$$

$$\text{cross: } L_1XL_2 \Leftrightarrow [C_i, C_j] \cap [C_k, C_l] \notin \{\emptyset, [C_i, C_j], [C_k, C_l]\}$$

$$\text{inverse parallel: } L_1P^{-1}L_2 \Leftrightarrow [C_k, C_l] \subset [C_i, C_j]$$

(1)

where L_1SL_2 denotes that L_1 is in series arrangement with L_2 (and similarly for P, X, and P^{-1}). As can be seen in Figure 1, such arrangements between each loop pair are analogous to the arrangements of elements in an electrical circuit; hence the name “circuit topology”. This definition applies to binary contacts (valency of binding sites is 1). Now if we allow for valency of 2, then we need to take into account the possibility of overlap between the connecting sites. Of course, we can slightly change the definitions by categorizing the shared binding site to allow for concerted parallel (CP), concerted series (CS), concerted inverse parallel (CP⁻¹), and concerted inverse series (CS⁻¹) (See Figure 2b). Valencies higher than 2 are less common in biology, although one can extend the definition to higher valency too. Thus, a topological space can



Figure 1. Configuration of a folded chain with 6 contact pairs is depicted in the left panel. Each contact along the chain is marked with different colors. The corresponding circuit topology of the folded chain is shown in the middle panel. The loop pairs are wired in series (S), parallel (P), or cross (X) topology. The inverse parallel topology (P^{-1}) is assumed to be the same as parallel topology (P). Since the arrangement of the contacts is analogous to the arrangement of elements in an electrical circuit, the mathematical framework is dubbed “circuit topology”. The circuit topology entropy of a folded chain configuration with 9-contact pairs is shown in right panel as a function of the series (n_s) and parallel (n_p) topological fractions.²² Reproduced from ref 22 2015 with permission from the Royal Society of Chemistry.

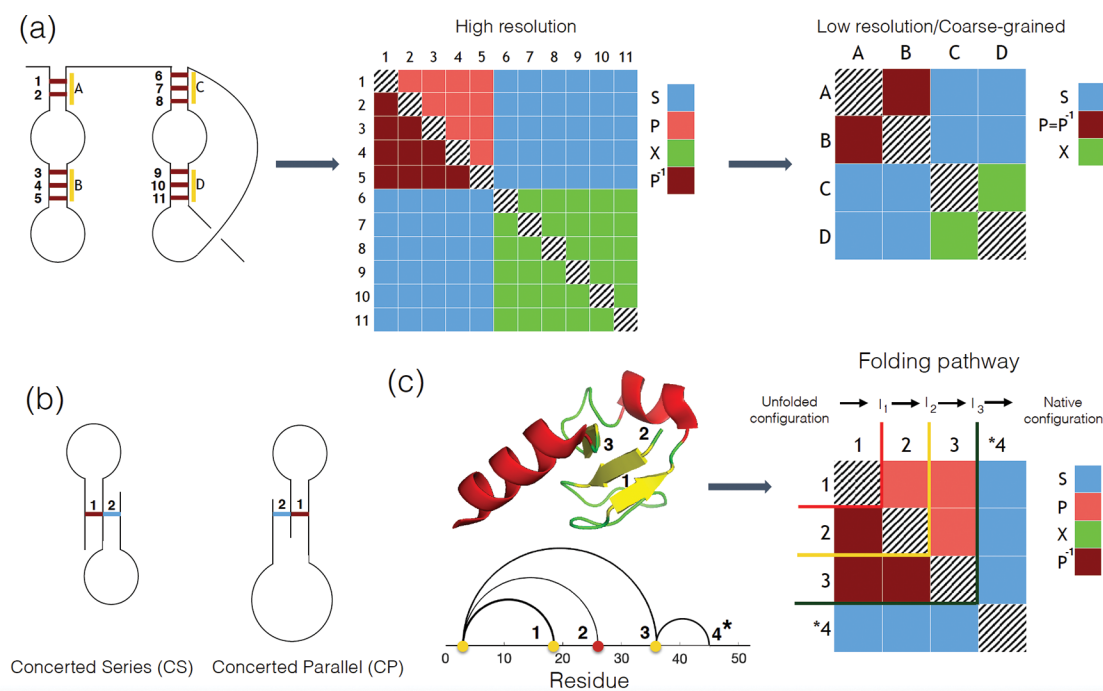


Figure 2. (a) Configuration of a folded chain with 11 contact pairs is illustrated. The contacts can exemplify local loops in the RNA molecules or β - β interactions in proteins. The circuit topology matrix built on the high resolution information, that is, the contact pairs, is shown in the top middle. The contact pairs are categorized in series (S), parallel (P), cross (X), and inverted parallel (P^{-1}). If the local contacts are grouped into single coarse-grained (CG) contact, the number of contacts reduce to four CG contacts, and the new low-resolution circuit topology is shown in the top right matrix. (b) Two molecular chains with contact pairs formed in the concerted series (CS) and concerted parallel (CP) arrangements are illustrated. In these cases, a single binding site on the chain forms two contacts and its valency rises up to two. By slight change of the definition, one can regard CP as P and CS as S, although a full consideration is also an option of course. (c) Snapshot and circuit topology of the folding trajectory of the N-terminal domain of ribosomal protein L9 (2HFV).²¹ First, the N-terminal hairpin forms. Next, the α helix forms a contact in concerted parallel with the hairpin, contacting the first strand. This second contact helps initiate the hydrogen bonding of the third β strand to the first, yielding three contacts in concerted parallel relation. Contact marked with * is not numbered in the molecular structure because the final α helix was excluded in folding simulations. The dashed and solid lines in the circuit topology matrix indicate the two folding steps.

