The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

2034-19

Advanced School and Conference on Knot Theory and its Applications to Physics and Biology

11 - 29 May 2009

DNA Topology

Dorothy Buck
Imperial College, Department of Mathematics
London SW7 2AZ
UK

Proceedings of Symposia in Applied Mathematics
Volume **66**, 2009

DNA Topology

Dorothy Buck

Abstract. This article briefly surveys the topological properties of DNA. We sketch how the helical structure of DNA intimately links the geometry and topology of the molecule. We then discuss the proteins responsible for regulating this DNA geometry and topology, and the mathematical questions that arise when understanding them. We next turn our attention to the site-specific recombinases, proteins that change DNA topology as a by-product of their primary action. We discuss one topological model that predicts the types of knots and links that arise from site-specific recombination. We conclude with a discussion of the tangle model, and give an extended example where it was useful in understanding the mechanism of a site-specific recombinase.

1. Introduction

Historically, knot theorists were motivated in part by chemical considerations. Lord Kelvin conjectured that different elements were comprised of different knotted vortices of the then-fashionable pervasive ether. (See [**85**] for a great introduction to the early history of knot theory.) His friend Peter Guthrie Tait, who was also interested in vortex rings, began to classify knots and links, and produced the first modern knot tables—on display at the British Library.

While the original applied motivation for understanding knots evaporated, knot theory is once again being utilized by biologists and chemists. Beginning in the mid-1980s, when linked and knotted DNA was first experimentally found, topologists have played an increasing role in exploring the ramifications. This brief article surveys some of the landscape charted out in the new interdisciplinary territory of DNA Topology.

1.1. What this article covers. This article is based on a lecture given at the 2008 AMS Short Course "Applications of Knot Theory", organized by Erica Flapan and myself, and so retains some of the informal tone of a lecture. The audience—then and (intended) now—is mathematicians who know some topology, and are interested in learning more about its interplay with molecular biology. Accordingly, here we give an introductory overview of this, focussing on knots in a fundamental biochemical setting—those formed by DNA. We begin in Section 2, by sketching the basic primary and secondary structure of DNA, and then shift our

2000 *Mathematics Subject Classification.* Primary 92C40, 92E10, 57M25.
Key words and phrases. Knot Theory, Molecular Biology.

focus to the more topologically relevant tertiary structure. In Section 3, we model DNA as a topological ribbon. We describe the Serret-Frenet framing, and introduce the fundamental relation $Lk = Tw + Wr$. In Section 4, we discuss the importance of DNA supercoiling for compactification, base-pair access and the free energy. In Section 5, we give a brief overview of the two families of proteins, type I and type II topoisomerases, that regulate DNA supercoiling and DNA knotting/linking, respectively. We touch on the topological question of type II topoisomerase unknotting, and its relation to the more general subject of knot adjacency. In Section 6, we discuss the primary biological techniques to separate and distinguish DNA knots and links. In Section 7, we introduce another family of DNA topology-changing proteins, the site-specific recombinases. And finally, in Section 8, we discuss the tangle model for site-specific recombination.

1.2. What this article doesn't cover, and where to read more. Both for space and expertise constraints, this article is a necessarily brief introduction to an extraordinarily rich interdisciplinary area. So unfortunately, there are a number of fascinating topics that we've only touched on or entirely neglected. For example, we only briefly sketch the basic background material for the energetics of DNA supercoiling. Similarly, our understanding of the structure and mechanism of type II topoisomerases has radically improved in the past 18 months, particularly with the emergence of the first crystal structure of a yeast-DNA co-complex [**29**]. Interested readers may also want to consult the Proceedings of the Topoisomerase conference held in Norwich in July 2008 for more details. Likewise, we will gloss over the many models of DNA migration through an electrophoretic gel. Also, we will not discuss the beautiful rod models for DNA.

There are a number of general expository treatments of (facets of) this area. We encourage the interested reader to consult, for example, the books of Vologodskii [**98**] and Frank-Kanemetskii [**37**], as well as overview articles by Pohl [**73**] and Weber [**103**] (the last in French, but with a nice historical overview), and the similar 1992 AMS short course [**28**]. From a more biological viewpoint, we recommend Maxwell and Bates' *DNA Topology* [**65**], and *Mobile DNA* for more on recombination. [**21**].

2. Structure of DNA

2.1. Primary and secondary structure. DNA, Deoxyribonucleic Acid, is the molecule (or group of molecules) responsible for encoding all genetic information and instructions in living organisms. The information stored in DNA determines all hereditary traits, for example, eye colour and susceptibility to inherited diseases. For this reason, DNA is often referred to as the blueprint (or more accurately the recipe book) for life.

The structure of DNA was famously pieced together by Francis Crick and James Watson, based on beautiful X-ray images of Rosalind Franklin [**102**]. DNA is composed of pairs of repeated units called nucleotides. Each nucleotide consists of a phosphate group, a (2′-deoxyribose) sugar, and one of four bases: Adenine, Thymine, Guanine or Cytosine.

These nucleotides form a twisted ladder, the famous 'double helix' [**102**]. The alternating sugar and phosphate groups form the backbones, or sides of the ladder, and stacked pairs of bases form the rungs of the ladder. These bases always pair

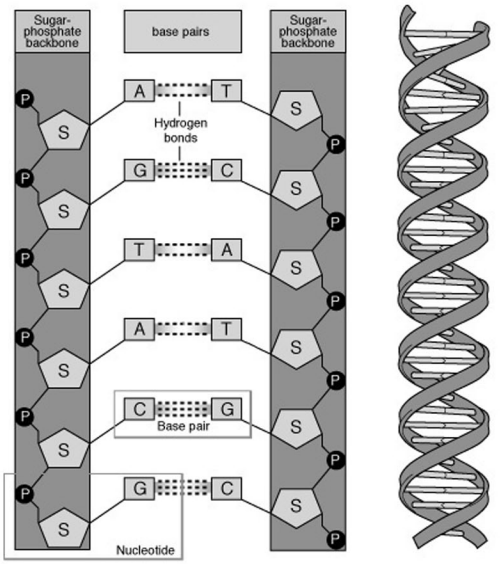

FIGURE 1. Primary and secondary structure of DNA.

either A with T or G with C. An AT 'rung' is formed via 2 hydrogen bonds via A and T, while the stronger GC rung consists of 3 hydrogen bonds between G and C.

The length of DNA is measured in terms of these base pairs (bp); for example, human genomic DNA is approximately 3 billion base pairs, while that of the most common bacteria, *E. coli*, is approximately 4.4 million base pairs. In more absolute measurements, each nucleotide is approximately 3 Angstroms ($= 3 \times 10^{-10}$m) in length. The average width of the double helix is approximately 25 Angstroms.

The sugar in DNA is a pentose sugar, with five carbons. Each sugar shares a (phosphodiester) bond with its two adjacent phosphates, one at its $3'$ hydroxyl (OH-group) and one at its $5'$ hydroxyl. These (phosphodiester) bonds then give each backbone an orientation, and the two strands are aligned antiparallel: the so-called Crick strand running $5'$ to $3'$ and the Watson strand running $3'$ to $5'$. Because of the monogamous base pairing, DNA sequences are typically written with respect to the $5'$ to $3'$ direction, (*i.e.* along the Crick backbone), for example, GATTACA.

The precise spelling of this sequence, the primary structure, encodes all hereditary information, as well as instructions for all cellular processes.

In vitro, the zoo of DNA forms is staggeringly diverse; for example, see Ned Seeman's companion article for some fascinating nanostructures, including truncated octrahedra and knots, synthesized from short (typically single-stranded) sequences [**82**]. Even *in vivo*, there is a wide menagerie—including left-handed helices, right-handed helices whose backbones zigzag, helices whose bases are on the outside of the sugar-phosphate backbone, or even short stem and loops. (See e.g. Maxwell and Bates [**65**] for a description of these forms.) In this article, we will focus our

attention on the most prevalent (and hence relevant) structure, B-form DNA. This is a regular right-handed helix with bases on the inside. Depending on its environment, there are approximately 10.5 base pairs per complete turn of the helix. We will discuss this in more detail below. There is both a major and minor groove in the helix. These grooves allow access to the base pairs.

