communications in mathematical
match
and in computer chemistry

no. 34, October 1996

ISSN 0340-6253

MATCDY (34) 51-76 (1996)

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

DE WITT SUMNERS
FLORIDA STATE UNIVERSITY

A central problem in molecular biology is understanding the mechanism by which enzymes carry out chemical transformations. The problem is challenging because most experimental techniques provide only a static snapshot, not a moving picture, of the sequence of molecular events that take place inside the catalytic core of the enzyme. For one class of enzymes, however, mathematics provides a powerful tool to the molecular biologist. These enzymes are the ones that perform topological reactions necessary for the winding, unwinding, recombination, and transposition of DNA. Using topological results about knots and tangles, one can peer into the reaction center and infer the mechanisms of action.

INTRODUCTION

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. The three-dimensional shape in solution is of great biological significance but is very difficult to determine (Wang, 1982).

Experimental techniques such as X-ray crystallography and nuclear magnetic resonance provide ways to infer precise distances between atoms. However, these methods are not well suited to studying the dynamic mechanism by which enzymes act. Interestingly, topology can

---

* Reprinted with permission from *Calculating the Secrets of Life: Contributions of the Mathematical Science to Molecular Biology*. © 1995 by the National Academy of Sciences. Courtesy of the National Academy Press, Washington, D.C.

** Electron micrograph on the title page courtesy of N.R. Cozzarelli and A. Stasiak

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) (Wasserman and Cozzarelli, 1986; Sumners, 1987a). Once the enzyme structure and mechanism are understood in a controlled laboratory situation, this knowledge can be extrapolated to enzyme mechanism in vivo, that is, in a living cell.

Topology is a branch of mathematics related to geometry. It is often characterized as "rubber-sheet geometry," because topological equivalence of spaces allows stretching, shrinking, and twisting of an object in order to make it congruent to another object. Topology is the study of properties of objects (spaces) that are unchanged by allowable elastic deformations. When a given topological property differs for a pair of spaces, then one can be sure that one space cannot be transformed into the other by elastic deformation. Changes that can produce non-equivalent spaces include cutting the space apart and reassembling the parts to produce another space. It is precisely this topological breakage and reassembly of DNA that characterizes the mechanism of some life-sustaining cellular enzymes, enzymes that facilitate replication, transcription, and transposition. Chapter 6 describes aspects of the geometry and topology of DNA and points out various topological transformations that must be performed on DNA by enzymes in order to carry out the life cycle of the cell. In the present chapter, we describe how recent results in three-dimensional topology (Culler et al., 1987; Ernst and Sumners, 1990; Sumners, 1990, 1992) have proven to be of use in the description and quantization of the action of these life-sustaining enzymes on DNA.

THE TOPOLOGY OF DNA

The DNA of all organisms has a complex and fascinating topology. It 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 up 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 sequence 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 classical 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 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. Duplex DNA can exist in nature in closed circular form, where the rungs of the ladder lie on a twisted cylinder. Circular duplex DNA exists in the mitochondria of human cells, for example. Duplex

DNA in the cell nucleus is a linear molecule, one that is topologically con-strained by periodic attachment to a protein scaffold in order to achieve efficient packing.

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 (Cozzarelli, 1992). For information retrieval and cell viability, some geometric and topological features must be introduced into the DNA, and others quickly removed (Wang, 1982, 1985). 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 introduction of these life-sustaining changes in DNA geometry and topology, and after the process that these changes make possible is finished, the original DNA conformation must be restored. Some enzymes maintain the 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, a move performed by topoisomerases. 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. Geometry and topology provide ways of inferring the dynamic process of topological transformation carried out by an enzyme. This use of mathematics as an analytic tool for the indirect determination of enzyme mechanism 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 (Wasserman and Cozzarelli, 1986; Sumners, 1987b). 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 (catenanes). Trapping such a topological signature is impossible if one uses linear DNA substrate. 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 1a gives the schematics of the topological enzymology protocol; the black box represents the dynamic reaction in which the enzyme attaches to the DNA substrate, breaks it apart and reconnects as necessary, and then releases the DNA products. Typical results of this experimental protocol are the reaction products displayed in Figure 1b and c. Figure 1b shows the electron micrograph of a DNA (+) figure eight catenane (Krasnow et al., 1983), and Figure 1c shows a micrograph of the DNA knot 6_2^* (Wasserman et al., 1985). Both are products of processive Tn3 recombination and are explained in detail below.

The topological approach to enzymology poses an interesting challenge for mathematics: from the observed changes in DNA geometry and topology, how can one mathematically 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. Mathematics is required to extract mechanism information from

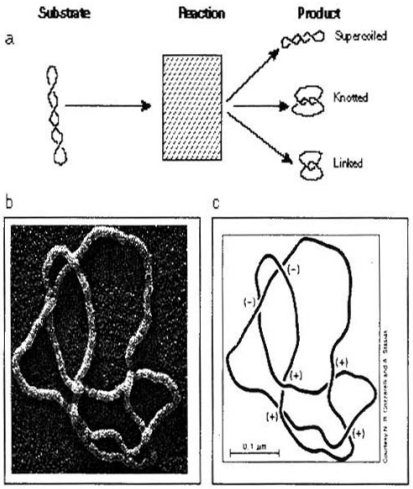

Figure 1 (a) Topological approach to enzymology. (b) DNA (+) figure eight catenane, (c) DNA knot 6_2^*. Figure 1b reprinted, with permission, from Krasnow et al. (1983). Copyright © 1983 by Macmillian Magazines Limited. Figure 1c reprinted, by permission, from Wasserman et al. (1985). Copyright © 1985 by the American Association for the Advancement of Science.

the topological structure of the reaction products. 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 (a move performed by transposase) (Sherratt et al., 1984) or by integrating a block of alien DNA into a host genome (a move performed by integrase). 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 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

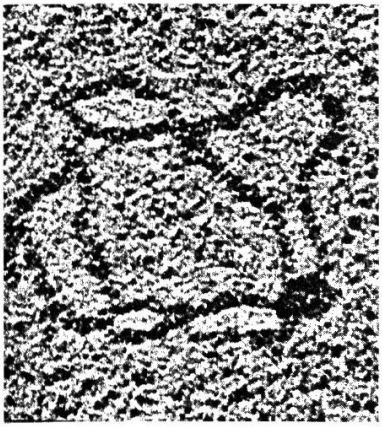

Figure 2 Tn3 synaptic complex. (Courtesy of N.R. Cozzarelli.)

