Linear biopolymers such as RNA, DNA, and proteins often need to fold stably or transiently in order to function [1–4]. Changes to the configuration of these folded biopolymeric chains may lead to changes in function, loss of function, or dysfunction (i.e., toxic effects). For instance, the misfolding of proteins has been implicated in many diseases including some cancers and neurodegenerative diseases such as Parkinson’s and Alzheimer’s disease. Sophisticated biomolecular machines known as “chaperones” help to fold biomolecules into their “correct” conformation. Intriguingly, tens of thousands of proteins are helped by a small number of chaperones, suggesting that chaperones sense some generic properties of their client proteins and are not sensitive to all molecular details. It is possible that chaperones detect certain generic chemical properties (e.g., stickiness or “hydrophobicity”), some generic geometric properties (e.g., diameter of the folded parts), or they may sense topological properties of the protein chains. Chemical and geometric properties have been extensively studied, yet the chaperone puzzle has not been resolved. The topology of folded molecular chains is a less well-understood aspect of their structure. Being able to describe, experimentally measure, and represent the topology of a molecular chain could provide additional information potentially useful in understanding biology and disease. In recent decades, we observed advances in both describing molecular fold topology and experimentally characterizing it: knot theory has been applied to classify biomolecules and their folding [5]; optical tweezers, a technology recognized by the 2018 Physics Nobel Prize, has enabled grabbing and pulling of a knotted molecular chain via its ends. The optical tweezers technology also enables monitoring chaperone-assisted protein folding in real time [6,7]. However, there is a key mismatch between the two approaches: optical tweezers typically target the molecular ends, measure end-to-end distance, and characterize the molecular folds by breaking intrachain interactions (i.e., the so-called “contacts” that represent constraints that keep two points on the chain in close physical proximity; represented by red dots in Figure 1(a), while knot theory ignores chain ends and intrachain contacts and focuses on entanglement. Molecular knot analysis often starts with connecting the chain ends through reasonable (albeit debatable) protocols. Many physical constraints on movement are ignored. In particular, knot isotopy allows a full range of flexibility and movement, but actual biomolecules can physically achieve only a subset of them. Furthermore, most biomolecules, including more than 97% of structurally identified proteins, are unknotted according to knot theory. Thus, describing the topology of a folded linear molecular chain requires new approaches [8]. My coworkers and I addressed this problem by introducing “circuit topology” [8]. It starts with the recognition that molecules ranging from proteins, to RNA, and complete chromosomes, are all folded linear chains held...
Keystones

We are extending the original notion of circuit topology to include the additional complexity of 3D fold conformations seen in natural and engineered molecular chains [9]. Here, (generalized) circuit topology refers to the arrangement of contacts along the chain was called circuit topology by analogy with the arrangement of elements in electrical circuits. To identify the arrangements, we choose two contacts and ignore the rest. For two contacts, three arrangements are possible: they are either in parallel (P), in series (S), or cross (X) each other (with respect to the “end-to-end axis”; see Figure 1(b)). This step allows one to represent complex topologies by a collection of simple relations, which are tabulated in the form of a topology matrix ($T_{ij}$: circuit topology relation between contacts $i$ and $j$). This simple definition of contact arrangement is based on physical insights: chain loops formed in series are independent while parallel loops are not and may facilitate the formation of each other (the so-called zipping effect). The patterns within the topology matrix can be readily quantified and related to experimental observations (e.g., folding rate). Another advantage is that the formal structure of this approach allows one to formalize topological constraints that folded chains must obey [8]. For example, the P relation is transitive (i.e., for three contacts A, B, and C, if A is parallel to B and B is parallel to C, then A is parallel to C) while the S and X relations are not. Such relations allow for the development of inference algorithms that help in extracting structural information from optical tweezers data. To summarize, circuit topology utilizes a “bottom-up” approach, ignores entanglement, and focuses on intrachain contacts. In contrast, knot theory utilizes a “top-down” approach where a chain is globally approximated as a knot, which may be split into smaller tangles.

We are extending the original notion of circuit topology to include the additional complexity of 3D fold conformations seen in natural and engineered molecular chains [9]. Here, (generalized) circuit topology refers to the arrangement of contacts within a folded chain, where contacts represent constraints of various origins, not limited to direct chemical bonds as described above. Physical entanglements, sometimes introduced by knotting, may also keep chain segments in “contact” (Figure 1(c)). Similar to contacts shown in Figure 1(b), physical entanglements can be demonstrably categorized using the language of...
circuit topology [9]. These extensions allow one to study both knotted and unknotted molecular chains all using one topological framework. Circuit topology thus provides a generic quantitative topological framework that can be readily used to study the physics of folded polymers.