The secondary structure of DNA, its helical geometry, also has deep biological consequences. For example, the double helix is more difficult to unzip than a straight ladder. This difficulty in accessing the base pair rungs results in fewer accidental changes to the DNA sequence, thus protecting the genetic code.

2.2. Tertiary structure. A decade after the double helix discovery, Vinograd found the fundamental tertiary structural feature: the axis of the DNA can also be coiled in space [**84**], leading to 'supercoiled' DNA. (To visualize this, imagine the handset cord of an often-used phone: it is highly twisted, possibly crossing itself and is under tension.)

The DNA axis is often linear, as in the case of human genomic DNA. Linear DNA is often attached to a protein scaffolding which constrains the topology, for example during nucleosomal compactification discussed in Section 4.1 below.

Perhaps more interestingly, the DNA axis is frequently circular—this occurs when the ends of the 2 backbone strands are covalently bonded. Bacterial genomic DNA, chloroplast DNA, and human mitochondrial DNA are all examples of DNA with a circular axis. Additionally, most DNA used in biochemical labs, plasmid DNA, is circular.

Even more interestingly, the central DNA axis can also be knotted or linked. For example, DNA knots and links (known as *catenanes* in the biological literature) occur during replication (DNA copying) and recombination (DNA rearranging). We will discuss DNA knots and links below in more detail. For now, we emphasize that the constrained or circular axis, coupled with the two levels of coiling (the double helix and supercoiling), already leads to the nontrivial topology of DNA.

3. DNA as a topological ribbon: $Lk = Tw + Wr$. The interplay between topology and geometry

From the outset, characterizing supercoiled DNA required a combination of mathematicians and biologists. Călugăreanu [**13**] and White [**105**] formulated a fundamental relation for space curves: $Lk = Tw + Wr$. Two geometric quantities, *Twist* and *Writhe* (that may change under deformations of the curves) sum to a topological quantity, *Linking Number* (which is invariant under such deformations). It took the combined efforts of Fuller (a mathematician) and Vinograd (a biologist) to translate Călugăreanu and White's and relation into biological terms and recognize its fundamental importance for DNA [**39**].

To understand this relation, we shall model a molecule of circular DNA as a twisted ribbon. For each of the two backbones, we associate a curve in 3-space (an edge of the ribbon). By convention we name the backbones C and W, after Crick and Watson. The orientation of these two ribbon edges will be taken to be parallel, and so not actually inheriting the natural orientation from the underlying chemical structure.

These curves C and W are linked around each other, and in fact the ribbon itself can be knotted. Likewise, if we consider more than one DNA molecule, the ribbons themselves can be linked. We formalize this in terms of the linking number, Lk:

DEFINITION 3.1. Given a projection of an oriented two-component link, to each crossing assign a ± 1, depending on the orientation of the two strands.[1] The *linking number, Lk* is one-half their sum:

$$Lk = \frac{1}{2} \sum_{crossings} \pm 1.$$

As discussed in Colin Adams' companion article, Lk is independent of the particular projection and is a topological invariant of the embedding of the ribbon [**1**]. (As before, we say two embedded ribbons are *equivalent* if there exists an orientation-preserving diffeomorphism of S^3 or $\mathbb{R}^3$ which sends one ribbon onto the other.

For a circular DNA molecule, the two DNA backbones, C and W, form a $(2, m)$-torus link. A $(2, m)$-torus knot or link is one which can be drawn so that all of its crossings occur as a row of m (positive or negative plectonemic) crossings, as illustrated in Figure 10. Note that if m is odd, then $T(2, m)$ is a knot and if m is even, then $T(2, m)$ is a link.

Lk can be defined in a variety of equivalent ways; see Rolfsen [**79**]. One alternate description of Lk is in terms of Gauss' integral:

$$Lk = \frac{1}{4\pi} \int_C \int_W \frac{(x'-x)(dydz' - dzdy') + (y'-y)(dzdx' - dxdz') + (z'-z)(dxdy' - dydx')}{[(x'-x)^2 + (y'-y)^2 + (z'-z)^2]^{3/2}} \tag{3.1}$$

where (x, y, z) ranges over C and (x', y', z') ranges over W. (Note: This is the pullback of the area form on S^2 in S^3.)

3.1. Frenet framing. Roughly, we construct the Frenet framing for a regular closed space curve $\gamma(s)$, as follows. (For a more detailed treatment, see [**74**].) Suppose we have a simple closed curve $\gamma(s)$, parametrized by arc length s. Then the tangent vector, T, has length 1: $||T|| = ||\dot{\gamma}(s)|| = 1$. Then $\dot{\gamma}(s).\dot{\gamma}(s) = 1$, and so $\ddot{\gamma}(s)$ is perpendicular to $\dot{\gamma}(s)$, i.e. $\ddot{\gamma}(s)$ is a normal vector. (Note: Since $\gamma(s)$ is regular, then $\dot{\gamma}(s) \neq 0$ by definition.) For simplicity, we will normalize this so that our normal vector, N, is also a unit vector: $N = \frac{\ddot{\gamma}(s)}{||\ddot{\gamma}(s)||}$ for all s such that $\ddot{\gamma}(s) \neq 0$. We then consider the binormal $B := T \times N$.

Since T, N and B are orthonormal, they form a basis so that any other vector can be described as a linear combination of these—in particular, each of the vectors $\dot{T}, \dot{N}$ and $\dot{B}$ can be described as such. Given $\dot{T} = aT + \kappa N + cB$, for some a, κ, c, then $a = 0$ since by above $\dot{T}$ is perpendicular to T. Also, since $N = \frac{\dot{T}}{||\dot{T}||}$, then $c = 0$ and so $\kappa = ||\dot{T}||$. (κ is often called the curvature; at a given point on $\gamma(s)$, since T and $\dot{T}$ span a plane, within that plane one can draw a circle tangent to $\gamma(s)$ with maximal radius $\frac{1}{\kappa}$.)

Similarly, $\dot{N} = -\kappa T + \tau B$, where τ is often called the torsion, a measure of how nonplanar the curve is. Finally, $\dot{B} = \dot{T} \times N + T \times \dot{N} = 0 + T \times (-\kappa T + \tau B) = -\kappa\tau T + \tau^2 B = \tau \dot{N}$.

[1] Unfortunately, the convention adopted by biologists in the field is opposite to the usual topologists' convention.

These 3 relations form the Serret-Frenet equations:

$$\begin{aligned} \dot{T} &= \kappa N, \\ \dot{N} &= -\kappa T + \tau B, \\ \dot{B} &= -\tau N. \end{aligned} \tag{3.2}$$

These crossings of the 2 backbone curves come in two flavours: local and nonlocal. Local crossings occur when viewing a segment of the ribbon almost edge-on, while nonlocal crossings occur when one segment of the ribbon crosses entirely over another. In the latter case the axis of the ribbon is also crossing over itself. (For an elegant description of this, and a rigorous characterization of Twist in terms of these local crossings; see [**25**].)

3.2. Twist. The local crossings reflect how tightly the ribbon wraps (or twists) around its central axis. In terms of our DNA double helix, these local crossings indicate the helical pitch (i.e. number of base pairs per complete revolution).

Intuitively, one can imagine stringing a penny onto the curve. If we marked the northern pole, then Tw would measure how this marked point changes as we slide the penny along the curve.

More precisely, we quantify this as *Twist*, Tw, using the Frenet framing, as the total torsion of the curve $\gamma(s)$:

$$Tw = \frac{1}{2\pi}\int_{\gamma(s)} (T \times N).\dot{N} ds = \frac{1}{2\pi}\int_{\gamma(s)} \tau ds.$$

Then in the Frenet framing, $(T \times N).\dot{N}$ is the projection of $\dot{N}$ in the direction of B, so by the Serret-Frenet equations, $= \tau$.

