Lifting the Curtain: Using Topology to Probe the Hidden Action of Enzymes

De Witt Sumners

One of the important issues in molecular biology is the three-dimensional structure (shape) of proteins and deoxyribonucleic acid (DNA) in solution in the cell and the relationship between structure and function. Ordinarily, protein and DNA structure is determined by X-ray crystallography or electron microscopy. Because of the close packing needed for crystallization and the manipulation required to prepare a specimen for electron microscopy, these methods provide little direct evidence for molecular shape in solution.

How is DNA wound around the enzyme, and what happens during recombination?

Topology can shed light on this key issue. The topological approach to enzymology is an experimental protocol in which the descriptive and analytical powers of topology and geometry are employed in an indirect effort to determine the enzyme mechanism and the structure of active enzyme-DNA complexes in vitro (in a test tube). We describe how recent results in 3-dimensional topology [3, 5, 9, 10, 11] have proven to be of use in the description and quantization of the action of cellular enzymes on DNA.

The Topology of DNA

DNA can be viewed as two very long curves that are intertwined millions of times, linked to other curves, and subjected to four or five successive orders of coiling to convert it into a compact form for information storage. If one scales the cell nucleus up to the size of a basketball, the DNA inside scales to the size of thin fishing line, and 200 km of that fishing line are inside the nuclear basketball. Most cellular DNA is double-stranded (**duplex**), consisting of two linear backbones of alternating sugar and phosphorus. Attached to each sugar molecule is one of the four bases (**nucleotides**): A = adenine, T = thymine, C = cytosine, G = guanine. A ladder whose sides are the backbones and whose rungs are hydrogen bonds is formed by hydrogen bonding between base pairs, with A bonding only with T, and C bonding only with G. The **base** pair sequences for a linear segment of duplex DNA is obtained by reading along one of the two backbones and is a word in the letters $\{A, T, C, G\}$. Due to the uniqueness of the bonding partner for each nucleotide, knowledge of the sequence along one backbone implies knowledge of the sequence along the other backbone. In the classic Crick-Watson double helix model for DNA, the ladder is twisted in a right-hand helical fashion, with an average and nearly constant pitch of approximately 10.5 base pairs per full

Adapted with permission from "Calculating the Secrets of Life". Copyright 1995 by the National Academy of Sciences. Courtesy of the National Academy Press, Washington, D.C. Electron micrographs provided courtesy of N. R. Cozzarelli and A. Stasiak.

De Witt Sumners is professor of mathematics at Florida State University in Tallahassee, FL. His e-mail address is sumners@gauss.math.fsu.edu.

helical twist. The local helical pitch of duplex DNA is a function of both the local base pair sequence and the cellular environment in which the DNA lives; if a DNA molecule is under stress, or constrained to live on the surface of a protein, or is being acted upon by an enzyme, the helical pitch can change.

The packing, twisting, and topological constraints all taken together mean that topological entanglement poses serious functional problems for DNA. This entanglement would interfere with, and be exacerbated by, the vital life processes of replication, transcription, and recombination. For information retrieval and cell viability, some geometric and topological features must be introduced into the DNA, and others quickly removed [12, 13]. For example, the Crick-Watson helical twist of duplex DNA may require local unwinding in order to make room for a protein involved in transcription to attach to the DNA. The DNA sequence in the vicinity of a gene may need to be altered to include a promoter or repressor. During replication, the daughter duplex DNA molecules become entangled and must be disentangled in order for replication to proceed to completion. After the process is finished, the original DNA conformation must be restored. Some enzymes maintain proper geometry and topology by passing one strand of DNA through another by means of a transient enzyme-bridged break in one of the DNA strands. Other enzymes break the DNA apart and recombine the ends by exchanging them, a move performed by **recombinases**. The description and quantization of the three-dimensional structure of DNA and the changes in DNA structure due to the action of these enzymes have required the serious use of geometry and topology in molecular biology. This use of mathematics as an analytic tool is especially important because there is no experimental way to observe the dynamics of enzymatic action directly.

In the experimental study of DNA structure and enzyme mechanism, biologists developed the topological approach to enzymology [15], shown schematically in Figure 1. In this approach, one performs experiments on circular substrate DNA molecules. These circular substrate molecules are genetically engineered by cloning techniques to contain regions that a certain enzyme will recognize and act upon. The circular form of the substrate molecule traps an enzymatic topological signature in the form of DNA knots and links (catananes). These DNA knots and links are observed by gel electrophoresis and electron microscopy of the reaction product DNA molecules. By observing the changes in geometry (supercoiling) and topology (knotting and linking) in DNA caused by an enzyme, the enzyme mechanism can be described and quantized.

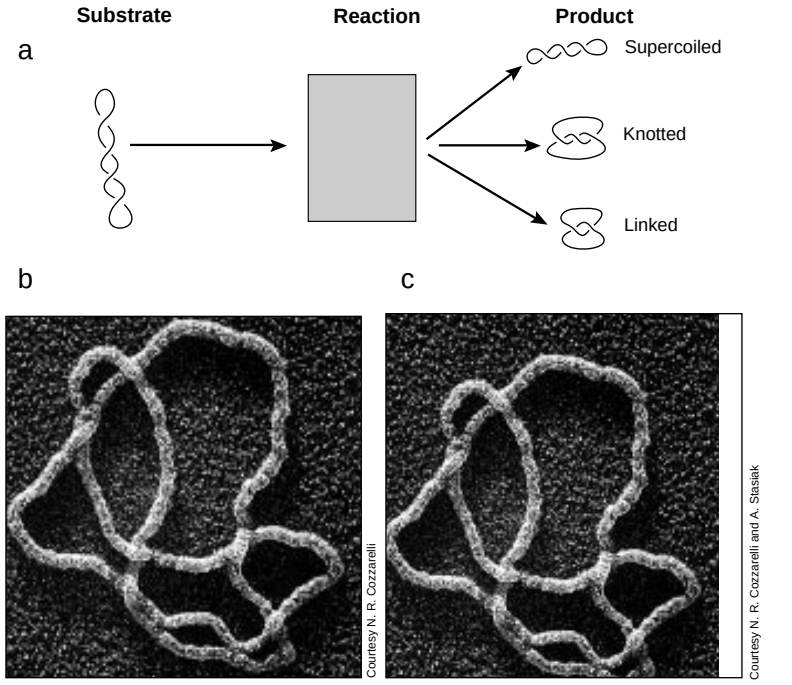

Figure 1
(a) Topological approach to enzymology, (b) DNA (+) Whitehead link, (c) DNA knot 6_2^*.