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 (Benjamin and Cozzarelli, 1990; Heichman and Johnson, 1990; Pollock and Nash, 1983; Griffith and Nash, 1985; Kim and Landy, 1992). 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 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 the 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 (the base pair sequence for the site read left-to-right is different from the base pair sequence read right-to-left), and hence induces a local orientation (arrow) on the substrate DNA circle. If these induced orien-

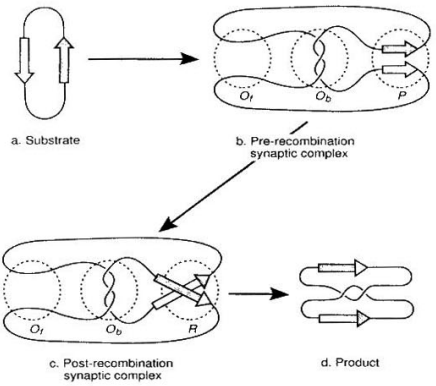

Figure 3 A single recombination event: direct repeats.

tations from a pair of sites on a single 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 catenane) (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 a 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 experimental strategy in the topological approach to enzymology is to observe the enzyme-caused changes in the geometry and topology of the DNA and to deduce the enzyme mechanism from these changes, as in Figure 1a. The geometry and topology of the 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 (usually the unknot, that is, a curve without knots). The amount of supercoiling of the substrate molecules (the supercoiling density) 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. DNA molecules are naturally.

negatively charged, with the amount of negative charge proportional to the molecular weight. A gel is a resistive medium through which the DNA molecules can be forced to migrate under the influence of an electric field. The DNA sample is placed at the top of a gel column, and similar molecules migrate through the gel with similar velocities, forming discrete DNA bands in the gel when the electric field is turned off. Normally, gel electrophoresis discrimi-

nates 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 (knot and link type) of the DNA molecules. For example, in unknotted DNA, gel electrophoresis discriminates on the basis of number of supercoils and can detect a difference of one in the number of supercoils. In gel electrophoresis of knotted and linked DNA, one must nick (break one of the two backbone strands of) the reaction products prior to electrophoresis in order to relax the supercoils in the DNA knots and links, because supercoiling confounds the gel migration of knotted and linked DNA. For nicked DNA knots and links, 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 (Dean et al., 1985); the higher the crossing number, the greater the gel mobility. After the gel is run, the gel bands are excised, and the DNA molecules are removed from the gel and coated with RecA protein. It is this new observation technique (RecA-enhanced electron microscopy) (Krasnow et al., 1983) that makes possible the detailed knot-theoretic analysis of reaction products. RecA is an *E. coli* protein that binds to DNA and mediates general recombination in *E. coli*. Naked (uncoated) duplex DNA is approximately 20 angstroms in diameter, and RecA-coated DNA is approximately 100 angstroms in diameter. The process of RecA coating fattens, stiffens, and stretches (untwists) the DNA. This fattening and stiffening facilitates the unambiguous determination of crossings (nodes) in an electron micrograph of a DNA knot or link and reduces the number of extraneous crossings. After RecA coating, the DNA is shadowed with platinum for viewing under the electron microscope. Electron micrographs of the reaction products (Figure 1b and c) are made, and frequency distributions of knot types of the products are prepared. This new precision in the determination of the topology of the reaction product spectrum opens the door for the building of detailed topological models for enzyme action.

TOPOLOGICAL TOOLS FOR DNA ANALYSIS

In this section, we will describe the parts of knot theory and tangle calculus of biological relevance. We give intuitive definitions that appeal to geometric imagination. For a rigorous mathematical treatment we refer the reader to Burde and Zieschang (1985), Kauffman (1987), and Rolfsen (1990) for knot theory and Ernst and Sumners (1990) for tangle calculus.

Knot theory is the study of the entanglement of flexible circles in 3-space. The equivalence relation between topological spaces is that of homeomorphism. A homeomorphism $h: X \to Y$ between topological spaces is a function that is one-to-one and onto, and both h and h^{-1} are continuous. An embedding of X in Y is a function $f: X \to Y$ such that f is a homeomorphism from X onto $f(X) \subset Y$. An embedding of X in Y is the placement of a copy of X into the ambient space Y. We will usually take Euclidean 3-space $\mathbf{R}^3$ (xyz-space) as our ambient space. A knot K is an embedding of a single circle in $\mathbf{R}^3$; a link L is an embedding of two or more circles in $\mathbf{R}^3$. For a link, each of the circles of L is called a component of L. In chemistry and biology a nontrivial link is called a catenane, from the Latin *cataena* for "chain," since the components of a catenane are topologically entangled with each other like the links in a chain. In this excursion, we will restrict attention to dimers, that is, links of two components, because dimers are the only links that turn up in topological enzymology experiments. We regard two knots (links) to be equivalent if it is possible to continuously and elastically deform one embedding (without breaking strands or passing strands one through another) until it can be superimposed upon the other. More precisely, if K_1 and K_2 denote two knots (links) in $\mathbf{R}^3$, they are equivalent (written $K_1 = K_2$) if and only if there is a homeomorphism of pairs $h: (\mathbf{R}^3, K_1) \to (\mathbf{R}^3, K_2)$ that preserves orientation on the ambient space $\mathbf{R}^3$. We take our ambient space $\mathbf{R}^3$ to have a fixed (right-handed) orientation, where the right-hand thumb corresponds to the X-axis, the right-hand index finger corresponds to the Y-axis, and the right-hand middle finger corresponds to the Z-axis. $\mathbf{R}^3$ comes locally equipped with this right-handed orientation at all points. A homeomorphism from $\mathbf{R}^3$ to $\mathbf{R}^3$ is orientation-preserving if the local right-handed frame at each point of the domain maps to a local right-handed frame in the range. Reflection in a hyperplane (such as reflection in the xy-plane by $f: (x,y,z) \to (x,y,-z)$) reverses the orientation of $\mathbf{R}^3$. We might also require that the circular subspace K come equipped with an orientation (usually indicated by an arrow). If so, we say that our knot or link K is oriented; if not, we say that it is unoriented. Unless otherwise specified, all of our knots will be unoriented. 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 1-parameter family of homeomorphisms $\{H_t\}_{t=0}^{1}$ of $\mathbf{R}^3$ 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. By a small rigid rotation of the knot (link) in 3-space, it can be arranged that no more than two strings cross at any point in the diagram. For short, crossing points in a diagram are called crossings. In the figures in