be defined by three independent elements (if we allow commutativity of the parallel loops, i.e., $P^{-1} = P$) whose values are set by the topological fraction of each category, that is, the number of loop pairs in that category divided by the total number of loop pairs. Higher order correlation patterns than binary ones can be expected in the contact matrix of macromolecules having hierarchical structures such as the genome. The detection and comparison of these patterns can be facilitated and quantified using machine learning as the size of the circuit topology matrix and the higher order correlations of the patterns grow. Figure 2a shows an example of a folded chain having 11 contacts. The contacts can exemplify local loops in RNA molecules or β - β interactions in proteins. The circuit topology matrix can be generated using the information on either all contacts (high resolution map) or coarse-grained (CG) contacts (low resolution map). In the latter, the local contacts are grouped into a single coarse-grained (CG) contact, which leads to four CG contacts. In Figure 2c, a folding pathway of N-terminal domain of ribosomal protein L9 (2HFV)²¹ is illustrated, and its corresponding circuit topology changes are shown.

In Figure 2 we show different levels of coarse graining and simplification. For CP and CS, by slight changes in definition one can regard CP as P and CS as S, although a full consideration is also an option of course.

The averaged contact orders of each topological set can also be calculated. The contact order of two loops with topology i is calculated by $CO_i = (1/2N_iL) \sum_i (\Delta L_i^1 + \Delta L_i^2)$, where N_i is the

number of double loops that are categorized in the topological state i , ΔL_i^1 and ΔL_i^2 are the monomer separation of each loop, and L is the total polymer length.

Having defined the topological circuit framework, one can introduce measures to quantify the distance between two distinct topologies. Here, analogous to reaction coordinates along a pathway between two thermodynamic states, it is possible to define measures to represent progress along a topological reaction pathway from initial topology to final topology.²² The developed measures on the topological space can also be employed to categorize and map molecular structures such as proteins and nucleic acids and compare the corresponding topological circuits. This allows translation of familiar molecular operations in biology, such as duplication, permutation, and elimination of contacts, into the language of circuit topology, which is based on a coherent algebraic framework.²³ Moreover, the statistical mechanics of the loops can be utilized to describe the statistical mechanics of networks with different circuit topologies as will be described in the following.²⁴

For a chain with N contacts, one can build a graph with a corresponding link configuration set defined between the connected monomer pairs, $\mathcal{L} = \{(i, j) \mid l = 1, \dots, N\}$. As prescribed in circuit topology, for any pair of links, three states are assigned, series (S), parallel (P), and cross (X). The topology matrix of links for each configuration is defined by $A_{l,l'} \in \{S, P, X\}$ as shown in Figure 1 for a chain having 6

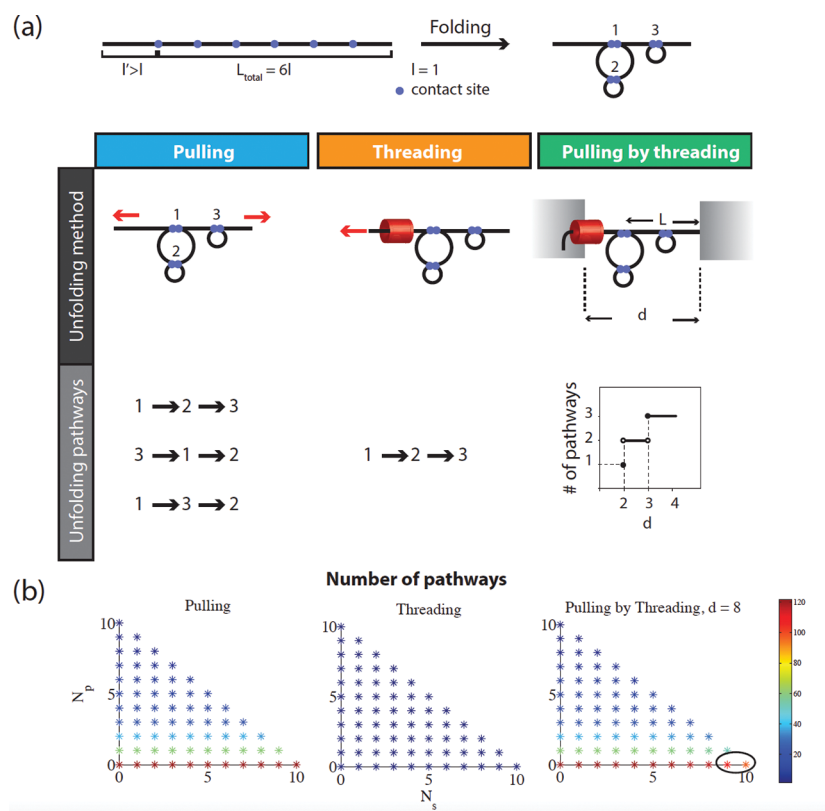


Figure 3. (a) Schematic representation of the three mechanical unfolding methods: pulling, threading, and pulling by threading. The number of pathways and efficiency of unfolding are listed for a 3-contact chain with a specific topology, L_1PL_2 , L_1SL_3 , L_2SL_3 (see eq 1). (b) Number of pathways for unfolding a 5-contact chain using pulling, pulling by threading, and threading methods. The total number of contact pairs in each topology is $N_s + N_p + N_x = 10$, where N_s , N_p , and N_x are the numbers of contact pairs in series, parallel and cross topologies, respectively. The plots in the figure show the number pathways as a function of N_s and N_p (N_x follows from these numbers). The color codes the number of pathways. Reproduced from ref 41. Copyright 2018 American Chemical Society.

