DNA Topology: Experiments and Analysis

De Witt SUMNERS

ABSTRACT

Cellular DNA is a long, thread-like molecule with remarkably complex topology. Enzymes that manipulate the geometry and topology of cellular DNA perform many important cellular processes (including segregation of daughter chromosomes, gene regulation, DNA repair, and generation of antibody diversity). Some enzymes pass DNA through itself via enzyme-bridged transient breaks in the DNA; other enzymes break the DNA apart and reconnect it to different ends. In the topological approach to enzymology, circular DNA is incubated with an enzyme, producing an enzyme signature in the form of DNA knots and links. By observing the changes in DNA geometry (supercoiling) and topology (knotting and linking) due to enzyme action, the enzyme binding and mechanism can often be characterized. This paper will discuss the tangle model for analysis of DNA site-specific recombination experiments.

Mathematical Subject Classification 2000: 57M25,92-08

Keywords: DNA, knot, link, enzyme, site-specific recombination, 2-string tangle, 4-plat

1. INTRODUCTION

Knot theory — a new kind of applicable mathematics! Some might scoff at the idea, believing knot theory to be a prime example of esoteric pure mathematics, or (worse yet) recreational doodling for terminally bored topologists. My aim is to convince you that knot theory is yet another example of what Edward E. David, Jr., [D] eloquently terms *the seemingly inevitable utility of mathematics conceived symbolically without reference to the real world*. I will freely admit that knot theory did not begin as pure mathematics. Knot theory began as applied mathematics, and now, in its second century, is enjoying new applications in science that go back to its beginnings. The family tree of modern knot theory has its roots in nineteenth-century physics, beginning with the work of Gauss on computing inductance (linking numbers) in a system of linked circular wires [Si]. Gauss's student Listing also thought about knots and was the person who coined the word *topology*. Things really got going in 1867 with Kelvin's vortex model of the atom [T]. In his model, an atom was a configuration of linked vortex tubes in the ether. Kelvin's idea was based on the work of Helmholtz, who proved that, once created, a vortex tube (an invariant solid torus) in the flow of a perfect fluid would be immortal, requiring external intervention to either create or annihilate the vortex tube. Kelvin was spurred on by seeing the smoke ring experiments of Tait, in which the smoke rings underwent elastic collisions, exhibiting interesting modes of vibration. Kelvin wanted to produce an atomic theory that would explain multiple lines in the emission spectrum of various elements. A swirling vortex tube

could absorb and emit energy at certain fundamental frequencies; linked vortex tubes would explain multiple spectral lines. The main advantage of the Kelvin atom over the Lucretius-Newton atom was that the indivisible bits of the Kelvin atom (the vortex tubes) would be held together by the forces of topology (linking), avoiding the problems inherent in devising forces to hold together an atom made up of little billiard balls. Tait soon discovered the main disadvantage of the Kelvin model–with no pure mathematics (algebraic topology) to help out, knots and links are impossible to characterize. In true pioneering spirit, Tait set out to build a table of the elements — a knot table — and the rest is history.

Knot theory is the study of entanglement of elastic graphs in 3-space. It is a study of embedding pathology and has proven to be fundamental as a laboratory for the development of algebraic-topological invariants and in the understanding of the topology of 3-manifolds. During the last 100 years, topologists have developed the discrete geometric language of knots to a fine mathematical art [BZ,K,R]. More recently, the unexpected (coming from Von Neumann algebras and quantum mechanics) discovery by Vaughan Jones [J] of new polynomial invariants that help with the knot classification problem has brought intense attention to the subject spawning that might be called the new *combinatorial* knot theory. Whatever it is, however, knot theory isn't just pure mathematics anymore. It is a prototype of what Lynn A. Steen [St] calls *the science of patterns* — theory built on relations among patterns and on applications derived from the fit between pattern and observation. The precise descriptive and calculational power of knot theory has been put to work in the description and computation of molecular configurations [WC1]. Entanglement in real physical systems has consequences. Leaving aside your frustration at resolving the entanglement in the spaghetti-like mass of computer wires under your desk, unresolved entanglement of DNA in cells is a death sentence for that cell [BZe,LZ,LM,DM]. Most drugs for the treatment of bacterial infections or cancer work by inhibiting cellular enzymes that resolve molecular entanglement in the cell, and the target cell (a pathogen or cancer cell) dies as a result. Knots in linear proteins have been discovered in the protein data base [VM], and give information about the time course of folding and the protein function, including the role of enzymes in protein degradation. In polymer science, macroscopic properties of polymer systems often depend on microscopic intermolecular entanglement [LB]; entanglement determines whether or not the polymer system is a gel or a polymer fluid, and if a solid, the entanglement has consequences in the strength of the material. In fluid dynamics, plasma and superfliud physics [Mof,Ric1,Ric2], entanglement of magnetic and vortex filaments have important consequences for the energy of the system.

2. TOPOLOGICAL ENZYMOLOGY

One of the important issues in molecular biology is the three-dimensional structure (shape) of proteins and their precursors (deoxyribonucleic acid (DNA) and ribonucleic acid (RNA)) in the cell, and the close relationship between macromolecular structure and function. Ordinarily, protein and DNA structure is determined by X-ray crystallography, electron and atomic force microscopy, and nuclear magnetic resonance imaging (NMR). Because of the close packing needed for crystallization, the manipulation required to prepare a specimen for electron or atomic force microscopy, and the lack of resolution of NMR,

these methods often do not provide conclusive evidence for molecular shape in solution. Moreover, some proteins (enzymes) function as molecular machines, changing their shape as they execute their function, so one static spatial snapshot may not tell the whole story. Topology can shed light on this key issue. The topological approach to enzymology (Fig. 5) is an experimental protocol in which the descriptive and analytical powers of topology and geometry are employed in an indirect effort to determine enzyme mechanism and the structure of active enzyme-DNA complexes *in vitro* (in a test tube) and *in vivo* (in the cell). This article will describe how recent results in 3-dimensional topology [C,CG,DS1-3,ES1-2,L,S1-4,SE,V,VS,VC] have proven to be of use in the description and quantization of DNA structure and the action of cellular enzymes on DNA, and will draw heavily on earlier articles [ES1,S1-4].

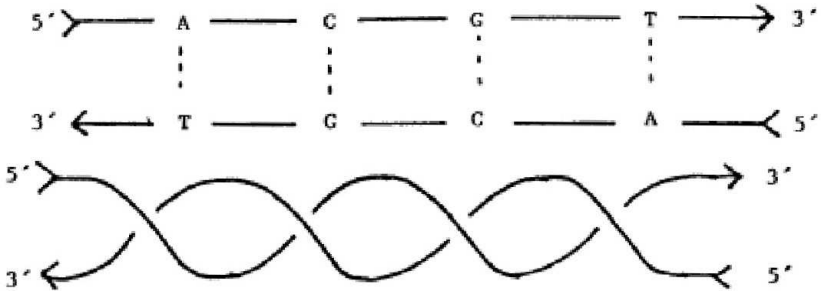

Figure 1: Duplex DNA (from [S3])

DNA molecules are long and thin, and the packing of DNA into the cell nucleus is very complex. 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 up the cell nucleus to the size of a basketball, the DNA inside scales up to the thickness of thin fishing line, with 200 km of that line inside the nuclear basketball. Most cellular DNA is double-stranded (duplex), consisting of two linear backbones of alternating sugar and phosphorus. Each 5-carbon (pentagonal) sugar molecule in the backbone is attached to the phosphate unit on one side by the carbon atom designated 5', and to the phosphate unit on the other side by the carbon atom designated 3'. Therefore, each backbone chain is endowed with a natural chemical orientation 5'→3'. Attached to each sugar molecule is one of the four nucleotide bases: A = Adenine, T = Thymine, C = Cytosine, G = Guanine. A ladder is formed by hydrogen bonding between base pairs, with A bonding with T via a double hydrogen bond, and C bonding with G via a triple hydrogen bond. The base-pair sequence (or code) 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,C,G,T}. One can model the duplex DNA molecule as a ribbon $\boldsymbol{R}$, in which the hydrogen bonds define the ruled surface of the ribbon. The edges of the ribbon R (the backbone strands) are chemically oriented opposite to each other, and the axis of the ribbon is unoriented. In the classical Crick-Watson double helix model for DNA, the backbone strands are twisted in a right-hand helical (plectonemic) fashion with an average and nearly constant pitch of approximately 10.5 base pairs per full helical twist (Fig. 1). Biologists often call the backbone edges of the DNA molecule C (for Crick) and W (for Watson)!