this chapter, at each crossing in a diagram, the undercrossing string is depicted with a break in it, so that the three-dimensional knot (link) type can be uniquely re-created from a two-dimensional diagram. Figure 4a-e shows standard diagrams (Rolfsen, 1990) 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. The mirror image of a knot (link) is the configuration obtained by reflecting the configuration in a plane in $\mathbf{R}^3$. Starting with a diagram for a knot (link), one can obtain a diagram for the mirror image by reversing each crossing; the underpass becomes the overpass and vice versa (compare Figure 4d and 4e). 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. For example, the Hopf link (Figure 4a) and the figure eight knot (Figure 4b) are achiral, and the (+) Whitehead link (Figure 4c) and the knot 6_2^* (Figure 4d) and its mirror image 6_2 (Figure 4e) are chiral. Moreover, all the knots and links in Figure 4 are prime, that is, they cannot be formed by the process of tying first one knot in a string and then another.

By moving the knot around in space and then projecting it, it is clear that every given knot (link) type admits infinitely many "different" diagrams, and so the task of recognizing that two completely different diagrams represent the same knot type can be exceedingly difficult. In order to make this job a bit easier, one usually seeks diagrams for the knot type with a minimal number of crossings. This minimal number is called the crossing number of the knot (link) type. The projections in Figure 4 are minimal. Crossing number is a topological invariant of knot type. A topological invariant is a number, algebraic group, polynomial, and so on that can be unambiguously attached to a knot (link) type. Most invariants can be algorithmically computed from diagrams (Burde and Zieschang, 1985; Crowell and Fox, 1977; Lickorish, 1988; Kauffman, 1987). If any invariant differs for two knots (links), then the two knots (links) are of different types. If all known invariants are identical, the only conclusion that can be reached is that all known invariants fail to distinguish the candidates. One must then either devise a new invariant that distinguishes the two or prove that they are of the same type by construction of the homeomorphism that transforms one to the other (often by direct geometric manipulation of the diagram or by manipulation of string models). Nevertheless, it is possible to devise invariants (algebraic classification schemes) that uniquely classify certain homologous subfamilies of knots and links, for example, torus knots, two-bridge knots (4-plats), and so on. The algebraic classification schemes for these homologous subfamilies can be used to describe and compute enzyme mechanisms in the topological enzymology protocol.

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.

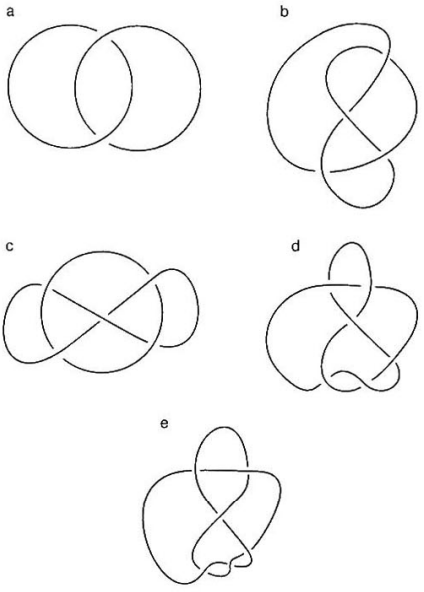

Figure 4 (a) Hopf link, (b) figure eight knot, (c) (+) Whitehead link, (d) 6_2^*, and (e) 6_2 (mirror image of 6_2^*).

All of the knots and links of 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, numbered 1 through 4 from top to bottom. The standard pattern of plectonemic interwinding for a 4-plat is encoded by an odd-length classifying vector with positive integer entries $<c_1,c_2,\ldots,c_{2k+1}>$, as shown in Figure 5. Beginning from the left, strings in positions 2 and 3 undergo c_1 left-handed plectonemic interwinds (half-twists), then strings in positions 1 and 2 undergo c_2 right-handed plectonemic interwinds, then strings in positions 2 and 3 undergo c_3 left-handed plectonemic interwinds, and this process continues until at the right the strings in positions 2 and 3 undergo c_{2k+1} left-handed plectonemic interwinds. In the standard diagram for a 4-plat, the string in position 4 is not involved in any crossing. The vector representation for the standard diagram of a 4-plat is unique up to reversal of the symbol. That is, the vector $<c_{2k+1},c_{2k},\ldots,c_1>$ represents the same type as the vector $<c_1,c_2,\ldots,c_{2k+1}>$, because turning the 4-plat 180° about the vertical axis reverses the pattern of supercoiling. The standard 4-plat diagram is alternating; that is, as one traverses any strand in the diagram, one alternately encounters over- and undercrossings. Also, standard 4-plat diagrams (with the exception of the unknot $<1>$) are minimal (Ernst and Sumners, 1987).