The topological approach to enzymology poses an interesting challenge for mathematics: from the observed changes in DNA geometry and topology, how can one deduce enzyme mechanisms? This requires the construction of mathematical models for enzyme action and the use of these models to analyze the results of topological enzymology experiments. The entangled form of the product DNA knots and links contains information about the enzymes that made them. In addition to utility in the analysis of experimental results, the use of mathematical models forces all of the background assumptions about the biology to be carefully laid out. At this point they can be examined and dissected, and their influence on the biological conclusions drawn from experimental results can be determined.

Site-Specific Recombination

Site-specific recombination is one of the ways in which nature alters the genetic code of an organism, either by moving a block of DNA to another position on the molecule or by integrating a block of alien DNA into a host genome. One of the biological purposes of recombination is the regulation of gene expression in the cell, because it can alter the relative position of the gene and its repressor and promoter sites on the

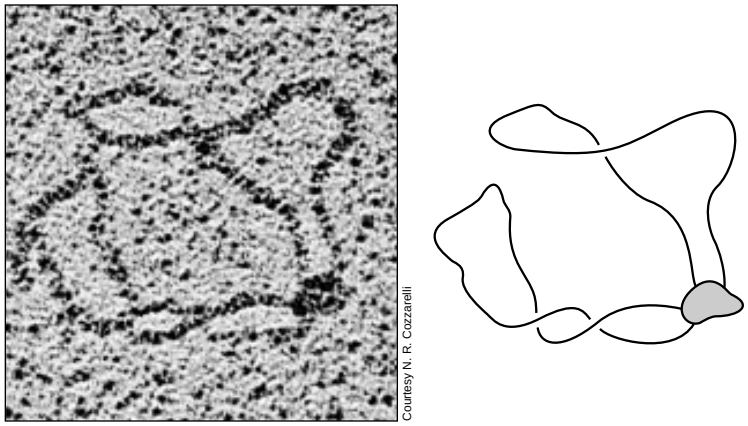

Figure 2
Tn3 synaptic complex

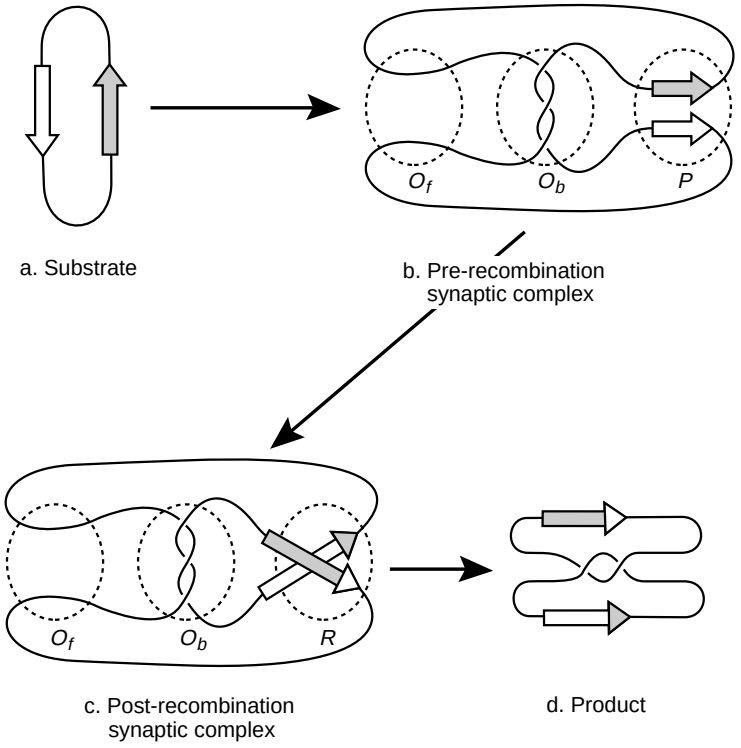

Figure 3
A single recombination event: direct repeats.

genome. Site-specific recombination also plays a vital role in the life cycle of certain viruses, which utilize this process to insert viral DNA into the DNA of a host organism. An enzyme that mediates site-specific recombination on DNA is called a **recombinase**. A recombination site is a short segment of duplex DNA whose sequence is recognized by the recombinase. Site-specific recombination can occur when a pair of sites (on the same or on different DNA molecules) become juxtaposed in the presence of the recombinase. The pair of sites is aligned through enzyme manipulation or random thermal motion (or both), and both sites (and perhaps some contiguous DNA) are then bound by the enzyme. This stage of the reaction is called **synapsis**, and we will call this intermediate protein-DNA complex formed by the part of the substrate that is bound to the enzyme together with the enzyme itself the **synaptosome**. We will call the entire DNA molecule(s) involved in synapsis (including the parts of the DNA molecule(s) not bound to the enzyme), together with the enzyme itself, the **synaptic complex**. The electron micrograph in Figure 2 (courtesy of N.R. Cozzarelli) shows a synaptic complex formed by the recombination enzyme Tn3 resolvase when reacted with unknotted circular duplex DNA. In the micrograph of Figure 2, the synaptosome is the black mass attached to the DNA circle, with the unbound DNA in the synaptic complex forming twisted loops in the exterior of the synaptosome. It is our intent to deduce mathematically the path of the DNA in the black mass of the synaptosome, both before and after recombination. We want to answer the question: How is DNA wound around the enzyme, and what happens during recombination?

After forming the synaptosome, a single recombination event occurs: the enzyme then performs two double-stranded breaks at the sites and recombines the ends by exchanging them in an enzyme-specific manner. The synaptosome then dissociates, and the DNA is released by the enzyme. We call the pre-recombination unbound DNA molecule(s) the substrate and the post-recombination unbound DNA molecule(s) the product. During a single binding encounter between enzyme and DNA, the enzyme may mediate more than one recombination event; this is called **processive recombination**. On the other hand, the enzyme may perform recombination in multiple binding encounters with the DNA, which is called **distributive recombination**. Some site-specific recombination enzymes mediate both distributive and processive recombination.