Moreover, since backbone bonds can only be formed 3' to 5', a covalently closed circular duplex DNA molecule is a twisted annulus, not a twisted Mobius band. We will ignore the natural antiparallel chemical orientation of the backbone strands, and adopt the orientation convention that the backbones of the circular DNA molecule are oriented in parallel with the axis, as in Fig. 2. The orientation of the axis itself is often biologically determined by short non-palindromic nucleotide sequences present on the molecules — for example, enzyme binding sites in site-specific recombination. If no biological orientation is determined, the orientation of the molecular axis or can be arbitrarily assigned in order to facilitate computations of molecular spatial structure.

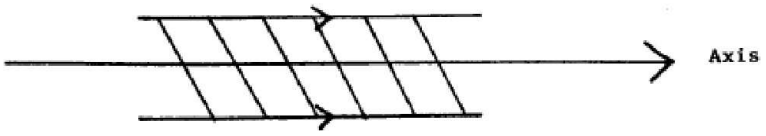

Figure 2: Orientation of DNA Ribbon R (from [S3])

The local helical pitch of duplex DNA is a function of the base-pair sequence, and if a DNA molecule is under stress or constrained to live on a surface (bound to a protein), 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. Duplex DNA in the cell nucleus is very long linear molecule, geometrically constrained by periodic attachment to a protein scaffold. The packing, twisting, and topological constraints all taken together mean that topological entanglement poses problems for the DNA molecules in the cell. This entanglement would interfere with, and be exacerbated by, the vital life processes of replication, transcription, translocation and recombination [DM, CC]. 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 requires 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 a metabolic 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 the DNA strands, a move performed by *topoisomerase* enzymes. Other enzymes break the DNA apart and recombine the ends by exchanging them, a move performed by *recombinase* enzymes. 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 directly observe the dynamics of enzymatic action.

One of the most useful descriptors of circular DNA ribbon structure comes from the conservation equation [Ca,F,W] relating the geometric quantities ***twist*** (of the ribbon) and ***writhe*** (of the ribbon axis) to the topological quantity ***link*** (of the ribbon boundary). The computation of linking number and writhe depend on the oriented skew lines sign

convention (right-hand rule) for a projected crossing of two oriented skew lines in space (Fig. 3):

Figure 3: Oriented skew lines sign convention

Given a regular projection (one in which all crossings are transverse intersections of exactly two projected arcs) of two disjoint oriented simple closed curves $\{C, W\}$ in R^3, the ***linking number*** $Lk(C, W)$ of C and W is the sum of signed crossings where C crosses over W. The linking number is a topological invariant. Given any regular projection of the ribbon axis, compute the sum of the signed self-crossings of the axis, obtaining the ***directional writhe*** of the ribbon axis (the writhe in the direction of the projection). By averaging the directional writhe over all projection directions (points on the unit 2-sphere), one obtains $Wr(R)$, the ***writhe*** of the ribbon axis. Although the directional writhe is an integer, upon averaging, one obtains a real number for the writhe; the writhe is not a topological invariant; it is a geometric measure of non-planarity of the ribbon axis, and can change when the ribbon is moved about in space. The other geometric quantity of interest is the winding of the ribbon boundary curves around the axis of the ribbon. Given a smooth closed ribbon (a twisted annulus), choose an origin on the axis, and one of the two boundary curves C, and parameterize the axis with arc length s. Let $\tau(s)$ denote the unit tangent vector to the axis at position s, and $v(s)$ denote the unit perpendicular on the ribbon to $\tau(s)$ at position s, pointing toward the boundary curve C. Let $\tau(s) \times v(s)$ denote the unit normal to the ribbon at position s. The ***twist*** of the ribbon R is a real number defined by the integral:

$$(1) \qquad Tw(R) = \frac{1}{2\pi} \oint_{\text{axis}} (\tau \times v) \cdot dv$$

The conservation equation [Ca,F,W] relating these quantities is:

$$(2) \qquad Lk(C, W) = Tw(R) + Wr(R)$$

Fig. 4 shows two isotopic configurations of a ribbon. The ribbon of Fig. 4A has a planar axis, so $Wr(R) = 0$, and one left handed full twist of the ribbon yields $Lk(C, W) = Tw(R) = -1$. In Fig. 4B, the local twist of the ribbon has been converted to a global writhe of the axis of the ribbon, and the ribbon has no twist, so $Tw(R) = 0$ and $Lk(C, W) = Wr(R) = -1$. The ribbon of Fig. 4A is said to be ***relaxed***, and the ribbon of Fig. 4B is said to be ***negatively supercoiled***. Most DNA extracted from living cells is negatively supercoiled [BC].

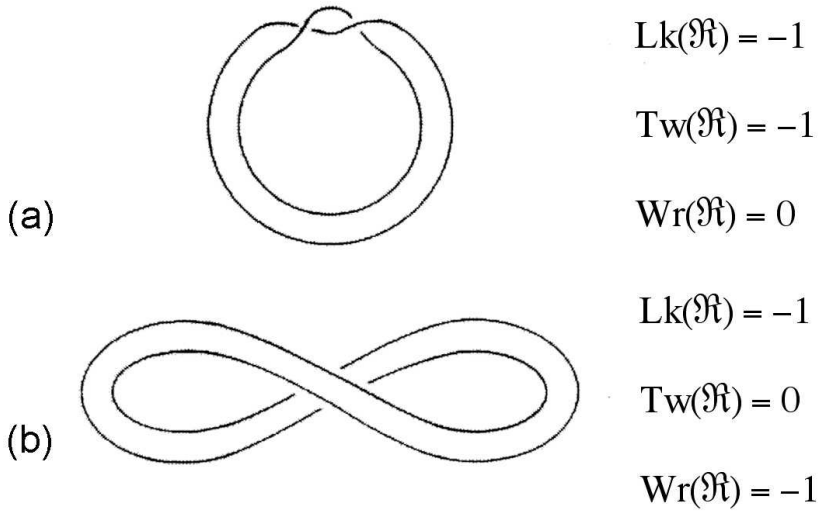

Figure 4: Link, Twist and Writhe of R (from [S4])

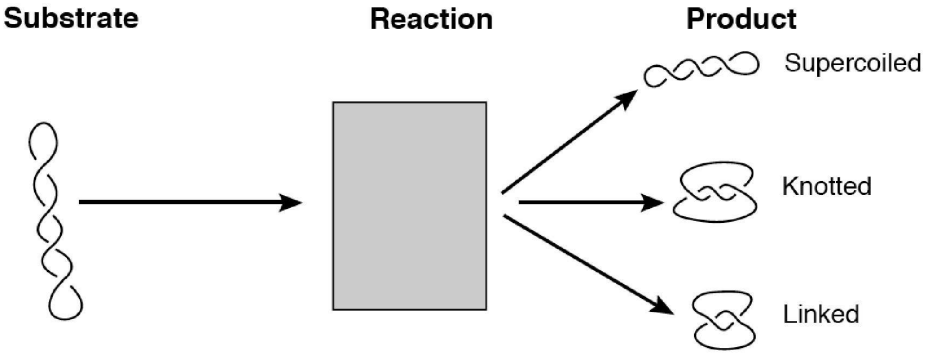

Figure 5. The Topological Approach to Enzymology (from [S2])

In order to describe and quantify DNA structure and its enzyme-mediated changes, it is clear that knot theory ought to be of some help. An interesting biological development for topology has been the recent (circa 1983) emergence of a new experimental protocol, the topological approach to enzymology [WC1], which aims to exploit knot theory directly to unravel the secrets of enzyme action. Here's how it works. Focus attention on an enzyme that mediates a local DNA interaction. Because there is at present no direct observational method (either in the cell or in a laboratory) for enzyme action, one must rely on indirect methods. One can use circular DNA as a probe to deduce facts about enzyme mechanism by detecting a topological enzyme signature, the change the enzyme causes in the topological state (embedding) of the molecule upon which it is acting. In many cases, the natural substrate for the enzyme action is linear DNA. The problem for the molecular detective is that linear DNA cannot trap topological changes caused by an enzyme — there can be no interesting (observable) topology (knots) in an unconstrained linear piece of DNA. The trick is to get a particular enzyme to act on circular DNA molecules. This can be done by manufacturing (via cloning) artificial circular DNA molecules on which the enzyme will act. When an enzyme acts on circular DNA molecules, some of the enzymatic

changes can be trapped in the form of DNA knots and links. One performs laboratory (*in vitro*) experiments, in which purified enzyme is reacted with a large collection of circular DNA molecules (the substrate). In such experiments, it is possible to control the amount of supercoiling (Fig. 4b), the knot type, and the linking of the family of substrate molecules. Using a new biological technique (rec A coating) to enhance viewing resolution under the electron microscope, one can observe the reaction products, an enzyme-specific family of DNA knots and links (Fig. 5).