contacts. The number of perfect matching configurations of the graph having $2N$ nodes reads $(2N - 1)!! = (2N - 1) \times (2N - 3) \dots \times 1$. Using Stirling's approximation for large N , the number of configurations grows as $e^{N \ln N}$. By imposing the perfect matching condition on the graph configurations, one can calculate the circuit topology entropy of the chain configurations normalized by the total number of possible configurations using $S = \frac{1}{N \ln N} \ln \mathcal{N}$, where \mathcal{N} is obtained by the enumeration of the configurations in each topology state. The exact entropy of a chain with nine contacts as a function of the series and parallel topology fractions is shown in the right panel of Figure 1. The cross topology fraction is not an independent variable and can be determined from the relation $(n_s + n_p + n_x) = 1$. The circuit topology entropy goes to zero for the all-(S, P, X) configurations in the corners of the entropy plot, while its maximum values appear at the position where the contacts are equally distributed within the topology states, that is, for $n_s = n_p = n_x = 1/3$.²⁴ We will discuss later how the topological state equipartition breaks when the chain is internally constrained.

As the circuit topology of a folded structure is defined by the arrangement of contacts and as it is invariant upon deformations of the loops formed between the contacts, several fundamental questions are raised: (i) How can a thermodynamic process, which is changing the configuration states of a chain, alter or pertain the circuit topology of a chain? (ii) Is a thermodynamic process, which might be the

consequence of an external interference such as chaperones involved in protein folding processes, able to smoothly deform a topological circuit of a chain into a different one? (iii) How does the dynamics and out-of-equilibrium nature of a process affect the final topological states? (iv) How is the circuit topology degeneracy associated with a topological reaction able to hamper evolution to desired topologies?

We address these questions by discussing results from simple and generic models prevailing in polymer physics and statistical mechanics in the following sections.

■ FOLDING/UNFOLDING PATHWAYS IN TOPOLOGICAL LANDSCAPES

Folding of molecular chains involves conformational searches within the free energy landscape. Since the conformational transitions are not always associated with topological transitions, finding the topological landscape and mapping the conformational transitions to their topological counterparts are fundamental questions in the field. Additionally, due to the cooperativity and nonlinearity of the interactions between the binding sites in the presence of solvent as well as the existence of many intermediate transition states, the folding and unfolding pathways are often irreversible.^{25,26} Thus, it is not obvious that the reverse unfolding process follows the same route as the folding process. In the following, we address the question of how folding and unfolding processes are mapped on the topological space by looking at different scenarios of mechanical folding and unfolding.

Folding. The physics of folding reactions can be mostly explained by conceptual frameworks including the nucleation–propagation mechanism and the diffusion–collision model.^{27–31} However, the topological arrangements and evolution of the contacts are not addressed in any of the theories. One way to monitor the real-time folding in a controlled manner is by restraining the ends of the linear (bio)molecules, for example, between two optically trapped beads (as is done for single-molecule mechanical interrogation studies). In silico modeling of end-restrained folding polymers revealed that the folding process starts with nucleation followed by growth. Before the onset of the nucleation, transient local entropic loops dominate leading to an increase in the number with series topology. After the nucleation, the circuit topology of the loops inside the nucleus leads to a drop in the topological fraction of the series topology while the other topologies, cross and parallel, grow. Such transient topological rearrangements converge to a steady-state, implying that the fold grows in a self-similar manner.³²

The circuit topology is a determinant of folding kinetics (and complements size and contact order). It is shown that the folding rate increases with the fractions of parallel and crossed relations.³³ The reason is backed by the zipping (nesting loops) effect in which the formation of contacts placed closely along the chain expedites formation of the contacts that are relatively far by bringing the contact sites closer together. The loops in parallel and cross topologies feature the zipping effect, while folding of the loops in series topology does not involve any nesting of contacts.

Unfolding. Many biological processes depend on the unfolding of biopolymers. For example, during translocation through a nanopore channel, degradation, and even folding of proteins and nucleic acids, it is required to partially or

completely unfold a biopolymer. The dependency of the unfolding pathways on the native state topology is under investigation. The unfolding of the biopolymers can be triggered either by a change in the thermodynamic conditions of the medium, such as a change in temperature²⁶ or ion concentration,³⁴ or by an externally applied mechanical force.^{35–40} Three different unfolding strategies can be anticipated for the mechanical unfolding: threading through a pore, pulling from the ends, and pulling by threading (see Figure 3). For a three-contact chain as shown in Figure 3, provided the contacts are likely to break under the same force, each unfolding route has a different number of unfolding pathways. In the pulling method, when the contacts are in series or cross relations relative to each other, any of the two contacts can be opened independently. Thus, there exist two unfolding pathways for series and cross topologies. However, for a 2-contact chain with parallel contact arrangement, there exists only one pathway. The contact nested inside the other contact cannot be opened unless the outer contact is opened first. Thus, for the pulling unfolding of the chain example in Figure 3, the contact number 2 cannot be opened prior to the other contacts. In the threading method, only the contact in the nanopore can be opened at each time. Therefore, there exists only one pathway for the three topologies in the threading method. In the pulling by threading method, which is a combination of the previous two methods, the ratio of the

Confinement may drastically change the configuration entropy of linear molecular chains and consequently, directs their folding towards a certain topology.