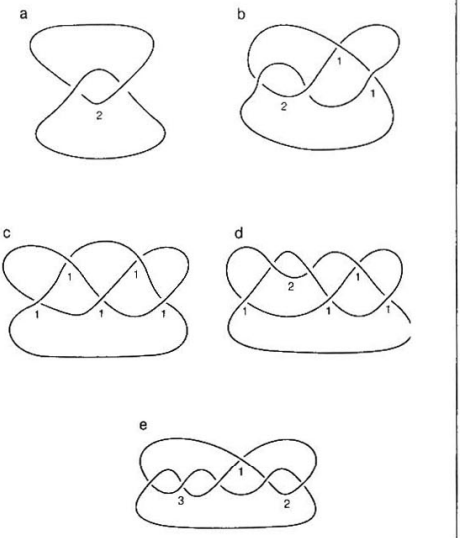

Figure 5. Standard 4-plats. (a) $< 2 >$ Hopf link, (b) $< 2,1,1 >$ figure eight knot, (c) $< 1,1,1,1,1 >$ (+) figure eight catenane, (d) $< 1,2,1,1,1 > 6_2^*$, and (e) $< 3,1,2 > 6_2$.

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. Tangles were introduced into knot theory by J.H. Conway (1970) in a seminal paper involving construction of enumeration schemes for knots and links. The unit 3-ball $\mathbf{B}^3$ in $\mathbf{R}^3$ is the set of all vectors of length ≤ 1. The boundary 2-sphere $\mathbf{S}^2 = \partial \mathbf{B}^3$ is the set of all vectors of length 1. The equator of this 3-ball is the intersection of the boundary $\mathbf{S}^2$ with the *xy*-plane; the equatorial disk is the intersection of $\mathbf{B}^3$ with the *xy*-plane. 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. Tangle theory

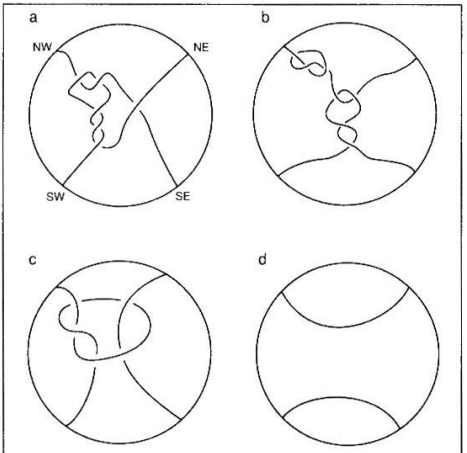

Figure 6. Tangles. (a) Rational, (b) locally knotted, (c) prime, and (d) trivial.

is knot theory done inside a 3-ball with the ends of the strings firmly glued down. 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 6a, 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 surface ($\mathbf{S}^2$) of the 3-ball, then the tangle of Figure 6a can be trans-formed into the trivial tangle of Figure 6d. This can be accomplished by rotating (on $\mathbf{S}^2$) the {NE,SE} endpoints one left half-turn (180°) about each other, then rotating the {SW,SE} endpoints three right half-turns about each other, and finally rotating the {NE,SE} endpoints two left half-turns about each other. 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 $\mathbf{S}^2$. The family of tangles that can be converted to the trivial tangle by moving the endpoints of the strings on $\mathbf{S}^2$ 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 $\mathbf{B}^3$ 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

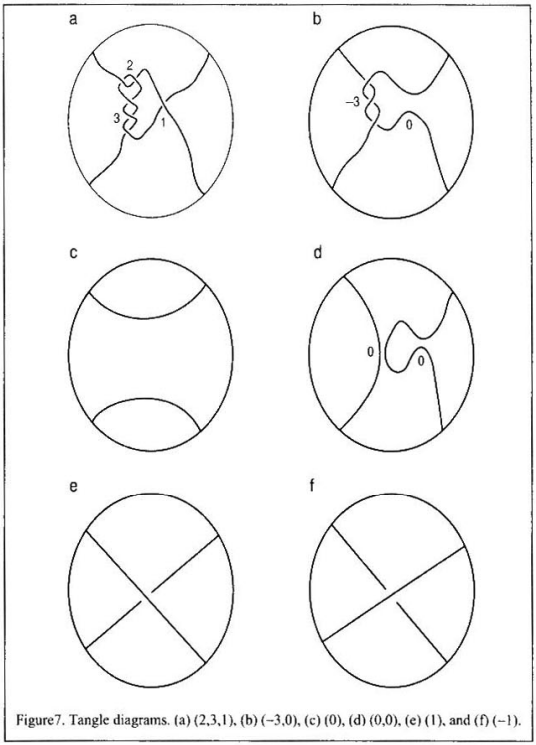

Figure7. Tangle diagrams. (a) (2,3,1), (b) (–3,0), (c) (0), (d) (0,0), (e) (1), and (f) (–1).

equivalent to our unit defining ball $\mathbf{B}^3$. 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 will admit 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.

A classification scheme for rational tangles 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

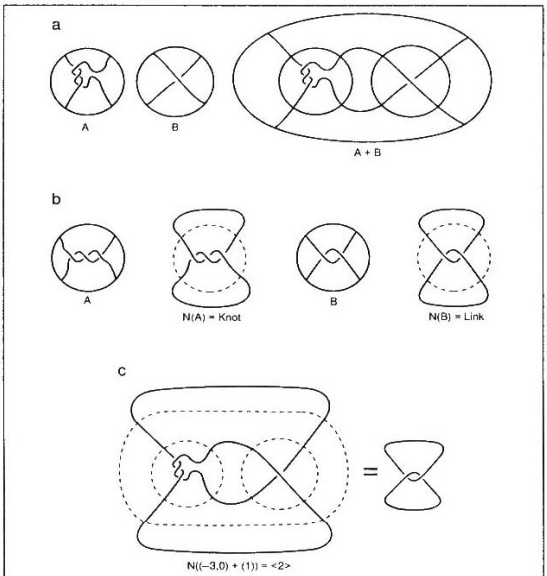

Figure 8. Tangle operations. (a) Tangle addition, (b) tangle closure, and (c) $N((-3,0)+(1)) = <2>$

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 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-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 (extended) rational numbers are equal (Conway, 1970), which is the reason for calling them "rational" tangles.

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 addition 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 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 $\mathbf{B}^3$ 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 $N((-3,0) + (1)) = <2>$, shown in Figure 8c. In general, if A and B are any two rational tangles, then $N(A+B)$ is a 4-plat. Given these constructions, rational tangles are summands for 4-plats.