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 binding and mechanism? 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 family of DNA knots and links contains information about the enzyme 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.

3. 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 recombination 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 and remove 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 (10-15 base pair) segment of duplex DNA whose base pair 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 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 Fig. 6 (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 Fig. 6, 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 look behind the curtain, to deduce mathematically the path of the DNA in the black mass of the globular protein, 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 performs two double-stranded breaks at the sites and recombines the ends by exchanging them in an enzyme-specific manner. The synaptosome can then dissociate, 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***.

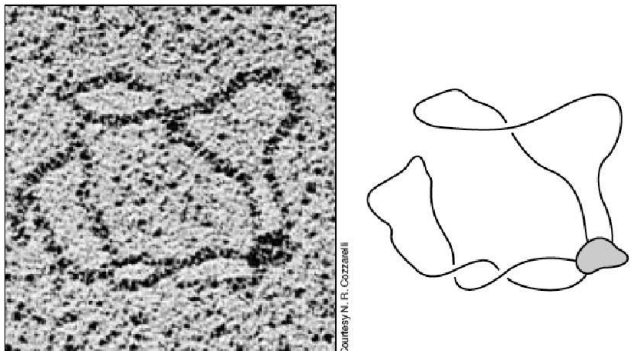

Figure 6. Tn3 Resolvase Synaptic Complex (from [S2])

During a single productive binding encounter between enzyme and DNA, the enzyme may mediate more than one recombination event before it dissociates; this is called ***processive recombination***. On the other hand, the enzyme may perform recombination in multiple productive binding encounters with the DNA, a scenario called ***distributive recombination***. Some site-specific recombination enzymes mediate both distributive and processive recombination; other site-specific recombination enzymes mediate only a single recombination event, and the changed genetic sequence of the product DNA molecules cannot support further recombination events. In addition to changing the genetic sequence of the substrate DNA molecule, site-specific recombination usually changes the topology of the substrate, producing knots, links and supercoiling in the circular molecules. 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 catenane) (Fig. 7). If the substrate is a pair of DNA circles with one site each, the product is a single DNA circle (Fig. 7 read in reverse) and can form a DNA knot (usually with direct repeats). In processive recombination on circular substrate with direct repeats, the product of an odd number of rounds of processive recombination is a pair of DNA circles, and the product of an even number of rounds of

processive recombination is a single DNA circle with the parental genotype. 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 iterated site-specific recombination on a single circle with inverted repeats, an odd number of rounds of recombination produces a single DNA circle of recombinant genotype, and an even number of rounds of recombination produces a DNA circle of parental genotype. In all figures where DNA is represented by a line drawing (such as Fig. 7), duplex DNA is represented by a single line, and supercoiling is omitted.

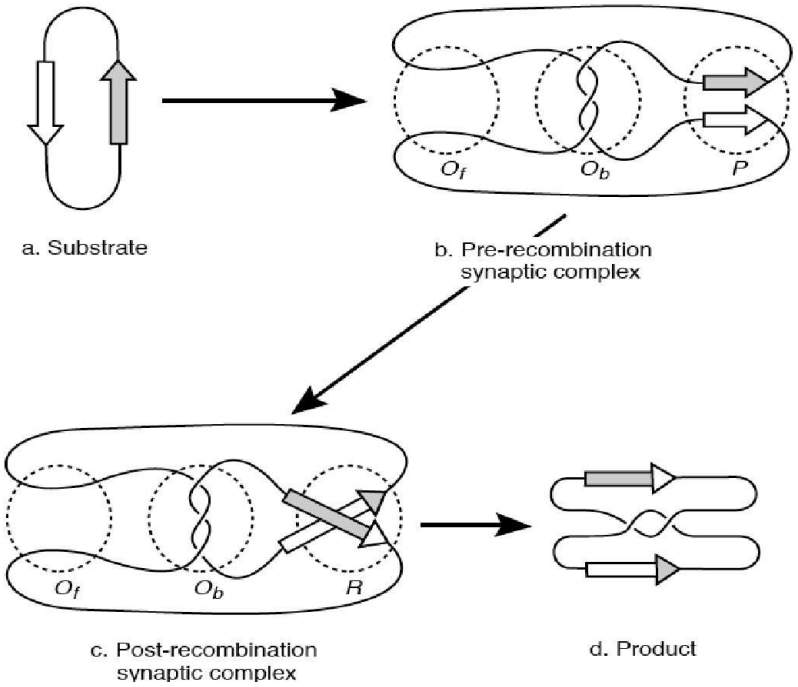

Figure 7: A Single Recombination Event: Direct Repeats (from [S2])

The geometry (supercoiling) and topology (knotting and linking) of circular DNA substrate molecules 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 negatively supercoiled unknot). The amount of supercoiling of the substrate molecules is also a control variable. The substrate molecules are reacted *in vitro* with purified enzyme, and the reaction products are fractionated by gel electrophoresis. DNA is a polyelectrolyte with negative charge proportional to molecular weight. An agarose gel is a sugar that provides a random obstruction field through the DNA molecules can be forced to migrate in the presence of an applied static electric field. Under the influence of the electric field, the DNA molecules migrate toward the positive electrode. Gel electrophoresis normally discriminates among DNA molecules on the basis of molecular weight, and can discriminate a difference of one base pair. Given that all molecules are exactly 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/or topology of the DNA molecules. Under the proper

conditions of gel density and applied voltage, circular DNA velocity in the gel is (perhaps surprisingly) determined by the crossing number of the knot or link [SK]; knots and links of the same crossing number migrate with the same gel velocities (Fig. 8). After running the gel, the DNA molecules can be removed from the gel and coated with Rec A protein. It is this observation technique (Rec A-enhanced electron microscopy) [KS] 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, but does not change the knot/link type of a DNA circles. Rec A coating facilitates the unambiguous determination of crossings (nodes) in an electron micrograph of DNA, allowing precise determination of the DNA knot/link type.

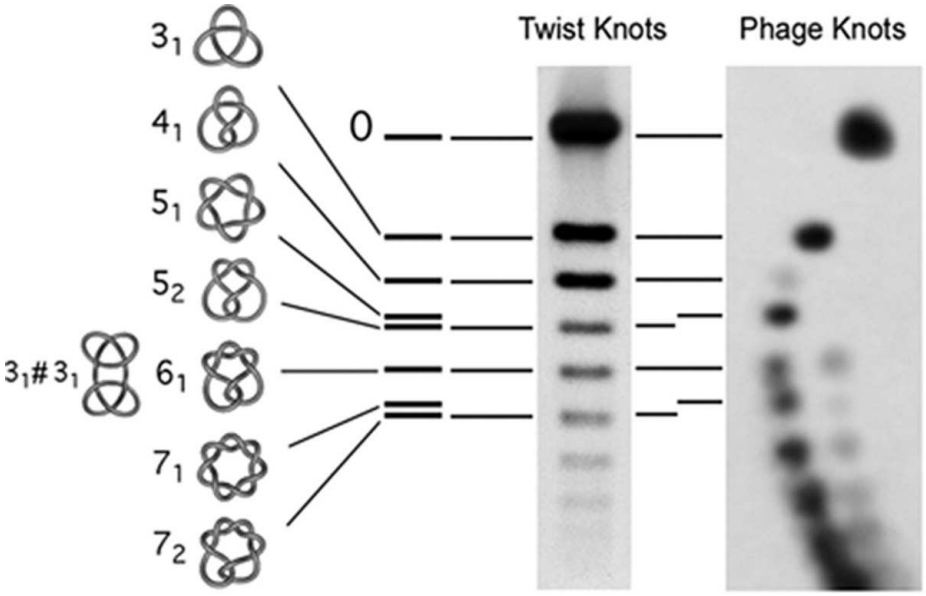

Figure 8: Gel Electrophoresis of DNA Knots (from [AV1])