3.3. Writhe. Writhe is an indicator of how the ribbon axis itself is contorted in space. If we consider the axis as inheriting the orientation of the two ribbon edges, we can sum the signed crossings of the axis with itself. This quantity is called the *writhe*, Wr. Note that Wr is projection-dependent, in particular a Reidemeister Type I move will contribute a ± 1 to Wr. Unsurprisingly then, Wr is not a topological invariant—it depends on the particular geometry of a given projection. (Note that this terminology varies in the literature. For example, Weber calls the signed sum of the crossings of a projection the *Tait number* of the projection. Then the writhe of a simple closed curve in S^3 or $\mathbb{R}^3$ is the average of the Tait numbers over all projections [**103**]. However, we will follow the (perhaps misleading) most common nomenclature used in the biological literature.)

More formally, we can consider writhe in terms of Gauss' integral, as in formula (3.1), with the axis representing both the C and W curve.

With this definition Wr can then be thought of in terms of the Gauss integral as Lk above, with both integrations along the same curve (the ribbon axis).

For a DNA molecule, the self-crossings of the DNA axis contribute to writhe. The closest macroscopic analogue is the handset of an old-fashioned telephone receiver, attached to the base by a helical cord. During normal use, the cord gains writhe, which is released by holding the cord near the base and letting the dangling handset spin.

As mentioned above, this relation between the topological and geometric properties of a ribbon was first discovered by Călugăreanu [**13**] and White [**105**]. Its relevance to DNA was first noted by Fuller [**39**].

THEOREM 3.2 (Călugăreanu-White-Fuller). $Lk = Tw + Wr$.

Biologists often visualize this correspondence using a length of Taigon clear plastic tubing, with two parallel curves drawn along the length, each representing one edge of the ribbon. By twisting and/or coiling the tubing before closing the ends, it helps to visualize the relation between Twist and Writhe.

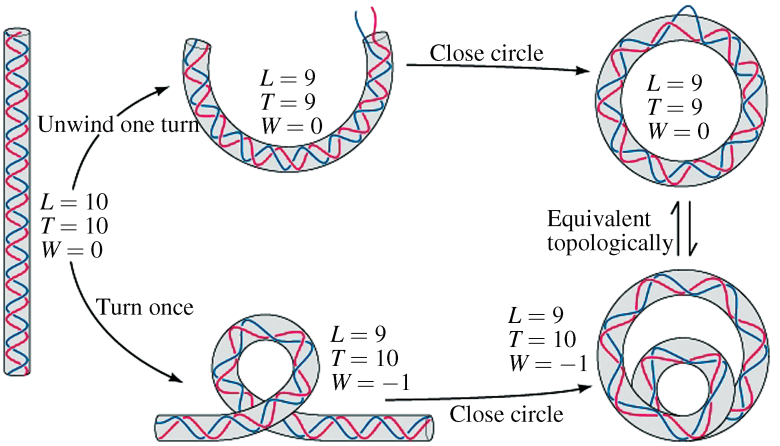

FIGURE 2. Twist and Writhe. From *Biochemistry* by Donald Voet and Judith G. Voet (2004). Reproduced with permission of John Wiley and Sons Ltd.

For a modern treatment of this, see Dennis and Hannay [**25**]. For a more detailed discussion of these differential geometric notions in terms of DNA, see Weber [**103**] or Pohl [**73**].

4. Biological importance of DNA supercoiling

Recall that within the context of circular DNA, nontrivial writhe occurs most commonly when the DNA axis wraps around itself in a helical manner, as discovered by Vinograd and colleagues [**84**]. This is more precisely referred to as *plectonemic supercoiling*, to reflect the coiled (writhed) axis of the already coiled (helical) ribbon. Throughout we will use the term 'supercoiling' to mean 'plectonemic supercoiling'. (DNA can also wrap around proteins, as discussed below, and this type of nontrivial writhe is referred to as *solenoidal supercoiling*.)

This 'supercoiled' DNA was found to be the native state in virtually all cells [**17**]. DNA extracted from cells from a wide variety of organisms, including cauliflower, humans, mice and monkeys, was found to be (plectonemically) supercoiled at approximately the same density.

We can quantify this density as follows. Given N, the number of base pairs of DNA and h, the number of base pairs per helical repeat in given experimental conditions, then the specific linking difference is σ:

$$\sigma = \frac{Lk - N/h}{N/h}.$$

DNA supercoiling is ubiquitous because supercoiling is crucial for three primary reasons. Firstly, it compactifies the DNA molecule. Secondly, the handedness of the supercoiling opens the major/minor grooves for easier access to the base pairs. And thirdly, supercoiled DNA provides an important source of free energy for cellular reactions.

4.1. Compactification. In all organisms, the physical length of genomic DNA is considerably larger than the diameter of the region it resides within. In prokaryotes, the DNA floats freely around the cytoplasm, while in eukaryotes the DNA is encapsulated within the cell nucleus. Eukaryotic genomic DNA is orders of magnitude longer, so while the typical compactification needed to store DNA in prokaryotes is significant, it is even more dramatic in eukaryotes. For example, the most common bacteria *E. coli* has genomic DNA of length approximately 1.5mm, and cell diameter less than 1 μ meter, requiring on the order of 10^3 compactification. Human genomic DNA is approximately a meter in length and a typical cell nucleus has diameter less than 1μm, so the compactification needed is on the order of 10^6. One of the most challenging packing problems is that of the South American lungfish, whose DNA is on the order of 35 meters.

Supercoiling is the first step of a hierarchy of compactification, the later stages of which are still poorly understood. The (plectonemically) supercoiled DNA molecule then wraps around large proteins called histones. Then the histones string together like pearls on a necklace. This chain of histones is then further compacted, possibly as suggested in Figure 3.

Thus supercoiling allows the DNA to be packaged in a tightly confined volume in a highly structured, organized manner. The orderliness of this process is crucial for efficient information retrieval—as (segments of) DNA must be accessed rapidly and constantly by proteins.

4.2. Access to the base pairs. The handedness of DNA supercoils *in vivo* is negative. That is, the helix formed by the DNA axis is opposite in handedness to the original double helix. If the circular molecule remains covalently closed while supercoiling, then the linking number Lk remains constant. Then $Lk = Tw + Wr$ implies that their must be a compensating change in Tw.

The handedness of the supercoils means that the molecule becomes undertwisted. Then the major and minor grooves, illustrated in Figure 1, open a small amount. This allows easier access to the corresponding base pairs, facilitating a wide variety of reactions. It also makes it easier to separate the DNA strands, necessary for example for copying DNA.

4.3. Free energy. As noted above, unzipping the DNA double helix can be difficult. But many essential reactions, including DNA copying, demand exactly this. Fortunately, Theorem 3.2 implies that the linking number of negatively supercoiled DNA (the natural state of DNA in cells) will be lower than that of relaxed DNA. It is this linking number 'deficit' of supercoiled DNA that provides the necessary energy for local strand separation.

In addition, for many biological processes involving DNA, including helical wrapping around histones and phage head packing, the DNA axis must be twisted or bent [**83**]. To understand the energetics of these processes, we briefly discuss how the free energy depends on linking difference.

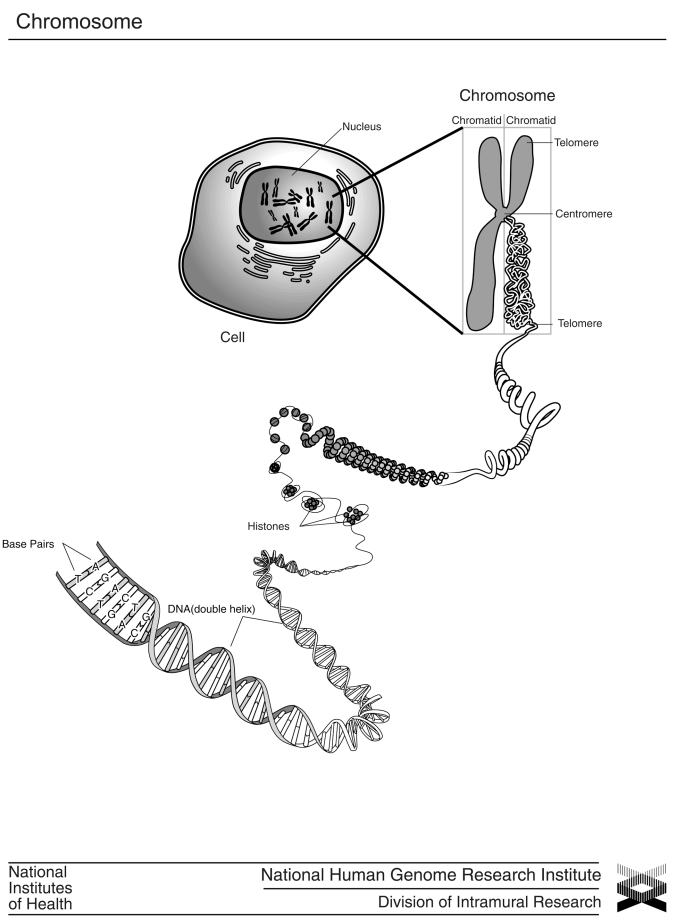

FIGURE 3. Compactifying DNA. Courtesy: National Human Genome Research Institute

For DNA rings larger than 2000bp, it has been experimentally shown in [**27**], [**75**], that the free energy of supercoiling, ΔG, obeys the quadratic relation