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 (Ernst and Sumners, 1990; Sumners, 1990, 1992; Sumners et al., 1994). 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.

The aim of our mathematical model is, given the observed changes in geometry and topology of the DNA, to compute the topology of the entire synaptic complex, both before and after enzyme action. 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 (Ernst and Sumners, 1990; Sumners, 1992; Sumners et al., 1994). Most of these assumptions are implicit in the existing analyses of the results of enzyme experiments on circular DNA (Cozzarelli et al., 1984; Stark et al., 1989; Spengler et al., 1985; Wasser-man and Cozzarelli, 1986; Wasserman et al., 1985; Kanaar et al., 1990; White et al., 1987; Kanaar et al., 1988; Abremski et al., 1986; Droge and Cozzarelli, 1986; Spengler et al., 1984).

We make the following biological assumption:

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 the 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 $O_b + P$ 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:

$$\text{pre-recombination synaptosome} = O_b + P$$

$$\text{post-recombination synaptosome} = O_b + R\,.$$

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 catenane 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:

$$\text{substrate equation:} \quad N(O_f + O_b + P) = \text{substrate}$$
$$\text{product equation:} \quad N(O_f + O_b + R) = \text{product}.$$

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:

$$\text{substrate equation:} \quad N((0) + (-3,0) + (0)) = <1>$$
$$\text{product equation:} \quad N((0) + (-3,0) + (1)) = <2>.$$

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 (Wasserman et al., 1985). 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 $<2>$ (the Hopf link, Figures 4a and 5a) (Wasserman and Cozzarelli, 1985). 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 20 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 $<2,1,1>$ (the figure eight knot, Figures 4b and 5b); three successive rounds of processive recombination produce the DNA 4-plat $<1,1,1,1,1>$ (the Whitehead link, Figures 4c and 5c), whose electron micrograph appears in Figure 1b; four successive rounds of recombination produce the DNA 4-plat $<1,2,1,1,1>$ (the knot 6_2^*, Figures 4d and 5d), whose electron micrograph appears in Figure 1c. The discovery of the DNA knot $<1,2,1,1,1>$ substantiated a model for Tn3 resolvase mechanism (Wasserman et al., 1985).

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: $N(O_f + O_b + rR) = r$th 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 (Sumners, 1990, 1992; Ernst and Sumners, 1990) 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 as extraneous solutions. The following theorems can be viewed as a mathematical proof of resolvase synaptic complex structure: the model proposed in (Wasserman et al., 1985) 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.

The process of obtaining electron micrographs of RecA-enhanced DNA knots and catenanes is technically difficult and requires a relatively large amount of product due to the extensive work-up required for RecA coating and microscopy. Gel electrophoresis is not only technically much easier to do, but it detects vanishingly small amounts of DNA product. For these reasons, biologists prefer to use gel electrophoresis as the assay from which experimental conclusions are to be drawn. For relaxed DNA knots and links, the gel determines the crossing number of the (relaxed) products, and comparison to gel ladders for known knot and catenane structures can be used to obtain more information than crossing number alone. As an aid to the analysis of topological enzymology experiments, a table of possible (and biologically reasonable!) tangle mechanisms has been prepared (Sumners et al., 1994) for each possible sequence of crossing numbers of reaction products that can be read from the gel. This tangle table should make the mathematical analysis of topological enzymology experiments easier to do.

We now come to the rigorous mathematical proof of Tn3 mechanism. The proofs of the following two theorems can be skipped without detriment to the continuity of the exposition.

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

(i) $N(O_b + P) = <1>$ *(the unknot)*

(ii) $N(O_b + R) = <2>$ *(the Hopf link)*

(iii) $N(O_b + R + R) = <2,1,1>$ *(the figure 8 knot).*

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

Proof: In this proof we use the following notation: $\mathbf{R}^n$ denotes Euclidean n-space, $\mathbf{B}^n$ denotes the unit ball in $\mathbf{R}^n$ (the set of all vectors in $\mathbf{R}^n$ of length ≤ 1), and $\mathbf{S}^{n-1}$ denotes the boundary of $\mathbf{B}^n$ (the set of all vectors in $\mathbf{R}^n$ of length 1). The first (and mathematically most interesting) step in the proof of this theorem is to argue that solutions $\{O_b, R\}$ must be rational tangles. Now O_b, R, and $(O_b + R)$ are locally unknotted, because $N(O_b + R)$ is the Hopf link, which has two unknotted components. Any local knot in a tangle summand would persist in the Hopf link. Likewise, P is locally unknotted, because $N(O_b + P)$ is the unknot. Let A' denote the 2-fold branched cyclic cover of the tangle A; then $\partial A' = \mathbf{S}^1 \times \mathbf{S}^1$. If A is a prime tangle, then the inclusion homeomorphism injects $\pi_1(\partial A') = \mathbf{Z} \oplus \mathbf{Z}$ into $\pi_1(A')$ (Lickorish, 1981). If both A and B are prime tangles, and $N(A+B)'$ denotes the 2-fold branched cyclic cover, then $\pi_1(N(A+B)')$ contains a subgroup isomorphic to $\mathbf{Z} \oplus \mathbf{Z}$. If K is any 4-plat, then $\pi_1(K')$ is a cyclic group, since K' is a lens space (Burde and Zieschang, 1985). Since no cyclic group contains $\mathbf{Z} \oplus \mathbf{Z}$, no 4-plat has two prime tangle summands. This means that if A and B are locally unknotted tangles, and $N(A+B)$ is a 4-plat, then *at least one* of A and B must be a rational tangle. From equation (ii) above, we conclude that at least one of $\{O_b, R\}$ is rational. Suppose that O_b is rational and that R is prime. Given that $N((O_b + R) + R)$ is a knot, one can argue (Lickorish, 1981) that $O_b + R$ is also a prime tangle. From equation (iii), we then have that the 4-plat $<2,1,1>$ admits two prime tangle summands, which is impossible. We therefore conclude that R must be a rational tangle.