4. KNOTS AND TANGLES

In this section, we will describe the parts of knot theory and tangle calculus that are of biological relevance. For a more rigorous mathematical treatment we refer the reader to [BZ,K,R] for knot theory and [ES1] 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 is endowed with a fixed orientation. Two knots (links) $\{K_1, K_2\}$ are ***equivalent*** (written $K_1 = K_2$) if there is an orientation-preserving (on R^3) homeomorphism of pairs $h : (R^3, K_1) \rightarrow: (R^3, K_2)$. The homeomorphism of pairs h superimposes K_1 on K_2; in this case the knots (links) can also be made congruent by a flexible motion or flow (ambient isotopy) of space. An ***ambient isotopy*** is a 1-parameter family of homeomorphisms $\{H_t\} 0 \leq t \leq 1$ 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) type 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, coded with breaks in the undercrossing strand so that the knot (link) type can be unambiguously reconstructed in 3-space from the 2-dimensional knot (link) diagram. The ***crossing number*** of a knot or link type is the smallest number of crossings possible in a planar diagram. A diagram that realizes the minimum number of crossings for a knot (link) type is called a ***minimal diagram***. A knot (link) diagram is alternating if, as one traverses any strand, the crossings encountered are alternately over and under. Fig. 9 shows minimal alternating diagrams 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 R^3. 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. One can convert a diagram of K to a diagram of K^* by reversing each of the crossings in the diagram (Fig. 9, d,e). If $K = K^*$, then we say that K is ***achiral***; if $K \neq K^*$, then we say that K is ***chiral***. In Fig. 9, (a) and (b) are achiral, and (c), (d) and (e) are chiral.

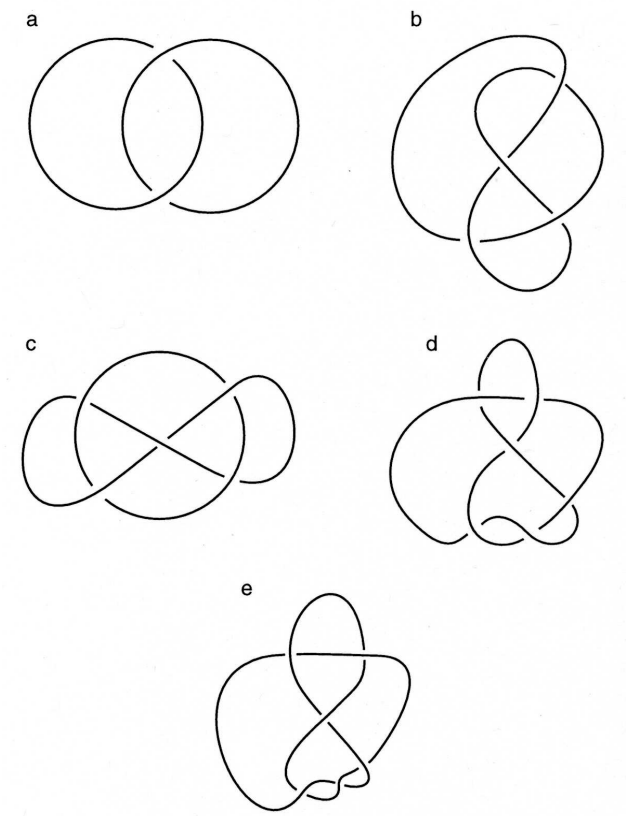

Figure 9: Tn3 Resolvase Products: (a) Hopf Link, (b) Figure 8 Knot, (c) (+) Whitehead Link, (d) 6_2^*, (e) 6_2 (from [S2]).

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

and computationally tractable family of ***4-plats*** (***2-bridge*** knots and links) [Sc,BZ]. This family consists of knot and link configurations produced by patterns of plectonemic interwinding of pairs of strands about each other, echoing the primary Crick-Watson helical winding in duplex DNA (Fig. 1). All small knots and links are members of the family of 4-plats — 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 Fig. 9 are 4-plats; their standard 4-plat minimal alternating diagrams are shown in Fig. 10. 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 Fig. 10. Two unoriented 4-plats are of the same knot (link) type if and only if their classifying vectors are identical, or identical upon reversal of one of the vectors. A classifying rational number can be computed from the classifying vector using the following extended fraction calculation:

$$\frac{\beta}{\alpha} = \frac{1}{c_1 + \frac{1}{c_2 + \ldots}}.$$

Given the classifying rational numbers β/α and β'/α' for a pair of unoriented 4-plats, the 4-plats are the same type if and only if $\alpha = \alpha'$ and $\beta^{\pm 1} \equiv \beta' (\mathrm{mod}\ \alpha)$ [Sc,BZ].

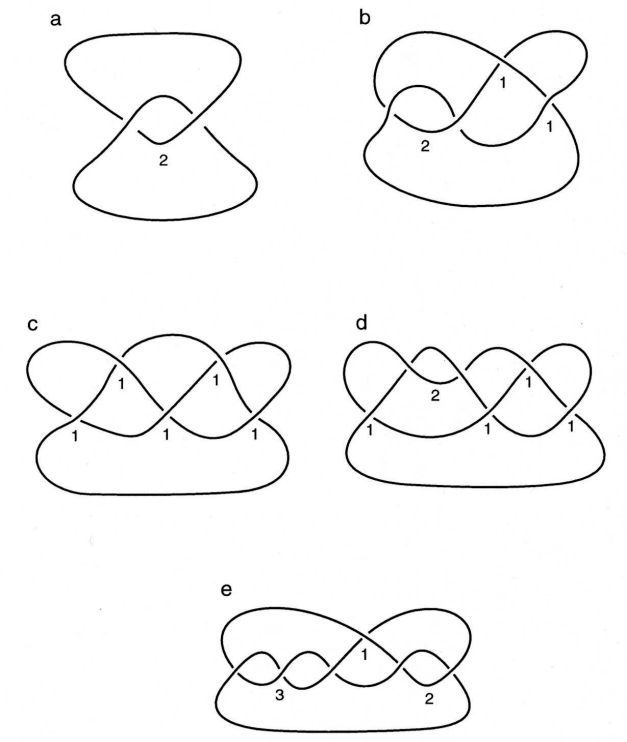

Figure 10. Tn3 Resolvase 4-Plats: (a) $\langle 2 \rangle$ Hopf Link, (b) $\langle 2, 1, 1 \rangle$ Figure 8 Knot, (c) $\langle 1, 1, 1, 1, 1 \rangle$ Whitehead Link, (d) $\langle 1, 2, 1, 1, 1 \rangle 6_2^*$, (e) $\langle 3, 1, 2 \rangle 6_2$ (from [S2]).

For *in vitro* topological enzymology experiments, we can regard the enzyme mechanism as a machine that transforms some DNA 4-plats into other DNA 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 (Fig. 6) to create the synaptosome. The enzyme then performs its topological moves, and the DNA is then released.

We need a mathematical language to describe configurations of linear strings (the segments of bound DNA) in a spatially confined region (the enzyme). As is evident from the enzyme-DNA complex in Fig. 6, the globular protein enzyme is homeomorphic to a 3-ball, and the 2 strands of bound DNA (which contains the recombination sites) forms a protein-DNA ***2-string tangle***. Tangles were introduced into knot theory by J.H. Conway [C]. Tangle theory is knot theory done inside a unit 3-ball (B^3) with the ends of the strings firmly glued down. On the unit 3-ball, select four points on the equator $\{NW; SW; SE; NE\}$. 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 distinguished points $\{NW; SW; SE; NE\}$. Two tangles in the unit 3-ball are ***equivalent*** if it is possible to ambient isotop in the interior of the 3-ball the strings of one tangle into the strings of the other while keeping the boundary sphere fixed. A class of equivalent tangles is called a ***tangle type***. Tangles are usually represented by their coded projections, called ***tangle diagrams***, onto the equatorial disk in the unit 3-ball, as shown in Fig. 11. In all figures containing tangles, we assume that the four boundary points $\{NW; SW; SE; NE\}$ are as in Fig. 11, and we suppress these labels.