We demonstrated the power of this approach by studying folding and unfolding dynamics of polymers [10]. For example, we showed (theoretically and through molecular simulations) that the circuit topology is a folding rate predictor [11]. In particular, the frequency of \( P \) relations within the topology matrix correlates with the rate. One can show that when geometric measures such as contact order or chain length fail, the circuit topology of a folded chain predicts the folding rate. Interestingly, using simple chain models and circuit topology analysis, one can see how chaperones may guide folding towards the desired topology [12,13]. Chaperones constrain (protein) chains and thereby enhance the formation likelihood of certain contact arrangements. Chaperones can even introduce transient contacts within the chain by bridging two parts of the polymer, and thus can directly alter the apparent topology of the intermediate fold. This is not only important from a biological point of view, but also opens up new avenues in molecular engineering [14]. In the field of molecular engineering, a major gap in our knowledge is how to synthesize folded molecular chains, such as easy-to-fold DNA or protein origami. This is challenging on multiple levels, including designing the topology, guiding the folding process towards the desired topology (increasing the yield of the reaction), and carrying out necessary purification and characterization approaches [10,14]. Our proof-of-concept studies show that circuit topology analysis can be readily applied to address these challenges.

Circuit topology rules place constraints on unfolding transitions, allowing one to interpret optical tweezers data [8]. By grabbing the ends of a folded (unknotted) molecular chain, one can measure the end-to-end distance of the unfolded conformation \( L_U \) (see Figure 2(a)). Obviously, a fully unfolded conformation of the chain will have an end-to-end distance \( L_{U1} \) larger than \( L_F \). Upon pulling the ends of the molecule, we will observe a sequence of lengths in between \( L_F \) and \( L_{U1} \). Not all values between \( L_F \) and \( L_{U1} \) are possible and often the sequence of events is not unique due to the stochastic nature of contact disruption. Furthermore, at each unfolding step, contacts belonging to the “shortest end-to-end path” will sense the mechanical force applied to the chain ends and thus break. In the special case shown in Figure 2(a), only one unfolding pathway is possible and two length changes will be experimentally observed (corresponding to breaking two contacts in steps 1 and 2). In this case, the process is deterministic. In a series conformation shown in Figure 2(b), two possibilities are available for the first unfolding step. Note that simultaneous disruption of two contacts is unlikely. In a case with a complex topology, many length transitions are possible. Allowed and forbidden transitions can be readily extracted from circuit topology of the chain. For example, in Figure 2(a), one cannot break loop \( L_2 \) before \( L_1 \) by pulling the ends of the molecule. This is trivial when we see the conformation in Figure 2(a). However, in an experimental setting one cannot “see” the contacts and in many cases the structure of the protein is not known. In this situation, circuit topology may allow one to extract structure (number and arrangement of contacts) from a series of measured lengths.

The use of circuit topology in molecular structure analysis leads to many open mathematical questions, including: What symmetries can be identified in the topology matrices? What is an appropriate metric for the comparative analysis of two distinct topologies? How do we account for uncertainties caused by a lack of resolution or noise in the experimental data? Answering these questions may require combinatorics, topology, statistical methods, machine learning techniques, and advanced concepts from programming and algorithms.

ACKNOWLEDGMENTS. The author thanks Erica Flapan and Mohamed Elhamdadi for helpful feedback.

References


**Credits**

Figures 1 and 2 are courtesy of the author. Author photo is courtesy of the author.

---

**Alireza Mashaghi**

---

MathSciNet is the authoritative gateway to the scholarly literature of mathematics. Containing information on more than 3 million articles and books, with direct links to over 2 million articles in more than 1,800 journals. MathSciNet includes expert reviews, customizable author profiles, and citation information on articles, books, journals, and authors.

MathSciNet’s extensive resources can help you throughout your entire math career. Use it to:
- Quickly get up to speed on a new topic
- Look up a researcher’s body of work
- Find an article or book and track its reference list
- Research a math department to prepare for a job interview or when applying to graduate school

**How to Subscribe**

Go to [www.ams.org/mathsciprice](http://www.ams.org/mathsciprice) to learn more about MathSciNet, including information about subscription rates, joining a consortium, and a 30-day free trial.