Site-specific recombination involves topological changes in the substrate. In order to identify these topological changes, one chooses to perform experiments on circular DNA substrate. One must perform an experiment on a large number of circular molecules in order to obtain an observable amount of product. Using cloning techniques, one can synthesize circular duplex DNA molecules, which contain two copies of a recombination site. At each recombination site, the base pair sequence is in general not palindromic hence induces a local orientation on the

substrate DNA circle. If these induced orientations from a pair of sites on a singular circular molecule agree, this site configuration is called **direct repeats** (or **head-to-tail**), and if the induced orientations disagree, this site configuration is called inverted repeats (or **head-to-head**). If the substrate is a single DNA circle with a single pair of directly repeated sites, the recombination product is a pair of DNA circles and can form a DNA link (or **catanane**) (Figure 3). If the substrate is a pair of DNA circles with one site each, the product is a single DNA circle (Figure 3 read in reverse) and can form a DNA knot (usually with direct repeats). In processive recombination on circular substrate with direct repeats, the products of an odd number of rounds of processive recombination are DNA links, and the products of an even number of rounds of processive recombination are DNA knots. If the substrate is a single DNA circle with inverted repeats, the product is a single DNA circle and can form a DNA knot. In all figures where DNA is represented by a line drawing (such as Figure 3), duplex DNA is represented by a single line, and supercoiling is omitted.

The geometry and topology of circular DNA substrate are experimental control variables. The geometry and topology of the recombination reaction products are observables. In vitro experiments usually proceed as follows: Circular substrate is prepared, with all of the substrate molecules representing the same knot type. The amount of supercoiling of the substrate molecules is also a control variable. The substrate molecules are reacted with a high concentration of purified enzyme, and the reaction products are fractionated by gel electrophoresis. Gel electrophoresis discriminates among DNA molecules on the basis of molecular weight; given that all molecules are the same molecular weight (as is the case in these topological enzymology experiments), electrophoresis discriminates on the basis of subtle differences in the geometry (supercoiling) and topology of the DNA molecules. Under the proper conditions gel velocity is (surprisingly) determined by the crossing number of the knot or link; knots and links of the same crossing number migrate with the same gel velocities [4]. After running the gel, the DNA molecules are removed from the gel and coated with Rec A protein. It is this new observation technique (Rec A-enhanced electron mocroscopy) [7] that makes possible the detailed knot-theoretic analysis of reaction products. Rec A is an E. coli protein that binds to DNA and mediates general recombination in E. coli. The process of Rec A coating fattens, stiffens, and stretches (untwists) the DNA. This facilitates the unambiguous determination of crossings (nodes) in an electron micrograph of DNA.

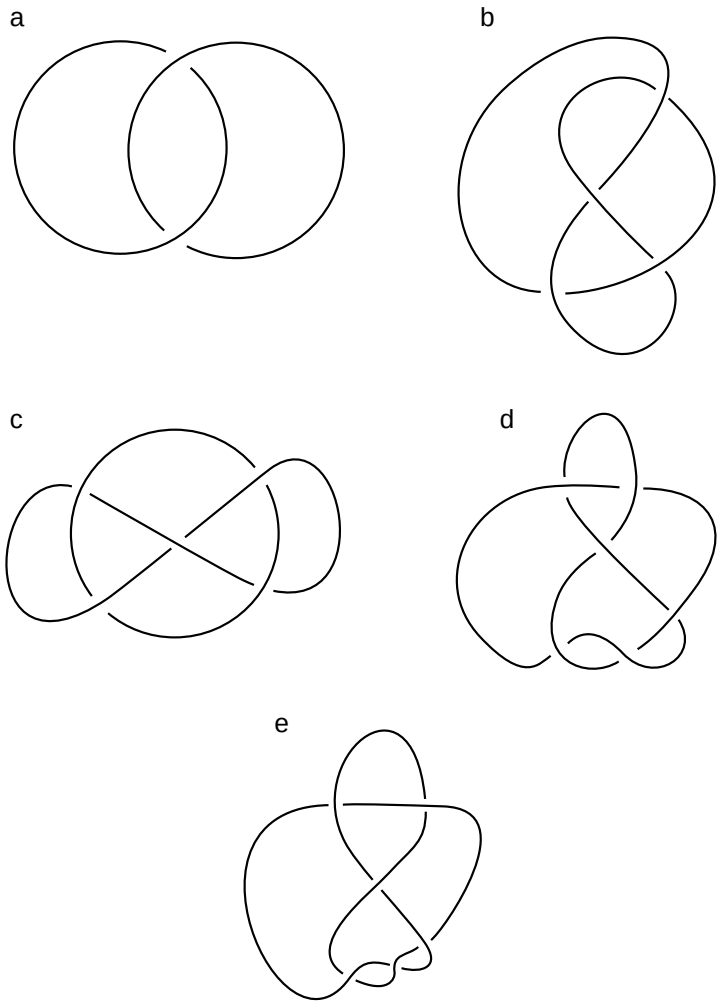

Figure 4
(a) Hopf link, (b) figure 8 knot, (c) (+) Whitehead link, (d) 6_2^* and (e) 6_2 mirror image of 6_2^*.

Topological Tools for DNA Analysis

In this section, we will describe the parts of knot theory and tangle calculus of biological relevance. For a rigorous mathematical treatment we refer the reader to [1, 6, 8] for knot theory and [5] for tangle calculus.