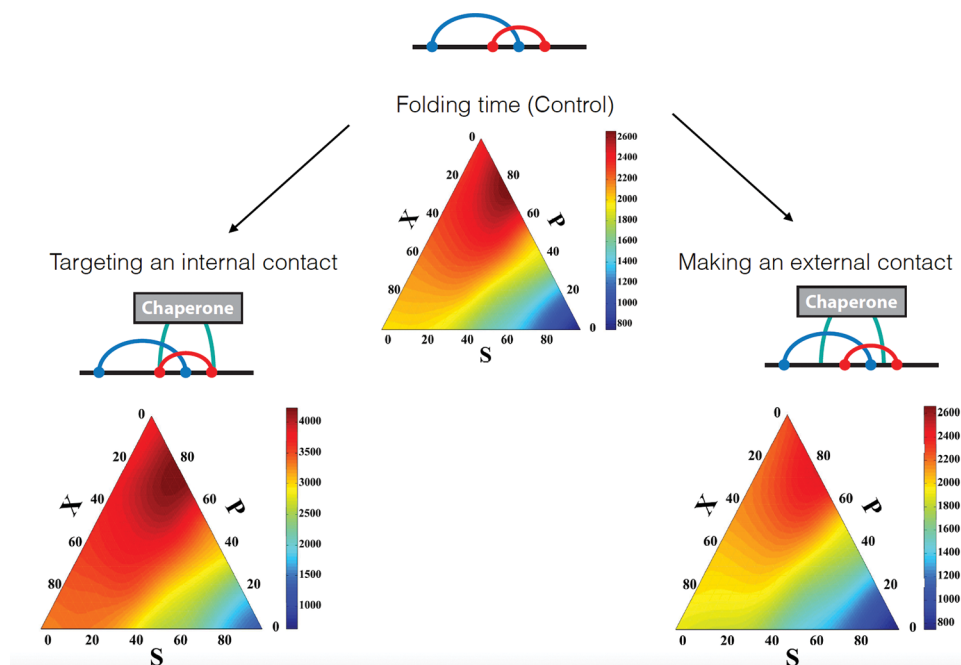


Figure 4. Folding time maps of a chain under different types of internal restraints. The folding time of a chain with no external perturbation (control case) is shown on the top. The folding time map in the presence of a chaperone is shown in the lower panels, where the chaperone binds to a native contact of the chain (lower left) and where it binds to the chain and forms an external contact (lower right). In both cases, the diffusion of the contact pairs of the chaperone is three times smaller than the diffusion of the chain's native binding sites. The axes of the ternary plots are topological fractions in series (S), parallel (P), and cross (X).⁴² Reproduced from ref 42 2017 with permission from the PCCP Owner Societies.

length of the released chain behind the nanopore and the distance between the nanopore and the tethered chain end, that is, L/d , determines whether pulling by threading primarily acts like pulling or threading. For $L < d$, the pulling component would be dominant, while for $L > d$, the unfolding method would be similar to simple threading.⁴¹ In Figure 3b, the number of pathways is shown for a 5-contact chain ($N = 5$) for the different unfolding methods. The total number of possible ways to pick a contact pair is $N(N - 1)/2 = 10$. This number is equal to the total number of contact pairs in each topology, $N_s + N_p + N_x = 10$, where N_s , N_p , and N_x are the number of contact pairs in series, parallel, and cross topologies, respectively. The plots in the figure show the number of unfolding pathways as a function of N_s and N_p (N_x follows from the other two numbers). As discussed earlier, the number of pathways in either pulling or pulling by threading methods, decreases upon increasing the number of parallel contact pairs in the system. On the other hand, the number of pathways is constant for the threading unfolding process.

CIRCUIT TOPOLOGY UNDER CONFINEMENT

Confinement may drastically change the configuration entropy of linear molecular chains and consequently affects their topology or directs their folding toward a certain topology. Here, two different scenarios can be considered: confinement can be introduced internally; alternatively the chain can be confined by enclosing confinements. Understanding the so-called confinement assisted folding is important both for biologists and chemists who wish to facilitate the synthesis of a desired molecular fold.

Internal Constraint. For linear chains with $2N$ intrachain binding sites, the number of possible ways to fold is given as $\sim(2N - 1)!!$. For example, for a chain having 10 binding sites, there are 1500 different folding pathways. Every binding pair can adopt different topologies with distinct transition rates between them. The total number of contact pairs that each can occupy either series, parallel, or cross topology grows as $N(N - 1)/2$. Using Monte Carlo simulations, it was shown that the topological dynamics of a simple linear chain is strongly affected by the presence of a “non-native” contact that is transiently introduced into the chain during the folding process.⁴² The role of an external molecule (chaperone) is schematically shown in Figure 4, where it can be seen that an external two point contact is sufficient to accelerate or slow down the formation of certain topologies. The presence of such contacts deforms the folding-time landscapes of the chains and hence alters the occupation probability of topological states. Examples of such mechanisms can be found in chaperone-assisted folding processes in cells. For example, trigger factor has finger like appendages that are able to touch a few sites on the unfolded chain and restrain those segments.^{43,44} Such internal confinements bias the conformational search of nonfolded molecular chains toward certain fold topologies.