$$\Delta G = K(\alpha - \alpha_0)^2 \tag{4.1}$$

where K is a proportionality constant (which depends on N, the number of base pairs), α is the linking number, and α_0 (for large DNA) is the average linking number. (Note that α_0 can be independently defined as N/h_0 where h_0 is the helical repeat, or average number of base pairs over which the helical twist of the DNA double helix increases by one.) For large N, the product NK has been experimentally shown to be independent of length [**49**].

5. Regulating DNA topology

DNA molecules that differ only in Lk (not, *e.g.* sequence or length) are called *topoisomers*. By standard thermodynamics, the concentration of DNA with linking

number α_i, $[\alpha_i]$, is

$$[\alpha_i] = \frac{1}{Z} \exp\left(\frac{-\Delta G}{RT}\right), \tag{5.1}$$

where Z is a normalization constant, T is the temperature in Kelvin, and R is Boltzmann's constant. Thus, the concentration of a given topoisomer with linking number α_i, is

$$[\alpha_i] = \frac{1}{Z} \exp\left(\frac{-K(\alpha_i - \alpha_0)^2}{RT}\right). \tag{5.2}$$

The graph of α_i v. $[\alpha_i]$ for equilibrium populations of topoisomers of large DNA rings is a normal distribution with center α_0 and standard deviation $\sqrt{\frac{RT}{2K}}$.

To maintain this equilibrium population, two families of proteins have evolved, the *topoisomerases*[2]. Topoisomerases are so-called because the primary function is to interconvert between topoisomers. Topoisomerases are classified into two families, depending on whether they cleave 1 DNA backbone and hence change the linking number by steps of 1 (Type I topoisomerases) or both backbones and hence change linking number in steps of 2 (Type II topoisomerases).

We briefly discuss these two families below, but emphasize that we are glossing over many of the subtleties of topoisomerases—in particular, the minor differences between family members of a given type. We refer the reader who is interested in exploring more in these remarkable proteins to look at, e.g. [**65**].

5.1. Regulating DNA supercoiling: Type I topoisomerases. Given the fundamental importance of DNA supercoiling, it is perhaps not surprising that there are proteins whose sole function is to regulate the amount of supercoiling—the Topoisomerase I family. Members of this family relax (negative, and in eukaryotes, also positive) supercoils (See Figure 4.)

Type I topoisomerases effectively harness Theorem 3.2, to convert a change in twist to a change in writhe (supercoiling). They bind (nonspecifically) to the DNA molecule, make a transient break in one of the DNA backbones and then pass the other backbone through before resealing and releasing the DNA. This changes the twist, and thus for a circular or topologically constrained molecule, changes the writhe.

One example where type I topoisomerases naturally come into play is when the DNA helix is unzipped, for example during DNA replication (copying). If the molecule is constrained at the end where the unzipping fork is heading, then supercoils build up in advance of the unzipped region. If a type I topoisomerase does not release the supercoils, then eventually the torsional strain becomes too extreme and the DNA molecule breaks.

Type I topoisomerases are found in all organisms studied thus far, and loss of these proteins is lethal to the cell.

[2]We will consider proteins, imprecisely, as small molecular machines that act on DNA. Like DNA, they are biopolymers. The fundamental unit is an amino acid (e.g. lysine, tyrosine or glutamine), rather than a nucleotide. There are 20 different flavours of amino acids, and the exact composition and ordering of these amino acids determines the shape and function of the protein.

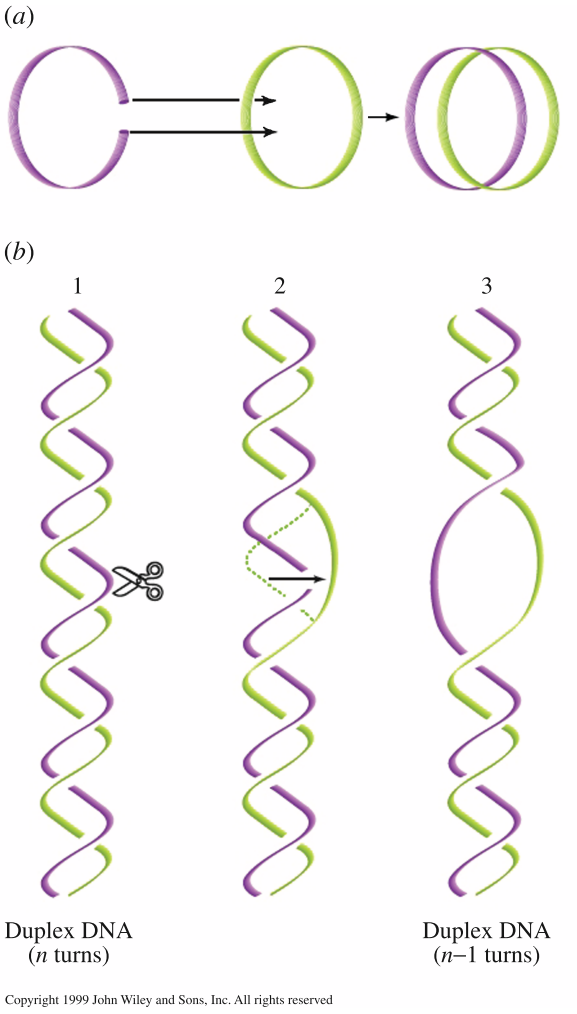

FIGURE 4. Action of Topoisomerase I. From *Biochemistry* by Donald Voet and Judith G. Voet (2004). Reproduced with permission of John Wiley and Sons Ltd.

5.2. Regulating DNA knotting and linking: Type II topoisomerases. Like type I topoisomerases, Type II topoisomerases are also both ubiquitous and essential. Their primary function is to change DNA knot or link type, but they can also remove or (in the case of DNA gyrase) add DNA supercoils.

One example where type II topoisomerases naturally come into play is when circular DNA (such as bacterial genomic DNA) is replicated. The end result of this process is two linked circular molecules. If a type II topoisomerase does not unlink the two daughter molecules, then the bacterial cell cannot divide properly and commits suicide.

Because of their crucial role in many cellular processes, type II topoisomerases have been drug targets for both human cancer and infectious diseases [**38, 61**]. For example, since bacterial type II topoisomerases differ from human, many antibiotics

work by inhibiting these proteins, killing the bacterial infection by inhibiting unlinking of daughter DNA as described above. Thus from both a basic biochemical and pharmaceutical perspective, understanding the mechanism by which type II topoisomerases act has been very important.

5.3. Mechanism of type II topoisomerases. One might assume that type II topoisomerases, as small molecules, act independently of the global DNA topology. Surprisingly, however, they act in a manner that preferentially *un*knots and *un*links DNA—this is known in the literatures as topological simplification. Furthermore, type II topoisomerases simplify DNA topology in an extraordinarily efficient manner [**80**]. Exactly how this is achieved has been intensely debated.

What is generally agreed on is that a generic type II topoisomerase binds to helix-helix juxtapositions, such as a supercoil or knot/link crossing (see Figure 5). The first helix, called the G (for Gate) segment is broken to allow the T (Transported) segment to pass in a unidirectional manner [**3**]. The G segment is then resealed, resulting in a crossing change (from $+1$ to -1 or *vice versa*). This process relies on ATP hydrolysis. [**3**]. A beautiful recent crystal structure of a type II topoisomerase bound to the G segment elegantly demonstrates that a type II topoisomerase severely bends the DNA (approximately 150°) [**29**]. This supports models that incorporate DNA bending [**11, 99**]. We discuss two in some detail below.