A knot K is an embedding of a single circle in R^3; a link L is an embedding of two or more circles in R^3. Unless otherwise specified, all knots and links will be unoriented; the ambient space R^3 has a fixed orientation. Two knots (links), $K-1$ and $K-2$, are equivalent if there is an orientation-preserving homeomorphism of pairs $h:(R^3;K_1)\to(R^3;K_2)$. The homeomorphism of pairs h superimposes K_1 on K_2; in this case the knots (links) can be made congruent by a flexible motion or flow (**ambient isotopy**) of space. An ambient isotopy is a l-parameter fam-

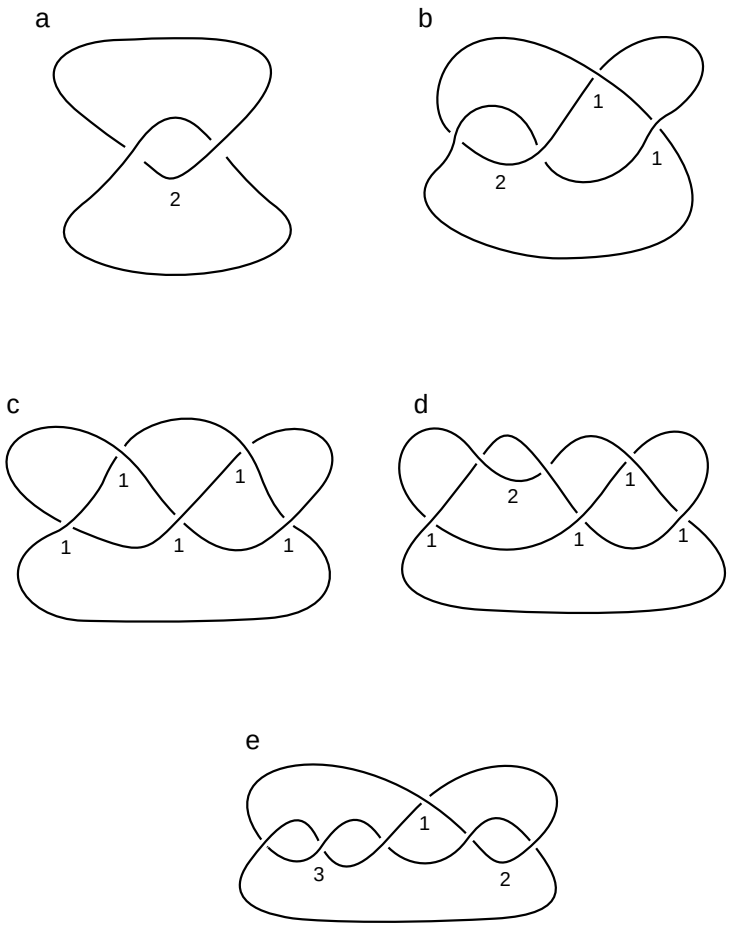

Figure 5: Standard 4-plats
(a) <2> Hopf link, (b) <2,1,1> figure 8 knot,
(c) <1,1,1,1,1> (+) Whitehead link, (d) <1,2,1,1,1> 6_2^*, and
(e) <3,1,2> 6_2.

ily of homeomorphisms $\{H_t\}_{t=0}^{1}$ of R^3 that begins with the identity and ends with the homeomorphism under consideration: H_0= identity and $H_1 = h$. An equivalence class of embeddings is called a **knot (link) type**.

A knot (link) is usually represented by drawing a diagram (projection) in a plane. This diagram is a shadow of the knot (link) cast on a plane in 3-space. Figure 4 shows standard diagrams [8] for the knots and links that turn up in Tn3 recombination experiments. In the definition of knot type, we insisted that the transformation that superimposes one knot on another must be orientation-preserving on the ambient space. This restriction allows us to detect a property of great biological significance: **chirality**. If K denotes a knot (link), let K^* denote the mirror image. If $K = K^*$, then we say that K is **achiral**; if $K \neq K^*$, then we say that K is **chiral**.

Fortunately for biological applications, most (if not all) of the circular DNA products produced by in vitro enzymology experiments fall into the mathematically well-understood family of 4-plats. This family consists of knot and link configurations produced by patterns of plectonemic supercoiling of pairs of strands about each other. All "small" knots and links are members of this family—more precisely, all prime knots with crossing number less than 8 and all prime (two-component) links with crossing number less than 7 are 4-plats. A **4-plat** is a knot or two-component link that can be formed by platting (or braiding) four strings. All of the knots and links in Figure 4 are 4-plats; their standard 4-plat diagrams are shown in Figure 5. Each standard 4-plat diagram consists of four horizontal strings, and the standard pattern of half-twists (plectonemic interwinds) of strings is encoded by an odd-length classifying vector with positive integer entries $\langle c_1; c_2; \ldots; c_{2k+1} \rangle$, as shown in Figure 5.

For in vitro topological enzymology, we can regard the enzyme mechanism as a machine that transforms 4-plats into other 4-plats. We need a mathematical language for describing and computing these enzyme-mediated changes. In many enzyme-DNA reactions, a pair of sites that are distant on the substrate circle are juxtaposed in space and bound to the enzyme. The enzyme then performs its topological moves, and the DNA is then released. We need a mathematical language to describe configurations of linear strings in a spatially confined region. This is accomplished by means of the mathematical concept of tangles, which were introduced into knot theory by J.H. Conway [2]. Tangle theory is knot theory done inside a 3-ball with the ends of the strings firmly glued down. On the unit 3-ball, select four points on the equator (called NW;SW;SE;NE). A **2-string tangle** in the unit 3-ball is a configuration of two disjoint strings in the unit 3-ball whose endpoints are the four special points {NW;SW;SE;NE}. Two tangles in the unit 3-ball are **equivalent** if it is possible to elastically transform the strings of one tangle into the strings of the other without moving the endpoints {NW;SW;SE;NE} and without breaking a string or passing one string through another. A class of equivalent tangles is called a tangle type. Tangles are usually represented by their projections, called **tangle diagrams**, onto the equatorial disk in the unit 3-ball, as shown in Figure 6. In all figures containing tangles, we assume that the four boundary points {NW;SW;SE;NE} are as in Figure 6, and we suppress these labels.

All four of the tangles in Figure 6 are pairwise inequivalent. However, if we relax the restriction that the endpoints of the strings remain fixed and allow the endpoints of the strings to move about on the surfaces (S^2) of the 3-ball, then the

tangle of Figure 6a can be transformed into the trivial tangle of Figure 6d. The tangles in Figures 6b and 6c cannot be transformed to the trivial tangle by any sequence of such turning motions of the endpoints on S^2. The family of tangles that can be converted to the trivial tangle by moving the endpoints of the strings on S^2 is the family of **rational tangles**. Equivalently, a rational tangle is one in which the strings can be continuously deformed (leaving the endpoints fixed) entirely into the boundary 2-sphere of the 3-ball, with no string passing through itself or through another string.