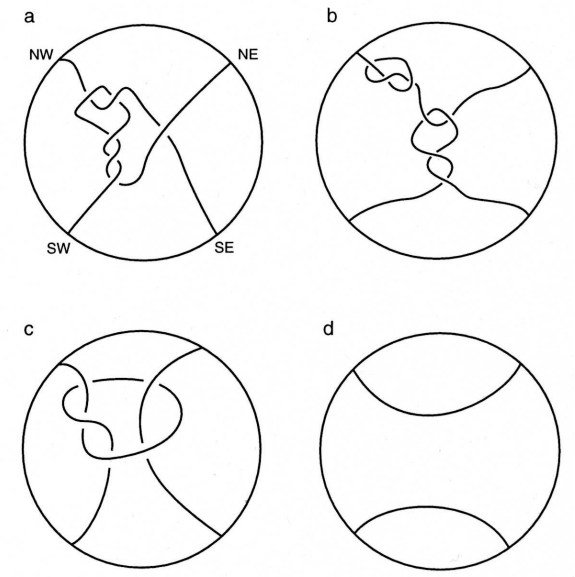

Figure 11. Tangles (a) Rational, (b) Locally Knotted, (c) Prime, (d) Trivial (from [S2]).

All four of the tangles in Fig. 11 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 boundary sphere of the 3-ball during the isotopy, then the

tangle of Fig. 11a can be transformed into the trivial tangle of Fig. 11d. The tangles in Figs. 11b and 11c cannot be transformed to the trivial tangle by any sequence of such turning motions of the endpoints on the boundary sphere. The family of tangles that can be converted to the trivial tangle by an ambient isotopy that is allowed to move the endpoints on the boundary sphere is the family of ***rational tangles***. Equivalently, a rational tangle is one in which the strings can be transformed by ambient isotopy entirely into the boundary 2-sphere of the 3-ball. Rational tangles form a homologous family of 2-string configurations in B^3, and are formed by a pattern of plectonemic interwinding of pairs of strings. Like 4-plats, rational tangles look like DNA configurations being built up out of successive plectonemic supercoiling of pairs of strings. More specifically, enzymes are often globular in shape and are topologically equivalent to our unit-defining ball. Furthermore, we assume that any two copies of the enzyme can be superimposed by rigid motion, and this spatial congruence of the protein also respects the atomic structure of the protein (it matches corresponding atoms and endpoints of corresponding DNA strands). Thus, in an enzymatic reaction between a pair of DNA duplexes, the pair {enzyme, bound DNA} forms a well-defined 2-string tangle.

How do we know the DNA tangles are rational? I will give three arguments: 1. Since the amount of bound DNA is small, the enzyme-DNA tangle so formed admits projections with only a few nodes and therefore is rational by default. For example, all locally unknotted 2-string tangles having less than five crossings are rational. 2. In all cases studied intensively, the 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. 3. We will give a mathematical proof: If the products of processive recombination are 4-plats, then one can often prove that the DNA tangles involved must be rational tangles.

There is a classification scheme for rational tangles which is based on a standard form that is a minimal alternating tangle 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 of the same sign, 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 super-coils and vertical right-handed plectonemic supercoils (Fig. 12). This sign convention is opposite to that of Conway [C], and was chosen to agree with existing sign conventions used by biologists.

Fig. 12 shows some standard tangle diagrams, which are minimal alternating 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 minimal tangle diagram must have at least two crossings. There are four rational tangles $\{(0); (0, 0), (1), (-1)\}$ that are exceptions to this convention

$|a_1| > 1$, and are displayed in Figs. 12c through 12f. The classifying vector $(a_1, a_2, \ldots, a_n)$ can be converted to an (extended) rational number by means of the following continued fraction calculation:

$$\frac{\beta}{\alpha} = \frac{1}{a_{n-1} + \frac{1}{a_{n-2}+\ldots}}.$$

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

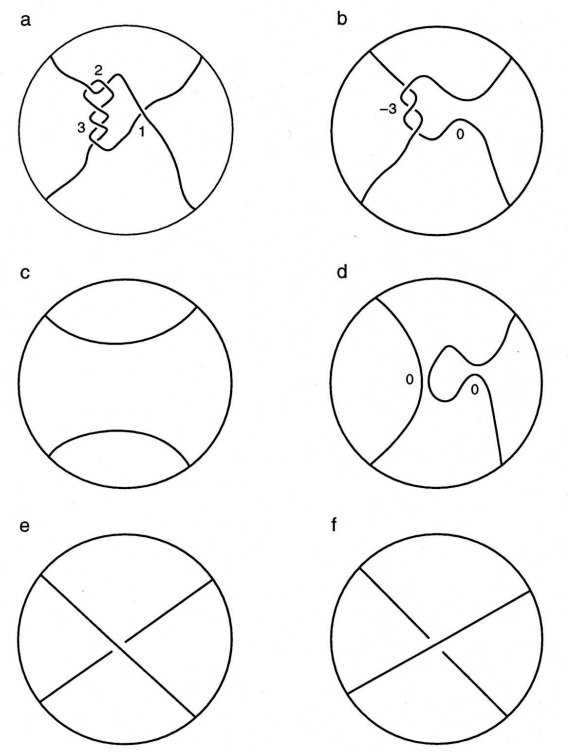

Figure 12. Rational Tangle Diagrams: (a) $(2,3,1)$, (b) $(-3,0)$, (c) (0), (d) $(0,0)$, (e) $(+1)$, (f) (-1) (from [S2]).

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 Fig. 13a. Equivalently, the tangle sum $(A + B)$ can be viewed as the decomposition of a complicated tangle into two simpler summands. The sum of two rational tangles need not be rational (Fig. 11c). Given any tangle A, one can form the tangle closure $N(A)$ as in Fig. 13b. In the closure operation on a 2-string tangle, ends NW and NE are connected outside the 3-ball, ends SW and SE are connected outside the 3-ball, and the tangle-defining ball is deleted, leaving a knot or a link of two components. Deletion of the tangle-defining 3-ball is the mathematical analogue of the biological action of deproteinization of the DNA that occurs 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 Fig. 13c. In general, if A and B are any two rational tangles, then $N(A+B)$ is a 4-plat.

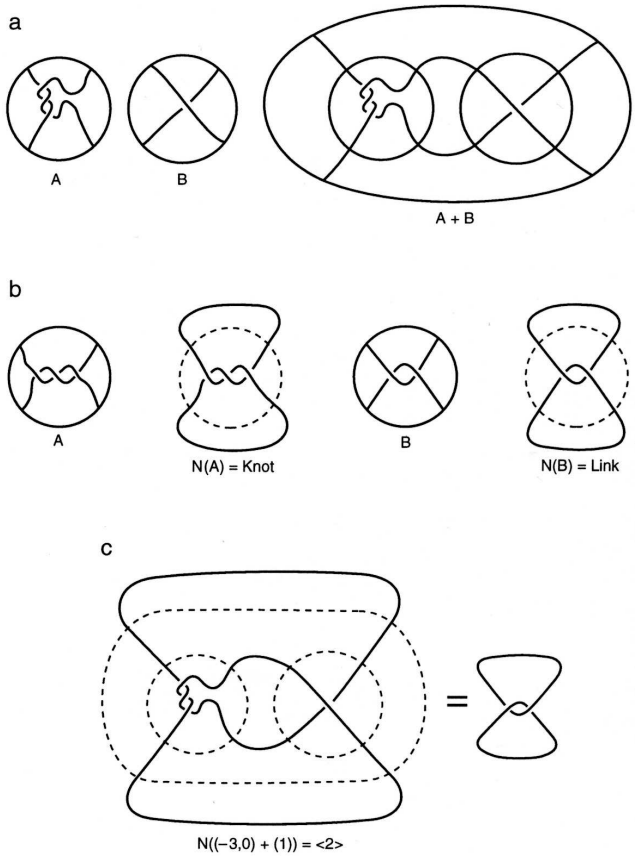

Figure 13: Tangle Operations: (a) Tangle Addition, (b) Tangle Closure, (c) Tangle Equation $N((-3,0)+(1)) = \langle 2 \rangle$ (from [S2]).

5. The Tangle Model for Site-Specific Recombination

The fundamental observations underlying this model are that a pair of sites bound by an enzyme (Fig. 6) forms a 2-string 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 performs a double-stranded break in 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 [ES1,S2,S4].

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-re-combination 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 consequently 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 that 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. If the enzyme must achieve a geometric footprint on the DNA in order to mediate productive synapsis, this footprint is captured in the tangle O_b. 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. The schematic of Fig. 7 shows the tangles involved in Tn3 Resolvase site-specific recombination.

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 (an experimentally controlled knot/link type); deproteinization of the post-recombination synaptic complex yields the product (an observable knot/link type). The topological structure (knot and link types) of the substrate and product yields equations in the four 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 Fig. 7. In Fig. 7, with the tangle unknowns evaluated as follows: $O_f = (0)$, $O_b = (-3, 0)$, $P = (0)$, $R = (1)$. With these values for the tangle variables, our recombination equations become:

Substrate Equation: $N((0) + ((-3, 0)) + (0))) = \langle 1 \rangle$,

Product Equation: $N((0) + ((-3, 0) + (1))) = \langle 2 \rangle$.