The first model, jointly proposed by a topologist (G. Buck) biologist (L. Zeichiedrich) [**11**], describes how the local information obtained at a crossing can be used to determine the global topology. They classify crossings in terms of several vector parameters, and show that certain crossing types, 'hooked junctions', are associated to nonsupercoiled crossings. Crossing changes at these hooked junctions are then more likely to lead to topological simplification.

The second model [**99**] by Vologodskii and co-workers, gives numerical (Monte Carlo simulations) and indirect experimental evidence that a type II topoisomerase actually bends the G segment into a hairpin conformation upon binding. There is some experimental evidence that DNA knots localize, which would lead to the T-segment inside the hairpin more often for knots than for unknots.

5.4. Topological and numerical strategies to understand type II topoisomerase unknotting. Topologically, understanding type II topoisomerase unknotting is part of a larger movement to understand what knots can be obtained from a given one via a single crossing change, that is, by a single round of topoisomerase relaxation. In general, this is very large question—for example, one area of knot theory research has been to (so far incompletely) classify all knots with 'unknotting number 1'—that is, all knots that become the unknot after a single crossing change.

More generally, the Gordian or strand passage distance between any 2 links or knots K and L, $d(K, L)$ is defined to be the minimum number of crossing changes ($+1$ to -1 or *v.v.*) needed to convert K into L, taken over all projections. It is a metric on the space of knot types—e.g. $d(K, L)$ is less than or equal to the sum of their unknotting numbers.

In this framework, we can consider which knots L are adjacent to a given knot K—i.e. L, s.t. $d(K, L) = 1$ or equivalently which knots L are a single TopoII move from K. Recently, there has been a flurry of activity in this area of 'knot

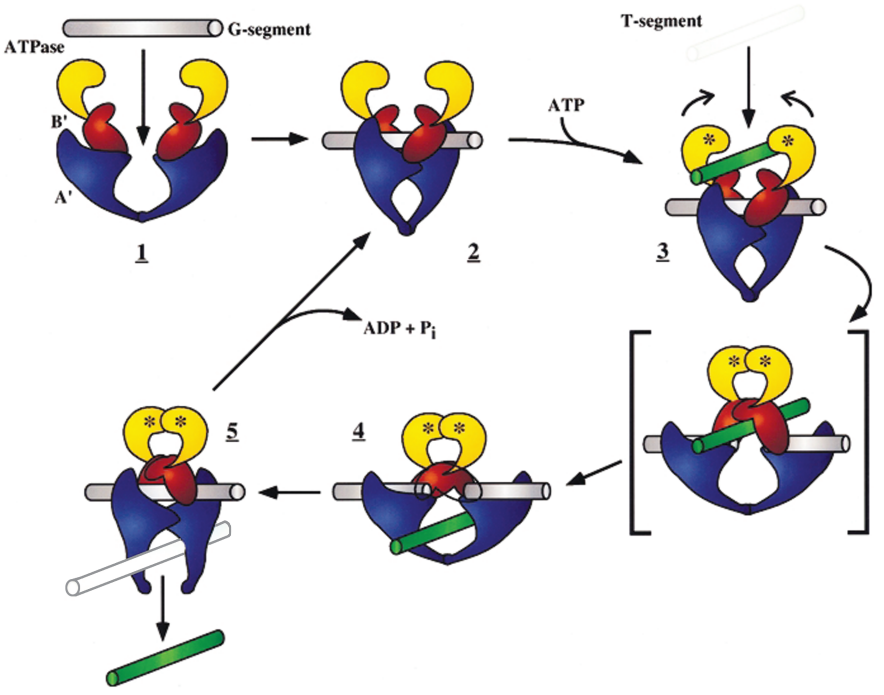

FIGURE 5. Rough mechanism of a type II topoisomerase. Reprinted by permission from Macmillan Publishers Ltd.: *Nature*, Structure and mechanism of DNA topoisomerase II. J. M. Berger, S. J. Gamblin, S. C. Harrison, J. C. Wang (1996).

adjacency'. The problem has been attacked using both finite-type invariants and Dehn surgery techniques. There has been some progress using both, in particular to determine bounds on adjacency number in terms of knot genus. (Recall that the genus g of a knot K is the minimal genus of an orientable spanning surface for K.) One example of this is via knot concordance, since distance g knots can be thought of as together bounding a smoothly embedded genus-g surface in 4-space. Applications to unknotting number and knot adjacency arise from the Heegaard-Floer knot homology of Ozsvath-Szabo and Rasmussen and from Khovanov-type knot homologies (due to work of Rasmussen [**76**] and Lobb [**64**]). Each of these theories are now combinatorial in nature, and so provide computable lower-bounds for knot adjacency. In particular, they have been used to classify all knots K with 9 or fewer crossings and $u(K) = 1$ [**71**].

Because DNA is plectonemically supercoiled, then some of the most common DNA knots and links are members of the *4-plats*: knots or 2-component links that admit a projection consisting of a braid on 4 strings, with one strand free of crossings[3]. Considering the smaller question for 4-plat knots and links has been more tractable, and there has been quite a bit of progress in this area already. For example, Kanenobu and Murakami classified all unknotting number one 4-plat knots [**52**], and Kohn classified all unlinking number one (necessarily 2-component) 4-plat

[3]We will discuss 4-plats in more detail in Section 8 below.

links [**55**]. More generally, a classification of adjacent 4-plats was given independently by Torisu, Darcy and Sumners, and John Berge [**93, 24**]. So in theory one could determine exactly when two knots were related by a single TopoII move, and the distance between any two 4-plats (i.e. the minimum number of times TopoII must act to convert K into a different given 4-plat type). For approximately 20 low-crossing knots and their mirror images, Darcy has a table determining these distances [**24**]. On the Heegaard-Floer homology side, although *a priori* the homologies are harder to compute the higher the crossing number of K, for 2-bridge knots the computations are easier (for example [**77**]).

In parallel, numerical simluations by Stasiak and collaborators have investigated the probability of K being converted into L (where possibly $K \simeq L$) via a single type II topoisomerase [**90**]. Loosely, they begin with a closed knotted polymer chain, randomize it via crookshank moves, and then allow 1 more move that could (or could not) involve a single strand passage. They then calculate the new knot type. Some interesting things emerge—*e.g.*, in 95% of the cases for the unknot, a single intersegmental passage returned the unknot again.

They (nicely) imagine the knot space as a foam, with a crossing change corresponding to moving between two adjacent bubbles. Then they consider the areas between different neighbouring knot spaces, in particular, whether there is a relation between the physical 3-D space available for configurations of a given knot and the high-dimensional configuration space occupied by the same knot type. They argue that only the probability to maintain the knot topology (*i.e.* the unknot remaining the unknot after a single crossing change) is related to the Length/Diameter ratio of the ideal configuration of the knot. This probability gave them a way to estimate the surface-to-volume ratio of configuration spaces of various knots.

6. Determining DNA topology

As discussed in Section 2.2, DNA knots and links (*aka* links) have been implicated in a number of cellular processes (see [**65, 18**] and references therein). They can occur as a result of replication (discussed above) and recombination (discussed below). Knots and links also arise as the *products* of, i.e. as the result of enzyme action. In addition to the topoisomerases, discussed above, knots and links also arise as products of reactions from two other protein families, the recombinases and transposases [**18, 72**]. Most prevalently, DNA knots and links arise as products of certain laboratory experiments (called 'topological enzymology' experiments) on artificially constructed small (5–10 kb) circular molecules [**101, 47, 66, 51, 22, 48, 23, 20, 86, 16**]. These product knots and links can help determine the binding and mechanism of the protein being studied. The variety of DNA knots and links observed has made biologically separating and distinguishing these molecules a critical issue.

Knotted DNA was first discovered in the lab in 1981 by Liu and Davis [**63**]. Experimentally, DNA knots and links can be resolved in two ways: via electron microscopy or electrophoretic migration [**56, 92, 107**].