External Confinement. An important class of confinement is external confinement imposed by an enclosing cavity. Here, the whole part or a large part of a chain is confined within a certain geometry such as a sphere of radius R_c .^{45–48} These external perturbations can alter the localization of the chain binding sites and accordingly enhance the binding probabilities and contact formations. It has been observed that the persistence length, which is determined by the stiffness of the chain, can influence the contact probabilities as well. Given

the confinement length scale R_c , and the gyration radius R_g of the chain, which depends on the persistence length as well as on the chain length, one can distinguish different regimes based on the ratio of the two scales. At large confining radius $R_c > R_g$, due to entropy, the linear chain remains in a coiled configuration in which the formation of independent contacts along the chain with small contact order are probable (see Figure 5). This accordingly increases the fraction of series

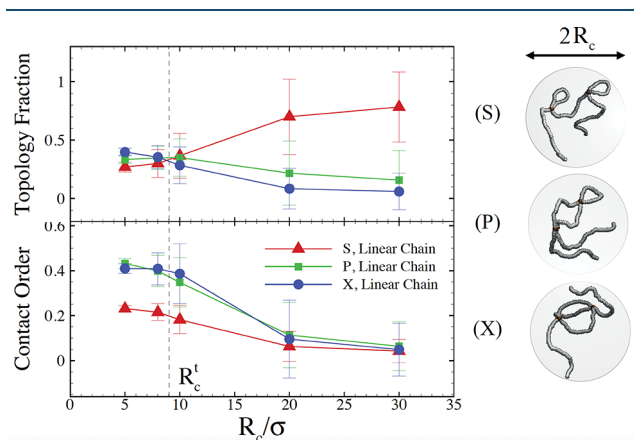


Figure 5. Fraction of topological circuits and corresponding averaged contact order (CO) as a function of the confining radius, R_c for a linear polymer chain (see eq 1). The dashed line indicates the transition confinement radius, R_c^t at which all topological fractions are equal.⁴⁵ Topological circuits of two intramolecular loops under spherical confinement with radius R_c are illustrated on the right side. Reproduced from ref 45 2017 with permission from the Royal Society of Chemistry.

topology with respect to other topologies. However, at small confining radius, $R_c < R_g$, the chain folds on itself. In this regime, the fractions of cross and parallel topologies are enhanced in the chain topological circuits with cross becoming predominant. At an intermediate confining regime, a critical radius can be expected at which all topological states have equal probability. This means that the confining radius can be considered as the topological reaction coordinate through which one can tune the occupancy of topological space. Furthermore, over a wide range of confining radii, loops arranged in parallel and cross topologies have nearly identical contact orders. The existence of such a degeneracy implies that the kinetics and transition rates between the topological states cannot be solely explained by contact order. In addition to spherical confinement, the topological circuit of the chain can be investigated under ellipsoidal confinement in which more topological reaction coordinates are introduced into the system. In this case, the change in the aspect ratio of the confinement can alter the contact probability and accordingly the circuit topology.

Nuclear Confinement. An extreme case of external confinement of a polymer is the organization of the DNA molecules inside cells⁴⁹ where macroscopic lengths of DNA, for example, 2 m in human cells, need to fit into micrometer-sized nuclei. As the cell needs access to genetic information, there are serious demands on the topology of DNA inside the nucleus. In fact, already in 1993,⁵⁰ it was speculated that the DNA cannot show an equilibrium conformation (a polymer inside a spherical compartment, similar to a polymer globule in a poor solvent) as this would lead to conformations too

entangled to be accessible. Instead a structure akin to a collapsed globule was proposed, the state into which a polymer folds when one quickly changes the solvent quality from good to poor (e.g., through a temperature jump). In 2009, experimental evidence for this claim was found through a then new method, chromosome conformation capture.⁹ This study suggested that the contact probability along the DNA (for human cells in interphase) decays as s^{-1} with genomic distance s , as opposed to $s^{-3/2}$ for the equilibrium globule.⁴⁹ Unlike for an equilibrium globule, the overall conformation is fractal. An interesting question is whether the hierarchical organization of such a nonequilibrium fractal globule would manifest itself in the context of circuit topology in different fractions of the three topological states. In this context, one might also study space filling curves like Peano and Hilbert curves that have been invoked as toy models for fractal globules.⁹

Chromosome capture experiments at higher resolution led in 2014⁵¹ to the discovery of contact domains (also called topologically associated domains), contiguous stretches of DNA of median length 185 kilobases, which have a substantially higher contact probability with themselves than with the rest of the genome. Boundaries of topologically associated domains are typically demarcated by a short base pair sequence to which insulator protein CCCTC-binding factor (CTCF) is bound. The two boundaries of a domain are in direct physical contact, and such a pair of nonpalindromic CTCF binding sites is always in convergent orientation. At first it appeared mysterious how pairs of DNA sites about 200 kilobases apart can find each other and then only bind if their sequences happen to be in convergent orientation. The solution to this puzzle lies in another protein that is bound to these locations: cohesin. It has been suggested that these molecules act as loop extruders, causing the formation of the contact domains.¹² Extrusion complexes contain two DNA binding subunits tethered together. Initially these two subunits bind nearby on the DNA. Then they move along the DNA in opposite directions while bridging these increasingly distant chromosomal sites, thereby increasing the size of the loop. The spooling of DNA into the loop continues until the subunits encounter CTCF proteins bound to flanking convergently arranged CTCF binding sites, which block further extrusion.

As a result, topological domains are dynamical systems of loops that are nonconcatenated with each other. As nonconcatenated polymer loops are known not to mix⁵² (unlike open polymers in solution¹⁸), such topological domains are separated spatially from each other. Circuit topology applied to such cohesin-induced loops would show that there are only series and parallel topological states but no cross topologies. If cross topologies would be present, especially beyond the boundaries of topological domains, the isolating effect of the ring topology would be destroyed and the effect of the cohesins would just be comparable to the effect of a poor solvent.