6. The Topologyof Tn3 Resolvase

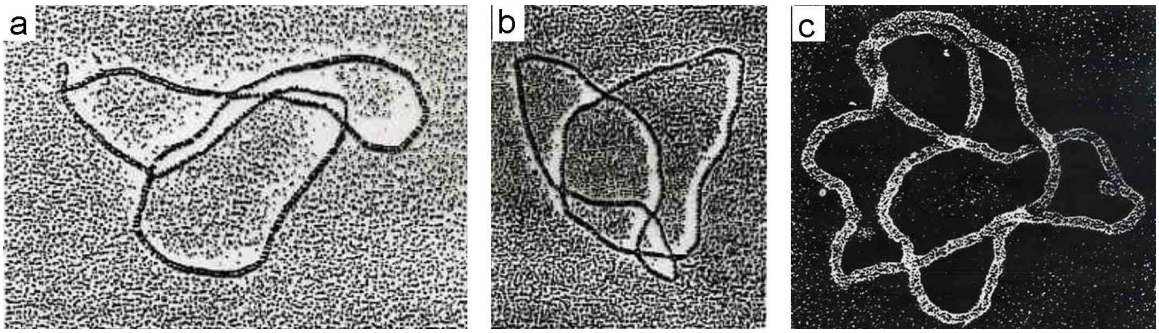

Figure 14: Electron Micrographs of RecA coated Tn3 Resolvase Recombination Products. (a) Figure 8 Knot (from [KS]), (b) +Whitehead Link (from [KS]), (c) 6_2^* (from [WD])

Tn3 resolvase is a site-specific recombination enzyme that reacts with negatively supercoiled circular duplex DNA substrate with directly repeated recombination sites [WD]. As substrate for the *in vitro* reaction, supercoiled unknotted DNA substrate is incubated with

resolvase. The principal product of this reaction is known to be the DNA 4-plat (2) (the Hopf link, Figs. 9a and 10a). Resolvase is known to act dispersively in this situation — to bind to the circular DNA substrate, to mediate a single recombination event, and then to release the linked product. It is also known that resolvase and free (unbound to protein) DNA links do not react. However, once in twenty encounters, resolvase acts processively-additional recombinant strand exchanges are mediated during the single binding encounter prior to the release of the product, with yield decreasing exponentially with increasing number of strand exchanges. Two successive rounds of processive recombination produce the DNA Figure 8 knot $\langle 2,1,1\rangle$, Fig. 9b and10b, whose electron micrograph appears in Fig. 14a. Three successive rounds of processive recombination produce the DNA link $\langle 1,1,1,1,1\rangle$ (the +Whitehead link, Fig. 9c and 10c, whose electron micrograph appears in Fig. 14b; four successive rounds of recombination produce the DNA knot $\langle 1,1,1,1,1\rangle$, the knot 6_2^*, Fig. 9d and 10d, whose electron micrograph appears in Fig. 14c.

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 procession recombination, each additional round of recombination adds a copy of the recombinant tangle R to the synaptosome.*

More precisely, p rounds of processive recombination at a single binding encounter generates the following system of $(p+1)$ tangle equations in the 4 tangle unknowns $\{O_f, O_b, P, R\}$:

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

N^{th} Round Product Equation: $N(O_f + (O_b + nR)) = n^{\text{th}}$ round product $1 \leq n \leq p$.

For resolvase, the electron micrograph of the synaptic complex in Fig. 6 reveals that $O_f = (0)$, since the DNA loops on the exterior of the synaptosome can be untwisted and are not entangled with each other. This observation from the micrograph reduces the number of variables in the tangle model by one, leaving us with three variables $\{O_b, P, R\}$.

Theorem 1 [ES1]: *Suppose that tangles $\{O_b, P, R\}$ satisfy the following equations:*

(i) $N(O_b + P) = \langle 1\rangle$ *(substrate = unknot)*

(ii) $N(O_b + R) = \langle 2\rangle$ *(1^{st} round product = Hopf link)*

(iii) $N(O_b + 2R) = \langle 2,1,1\rangle$ *(2^{nd} round product = Figure 8 knot).*

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

Theorem 1 gives an even number of tangle solution pairs for $\{O_b, R\}$ because each of the knots and links involved in the three equations is achiral (the unoriented Hopf link is

achiral); therefore the mirror image of any solution pair is also a solution pair. Theorem 1 says that there are two mirror image solution pairs for $\{O_b, R\}$. Notice that Theorem 1 says nothing about the parental tangle P. This is because P is involved in only one tangle equation. It is known [ES1] that if the tangle X is a solution to an equation of the form $N(A + X) = K$, where A is a rational tangle and K is a 4-plat, then there are infinitely many rational solutions for X. Biologists believe that $P = (0)$; since the recombination sites are short ($\sim$15 base pairs), and duplex DNA is fairly stiff, then the DNA axis at the sites can be represented by oriented straight line segments. Given two spatially juxtaposed straight line segments in space, one can choose a projection where the segments project to a pair of parallel oriented line segments [SE].

The first (and most mathematically interesting) step in the proof of Theorem 1 is to argue that the solutions $\{O_b, R\}$ must be rational tangles. We have the following facts about tangles [L]: tangles come in three classes: locally knotted, prime and rational; if a tangle is locally unknotted, then it must be either prime or rational. A tangle A is rational if and only if its 2-fold branched cyclic cover A' is a solid torus. If A is a prime tangle, and A' is its 2-fold branched cyclic cover, then the inclusion induced homomorphism injects $\pi_1(\partial A') = Z \oplus Z$ into $\pi_1(A')$. Now O_b and R are locally unknotted because of equation (ii), since the 1^{st} round recombination product Hopf link has two unknotted components, and any local knot in any of the tangles would persist as a local knot in the recombination product. Also, at least one of $\{O_b, R\}$ must be rational; otherwise they are both prime, which means that the 2-fold branched cover of $N(O_b + R) = \langle 2 \rangle$ (the lens space L(2,1)) is obtained from O_b' and R' , glued together along their common incompressible torus boundary. This means that L(2,1) contains an incompressible torus, which is impossible.

Suppose now that O_b is rational and R is prime. Given that $N((O_b + R) + R)$ is a knot (equation (iii)), one can argue [ES1] that $(O_b + R)$ is also a prime tangle, which means that the 2-fold branched cyclic cover of the Figure 8 knot (the lens space L(5,3)) contains an incompressible torus, also impossible. We conclude that R must be a rational tangle.

The next step is to argue that O_b is rational. Otherwise, O_b is prime; in this case, since $N(O_b + P) = \langle 1 \rangle$, P must be locally unknotted and rational. The 2-fold branched cyclic cover $N(O_b + P)' = [(O_b' \cup P')/(\partial O_b' = \partial P' = S^1 \times S^1)] = S^3$. Since P' is a solid torus, this means that O_b' is a bounded knot complement in S^3. We have that R is rational, and can argue that equation (iii) implies that $(R + R)$ is also rational. Passing to the 2-fold branched cyclic covers of equations (ii) and (iii), we obtain the equations $N(O_b + R)' =$ L(2,1) and $N(O_b + (R) + R))' =$ L(5,3). Because R' and $(R + R)'$ are each a solid torus, this means that there are two Dehn fillings of the knot complement O_b', resulting in the lens spaces L(2,1) and L(5,3). The cyclic surgery theorem [CG] now applies to argue that, since the orders of the cyclic fundamental groups of the lens spaces differ by more than one, this means that O_b' must be a Seifert fiber space, and that O_b' is a torus knot complement. The results of Dehn surgery on torus knots is well understood [M], and one can show that in fact the torus knot in question must be the unknot, and that O_b' is a solid torus, hence O_b is a rational tangle.