DNA molecules have been visualized beautifully by electron microscopy. (See Figure 6.) In this process, the entire DNA molecule is ensheathed in a (RecA) coating, which thickens and stiffens the molecule. It is then possible to determine the sign of the axis crossings and thus definitively determine the precise knot or link

type. However, this process can be labourious and difficult, particularly producing a relatively large amount of knotted/linked DNA and deciphering the sign of crossings. Perhaps unsurprisingly then, there are only a few dozen such published electron microscopy images of DNA knots and links, and only a few labs (most notably Andrzej Stasiak's) that have developed the necessary expertise.

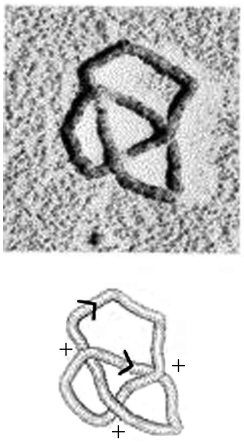

FIGURE 6. Electron microscope image of knotted DNA. Courtesy: Shailja Pathania

A much more widespread (but incomplete) technique to separate DNA knot and link types is that of *agarose gel electrophoresis*. To do this, the DNA is first nicked—that is, one backbone strand is cut to release any supercoils. Next, one prepares a slab of jello-like substance from agarose, a seaweed derivative. This gel slab is then placed in an aqueous bath, a current is run through the bath, and DNA is pipetted into one end of the gel. Recall from Section 2.1 that the DNA backbones are composed of alternating sugar and phosphate groups. These phosphate groups are negatively charged, and so the DNA will migrate to the positive end of the gel. If all the nicked DNA pipetted in is of the same molecular mass and sequence, then this process, *gel electrophoresis*, will stratify DNA knots and links. Under UV light, this appears as a series of dark bands marching down the white gel. Each band corresponds to many DNA molecules, and its intensity is proportional to the amount of DNA present.

Gel electrophoresis is straightforward and requires relatively small amounts of DNA. Typically the distance a given knot or link migrates through the gel is proportional to the minimal crossing number (MCN, the fewest number of crossings with which it can be drawn). Under standard conditions, knots of greater MCN migrate more rapidly than those with lesser MCN [**53, 91, 60**][4]

However, there are 1,701,936 knots with MCN ≤ 16, so a better stratification is needed to positively identify a particular knot [**50**]. By cleverly running the DNA in a second dimension, 2-dimensional gel electrophoresis can separate some prime knots with the same MCN [**92**]. Unfortunately, there is no clear relationship between relative migration of knots with the same MCN in the second dimension.

[4]However, there are gel conditions where, for example, the unknot will migrate ahead of the trefoil.

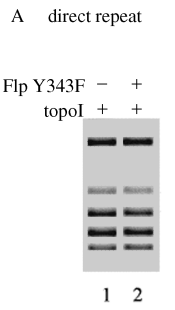

FIGURE 7. Gel electrophoresis of linked DNA. In this image, the DNA was travelling towards the bottom of the page. (Author's Gel.)

In some cases, for DNA of a given length, (1-dimensional) gel electrophoresis can separate some knots with the same MCN. For example, the five and seven-crossing torus knots migrate more slowly than the corresponding five and seven-crossing twist knots [**53, 97**]. This has not generalized, although recent experiments tantalizingly indicate that knots and links may migrate linearly with respect to the average crossing number of a particular conformation—the *ideal configuration*[5] of the knot or link [**97, 58, 54**].

To further complicate matters, gel electrophoresis is incomplete in that adjacent bands determine only relative MCN or average crossing number, not precise values. That is, give two DNA molecules A and B, if A runs further through the gel than B, then A has a higher MCN than B, but one cannot immediately generalize whether MCN(A) = MCN(B) + 1 or MCN(A) = MCN(B) + 2, etc. Nor can one determine if MCN(B) = 3, say, or 4 without further analysis.

Thus gel electrophoresis and electron microscopy are often used in tandem. First DNA is run through a gel to isolate a knot with a given MCN, and the particular band of interest is physically cut out of the gel. This DNA is further purified, and then examined via electron microscopy to determine the precise knot or link type.

To attempt to determine the exact knot or link type without resorting to electron microscopy, one must also construct an appropriate control: a knot ladder. This is a solution of DNA molecules whose precise MCN are known, which is then pipetted into an adjacent well to the unknown DNA in the gel. These known knots or links then serve as markers in the gel in which to calibrate the DNA molecules of interest. While this can be done in some cases (*e.g.* T4 topoisomerase will produce a ladder of twist knots [**100**]), generating such a ladder of known knots and links from DNA of the same length and similar sequence as the unknown knots is highly nontrivial.

Because of the challenges inherent in both electron microscopy and gel electrophoresis, new methods for determining (or predicting) the precise DNA knot or

[5] *Ideal geometric configurations* of knots are the conformations that allow maximal radial expansion of a virtual tube of uniform diameter centered around the knot [**45, 54**]. There is a growing amount of mathematical literature on these ideal configurations. Originally they were determined computationally, but recent analytic work ([**15**] and separately [**41**]) has proven the existence and uniqueness of these solutions.

link type has been an active arena for both mathematicians and experimentalists (see below). More generally, modelling DNA migration through a gel is an active area of research.

7. Changing DNA topology: Site-specific recombination

As discussed above, the sole function of type II topoisomerases is to change DNA knot or link type. We now turn our attention to knots and links that arise from the action of another family of proteins, the *site-specific recombinases*. These proteins mediate *site-specific recombination*, the reshuffling of the genetic sequence, for example changing GATTACA into ACATTAG.

Site-specific recombination is important because of its key role in a wide variety of biological processes. The result of site-specific recombination is the deletion, insertion or inversion of a DNA segment. (See Figure 8.) This corresponds to a wide variety of physiological processes, including crucial steps in viral infections.

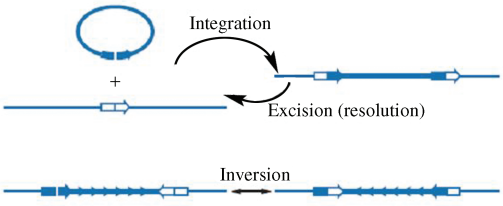

FIGURE 8. Outcomes of site-specific recombination.

In addition to their inherent biochemical interest, pharmaceutical and agricultural industries have become increasingly involved in genetically modifying organisms or testing whether a mutation in a particular gene leads to a disease. As a result, these industries are now interested in site-specific recombinases as tools for precisely manipulating DNA (*e.g.* [**36**]).

While changing DNA topology is not the primary function of these site-specific recombinases, it can be a byproduct of the reaction. If the original circular DNA is supercoiled, the supercoils can be converted into knot or link crossings during the process of recombination. (See [**7**] or [**21**] for more information.)

7.1. The mechanics of site-specific recombination. Minimally, site-specific recombination requires both the site-specific recombinase and two short (30–50bp) DNA segments, the *crossover sites*, inserted into 1 or 2 small circular DNA molecules [**44**]. If there are 2 crossover sites on a single molecule of circular DNA, they can be in either *direct* orientation (head-to-tail, *e.g.*, ...ATGC...ATGC) or *inverted* orientation (head-to-head *e.g.* ...ATGC...CGTA). (See Figure 9.) Larger site-specific recombination systems can also require additional proteins and DNA sites.

During site-specific recombination, two recombinase molecules first bind to each crossover site. The two crossover sites are then brought together within a *recombinase complex*, B: the smallest convex region containing the four bound recombinase molecules and the two crossover sites [**88**]. So B is a *topological ball* (*i.e.*, it can be deformed to a round ball). The crossover sites can be located either on the outside or inside the 4 recombinase subunits [**59, 70, 66, 62, 30, 46, 78, 4**].

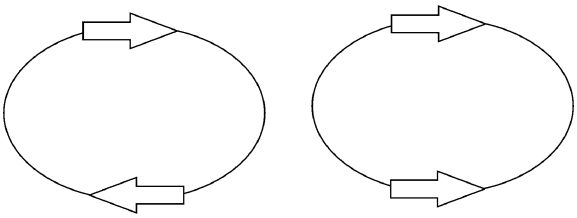

FIGURE 9. On a circular DNA molecule, the two crossover sites can be in direct (left image) or inverted orientation.

