Knaughty Biology

An Intersection of Topology and Molecular Biology

By Sana Raoof

In living systems, diversity safeguards species against extinction, allows for advantageous variations to accumulate, and makes our biosphere the rich, functionally interdependent system we thrive in today. At the molecular level, the miracle of biodiversity abounds: 20 amino acids polymerize in delicate helices and pleated sheets to form everything from hair and nails to gastric enzymes and silk thread; four nucleic acids arrange in compact, lengthy chains, a 1% divergence in which encodes the difference between Einstein and a chimpanzee; and carbon, hydrogen, and oxygen form carbohydrates, giving us sugars and starches, or lipids, giving us steroids and hydrophobic membrane bilayers. While we appreciate the essential condition of diversity for our survival as living beings, we recognize the daunting challenge it poses: the classification and study of nature's incredibly vast number of organic compounds.

Fortunately, a fundamental law of molecular biology is that the structure and function of organic systems are intimately connected. Therefore, the problem of classifying unidentified organic molecules can be reduced, with reasonable simplifications and exceptions in mind, to a problem of discriminating compounds based on structural conformation. As organic molecules are found heavily tangled and linked in nature (e.g. six meters of cellular DNA are compacted into a single nucleus in the form chromatin) (Figure 1), the complete and unique characterization of organic compounds on a purely structural basis constitutes a tantalizingly difficult and valuable goal of classic biochemistry (1).

A Knotty Problem

Lying at the other end of the scientific spectrum, the mathematical field of topology continues to twist, stretch, and tangle the imaginations of mathematicians. Formally, topologists study which properties of objects are preserved under deformations in space (2, 10). In the particularly loopy subfield of topology known as knot theory, these objects are restricted to knots. Mathematical knots can be thought of as everyday knots, as we see in shoelaces, iPod earphones, and electrical cables, except their ends are "glued together" to form a closed loop and are infinitely thin. Given a single knot, we can wrangle it in an infinite number of ways by deforming it freely in space, limited only by the rules that no region can shrink to a single point or cross through another region (4). If two knots can be wrangled into one another by a finite series of such deformations, we call them "equivalent." For any knot in 3-dimensional space, we can draw an infinite number of equivalent knots on paper by applying our favorite deformations (1, 4). Therefore, even in this naughtily abstract subfield of topology, a central objective is to be able to practically classify knots. Mathematicians have developed clever tools to distinguish knots, although every easily-computable such tool has generated false positive results, wrongly indicating that two different knots are equivalent. The quest to find a perfect tool to distinguish between knots (an easily-computable "complete invariant") continues to tease mathematicians to this day (6, 7, 8, 9).

"Four nucleic acids arrange in compact, lengthy chains, a 1% divergence in which encodes the difference between Einstein and a chimpanzee."