Finally, before cell division, there is a spectacular polymer physics problem to overcome. Each DNA double helix (chromosome) is copied, and one has two identical DNA copies entangled with each other. In order to propagate the genome to the two daughter cells, these molecules need to be neatly separated. The result, the X-shaped mitotic chromosome, is well-known but how does the cell arrive at that state? Polymer physics teaches us that the gain in the separation of two overlapping chains is only on the order of the thermal

energy.⁵³ What is needed are motor proteins that pull the chains apart, but the challenge lies in how these motors can distinguish the two identical chains. We know now that this is achieved by another loop extruder, condensin (see ref 49 for a historical overview on how this insight has been reached). When condensin molecules start to act on the DNA molecules, they create loops, shortening each chromosome lengthwise. As the loops are nonconcatenated, they repel each other. On one hand, this creates the desired repulsion between the chromosome pairs, which are only kept together at the centrosome. On the other hand, the loops along each chromosome stiffen the complex. The result of this process is the mitotic X-shaped chromosome, as beautifully demonstrated in a computer simulation⁵⁴ (see ref 13 for a more detailed study into mitotic chromosome formation). Also here one finds in terms of circuit topology only series and parallel topological states but no cross topology. Interestingly, despite the absence of cross topology, the two DNA copies still suffer from entanglements as they are driven apart from each. A specialized protein, topoisomerase II, resolves these entanglements by letting the DNA double helices pass through each other.

In summary, the three-dimensional organization of extremely long DNA molecules in tiny nuclear compartments leads to various challenges. Nature overcomes some of these challenges by making use of topology. Typically these topological states cannot be seen directly but leave their footprints in circuit topology by showing unusual distributions of the different topological states.

■ TOPOLOGY SORTING AND ENRICHMENT

In the previous sections, we discussed how a chain folds into its “native” circuit topology and how this process can be guided (e.g., by confinement). Synthesis is a key process both in biology and in polymer chemistry. In chemistry, synthesis is typically followed by purification and enrichment and characterization. This can be achieved, for example, by passing the synthesized chains through nanopores.⁵⁵ This technique has been used to identify the contacts and loops in nucleic acids. Under constant pulling force, one can identify the passage of loops with different topologies by examining the passage time. If the intramolecular contacts remain intact during the passage, the topology of the chain dramatically affects the dynamics of passage through the nanopore. This phenomenon can be exploited to separate or enrich a certain topology in a mixture. If the chain contacts are lost during passage, the discrimination between pure states, in which the majority of contacts are arranged identically, becomes possible.⁵⁵

■ CONCLUSIONS AND FUTURE PERSPECTIVES

The topological description of a molecular chain ignores geometric and chemical details but keeps the contact

The topological description of a molecular chain ignores geometric and chemical details, but keeps the contact arrangements and knotted structures.

arrangements and knotted structures. This allows us to disentangle the contribution of topology to folding processes from chemistry and geometry. In this Outlook, we discussed the recent studies on the circuit topology of polymer folding and unfolding reactions occurring in different circumstances, such as confinements, internal constraints, and nanopore translocation. The circuit topology framework can be readily extended to include additional complexities seen in nature or designed in an engineered setting. Contacts with higher valencies, transient contacts, and crossing can be included. Geometric and mechanical information can also be combined with the topological description to address a given physicochemical problem. Among these possible extensions, merging knot theory and circuit topology is a key step. Some initial progress has been made in this direction. Adams et al. have recently added intrachain contacts to knot theory.⁵⁶ Future developments are needed to provide a generalized circuit topology that can be generically applied to describe topology of any given molecular fold and can provide measures that are readily observable in experiments. We envision that the topology approaches discussed here open up new research directions in polymer chemistry, genome biology, and protein biophysics.

AUTHOR INFORMATION

Corresponding Author

Alireza Mashaghi – Leiden Academic Centre for Drug Research, Faculty of Science, Leiden University, Leiden 2300 RA, The Netherlands; orcid.org/0000-0002-2157-1211; Email: a.mashaghi.tabari@lacdr.leidenuniv.nl

Authors

Maziar Heidari – Leiden Academic Centre for Drug Research, Faculty of Science, Leiden University, Leiden 2300 RA, The Netherlands; Laboratoire Gulliver, UMR 7083, ESPCI Paris and PSL University, 75005 Paris, France; orcid.org/0000-0002-8081-6602

Helmut Schiessel – Institute Lorentz for Theoretical Physics, Faculty of Science, Leiden University, Leiden 2333 CA, The Netherlands

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acscentsci.0c00308>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Dr. Jaie Woodard for help with Figure 2c. A.M. acknowledges funding support from Muscular Dystrophy Association (USA), Grant Number MDA628071.

REFERENCES

- (1) Flapan, E.; He, A.; Wong, H. Topological descriptions of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 9360–9369.
- (2) Bustamante, A.; Sotelo-Campos, J.; Guerra, D. G.; Floor, M.; Wilson, C. A.; Bustamante, C.; Báez, M. The energy cost of polypeptide knot formation and its folding consequences. *Nat. Commun.* **2017**, *8*, 1581.
- (3) Kluber, A.; Burt, T. A.; Clementi, C. Size and topology modulate the effects of frustration in protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 9234–9239.
- (4) Sulkowska, J. I.; Noel, J. K.; Onuchic, J. N. Energy landscape of knotted protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17783–17788.