Site-specific recombination roughly has three stages. First, two recombinase molecules bind to each of two specific sites on one or two molecules of circular DNA (known as the *substrate*) and then bring them close together. The sites and bound proteins together are called the *recombinase complex*. Next, the sites are cleaved, exchanged and resealed. The precise nature of this intermediary step is determined by which of the two recombinase subfamilies the particular protein belongs to. And finally, the rearranged DNA, the *product*, is released.

Multiple rounds of strand exchange can occur before releasing the DNA—this process is known as *processive* recombination. This is in contrast to *distributive* recombination, where multiple rounds of the entire process of recombination (including releasing and rebinding) occurs.

7.2. The two families of site-specific recombinases. Site-specific recombinases fall into two families—the serine (also known as the resolvase) and tyrosine (also known as the integrase) recombinases—based on the particular amino acid in the protein polymer that catalyzes the cleaving reaction [**44**]. The serine and tyrosine recombinases also differ in their mechanism of cutting and rejoining DNA at the crossover sites. Both families are large: an evolutionary analysis has been performed on 72 serine recombinases [**87**] and a recent sequence search documents approximately 1000 related sequences of putative tyrosine recombinases [**2**].

The diverse family of serine recombinases is comprised of four subfamilies[6] [**87**]. These recombinases may trap a fixed number of supercoils before initiating recombination [**44**]. For example, Tn3 resolvase requires three negative supercoils to be trapped by the binding of (nonactive) resolvase molecules. These trapped supercoils (outside of the recombinase complex) together with the recombinase complex itself are known as the *synaptic complex* [**101, 88**]. Likewise, the invertases also require a fixed number of supercoils trapped outside the recombinase complex. Rather than using additional recombinase molecules, they rely on additional proteins (called accessory proteins) and DNA sites (called enhancer sequences), which facilitate the organization of a unique synaptic complex that promotes DNA cleavage.

With serine recombinases, recombination proceeds through a concerted 4-backbone cleaving and rejoining reaction [**44**]. Serine recombinases can perform processive recombination.

In contrast, tyrosine recombinases first cleave, exchange and reseal two sugar-phosphate backbones. The DNA-protein complex then proceeds through an intermediary structure (a Holliday junction) before repeating the process with the other

[6]resolvases (such as Tn3 and $\gamma\delta$), invertases (such as Gin, Hin, Pin, and Min), large serine recombinases (also called large resolvases) and IS elements.

two DNA backbones [**43, 44**]. Most tyrosine recombinases, including the most famous members Flp, λ Int and Cre, tolerate varying numbers of supercoils outside of the recombinase complex. (However, there are exceptions, most notably XerCD, which trap a fixed number of supercoils using accessory proteins before initiating cleavage [**16**]). Like serine recombinases, tyrosine recombinases can also employ accessory proteins to help assemble the synaptic complex, and to drive the overall reactions (*e.g.* λ Int and XerCD) [**16, 4**].

7.3. DNA knots and links as recombination products. As mentioned above, site-specific recombination can transform supercoiled circular DNA into a knot or link. Given a substrate with crossover sites in a particular orientation, a single round of recombination mediated by a serine recombinase yields a unique topological product. For example, Tn3 acting on an unknot with direct sites yields exclusively the (2,2)-torus link (the Hopf link). Multiple rounds of processive recombination successively lead to the figure eight knot 4_1, the Whitehead link 5_1^2 and the six-crossing knot 6_2. By contrast, a single round of recombination mediated by a generic tyrosine recombinase yields a spectrum of DNA knots or links[7]. For example, Flp acting on an unknot with direct sites yields the unlink and a variety of 2-torus links: $2_1^2, 4_1^2, 6_1^2, 8_1^2, 10_1^2, 12_1^2$, and higher-crossing links as well. This distribution of knot/link products should reflect the supercoiling density of the substrate DNA.

Understanding precisely which knots and links arise during site-specific recombination can help understand the details of the process, *e.g.* [**42**]. Topological techniques have played a significant role in characterizing knotted and linked products of site-specific recombination. For example, several approaches have been developed to determine a particular DNA knot or catenane type, including utilizing the node number for knots [**18**], the Jones polynomial for catenanes [**6**], Schubert's classification of 4-plats [**104**] and the HOMFLY polynomial [**106**]. Below, we discuss two large topological contributions to our understanding of site-specific recombination.

7.4. Predicting DNA knot and link products. Recently, Buck and Flapan [**8, 7**] developed a predictive model for DNA knots and links that arise as products of site-specific recombination. More specifically, rather than focusing on a specific recombinase as many earlier studies have done, the authors presented a topological model that predicts which knots and links can occur as products of site-specific recombination *in general*. They do this by describing the topology of how DNA knots and links are formed as a result of a single—or multiple rounds of processive—recombination event(s), given a plectonemically supercoiled unknot, unlink, or $T(2, m)$ torus knot or catenane substrate. (See Figure 10 for illustrations of each substrate type.)[8]

The general idea is to consider a ball B containing the convex hull of the four recombinase molecules, and a spanning surface D (consisting of a disc for the unknot, two disjoint discs for the unlink, or twisted band for a torus knot or link) whose boundary is the unrecombined DNA axis. Then consider $D \cap B$ pre-recombination and post-recombination. Similarly, characterize $D \cap cl(S^3 \setminus B)$ for each substrate. Restrict these possibilities using topological arguments. Then glue

[7]There are a few exceptions to this, for example the tyrosine recombinase XerCD. However, in this introductory article we will consider only the generic case.

[8]Note: *All* figures represent the axis of duplex DNA.

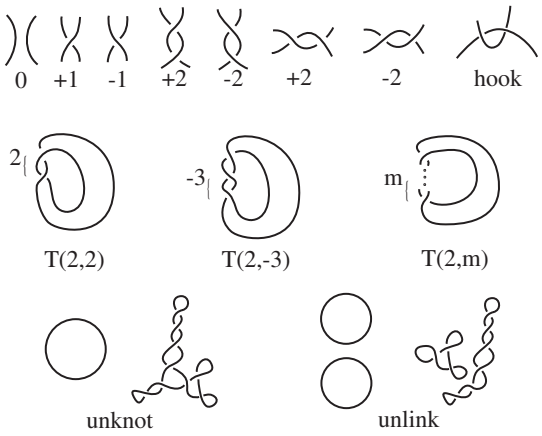

FIGURE 10. *Above:* Possible conformations of the recombinase complex. *Below:* Possible substrates: the unknot, unlink, or torus knot or link.

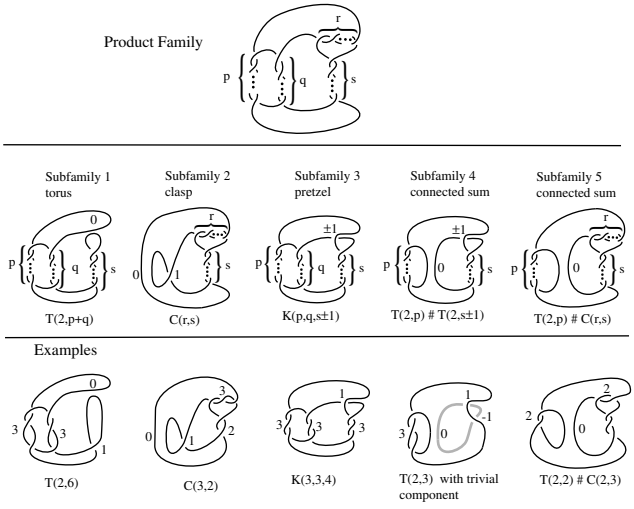

FIGURE 11. Product family, and subfamilies. Note that this family contains the same torus knots and links as subfamily 1.

each of the post-recombinant forms of $D \cap B$ to each form of $D \cap cl(S^3 \setminus B)$ to classify possible product knots and links.

The model is independent of the size of the substrate, the sign of the supercoils and the site orientation. The model relies on three assumptions, each supported by biological evidence. Given these assumptions, the authors predict that products arising from site-specific recombination must be members of a single family of products (illustrated in Figure 11). This is a specific family of Montesinos knots and links, described in more detail in Section 8.1, given by $(\frac{1}{p}, \frac{1}{q}, \frac{r}{rs+1})$, with $p, q, r, s \in \mathbb{Z}$.