Rational tangles form a homologous family of 2-string configurations in B^3 and are formed by a pattern of plectonemic supercoiling of pairs of strings. Like 4-plats, rational tangles look like DNA configurations being built up out of plectonemic supercoiling of pairs of strings. More specifically, enzymes are often globular in shape and are topologically equivalent to our unit-defining ball. Thus, in an enzymatic reaction between a pair of DNA duplexes, the pair {enzyme, bound DNA} forms a 2-string tangle. Since the amount of bound DNA is small, the enzyme-DNA tangle so formed admits projections with few nodes and therefore is very likely rational. For example, all locally unknotted 2-string tangles having less than five crossings are rational. There is a second, more natural argument for rationality of the enzyme-DNA tangle. In all cases studied intensively, DNA is bound to the surface of the protein. This means that the resulting protein-DNA tangle is rational, since any tangle whose strings can be continuously deformed into the boundary of the defining ball is automatically rational.

There is a classification scheme for rational tangles which is based on a standard form that is a minimal alternating diagram. The classifying vector for a rational tangle is an integer-entry vector $(a_1;a_2;\ldots;a_n)$ of odd or even length, with all entries (except possibly the last) nonzero and having the same sign and with $|a_1| > 1$. The integers in the classifying vector represent the left-to-right (west-to-east) alternation of vertical and horizontal windings in the standard tangle diagram, always ending with horizontal windings on the east side of the diagram. Horizontal winding is the winding between strings in the top and bottom (north and south) positions; vertical winding is the winding between strings in the left and right (west and east) positions. By convention, positive integers correspond to horizontal plectonemic right-handed supercoils and vertical left-handed plectonemic supercoils; negative integers correspond to horizontal left-handed plectonemic supercoils and vertical right-handed plectonemic supercoils. Figure 7 shows some standard tangle

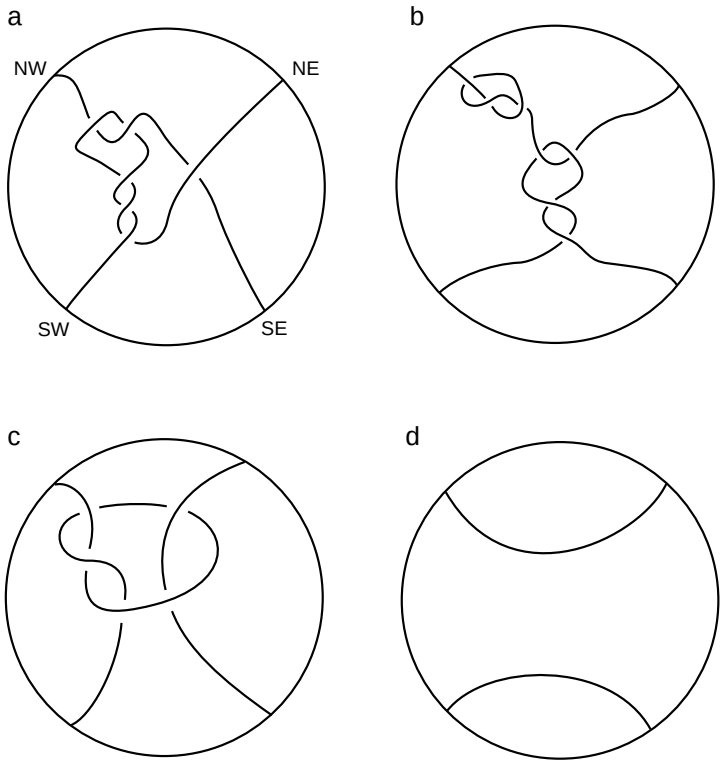

Figure 6: Tangles
(a) Rational, (b) locally knotted, (c) prime, and (d) trivial.

diagrams. Two rational tangles are of the same type if and only if they have identical classifying vectors. Due to the requirement that $|a_1| > 1$ in the classifying vector convention for rational tangles, the corresponding tangle projection must have at least two nodes. There are four rational tangles $\{(0);(0;0);(1);(-1)\}$ that are exceptions to this convention ($|a_1| = 0$ or 1), and are displayed in Figure 7c through f. The classifying vector $(a_1;a_2;\ldots;a_n)$ can be converted to an (extended) rational number $b/a \in Q \cup \infty$ by means of the following continued fraction calculation:

$$b/a = a_n + 1/(a_{n-1} + (1/(a_{n-2} + \cdots))):$$

Two rational tangles are of the same type if and only if these rational numbers are equal [2].

In order to use tangles as building blocks for knots and links and mathematically to mimic enzyme action on DNA, we now introduce the geometric operations of **tangle additional** and **tangle closure**. Given tangles A and B, one can form the tangle $A + B$ as shown in Figure 8a. The sum of two rational tangles need not be rational. Given any tangle C, one can form the closure $N(C)$ as in Figure 8b. In the closure operation on a 2-string tangle, ends NW and NE are