- (5) Niewiczzerzal, S.; Sulkowska, J. I. Knotting and unknotting proteins in the chaperonin cage: Effects of the excluded volume. *PLoS One* **2017**, *12*, e0176744.

- (6) Virnau, P.; Mirny, L. A.; Kardar, M. Intricate knots in proteins: Function and evolution. *PLoS Comput. Biol.* **2006**, *2*, e122.

- (7) Einert, T. R.; Näger, P.; Orland, H.; Netz, R. R. Impact of loop statistics on the thermodynamics of RNA folding. *Phys. Rev. Lett.* **2008**, *101*, 048103.

- (8) Bois, J. S.; Venkataraman, S.; Choi, H. M.; Spakowitz, A. J.; Wang, Z.-G.; Pierce, N. A. Topological constraints in nucleic acid hybridization kinetics. *Nucleic Acids Res.* **2005**, *33*, 4090–4095.

- (9) Lieberman-Aiden, E.; Van Berkum, N. L.; Williams, L.; Imakaev, M.; Ragozcy, T.; Telling, A.; Amit, I.; Lajoie, B. R.; Sabo, P. J.; Dorschner, M. O.; et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **2009**, *326*, 289–293.

- (10) Simonis, M.; Klous, P.; Splinter, E.; Moshkin, Y.; Willemsen, R.; De Wit, E.; Van Steensel, B.; De Laat, W. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). *Nat. Genet.* **2006**, *38*, 1348–1354.

- (11) Nuebler, J.; Fudenberg, G.; Imakaev, M.; Abdennur, N.; Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E6697–E6706.

- (12) Fudenberg, G.; Imakaev, M.; Lu, C.; Goloborodko, A.; Abdennur, N.; Mirny, L. A. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **2016**, *15*, 2038–2049.

- (13) Gibcus, J. H.; Samejima, K.; Goloborodko, A.; Samejima, I.; Naumova, N.; Nuebler, J.; Kanemaki, M. T.; Xie, L.; Paulson, J. R.; Earnshaw, W. C.; et al. A pathway for mitotic chromosome formation. *Science* **2018**, *359*, No. eaa06135.

- (14) Patil, V. P.; Sandt, J. D.; Kolle, M.; Dunkel, J. Topological mechanics of knots and tangles. *Science* **2020**, *367*, 71–75.

- (15) Polles, G.; Marenduzzo, D.; Orlandini, E.; Micheletti, C. Self-assembling knots of controlled topology by designing the geometry of patchy templates. *Nat. Commun.* **2015**, *6*, 6423.

- (16) Marena, M.; Orlandini, E.; Micheletti, C. Discovering privileged topologies of molecular knots with self-assembling models. *Nat. Commun.* **2018**, *9*, 3051.

- (17) Mashaghi, A.; van Wijk, R. J.; Tans, S. J. Circuit topology of proteins and nucleic acids. *Structure* **2014**, *22*, 1227–1237.

- (18) De Gennes, P.-G.; Gennes, P.-G. *Scaling concepts in polymer physics*; Cornell university press, 1979.

- (19) Mirny, L. A. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.* **2011**, *19*, 37–51.

- (20) Duplantier, B. Polymer network of fixed topology: renormalization, exact critical exponent γ in two dimensions, and $d = 4 - \epsilon$. *Phys. Rev. Lett.* **1986**, *57*, 941.

- (21) Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S. Molecular simulation of ab initio protein folding for a millisecond folder NTL9 (1–39). *J. Am. Chem. Soc.* **2010**, *132*, 1526–1528.

- (22) Mashaghi, A.; Ramezanpour, A. Distance measures and evolution of polymer chains in their topological space. *Soft Matter* **2015**, *11*, 6576–6585.

- (23) Schullian, O.; Woodard, J.; Tirandaz, A.; Mashaghi, A. A Circuit Topology Approach to Categorizing Changes in Biomolecular Structure. *Front. Phys.* **2020**, *8*, 5.

- (24) Mashaghi, A.; Ramezanpour, A. Circuit topology of linear polymers: a statistical mechanical treatment. *RSC Adv.* **2015**, *5*, 51682–51689.

- (25) Strucksberg, K.; Rosenkranz, T.; Fitter, J. Reversible and irreversible unfolding of multi-domain proteins. *Biochim. Biophys. Acta, Proteins Proteomics* **2007**, *1774*, 1591–1603.

- (26) Mallamace, F.; Corsaro, C.; Mallamace, D.; Vasi, S.; Vasi, C.; Baglioni, P.; Buldyrev, S. V.; Chen, S.-H.; Stanley, H. E. Energy landscape in protein folding and unfolding. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 3159–3163.