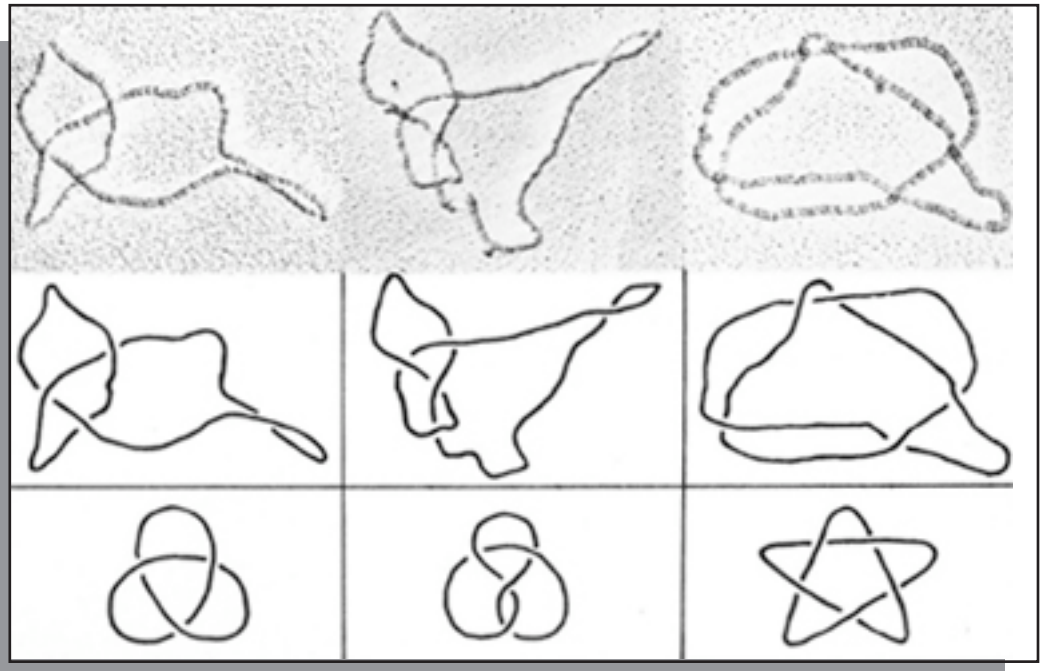
Math and Biology Intertwine

Could it be that research in the fields of mathematics and molecular biology, faced with the common quandary of classifying dynamic tangles, exist in a symbiotic relationship in which insights in one field can complement progression towards a solution in the other? To answer this question, we first take a brief trip through the central concepts and current research in the two respective fields.

To review, there are an infinite number of configurations a knot may assume due to trivial deformations. We therefore appreciate the value of knot invariants, the "tools" that mathematicians use to tell knots apart. A knot invariant is a function which maps the set of knots in 3-dimensional space to some other mathematical objects, which may be numbers, polynomials, graphs, etc (1, 2, 3, 4). Very importantly, knot invariants compute the

credit: 6

same value for all projections of a particular knot, no matter how deformed they are. So, if two knots yield different values under a knot invariant, then the knots must be distinct (4, 11). However, the converse statement isn't true: no known invariant can guarantee knot equivalence, because every known easily-computable invariant generates false positive results in certain cases (4). (We use the qualification "easily-computable" because there does exist a complete invariant: the fundamental group. Roughly speaking, the fundamental group is almost as complicated to deal with as a knot itself, and therefore does not offer a practical means of classifying unidentified knots) (1, 2, 4). The Alexander-Conway polynomial is a famous example of a knot invariant; as the name suggests, this invariant maps a knot to a specific polynomial (1, 2, 4).



▲ Figure 1. Electron micrographs of knots in DNA, line diagrams of the same knots, and simplified representations of the same knots.