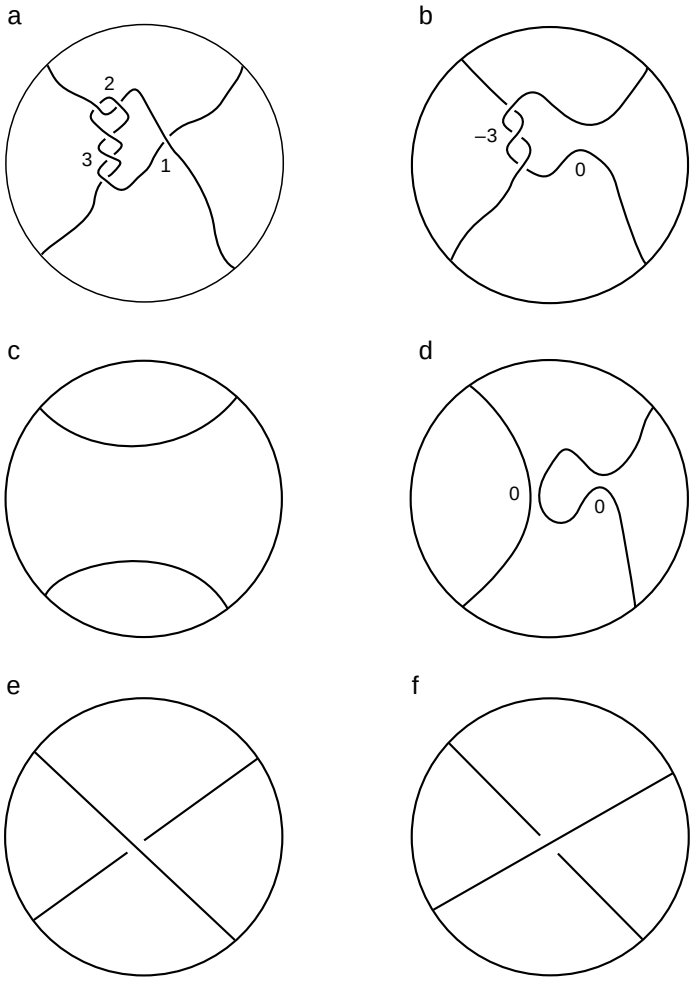

Figure 7: Tangle Diagrams
(a) (2, 3, 1), (b) (-3, 0), (c) (0), (d) (0,0), (e) (1), and (f) (-1).

connected, ends SW and SE are connected, and the defining ball is deleted, leaving a knot or a link of two components. Deletion of the defining B^3 is analogous to deproteinization of the DNA when the synaptosome dissociates. One can combine the operations of tangle addition and tangle closure to create a **tangle equation** of the form N(A + B) = knot (link). In such a tangle equation, the tangles A and B are said to be **summands** of the resulting knot (link). An example of this phenomenon is the tangle equation $N((-3;0)+(1)) = \langle 2 \rangle$ shown in Figure 8c. In general, if A and B are any two rational tangles, then N(A + B) is a 4-plat.

The Tangle Model for Site-Specific Recombination

The fundamental observations underlying this model are that a pair of sites bound by an enzyme forms a tangle and that most of the products of recombination experiments performed on unknotted substrate are 4-plats. We will use tangles to build a model that will compute the topology of the pre- and post-recombination synaptic complex in a single recombination event, given knowledge of the topology of the substrate and product [5, 10, 9, 11]. In site-specific recombination on circular DNA substrate, two kinds of geometric manipulation of the DNA occur. The first is a global ambient isotopy, in which a pair of distant recombination sites are juxtaposed in space and the enzyme binds to the molecule(s), forming the synaptic complex. Once synapsis is achieved, the next move is local and due entirely to enzyme action. Within the region occupied by the enzyme, the substrate is broken at each site, and the ends are recombined. We will model this local move.

Within the region controlled by the enzyme, the enzyme breaks the DNA at each site and recombines the ends by exchanging them. We model the enzyme itself as a 3-ball. The synaptosome consisting of the enzyme and bound DNA forms a 2-string tangle.

What follows is a list of biological and mathematical assumptions made in the tangle model [5, 10, 11].

Assumption 1. *The enzyme mechanism in a single recombination event is constant, independent of the geometry (supercoiling) and topology (knotting and catenation) of the substrate population. Moreover, recombination takes place entirely within the domain of the enzyme ball, and the substrate configuration outside the enzyme ball remains fixed while the strands are being broken and recombined inside and on the boundary of the enzyme.*

That is, we assume that any two pre-recombination copies of the synaptosome are identical, meaning that we can by rotation and translation superimpose one copy on the other, with the congruence so achieved respecting the structure of both the protein and the DNA. We likewise assume that all of the copies of post-recombination synaptosome are identical.

In a recombination event, we can mathematically divide the DNA involved into three types: (1) the DNA at and very near the sites where the DNA breakage and reunion are taking place; (2) other DNA bound to the enzyme, which is unchanged during a recombination event; and (3) the DNA in the synaptic complex that is not bound to the enzyme and that does not change during recombination. We make the following mathematical assumption about DNA types (1) and (2):

Assumption 2. *The synaptosome is a 2-string tangle and can be mathematically subdivided into the sum* $0_b + P$ *of two tangles.*

One tangle, the parental tangle P, contains the recombination sites where strand breakage and reunion take place. The other tangle, the outside bound tangle O_b, is the remaining DNA in the synaptosome outside the P tangle; this is the DNA that is bound to the enzyme but that remains unchanged during recombination. The enzyme mechanism is modeled as tangle replacement (surgery) in which the parental tangle P is removed from the synaptosome and replaced by the recombinant tangle R. Therefore, our model assumes the following:

pre-recombination synaptosome $= O_b + P$,

post-recombination synaptosome $= O_b + R$.

In order to accommodate nontrivial topology in the DNA of type (3), we let the outside free tangle O_f denote the synaptic complex DNA that is free (not bound to the enzyme) and that is unchanged during a single recombination event. We make the following mathematical assumption:

Assumption 3. *The entire synaptic complex is obtained from the tangle sum (O_f + synaptosome) by the tangle closure construction.*

If one deproteinizes the pre-recombination synaptic complex, one obtains the substrate; deproteinization of the post-recombination synaptic complex yields the product. The topological structure (knot and catanane types) of the substrate and product yields equations in the recombination variables $\{O_f; O_b; P; R\}$. Specifically, a single recombination event on a single circular substrate molecule produces two recombination equations in four unknowns:

substrate equation: $N(O_f + O_b + P) =$ substrate,

product equation: $N(O_f + O_b + R) =$ product.

The geometric meaning of these recombination equations is illustrated in Figure 3. In Figure 3, $O_f = (0); O_b = (-3;0); P = (0)$, and $R = (1)$. With these values for the variables, our recombination equations become:

substrate equation: $N((0) + (-3;0) + (0)) = \langle 1 \rangle$,

product equation: $N((0) + (-3;0) + (1)) = \langle 2 \rangle$:

The Topology of Tn3 Resolvase

Tn3 resolvase is a site-specific recombinase that reacts with certain circular duplex DNA substrate with directly repeated recombination sites [14]. One begins with supercoiled unknotted DNA substrate and treats it with resolvase. The principal product of this reaction is known to be the DNA 4-plat $\langle 2 \rangle$ (the Hopf link, Figures 4a and 5a). Resolvase is known to act dispersively in this situation to bind to the circular DNA, to mediate a single recombination event, and then to release the linked product. It is also known that resolvase and free (unbound) DNA links do not react. However, once in twenty encounters, resolvase acts processively—additional recombinant strand exchanges are promoted prior to the release of the product, with yield decreasing exponentially with increasing number of strand exchanges at a single binding encounter with the enzyme. Two successive rounds of processive recombination produce the DNA 4-plat $\langle 2;1;1 \rangle$ (the figure eight knot, Figures 4b and 5b); three successive rounds of processive recombination produce the DNA 4-plat $\langle 1;1;1;1;1 \rangle$ (the Whitehead link, Figures 4c and 5c), whose electron micrograph appears in Figure 1b; four successive rounds of recombination produce the DNA 4-plat $\langle 1;2;1;1;1 \rangle$ (the knot 6_2^*, Figures 4d and 5d), whose electron micrograph appears in Figure 1c. The discovery of the

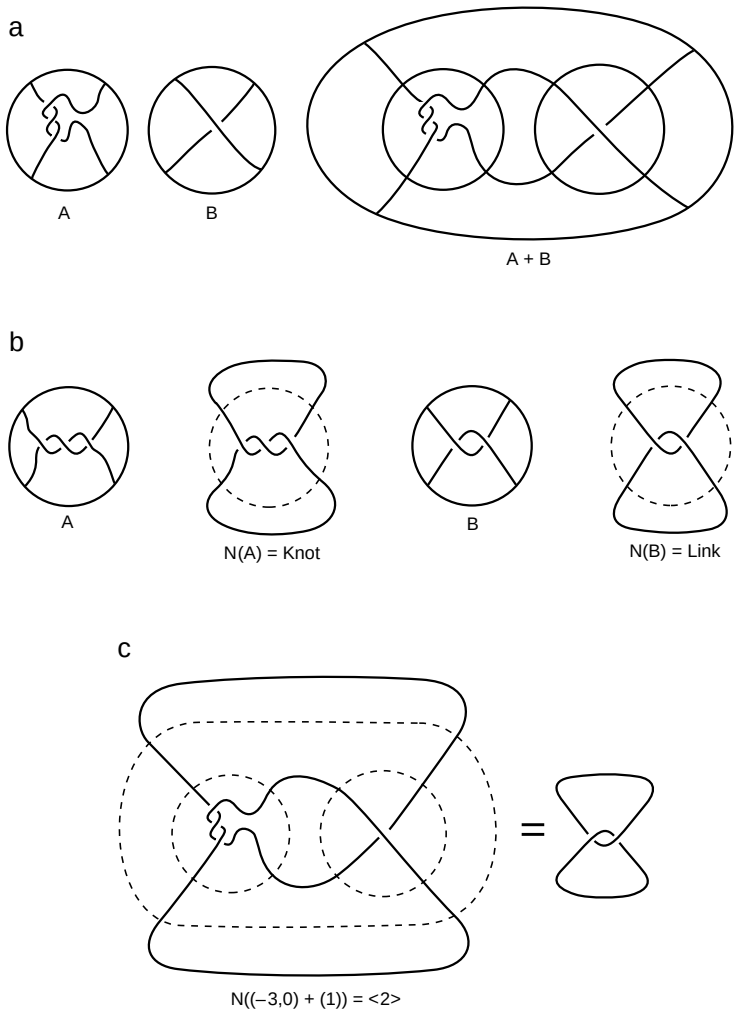

Figure 8: Tangle Operations
(a) Tangle addition, (b) tangle closure, and (c) N ((-3,0) + (1)) = < 2 >.

DNA knot $\langle 1;2;1;1;1\rangle$ substantiated a model for Tn3 resolvase mechanism [14].

In processive recombination, it is the synaptosome itself that repeatedly changes structure. We make the following biologically reasonable mathematical assumption in our model:

Assumption 4. *In processive recombination, each additional round of recombination adds a copy of the recombinant tangle* R *to the synaptosome.*

More precisely, n rounds of processive recombination at a single binding encounter generate the following system of $(n+1)$ tangle equations in the unknowns $\{O_f;O_b;P;R\}$:

substrate: $N(O_f+O_b+P)=$ substrate

rth round: $N(O_f+O_b+rR)=$ rth round product, $1\le r\le n$.

For resolvase, the electron micrograph of the synaptic complex in Figure 2 reveals that $O_f=(0)$, since the DNA loops on the exterior of the synaptosome can be untwisted and are not entangled. This observation from the micrograph reduces the number of variables in the tangle model by one, leaving us with three variables $\{O_b;P;R\}$. One can prove [10, 9, 5] that there are four possible tangle pairs $\{O_b;R\}$ which can produce the experimental results of the first two rounds of processive Tn3 recombination. The third round of processive recombination is then used to discard three of these four pairs of extraneous solutions. The following theorems can be viewed as a mathematical proof of resolvase synaptic complex structure: the model proposed in [14] is the unique explanation for the first three observed products of processive Tn3 recombination, assuming that processive recombination acts by adding on copies of the recombinant tangle R.

Mathematical analysis makes feasible the reconstruction of DNA topology from gel electrophoresis, avoiding the technically difficult electron microscopy of Rec A-enhanced DNA.

Theorem 1. *Suppose that tangles* $O_b;P$, *and* R *satisfy the following equations:*

1. $N(O_b+P)=\langle 1\rangle$ (the unknot),
2. $N(O_b+R)=\langle 2\rangle$ (the Hopf link),
3. $N(O_b+R+R)=\langle 2;1;1\rangle$ (the figure 8 knot).

Then $\{O_b;R\}=\{(-3;0);(1)\}$, $\{(3;0);(-1)\}$, $\{(-2;-3;-1);(1)\}$, *or* $\{(2;3;1);(-1)\}$:

In order to decide which is the biologically correct solution, we must utilize more experimental evidence. The third round of processive resolvase recombination determines which of these four solutions is the correct one.

Theorem 2. *Suppose that tangles* $O_b;P$ *;and* R *satisfy the following equations:*

1. $N(O_b+P)=\langle 1\rangle$ (the unknot),
2. $N(O_b+R)=\langle 2\rangle$ (the Hopf link),
3. $N(O_b+R+R)=\langle 2;1;1\rangle$ (the figure 8 knot),
4. $N(O_b+R+R+R)=\langle 1;1;1;1;1\rangle$ (the (+) Whitehead link).