The next step is to argue that O_b is a rational tangle. Suppose that O_b is a prime tangle. Then P must be a rational tangle, because $N(O_b + P)$ is the unknot (equation (i)). Passing to 2-fold branched cyclic covers, we have that $N(O_b + P)' = \mathbf{S}^3$, and P' is homeomorphic to $\mathbf{S}^1 \times \mathbf{B}^2$ (since P is rational), so O_b' is a bounded knot complement in $\mathbf{S}^3$. We know that R is a rational tangle and can argue that equation (iii) implies that $(R+R)$ is likewise rational. Again passing to the 2-fold branched cyclic covers of equations (ii) and (iii), we obtain the equations $N(O_b + R)' =$ the lens space $L(2,1)$ and $N(O_b + (R+R))' =$ the lens space $L(5,3)$. Since R' and $(R+R)'$ are each homeomorphic to a solid torus $\mathbf{S}^1 \times \mathbf{B}^2$, this means that there are two attachments of a solid torus to O_b' along $\partial O_b' = \mathbf{S}^1 \times \mathbf{S}^1$, yielding the lens spaces $L(2,1)$ and $L(5,3)$. The process of adding on a solid torus along its boundary is called Dehn surgery, and the Cyclic Surgery Theorem (Culler et al., 1987) now applies to this situation to imply that, since the orders of the cyclic fundamental groups of the lens spaces differ by more than one, the only way this can happen is for O_b' to be a Seifert fiber space and hence a torus knot complement. Fortunately, the results of Dehn surgery on torus knot complements are well understood, and one can argue that in fact O_b' must be a complement of the unknot (a solid torus) (Ernst and Sumners, 1990), which means that O_b is a rational tangle.

The proof now amounts to computing the rational solutions to equations (ii) and (iii), exploiting the classifying schemes for rational tangles and 4-plats. In Ernst and Sumners (1990), a "calculus for rational tangles" was developed to perform such calculations. One can use this calculus of classifying vectors to solve equations (ii) and (iii), obtaining the four solution pairs $\{O_b, R\} = \{(-3,0),(1)\}$, $\{(3,0),(-1)\}$, $\{(-2,-3,-1),(1)\}$, and $\{(2,3,1),(-1)\}$. Because each of the unoriented 4-plat products in equations (ii) and (iii) is achiral, given any solution set $\{O_b, R\}$ to equations (ii) and (iii), its mirror image $\{-O_b, -R\}$ must also be a solution. So the mathematical situation, given equations (i) through (iii), is that we have two pairs of mirror image solution sets for $\{O_b, R\}$.

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*:

(i) $N(O_b + P) = <1>$ *(the unknot)*

(ii) $N(O_b + R) = <2>$ *(the Hopf link)*

(iii) $N(O_b + R + R) = <2,1,1>$ *(the figure 8 knot).*

(iv) $N(O_b + R + R + R) = <1,1,1,1,1>$ *(the* (+) *Whitehead link).*

Then $O_b = -3,0$, $R = (1)$, *and* $N(O_b + R + R + R + R) = <1,2,1,1,1>$.

Proof: The unoriented (+) Whitehead link is chiral and $\{O_b R\} = \{(-3,0),(1)\}$ is the unique solution to equations (i) and (iv).

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 $<1,2,1,1,1>$, 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 (Ernst and Sumners, 1990). Most biologists believe that $P = (0)$, and a biomathematical argument exists for this claim (Sumners et al., 1994).

SOME UNSOLVED PROBLEMS

1. *How does TOPO II recognize knots? E. coli* contains circular duplex DNA molecules. In wild-type *E. coli*, no knotting has been observed for these molecules. However, in a mutant strain of *E. coli* where the production of Topoisomerase II (the enzyme that performs strand passage via an enzyme-bridged transient double-stranded break in the DNA) can be blocked by heat shock, a small fraction (about 7 percent) of knotted DNA has been observed (Shishido et al., 1987). All observed knots have the gel mobility of trefoil knots. The observed knots are presumably the by-products of other cellular processes (such as recombination). This experiment shows that TOPO II is able to detect DNA knots and kill them in wild-type *E. coli*. How does the enzyme (which can act only locally) detect the global topology of a DNA knot and then make just the right combination of passages to kill the knot? It must be the energy minimization of the DNA itself that detects the knotting. The enzyme has only to detect when two DNA strands are being pushed together in space by the DNA itself in an effort to attain a lower energy state, whence the enzyme can operate, allowing one DNA strand to pass through another to reach a lower-energy configuration. If one ties a knot in a short stiff rubber tube (an elastic tube) and seals up the ends to form a knotted circle, the tube will touch itself, trying to pass through itself to relieve strain and minimize energy. For circular elastica in $\mathbf{R}^3$, minimization of the bending energy functional occurs when the elasticum is the round planar unknot (Langer and Singer, 1984, 1985). This means that a knotted elasticum has at least one point of self-contact and that the elasticum is pushing at that point of self-contact to get through to a lower energy state. Does this elasticum model adequately explain the ability of Topoisomerase II to detect and selectively kill DNA knots in vivo?

2. *What is the topology of the kDNA network?* The kinetoplast DNA (kDNA) of the parasite *trypanosome* forms a link of some 5,000 to 10,000 unknotted DNA circles—the DNA equivalent of chain mail (Marini et al., 1980; Englund et al., 1982; Rauch et al., 1994). Work is ongoing (Rauch et al., 1994) in which the topological structure of kDNA is being studied by means of partial digest of the network, electrophoresis, and electron microscopy of the characteristic fragments, in which the large kDNA link is being randomly broken up into small sublinks, and the frequency of occurrence of these sublink units is being used (statistically) to reconstruct the large link itself. The kDNA network consists of small minicircles and a few large maxicircles. The minicircles are known to be unknotted, and it is known that neighbors link in the fashion of the Hopf link (Figure 4a) (like the links in a chain). Moreover, it is believed that the kDNA network has a fundamental region that is repeated in space to generate the entire structure. This gives rise to a knot theory problem: classify the links that allow a diagram in which each component has no self-crossings (and hence is unknotted) and in which each component links another component simply (like the links in a chain) or not at all, and in which the linking structure is periodic in space. The spatial periodicity amounts to drawing the link diagram on a torus (or some other compact, orientable 2-manifold), from whence the entire diagram is reproduced by taking the universal cover. The algebraic classification of such "chain mail links" should be interesting and obtainable with off-the-shelf topological invariants. Another topological problem has arisen in this biological system. A trefoil knotted minicircle has been observed as an intermediate to the replication process on the kDNA network (Ryan et al., 1988). What is the mechanism that

produces this knotted minicircle? Does the topology of the network naturally generate knots as replication intermediates?