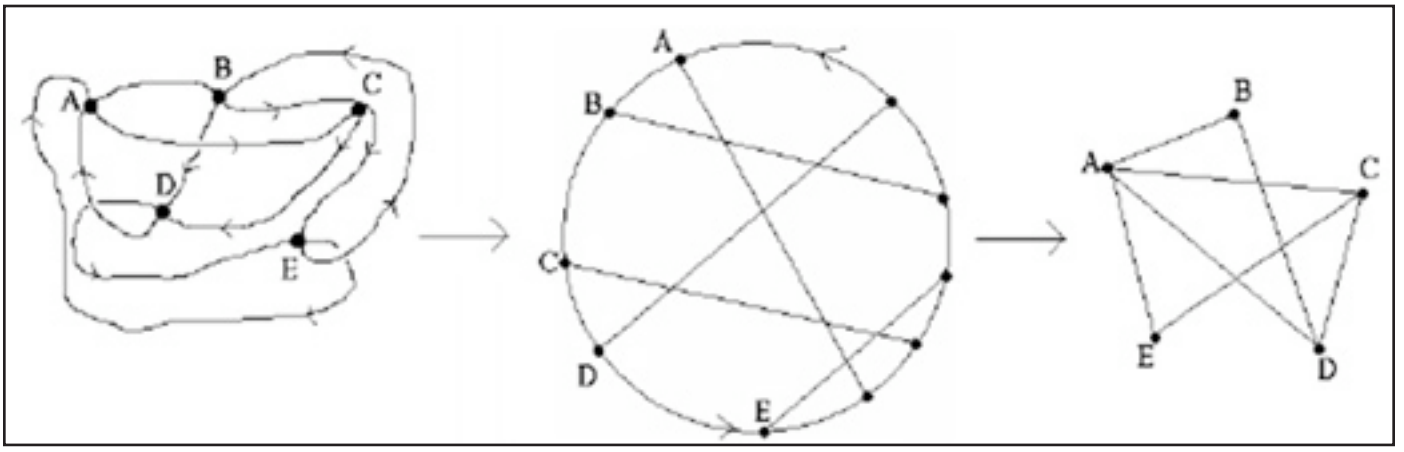
The Elusive Complete Invariant

A singularity is a point where two regions of a knot intersect. Hence, an n-singular knot is a knot with n points of intersection. Singular knots can be mapped to chord diagrams. A chord diagram of degree n is an oriented circle with n chords connecting 2n distinct points on the circumference of the circle. The relative order of chords around the chord diagram corresponds to the relative positions of singularities in the knot (Figure 2) (4).

Each chord diagram is associated to a combinatorial graph. A graph is a set of points connected by edges. For each chord in the chord diagram, we put a vertex on the graph. If any two chords intersect, then their vertices are connected by an edge. A graph constructed in this manner is called an intersection graph (Figure 2). The Intersection Graph Conjecture, which roughly proposed a one-to-one relation between the chord diagrams of singular knots and intersection graphs, was proven to be false.

Why would mathematicians want to map singular knots to chord diagrams or combinatorial graphs (4, 6, 7, 8, 9)? The motivation for such mappings is to simplify the object being studied; 2-dimensional pictures of singular knots can be quite unwieldy, whereas their corresponding chord diagrams and intersection graphs are relatively easy to study. One may wonder, therefore, if applying certain invariant functions to chord diagrams or graphs on behalf of their corresponding singular knots may be a much more practical, constructive task (1, 2, 3, 4). In fact, invariants for chord diagrams do exist; there is an analog of the Alexander-Conway polynomial which is applied to chord diagrams, the □ invariant. Since the □ invariant inputs chord diagrams rather than n-singular knots, it is considerably simpler to compute (4, 9). Experimenting with the □ invariant is therefore incredibly alluring: could such an easy-to-use tool provide a guarantee of knot equivalence?

Since organic molecules like DNA are found heavily tangled in nature, it is natural to model them as knots (Figure 1)(12); because the ends of molecules are free, we say their knots are "linked at infinity," unless the molecule forms a closed loop such as bacterial plasmid DNA). In order to carry out vital biological functions, such as replication, transcription, and recombination, certain enzymes manipulate the topology of organic molecules. We are therefore presented with a practical problem: it is hard enough to characterize a knot's structure, so how can we deal with knots whose structure is continuously changing? Luckily, many intermolecular interactions, such as salt bridges, the hydrophobic effect, and van der Waals attraction, only perform a simple deformation through space, and therefore they do not change the invariant value corresponding to a molecule's knot model (11)! Certain types of topoisomerases do, however, change the invariant values of knots corresponding to their substrate molecules by cutting the double helix, allowing DNA to relax and partially unwind, and then reattaching the two ends. To be specific, toposisomerase 1A changes the linking number of its substrate by 1, IIA and IIB change the linking



▲ Figure 1. A singular knot and its corresponding chord diagram and intersection graph. Starting at any point, follow the orientation to understand each mapping.

number by 2, and 1c changes the linking number by any integer (8).

In order to judge the knottedness of a particular molecule, we can use knot invariants to quantify the crossing number of a given knot. Interestingly, the gel velocity during electrophoresis is a function of how knotted a molecule is, not simply molecular weights.