The proof now amounts to computing the rational solutions to the equations in Theorem

1. Claus Ernst and I developed a "rational tangle calculus [ES1] which uses the classifying symbols for rational tangles and 4-plats to do these calculations, obtaining the four solutions in Theorem 1. In order to decide which of these four tangle pair solutions is the biologically correct one, we must utilize more experimental evidence, and to get to a unique solution, must have a chiral recombination product. The result of 3 rounds of 3 rounds of recombination is the unoriented (+) Whitehead link, which is chiral. Using this information, we have the following:

Theorem 2 [ES1]: *Suppose that tangles* $\{O_b, P, R\}$ *satisfy the following equations:*

(i) $N(O_b + P) = \langle 1 \rangle$ *(substrate = unknot)*

(ii) $N(O_b + R) = \langle 2 \rangle$ *(1^{st} round product = Hopf link)*

(iii) $N(O_b + 2R) = \langle 2, 1, 1 \rangle$ *(2^{nd} round product = Figure 8 knot)*

(iv) $N(O_b + 3R) = \langle 1, 1, 1, 1, 1 \rangle$ *(3^{rd} round product = +Whitehead link).*

Then $\{O_b, R\} = \{(-3, 0), (1)\}$ and the result of 4 rounds of processive recombination is the 4-plat $6_2^* = \langle 1, 2, 1, 1, 1 \rangle$.

Of the four solutions produced in Theorem 1, only $\{O_b, R\} = \{(-3, 0), (1)\}$ is a solution to equation (iv). The correct global topology of the first round of iterated processive Tn3 site-specific recombination is shown in Fig. 7. Moreover, the first 3 rounds of recombination uniquely determine the result of 4 rounds of recombination, the observed DNA knot 6_2^*(Fig. 14(c)).

Is it possible to use the first two rounds of recombination to uniquely determine the enzyme binding and mechanism, and to correctly predict the results of 3 and 4 rounds of recombination? In order to do this, we need a chiral product, and fortunately experimental evidence exists which allows us to unambiguously put a recombination-induced orientation on the Hopf link product, making it chiral.

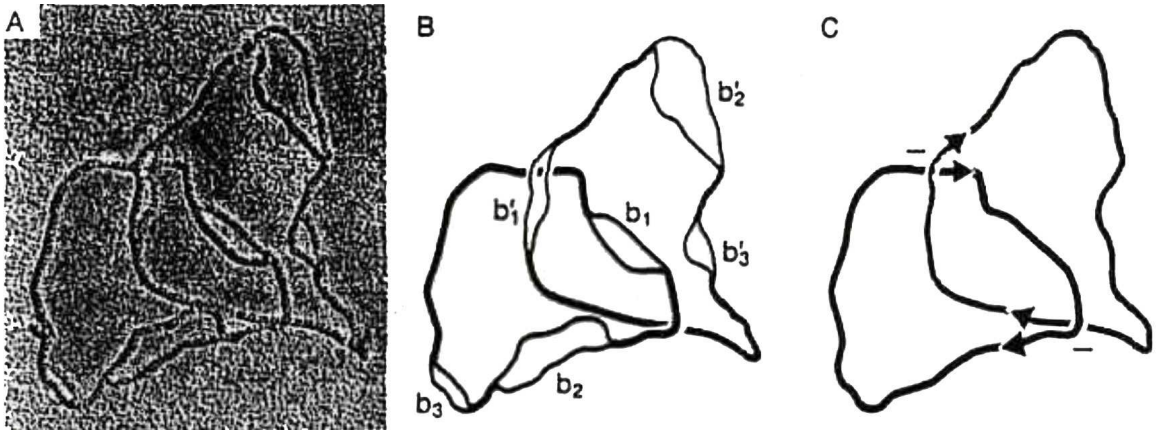

Figure 15. The Biologically Oriented DNA Hopf Link (from [WC2])

In a remarkable experiment, Steve Wassserman and Nick Cozzarelli [WC2] were able to determine the orientation induced on the 1^{st} round recombination product of resolvase acting on supercoiled unknotted substrate with directly repeated sites. In duplex DNA, the AT base pairs have a double hydrogen bond, and the CG base pairs have a triple hydrogen bond, so the AT bonding is weaker than the CG bonding. When duplex DNA is partially denatured by heating, the AT bonds break before the CG bonds. On circular substrate with directly repeated sites, the recombination sites divide the circular substrate into two domains. Using cloning techniques, one can install 3 AT-rich regions into each domain; the AT-rich regions are of 3 different lengths. Upon recombination, one obtains a linked pair of circles, each inheriting 3 AT-rich regions from the parental DNA substrate circle. In order to visualize these regions, partial denaturation of the Hopf link product revels 3 bubbles on each component circle, allowing us to unambiguously determine the orientation of each circle. With this induced orientation, the linking number of the oriented DNA Hopf link is -1 (Fig. 7 and Fig. 15), making it a chiral product. Assuming again that $P = (0)$, we have the following:

Theorem 3 [ES1]: *Suppose that tangles $\{O_b, R\}$ satisfy the following equations:*

(i) $N(O_b) = \langle 1 \rangle$ *(substrate = unknot)*

(ii) $N(O_b + R) = \langle 2 \rangle$ *(1^{st} round product = Hopf link, with linking number = -1)*

(iii) $N(O_b + 2R) = \langle 2, 1, 1 \rangle$ *(2nd round product = Figure 8 knot).*

Then $\{O_b, R\} = \{(-3, 0), (1)\}$ and the result of 3 rounds of processive recombination is the +Whitehead link $\langle 1, 1, 1, 1, 1 \rangle$, and the result of 4 rounds of processive recombination is the 4-plat $6_2^* = \langle 1, 2, 1, 1, 1 \rangle$.

So the first two rounds of recombination determine enzyme mechanism and binding, and correctly predict the observed result of 3 and 4 rounds of recombination. The above analysis amounts to a mathematical proof of enzyme binding and mechanism, and is a mathematical model that is of utility in many other situations where circular DNA is used as a probe for biological activity. For example, the tangle model is useful when electron microscopy is not available. Gel electrophoresis (enhanced by radiolabelling of DNA) can be done to detect vanishingly small amounts of DNA product, and the gel velocity of relaxed circular DNA tells us the crossing number of the DNA product. Knowing only the crossing number is very useful; one can for example use crossing number information to help characterize the geometry of packing of viral DNA in phage capsids [AT,AV1-2,TA,MM]. Since all knots and links of small crossing numbers are known, one can write down tangle equations and solve them, knowing only the crossing number of the DNA products. One goes to the knot and link tables, and for the right-hand sides of each tangle equation, plugs in all the possible knot (link) products of a given crossing number [V]. Moreover, computer programs exist to solve systems of tangle equations and visualize the answers [SV], [`http://bio.math.berkeley.edu/TangleSolve/`; `http://www.math.uiowa.edu/~idarcy/`]. More generally, the existence of a mathematical model allows one to answer "what if questions, and carefully investigate the utility of

the assumptions that go into the model.

I gratefully acknowledge research support from the National Science Foundation and the Burroughs Wellcome Fund Interfaces Grant to the Program in Mathematics and Molecular Biology at Florida State University. I would also like to thank Prof. Akio Kawauchi and the Osaka City University Advanced Mathematical Institute for hospitality shown me during February and March of 2006.

References

[AT] Arsuaga, J., Tan, R.K-Z., Vazquez, M., Sumners, D.W., Harvey, S.C.: Investigation of viral DNA packing using molecular mechanics models, Biophysical Chemistry **101–102** (2002), 475–484.

[AV1] Arsuaga, J., Vazquez, M., McGuirk, P., Sumners, D.W., Roca, J.: DNA Knots Reveal Chiral Organization of DNA in Phage Capsids, Proc. National Academy of Sciences USA **102**(2005), 9165–9169.

[AV2] Arsuaga, J., Vazquez, M., Trigueros, S., Sumners, D.W., Roca, J.: Knotting probability of DNA molecules confined in restricted volumes: DNA knotting in phage capsids, Proc. National Academy of Sciences USA **99**(2002), 5373–5377.

[BC] Bauer, W.R., Crick, F.H.C., White, J.H.: Supercoiled DNA. Scientific American **243** (1980), 100–113.

[BZe] Buck, G, Zechiedrich, E.L.: DNA disentangling by type-2 topoisomerases. J. Mol Biol. **340** (2004), 933–939.

[BZ] Burde, G., Zieschang, H.: Knots. Berlin, De Gruyter (1985).

[Ca] Călugareănu, G.: Sur les classes disotope des noeuds tridimensionnels et leurs invariants. Czechoslovak Mathematical Journal **11** (1961), 588–625.

[C] Conway, J.H.: On enumeration of knots and links, and some of their algebraic properties. Computational Problems in Abstract Algebra, Proc. Conf. Oxford 1967, Pergamon (1970), 329–358.