3. *Why is the figure eight knot faster than the trefoil knot?* The phenomenon of gel mobility of relaxed knotted duplex DNA circles (Dean et al., 1985) has no adequate theoretical explanation. The gel velocity of relaxed DNA knots is determined by crossing number; the larger the crossing number, the faster the migration. Perhaps this is because among knots of the same length with small crossing numbers, the average value of the radius of gyration (a measure of the average size) correlates strongly with crossing number. It is very curious that the crossing number, clearly an artifact of planar diagrammatic representation of knots, would have anything at all to do with the three-dimensional average knot conformation. What is the relationship (if any) between radius of gyration of DNA circles of fixed molecular weight and fixed knot type, crossing number, and the gel mobility of these knotted DNA circles?

ANNOTATED BIBLIOGRAPHY

Knot Theory

Adams, C., 1994, *The Knot Book: An Elementary Introduction to Mathematical Theory of Knots,* New York: W.H. Freeman.

Kauffman, L.H., 1987, *On Knots,* Princeton, N.J.: Princeton University Press.

Livingston, C., 1994, *Knot Theory, Carus Mathematical Monograph,* Vol. 24, Washington, D.C.: Mathematical Association of America.

Rolfsen, D., 1990, *Knots and Links,* Berkeley, Calif.: Publish or Perish, Inc.

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

Application of Geometry and Topology to Biology

Bauer, W.R., F.H.C. Crick, and J.H. White, 1980, "Supercoiled DNA," *Scientific American* **243**, 100-113.

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

Sumners, D.W., 1987, "The role of knot theory in DNA research," pp. 297-318 in *Geometry and Topology,* C. McCrory and T. Shifrin (eds.), New York: Marcel Dekker.

Sumners, D.W., 1990, "Untangling DNA," *The Mathematical Intelligencer* **12**, 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.

Sumners, D.W. (ed.), 1994, *New Scientific Applications of Geometry and Topology, Proceedings of Symposia in Applied Mathematics,* Vol. 45, Providence, R.I.: American Mathematical Society.

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.

Walba, D.M., 1985, "Topological stereochemistry," *Tetrahedron* **41**, 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.

Wang, J.C., 1982, "DNA topoisomerases," *Scientific American* **247**, 94-109.

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

Wasserman, S.A., and N.R. Cozzarelli, 1986, "Biochemical topology: Applications to DNA recombination and replication," *Science* **232**, 951-960.

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

White, J.H., 1989, "An introduction to the geometry and topology of DNA structure," pp. 225-253 in *Mathematical Methods for DNA Sequences,* M.S. Waterman (ed.), Boca Raton, Fla.: CRC Press.

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

REFERENCES

Abremski, K., B. Frommer, and R.H. Hoess, 1986, "Linking-number changes in the DNA substrate during Cre-mediated loxP site-specific recombination," *Journal of Molecular Biology* **192**, 17-26.

Benjamin, H.W., and N.R. Cozzarelli, 1990, "Geometric arrangements of Tn3 resolvase sites," *J. Biol. Chem.* **265**, 6441-6447.

Burde, G., and H. Zieschang, 1985, *Knots,* New York: W. De Gruyter.

Conway, J.H., 1970, "An enumeration of knots and links and some of their related properties," pp. 329-358 in *Computational Problems in Abstract Algebra; Proceedings of a Conference at Oxford 1967,* Oxford: Pergamon Press.

Cozzarelli, N.R., 1992, "The biological roles of DNA topology," in *New Scientific Applications of Geometry and Topology, Proceedings of Symposia in Applied Mathematics,* D.W. Sumners (ed.), Providence, R.I.: American Mathematical Society.

Cozzarelli, N.R., M.A. Krasnow, S.P. Gerrard, and J.H. White, 1984, "A topological treatment of recombination and topoisomerases," *Cold Spring Harbor Symp. Quant. Biol.* **49**, 383-400.

Crowell, R.H., and R.H. Fox, 1977, *Introduction to Knot Theory. Graduate Texts in Mathematics* **57**, New York: Springer-Verlag.

Culler, M.C., C.M. Gordon, J. Luecke, and P.B. Shalen, 1987, "Dehn surgery on knots," *Ann. Math.* **125**, 237-300.

Dean, F.B., A. Stasiak, T. Koller, and N.R. Cozzarelli, 1985, "Duplex DNA knots produced by *Escherichia coli* topoisomerase I," *J. Biol. Chem.* **260**, 4975-4983.

Droge, P., and N.R. Cozzarelli, 1989, "Recombination of knotted substrates by Tn3 resolvase," *Proceedings of the National Academy of Sciences USA* **86**, 6062-6066.

Englund, P.T., S.L. Hajduk, and J.C. Marini, 1982, "The molecular biology of trypanosomes," *Annu. Rev. Biochem.* **51**, 695-726.

Ernst, C., and D.W. Sumners, 1987, "The growth of the number of prime knots," *Math. Proc. Cambridge Philos. Soc.* **102**, 303-315.

Ernst, C., and D.W. Sumners, 1990, "A calculus for rational tangles: Applications to DNA recombination," *Math. Proc. Cambridge Philos. Soc.* **108**, 489-515.

Griffith, J.D., and H.A. Nash, 1985, "Genetic rearrangement of DNA induces knots with a unique topology: Implications for the mechanism of synapsis and crossing-over," *Proceedings of the National Academy of Sciences USA* **82**, 3124-3128.