If it is ever possible to guarantee knot equivalence, an atlas of invariant values corresponding to the mathematical models of known organic molecules, including DNA, complex enzymes, and other tertiary proteins, may eventually be created, offering an incredibly useful way to completely and uniquely characterize unidentified molecules on a purely structural basis. Given the close relationship between the structure and function of organic molecules, our simplified knot model of a molecule should offer a strong, precise approach towards solving the age-old biochemical problem of classifying unidentified, natural structures.

A Classic Biochemical Problem

We return our attention to the □ invariant. The □ invariant requires mathematicians to decompose chord diagrams into a signed sum of degree 1 or 2 chord diagrams (visualize a degree 1 chord diagram by drawing a circle with 1 chord within it), whose invariant values are stated as initial conditions (1, 2, 3, 4). Computing the □ invariant on chord diagrams of degree>12 (corresponding to >12-singular knots) is therefore rather tedious. Fortunately, such a complicated knot is not encountered in organic systems; researchers at Harvard/MITs school of Health Sciences and Technology have discovered the most complicated knot in the human body has a crossing number of 5 and belongs to the enzyme Human ubiquitin hydrolase.

Over the past five years, it was proven that the □ invariant applied to chord diagrams with complete intersection graphs computes (the coefficients of the Taylor series expansion for the tangent function)(n!) for chord diagrams. It was later proven that the □ invariant computes n!(n-1)! for chord diagrams corresponding to complete bipartite graphs (which have lattice chord diagrams) (4). Notice that, given a unique value of n, the general values of the □ invariant are also unique; hence, □ does not generate false positive results for chord diagrams in the mentioned families! Although we only have generalized formulas for the value of □ in specific cases, we are on a hot track—future research generalizing the value of the □ invariant on other families of chord diagrams will bring us closer to our goal of guaranteeing knot equivalence in as many cases as possible, an exciting prospect in both knot theory and molecular biology.

The Easily-Computable Complete Invariant: A Solution to Two Mysteries?

Our biological imaginations run wild with the idea of characterizing organic molecules structurally. The identification of pathogens in the human body (based on structural similarity to known molecules), classification of newly discovered microorganisms, and a determination of catalytic mechanisms are a few of the many natural applications extending from a system of uniquely classifying tangled molecules!

Readers interested in learning more about knot theory and its open problems are referred to (1). H

—Sana Raoof '12 is a prospective Biophysics and Math concentrator in Grays.

Reference List
[1] Adams, Colin C.: The Knot Book American Mathematical Society, 2004.
[2] Armstrong, M.A.: Basic Topology Springer-Verlag, 1983.
[3] Schneider, Hans.: Matrices and Linear Algebra New York: Dover Publications, 1973.
[4] Jordan, David.: Knot Invariants, Chord Diagrams, Lie Algebras, and the Intersection Graph Property for the Universal Invariants. University of Oregon, Department of Mathematics. (2006) (Thesis)
[5] Mellor, B.:The Intersection Graph Conjecture for Loop Diagrams. J. Knot Theory Ramif 9, NO. 2. (2000)
[6] Bar-Natan, D. Topology 34, No. 2, pp. 423-472, 1995
[7] Chmutov, et al. Advances in Soviet Mathematics 21, pp.117-145, 1994
[8] Champoux JJ (2001). Annu. Rev. Biochem. 70: 369-413.
[9] "Duplex DNA knots produced by Escherichia coli topo...[J Biol Chem. 1985]
[10] Kontsevich. Advances in Soviet Mathematics Vol. 16, Part 2, pp. 137-150, 1993
[11] Vaintrob, A. Math Research Letters 1 (1994)
[12] Weisstein, Eric W.:Knot. MathWorld-A Wolfram Web Resource, 2003.
[13] Nielsen, Bjrn G.:Gauss-integral based representation of protein structure for predicting the fold class from the sequence. Department of Mathematics, Technical University of Denmark. (2004)

credit: Sana Raoof