Then $O_b=(-3;0);R=(1)$, *and* $N(O_b+R+R+R+R)=\langle 1;2;1;1;1\rangle$:

The correct global topology of the first round of processive Tn3 recombination on the unknot is shown in Figure 3. Moreover, the first three rounds of processive Tn3 recombination uniquely determine $N(O_b+R+R+R+R)$, the result of four rounds of recombination. It is the 4-plat knot $\langle 1;2;1;1;1\rangle$, and this DNA knot has been observed (Figure 1c). We note that there is no information in either Theorem 1 or Theorem 2 about the parental tangle P. Since P appears in only one tangle equation (equation (i)), for each fixed rational tangle solution for O_b there are infinitely many rational tangle solutions to equation (i) for P [5]. Most biologists believe that $P=(0)$, and a biomathematical argument exists for this claim [11].

Research partially supported by NSF DMS-9024995 and NSF DMS-8720208 (Program in Mathematics and Molecular Biology).

Annotated Bibliography

Knot Theory

C. ADAMS, *The knot book: An elementary introduction to mathematical theory of knots*, W. H. Freeman, New York, 1994.

L. H. KAUFFMAN, *On knots*, Princeton Univ. Press, Princeton, NJ, 1987.

C. LIVINGSTON, *Knot theory*, Carus Mathematical Monograph, vol. 24, Mathematical Association of America, Washington, DC, 1994.

D. ROLFSEN, *Knots and links*, Publish or Perish, Inc., Berkeley, CA, 1976.

Each of these mathematics books has an easygoing, reader-friendly style and lots of pictures, a very important commodity when one is trying to understand knot theory.

Application of Geometry and Topology to Biology

W. R. BRAUER, F. H. C. CRICK, and J. H. WHITE, *Supercoiled DNA*, Sci. Amei. **243** (1980), 100–113.

This paper is a very nice introduction to the description and measurement of DNA supercoiling.

D. W. SUMNERS, *The role of knot theory in DNA research* (C. McCrory and T. Shifrin, eds.), Geometry and Topology, Marcel Debber, New York, 1987, pp. 297–318.

———, *Untangling DNA*, Math. Intelligencer **12** (1990), 71–80.

These papers are expository articles written for a mathematical audience. The first gives an overview of knot theory and DNA, and the second describes the tangle model.

D. W. SUMNERS (ed.), *New scientific applications of geometry and topology*, Proc. Sympos. Appl. Math., vol. 45, Amer. Math. Soc., Providence, RI, 1994.

This volume contains six expository papers outlining new applications of geometry and topology in molecular biology, chemistry, polymers, and physics. Three of the papers concern DNA applications.

D. M. WALBA, *Topological stereochemistry*, Tetrahedron **41** (1985), 3161–3212.

This paper is written by a chemist and describes topological ideas in synthetic chemistry and molecular biology. It is a good place to witness the translation of technical terms of science to mathematical concepts and vice versa.

J. C. WANG, *DNA topoisomerases*, Sci. Amer. **247** (1982), 94–109.

This paper describes how topoisomerases act to control DNA geometry and topology in various life processes in the cell.

S. A. WASSERMAN and N. R. COZZARELLI, *Biochemical topology: Applications to DNA recombination and replication*, Science **232** (1986), 951–960.

This paper describes the topological approach to enzymology protocol and reviews the results of various experiments on topoisomerases and recombinases.

J. H. WHITE, *An introduction to the geometry and topology of DNA structure*, Mathematical Methods for DNA Sequences (M.S. Waterman, ed.), CRC Press, Boca Raton, FL, 1989, pp. 225–253.

This is a very nice introductory mathematical treatment of linking number, twist, and writhe, with DNA applications.

References

1. G. BURDE and H. ZIESCHANG, *Knots*, deGruyter, New York, 1985.
2. J. H. CONWAY, *An enumeration of knots and links and some of their related properties*, Computational Problems in Abstract Algebra (Proc. Conf. Oxford, 1967), Pergamon Press, Oxford, 1970, pp. 329–358.
3. M. C. CULLER, C. M. GORDON, J. LUECKE, and P. B. SHALEN, *Dehn surgery on knots*, Ann. of Math. **125** (1987), 237–300.
4. F. B. DEAN, A. STASIAK, T. KOLLER, and N. R. COZZARELLI, *Duplex DNA knots produced by Escherichia coli topoisomerase* I, J. Biol. Chem. **260** (1985), 4795–4983.
5. C. ERNST and D. W. SUMNERS, *A calculus for rational tangles: Applications to DNA recombination*, Math. Proc. Cambridge Philos. Soc. **108** (1990), 489–515.
6. L. H. KAUFFMAN, *On knots*, Princeton Univ. Press, Princeton, NJ, 1987.
7. M. A. KRASNOW, A. STASIAK, S. J. SPENGLER, F. DEAN, T. KOLLER, and N. R. COZZARELLI, *Determination of the absolute handedness of knots and catenanes of DNA*, Nature **304** (1983), 559–560.
8. D. ROLFSEN, *Knots and links*, Publish or Perish, Inc., Berkeley, CA, 1990.
9. D. W. SUMNERS, *Untangling DNA*, Math. Intelligencer **12** (1990), 71–80.
10. ____ *Knot theory and DNA*, New Scientific Applications of Geometry and Topology (D. W. Sumners, ed.), Proc. Sympos. Appl. Math., vol. 45, Amer. Math. Soc., Providence RI, 1992.
11. D. W. SUMNERS, C. E. ERNST, N. R. COZZARELLI, and S. J. SPENGLER, *Analysis of the mechanism of DNA recombination*, Quart. Rev. of Biophys. (to appear).
12. J. C. WANG, *DNA topoisomerases*, Sci. Amer. **247**, (1982), 94–109.
13. ____, DNA topoisomerases, Ann. Rev. Biochem. **54** (1985), 665–697.
14. S. A. WASSERMAN, J. M. DUNGAN, and N. R. COZZARELLI, *Discovery of a predicted DNA knot substantiates a model for site-specific recombination*, Science **229** (1985), 171–174.
15. S. A. WASSERMAN and N. R. COZZARELLI, *Biochemical topology: Applications to DNA recombination and replication*, Science **232** (1986), 951–960.