Heichman, K.A., and R.C. Johnson, 1990, "The Hin invertasome: Protein-mediated joining of distant recombination sites at the enhancer," *Science* **249**, 511-517.

Kanaar, R., P. van de Putte, and N.R. Cozzarelli, 1988, "Gin-mediated DNA inversion: Product structure and the mechanism of strand exchange," *Proceedings of the National Academy of Sciences USA* **85**, 752-756.

Kanaar, R., A. Klippel, E. Shekhtman, J.M. Dungan, R. Kahmann, and N.R. Cozzarelli, 1990, "Processive recombination by the phage Mu gin system: Implications for mechanisms of DNA exchange, DNA site alignment, and enhancer action," *Cell* **62**, 353-366.

Kauffman, L.H., 1987, *On Knots,* Princeton, N.J.: Princeton University Press.

Kim, S., and A. Landy, 1992, "Lambda Int protein bridges between higher order complexes at two distant chromosomal loci *att*L and *att*R," *Science* **256**, 198-203.

Krasnow, M.A., A. Stasiak, S.J. Spengler, F. Dean, T. Koller, and N.R. Cozzarelli, 1983, "Determination of the absolute handedness of knots and catenanes of DNA," *Nature* **304**, 559-560.

Langer, J., and D.A. Singer, 1984, "Knotted elastic curves in R^3," *J. London Math. Soc.* **30**, 512-520.

Langer, J., and D.A. Singer, 1985, "Curve straightening and a minimax argument for closed elastic curves," *Topology* **24**, 75-88.

Lickorish, W.B.R., 1981, "Prime knots and tangles," *Trans. Am. Math. Soc.* **267**, 321-332.

Lickorish, W.B.R., 1988, "Polynomials for links," *Bull. London Math. Soc.* **20**, 558-588.

Marini, J.C., K.G. Miller, and P.T. Englund, 1980, "Decatenation of kinetoplast DNA by topoisomerases," *J. Biol. Chem.* **255**, 4976-4979.

Pollock, T.J., and H.A. Nash, 1983, "Knotting of DNA caused by genetic rearrangement: Evidence for a nucleosome-like structure in site-specific recombination of bacteriophage lambda," *Journal of Molecular Biology* **170**, 1-18.

Rauch, C.A., P.T. Englund, S.J. Spengler, N.R. Cozzarelli, and J.H. White, 1994, "Kinetoplast DNA: Structure and replication," (in preparation).

Rolfsen, D., 1990, *Knots and Links,* Berkeley, Calif.: Publish or Perish, Inc.

Ryan, K.A., T.A. Shapiro, C.A. Rauch, J.D. Griffith, and P.T. Englund, 1988, "A knotted free minicircle in kinetoplast DNA," *Proceedings of the National Academy of Sciences USA* **85**, 5844-5848.

Sherratt, D., P. Dyson, M. Boocock, L. Brown, D. Summers, G. Stewart, and P. Chan, 1984, "Site-specific recombination in transposition and plasmid stability," *Cold Spring Harbor Symp. Quant. Biol.* **49**, 227-233.

Shishido, K., N. Komiyama, and S. Ikawa, 1987, "Increased production of a knotted form of plasmid pBR322 DNA in *Escherichia coli* DNA topoisomerase mutants," *Journal of Molecular Biology* **195**, 215-218.

Spengler, S.J., A. Stasiak, and N.R. Cozzarelli, 1984, "Quantitative analysis of the contributions of enzyme and DNA to the structure of lambda integrative recombinants," *Cold Spring Harbor Symp. Quant. Biol.* **49**, 745-749.

Spengler, S.J., A. Stasiak, and N.R. Cozzarelli, 1985, "The stereostructure of knots and catenanes produced by phage lambda integrative recombination: Implications for mechanism and DNA structure," *Cell* **42**, 325-334.

Stark, W.M., D.J. Sherratt, and M.R. Boocock, 1989, "Site-specific recombination by Tn3 resolvase: Topological changes in the forward and reverse reactions," *Cell* **58**, 779-790.

Sumners, D.W., 1987a, "Knots, macromolecules and chemical dynamics," pp. 297-318 in *Graph Theory and Topology in Chemistry*, King and Rouvray (eds.), New York: Elsevier.

Sumners, D.W., 1987b, "The role of knot theory in DNA research," pp. 297-318 in *Geometry and Topology*, C. McCrory and T. Shifrin (eds.), New York: Marcel Dekker.

Sumners, D.W., 1990, "Untangling DNA," *The Mathematical Intelligencer* **12**, 71-80.

Sumners, D.W., 1992, "Knot theory and DNA," in *New Scientific Applications of Geometry and Topology, Proceedings of Symposia in Applied Mathematics*, Vol. 45, D.W. Sumners (ed.), Providence, R.I.: American Mathematical Society.

Sumners, D.W., C.E. Ernst, N.R. Cozzarelli, and S.J. Spengler, 1994, "The tangle model for enzyme mechanism," (in preparation).

Wang, J.C., 1982, "DNA topoisomerases," *Scientific American* **247**, 94-109.

Wang, J.C., 1985, "DNA topoisomerases," *Annu. Rev. Biochem.* **54**, 665-697.

Wasserman, S.A., and N.R. Cozzarelli, 1985, "Determination of the stereostructure of the product of Tn3 resolvase by a general method," *Proceedings of the National Academy of Sciences USA* **82**, 1079-1083.

Wasserman, S.A., J.M. Dungan, and N.R. Cozzarelli, 1985, "Discovery of a predicted DNA knot substantiates a model for site-specific recombination," *Science* **229**, 171-174.

Wasserman, S.A., and N.R. Cozzarelli, 1986, "Biochemical topology: Applications to DNA recombination and replication," *Science* **232**, 951-960.

White, J.H., K.C. Millett, and N.R. Cozzarelli, 1987, "Description of the topological entanglement of DNA catenanes and knots by a powerful method involving strand passage and recombination," *Journal of Molecular Biology* **197**, 585-603.