- (27) Lifson, S.; Roig, A. On the theory of helix—coil transition in polypeptides. *J. Chem. Phys.* **1961**, *34*, 1963–1974.
- (28) Zimm, B. H.; Bragg, J. Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* **1959**, *31*, 526–535.
- (29) Zwanzig, R. Simple model of protein folding kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 9801–9804.
- (30) Karplus, M.; Weaver, D. L. Protein-folding dynamics. *Nature* **1976**, *260*, 404.
- (31) Karplus, M.; Weaver, D. L. Protein folding dynamics: The diffusion-collision model and experimental data. *Protein Sci.* **1994**, *3*, 650–668.
- (32) Heidari, M.; Satarifard, V.; Mashaghi, A. Mapping a single-molecule folding process onto a topological space. *Phys. Chem. Chem. Phys.* **2019**, *21*, 20338–20345.
- (33) Mugler, A.; Tans, S. J.; Mashaghi, A. Circuit topology of self-interacting chains: implications for folding and unfolding dynamics. *Phys. Chem. Chem. Phys.* **2014**, *16*, 22537–22544.
- (34) Otzen, D. E. Protein unfolding in detergents: effect of micelle structure, ionic strength, pH, and temperature. *Biophys. J.* **2002**, *83*, 2219–2230.
- (35) Jakobi, A. J.; Mashaghi, A.; Tans, S. J.; Huizinga, E. G. Calcium modulates force sensing by the von Willebrand factor A2 domain. *Nat. Commun.* **2011**, *2*, 385.
- (36) Hansen, T. M.; Reihani, S. N. S.; Oddershede, L. B.; Sørensen, M. A. Correlation between mechanical strength of messenger RNA pseudoknots and ribosomal frameshifting. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 5830–5835.
- (37) Saibil, H. Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 630–642.
- (38) Lin, Z.; Madan, D.; Rye, H. S. GroEL stimulates protein folding through forced unfolding. *Nat. Struct. Mol. Biol.* **2008**, *15*, 303.
- (39) Schneider, S.; Nuschele, S.; Wixforth, A.; Gorzelanny, C.; Alexander-Katz, A.; Netz, R.; Schneider, M. F. Shear-induced unfolding triggers adhesion of von Willebrand factor fibers. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 7899–7903.
- (40) Alexander-Katz, A.; Schneider, M.; Schneider, S.; Wixforth, A.; Netz, R. Shear-flow-induced unfolding of polymeric globules. *Phys. Rev. Lett.* **2006**, *97*, 138101.
- (41) Nikoofard, N.; Mashaghi, A. Implications of Molecular Topology for Nanoscale Mechanical Unfolding. *J. Phys. Chem. B* **2018**, *122*, 9703–9712.
- (42) Heidari, M.; Satarifard, V.; Tans, S. J.; Ejtehadi, M. R.; Mashaghi, S.; Mashaghi, A. Topology of internally constrained polymer chains. *Phys. Chem. Chem. Phys.* **2017**, *19*, 18389–18393.
- (43) Mashaghi, A.; Kramer, G.; Bechtluft, P.; Zachmann-Brand, B.; Driessen, A. J.; Bukau, B.; Tans, S. J. Reshaping of the conformational search of a protein by the chaperone trigger factor. *Nature* **2013**, *500*, 98–101.
- (44) Singhal, K.; Vreede, J.; Mashaghi, A.; Tans, S. J.; Bolhuis, P. G. The trigger factor chaperone encapsulates and stabilizes partial folds of substrate proteins. *PLoS Comput. Biol.* **2015**, *11*, e1004444.
- (45) Satarifard, V.; Heidari, M.; Mashaghi, S.; Tans, S. J.; Ejtehadi, M. R.; Mashaghi, A. Topology of polymer chains under nanoscale confinement. *Nanoscale* **2017**, *9*, 12170–12177.
- (46) Sakaue, T.; Raphaël, E. Polymer chains in confined spaces and flow-injection problems: some remarks. *Macromolecules* **2006**, *39*, 2621–2628.
- (47) Fathizadeh, A.; Heidari, M.; Eslami-Mossallam, B.; Ejtehadi, M. R. Confinement dynamics of a semiflexible chain inside nano-spheres. *J. Chem. Phys.* **2013**, *139*, 044912.
- (48) Emanuel, M.; Radja, N. H.; Henriksson, A.; Schiessel, H. The physics behind the larger scale organization of DNA in eukaryotes. *Phys. Biol.* **2009**, *6*, 025008.
- (49) Sazer, S.; Schiessel, H. The biology and polymer physics underlying large-scale chromosome organization. *Traffic* **2018**, *19*, 87–104.
- (50) Grosberg, A.; Rabin, Y.; Havlin, S.; Neer, A. Crumpled globule model of the three-dimensional structure of DNA. *Europhys. Lett.* **1993**, *23*, 373–378.
- (51) Rao, S. S. P.; Huntley, M. H.; Durand, N. C.; Stamenova, E. K.; Bochkov, I. D.; Robinson, J. T.; Sanborn, A. L.; Machol, I.; Omer, A. D.; Lander, E. S.; Lieberman-Aiden, E. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **2014**, *159*, 1665–1680.
- (52) Halverson, J. D.; Lee, W. B.; Grest, G. S.; Grosberg, A. Y.; Kremer, K. Molecular dynamics simulation study of nonconcatenated ring polymers in a melt. I. Statics. *J. Chem. Phys.* **2011**, *134*, 204904.
- (53) Grosberg, A. Y.; Khalatur, P. G.; Khokhlov, A. R. Polymer coils with excluded volume in dilute solution: the invalidity of the model of impenetrable spheres and the influence of excluded volume on the rates of diffusion-controlled intermacromolecular interactions. *Makromol. Chem., Rapid Commun.* **1982**, *3*, 709–713.
- (54) Goloborodko, A.; Imakaev, M. V.; Marko, J. F.; Mirny, L. Compaction and segregation of sister chromatids via active loop extrusion. *eLife* **2016**, *5*, No. e14864.
- (55) Nikoofard, N.; Mashaghi, A. Topology sorting and characterization of folded polymers using nano-pores. *Nanoscale* **2016**, *8*, 4643–4649.
- (56) Adams, C.; Devadoss, J.; Elhamdadi, M.; Mashaghi, A. Knot Theory for Proteins: Gauss Codes, Quandles and Bondles. *arXiv arXiv:1912.09353* 2019, <https://arxiv.org/abs/1912.09353>.