[CC] Cozzarelli, N.R., Cost, G.J., Mollman, M., Viard, T., Stray, J.E.: Giant proteins that move DNA: bullies of the genomic playground. Nature Reviews Molecular Cell Biology **7** (2006), 580–588.

[CW] N.J. Crisona, R.L. Weinberg, B.J. Peter, D.W. Sumners, and N.R. Cozzarelli, The Topological Mechanism of Phage Lambda Integrase, Journal of Molecular Biology **289** (1999), 747–775.

[CG] Culler, M.C., Gordon, C.M., Luecke, J., Shalen, P.B.: Dehn surgery on knots. Ann. of Math. **125** (1987), 237–300.

[DS1] Darcy, I.K., Sumners, D.W.: A strand passage metric for topoisomerase action. Knots 96, Proceedings of the Fifth International Research Institute of the Mathematical Society of Japan, S. Suzuki, ed., World Scientific (1997), 267–278.

[DS2] Darcy, I.K., Sumners, D.W.: Applications of Topology to DNA, Banach Center Publications **42**, Polish Academy of Sciences, Warsaw (1998), 65–75.

[DS3] Darcy, I.K., Sumners, D.W.: Rational Tangle Distances on Knots and Links, Math. Proc. Camb. Phil. Soc. **128** (2000), 497–510.

[D] David, E.E.Jr.: Renewing US mathematics: an agenda to begin the second century. Notices of the A.M.S. **35** (1988), 1119–1123.

[DM] Diebler, R., Mann, J.K., Zechiedrich, E.L.: Hin-Mediated DNA Knotting and Recombination Promote Replicon Dysfunction and Mutation. In preparation.

[ES1] Ernst, C., Sumners, D.W.: A calculus for rational tangles: applications to DNA recombination. Math. Proc. Camb. Phil. Soc. **108** (1990), 489–515.

[ES2] Ernst, C., Sumners, D.W.: Solving Tangle Equations Arising in a Recombination Model, Math. Proc. Camb. Phil. Soc. **126** (1999), 23–36.

[F] Fuller, B.: The writhing number of a space curve. Proc. Nac. Acad. Sci. USA **68** (1971), 815–819.

[J] Jones, V.F.R.: A polynomial invariant for knots an links via Von Neumann Algebras. Bull. A.M.S. **12** (1985), 103–111.

[K] Kauffman, L.H.: On knots. Ann. of Math. Studies **115**, Princeton U. Press (1987).

[KS] Krasnow, M.A., Stasiak, A., Spengler, S.J., Dean, F., Koller, T., Cozzarelli, N.R.: Determination of the absolute handedness of knots and catenanes of DNA. Nature **304** (1983), 559–560.

[LB] Lacher R.C., Bryant J.L., Howard L., Sumners D.W.: Linking phenomena in the amorphous phase of semicrystalline polymers. Macromolecules **19**, 2639–2643 (1986).

[L] Lickorish, W.B.R.: Prime knots and tangles. Trans. A.M.S. **267** (1981), 321–332.

[LZ] Liu, Z, Zechiedrich, E.L., Chan, H.S.: Inferring global topology from local juxtaposition geometry: interlinking polymer rings and ramifications for topoisomerase action. Biophys. J., **90**, 2344–2355 (2006).

[LM] Liu, Z, Mann, J.K., Zechiedrich, E.L., Chan, H.S.: Topological information embodied in local juxtaposition geometry provides a statistical mechanical basis for unknotting by type-2 DNA topoisomerases. J. Mol. Bio. **361** (2006), 268–285.

[M] Moser, L.: Elementary surgery along a torus knot. Pac. J. Math. **38** (1971), 737–745.

[Mof] Moffatt, H.K.: Knots and Fluid Dynamics. In Ideal Knots, Series on Knots and Everything, eds. Stasiak, A., Katritch, V., Kauffman, L.H. (World Scientific, Singapore), Vol. **19**, 223–233 (1999).

[Ric1] Ricca, R.L.: New Developments in Topological Fluid Mechanics: From Kelvin's Vortex Knots to Magnetic Knots. In Ideal Knots, Series on Knots and Everything, eds. Stasiak, A., Katritch, V., Kauffman, L.H. (World Scientific, Singapore), Vol. **19**, 255–273 (1999).

[Ric2] Ricca, R.L.: Tropicity and Complexity Measures for Vortex Tangles, Lecture Notes in Physics v. 571. Springer, Berlin Heidelberg New York, 366–372 (2001).

[R] Rolfsen, D.: Knots and links. Publish or Perish, Berkeley, CA (1976).

[SV] Saka, Y., Vazquez, M.: TangleSolve: topological analysis of site-specific recombination. Bioinformatics **18** (2002) 1011–1012.

[Sc] Schubert, H.: Knoten mit zwei brücken. Math. Z. 65 (1956), 133–170.

[Si] Silver, D.: Knot theorys odd origins. American Scientist **94** (2006), 158–165.

[SK] Stasiak, A., Katrich, V., Bednar, J., Michoud, D., Dubochet, J.: Electrophoretic mobility of DNA knots. Nature **384**, 122 (1996).

[St] Steen, L.A.: The science of patterns. Science **240** (1988), 611–616.

[S1] Sumners, D.W.: Knot theory and DNA. In New Scientific Applications of Geometry and Topology, Proceedings of Symposia in Applied Mathematics **45**, D.W. Sumners, ed., Am. Math. Soc. (1992), 39–72.

[S2] Sumners, D.W.: Lifting the Curtain: Using topology to probe the hidden action of enzymes, Notices of the AMS **42** (1995), 528–537.

[S3] Sumners, D.W.: The role of knot theory in DNA research. Ch. 24 of Geometry and Topology, C. McCrory and T. Schifrin, eds., Marcel Dekker (1987), 297–318.

[S4] Sumners, D.W.: Untangling DNA. The Mathematical Intelligencer **12** (1990), 71–80.

[SE] Sumners, D.W., Ernst, C., Spengler, S.J., Cozzarelli, N.R.: Analysis of the mechanism of DNA recombination using tangles, Quarterly Reviews of Biophysics **28** (1995), 253–313.

[T] Thompson, W.T.: Mathematical and physical papers, III. Cambridge U. Press (1890).

[TA] Trigueros,S., Arsuaga, J., Vazquez, M., Sumners, D.W., Roca, J.: Novel display of knotted DNA molecules by two-dimensional gel electrophoresis, Nucleic Acids Research **29** (2001), 67–71.

[VS] Vazquez, M., Sumners, D.W.: Tangle Analysis of Gin Site-Specific Recombination, Math. Proc. Camb. Phil. Soc. **136** (2004), 565–582.

[VC] Vazquez, M., Colloms, S.D., Sumners, D.W.: Tangle Analysis of Xer Recombination Reveals only Three Solutions, all Consistent with a Single Three-Dimensional Topological Pathway, Journal of Molecular Biology **346** (2005), 493–504.

[MM] Micheletti, C., Marenduzzo, D., Orlandini, E., Sumners, D.W.: Knotting of Random Ring Polymers in Confined Spaces, J. Chem. Phys. **124** (2006), 064903 (1–10).

[WC1] Wasserman, S.A., Cozzarelli, N.R.: Biochemical topology: applications to DNA recombination and replication. Science **232** (1986), 951–960.

[WC2] Wasserman, S.A., Cozzarelli, N.R.: Determination of the stereostructure of the product of Tn3 resolvase by a general method. Proc. National Academy of Sciences USA **82** (1985), 1079–1083.

[WD] Wasserman, S.A., Dungan, J.M., Cozzarelli, N.R.: Discovery of a predicted DNA knot substantiates a model for site-specific recombination. Science **229** (1985), 171–174.

[W] White, J.H.: Self-linking and the Gauss integral in higher dimensions. Am. J. Math. **91** (1969), 693–728.

[V] Vazquez, M. Tangle analysis of site-specific recomgination: Gin and Xer systems. PhD dissertation, Florida State University, Tallahassee, FL (2000).

[VM] Virnau, P., Mirny, L.A., Kardar, M.: Intricate knots in proteins: function and evolution. PLOS Comp. Bio. **2** (2006), e122.

Department of Mathematics, Florida State University, Tallahassee, FL 32306-4510
`sumners@math.fsu.edu`

Reference for this paper:

D.W. Sumners. DNA Topology: Experiments and Analysis. In Knot Theory for Scientific Objects, Proceedings of the International Workshop on Knot Theory for Scientific Objects, A. Kawauchi, ed., OCAMI Studies 1 (2), 213-237 (2007).