One comforting thing about this family is its small size. As mentioned above, the minimal crossing number (MCN) of a DNA knot or catenane can be determined experimentally [**60**]. For small values of the MCN there are not many knots or links with a given value. However, the number of knots and links with MCN $= n$ grows exponentially as a function of n [**33**], and there are 1,701,936 knots with MCN $\leq$ 16 [**50**]. So knowing the MCN is not sufficient to determine the knot or catenane.

However, as proved in [**8**], the total number of knots and links in the family $F(p, q, r, s)$ of Figure 11 grows linearly with n^3. So the proportion of all knots and links which are contained in our family decreases exponentially as n increases. Thus, knowing the MCN of a product and knowing that the product is in one of this families allows us to significantly narrow the possibilities for its knot or link type.

There are several applications of this work. Firstly, it can predict knot and link products for previously uncharacterized data. Secondly, it can determine pathway of recombination: processive (multiple strand exchanges before releasing) versus distributive (entire recombination process repeats). In particular, it predicts that any knot or link product that is not of the family above must arise distributively. And finally, this work can help determine sequence of products of processive recombination.

This recent work complements earlier results of [**89**], which used the tangle model described below and several biologically reasonable assumptions to solve tangle equations. They then determined which (4-plat) knots and links arise as a result of (possibly processive) site-specific recombination on the unknot for the serine subfamily of recombinases. In addition to an unknotted substrate for a generic recombinase, the more recent work allows substrates that are unlinks with one site on each component, as well as $(2, m)$-torus knots and links. The more recent work of [**8, 7**] also does not assume the tangle model holds or that the products must be 4-plats. This is particularly important as (distributive) recombination has been seen to produce knots and links which are connected sums, and thus not 4-plats.

8. The tangle model

We now consider the alternative scenario. Suppose we know the substrate and product topologies for a given site-specific recombinase. For example, as above, we have determined that Flp acting on an unknot with direct sites yields the unlink and a variety of 2-torus links: $2_1^2, 4_1^2, 6_1^2, 8_1^2, 10_1^2, 12_1^2$. Our goal is now to use this input, together with topological arguments to determine details of the pathway and/or mechanism. For example in [**42**], using the topological products as probes, it was determined that Flp aligns the crossover sites so that their sequences are aligned antiparallel to each other.

In the tangle model different regions of the DNA molecule are represented by tangles, and the action of the recombinase as a change in one of the tangles. Given the substrate and product topologies, mathematicians can help biologists by finding all tangle combinations that may explain the recombinase's action. Some of these mathematically possible solutions can then be eliminated through biological considerations. Here, the input is the knot or link type of both the substrate and product DNA. The output of the tangle model is a list of solutions to the

corresponding tangle equations, which yield a corresponding set of solutions for the recombination pathway and/or mechanism.

The tangle model was originally developed by Ernst and Sumners to describe the action of particular site-specific recombinases in terms of tangle sums [**34**]. Building on the experimental work of Wasserman and Cozzarelli [**101**] and Conway's theory of tangles [**19**], they used their tangle model to make predictions—later experimentally verified—about how the recombinase Tn3 resolvase interacts with DNA [**34**]. The tangle model has since been used to determine various features of protein-DNA interactions for a number of specific proteins [**9, 10, 23, 24, 96, 42, 35, 89, 95, 94**].

8.1. Tangles, tangle operations and double branch covers. We begin by stating a few elementary facts about tangles. (For a more comprehensive introduction to tangles, see [**9**] and [**40**].) A *tangle* T is (the isotopy (rel ∂) class of) a pair (B^3, t), where B^3 is a 3-ball with a given boundary parametrization with four distinguished boundary points labelled NW, NE, SW, SE, and t consists of a pair of properly embedded unoriented arcs with endpoints NW, NE, SW and SE. We say two tangles C and D are *equivalent* if there exists an isotopy taking C to D, which remains the identity on ∂C.

Tangles can be divided into three mutually exclusive families: locally knotted, rational and prime. See Figure 12.

FIGURE 12. From left to right, a rational, prime and locally knotted tangle.

A tangle is *locally knotted* if there exists a sphere in B^3 meeting t transversely in 2 points such that the 2-ball bounded by the sphere intersects t in a knotted spanning arc.

Rational tangles are the second family; they are so-called because their isotopy classes are in one-to-one correspondence with the extended rational numbers $(\mathbb{Q} \cup \{\infty\})$ via a continued fraction expansion, as first constructed by Conway [**19**]. (See [**12**] for an outline of the original proof, based on branched covering spaces. Alternatively see [**40**] for subsequent nice classifications that utilize less machinery.) A tangle whose corresponding rational number is $\frac{p}{q}$ will be denoted by $\left(\frac{p}{q}\right)$. Rational tangles are formed by an alternating series of horizontal and vertical half-twists of two (initially untwisted) parallel arcs (and hence are freely isotopic to them). Any continued fraction decomposition of $\frac{p}{q} = a_n + 1/(a_{n-1} + ... \ (1/a_1))$ yields a finite list of integers $[a_1, \ldots, a_n]$ which tell us how to twist the strands around each other to get a diagram of the tangle. The (0) tangle corresponds to two untwisted horizontal arcs (one joining NE to NW and one joining SE to SW), whereas the (∞) tangle corresponds to two untwisted vertical arcs. The double branch cover of a rational tangle is a solid torus.

All locally unknotted, nonrational tangles are *prime*. Bleiler proved that the minimal prime tangle has a minimal projection with five crossings [**5**].

There are several operations one can perform on tangles. We concentrate on two. (See Figure 13.)

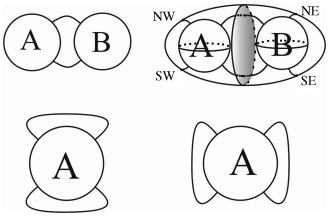

FIGURE 13. Two tangle operations: sum and numerator closure.

The first operation forms a knot or 2-component link from a given tangle A: the *numerator closure*, $N(A)$. This adds an unknotted arc joining the northern endpoints, and another unknotted arc joining the southern endpoints, or equivalently, the boundary of A and (0) are identified so that E_A is identified with $\mathrm{E}_{(0)}$ for $E \in \{NE, NW, SE, SW\}$.

The numerator closure of a rational tangle yields a *4-plat* or *rational knot*: a knot or 2-component link that admits a projection consisting of a braid on 4 strings, with one strand free of crossings [**12**]. Given two rational tangles $\left(\frac{p}{q}\right)$ and $\left(\frac{p'}{q'}\right)$, then $N\left(\frac{p}{q}\right) = N\left(\frac{p'}{q'}\right)$ iff $p = p'$ and $q^{\pm 1} \equiv q \mod p$. Thus given a 4-plat, we can write it as the numerator closure of a rational tangle that is unique up to the relationship above. Schubert showed that all 4-plats are prime knots [**81**]. We will denote the 4-plat obtained by the numerator closure of $\left(\frac{p}{q}\right)$ as $b(p, q)$. For example, we can write the unknot as $b(1, 1)$, and the trefoil as $b(3, 1)$.

The second operation, *tangle sum*, takes a pair of tangles A, B, and, under certain restrictions, yields a third tangle, $A + B$, by identifying the eastern hemispheric boundary disk of A with the western one of B in such a way that NE_A is identified with NW_B and SE_A is identified with SW_B. Note that the (0) tangle is the identity under this operation: $A + (0) = A$. Beware that under tangle addition, we cannot distinguish between $A + (p)$ added to $(-p) + B$ and A added to B. Thus although tangle summands are written in their simplest form, they are unique only up to an arbitrary number of compensating positive and negative horizontal twists. For example, given a solution $P = (p), R = (r)$ and $O^k = (s)$, then $P = (p + n), R = (r + n)$ and $O^k = (s - n)$ is another solution which is not a minimal projection.

A particular class of prime tangles is obtained by tangle sum of rational tangles *Montesinos* tangles [**68, 67, 69**]. We will use the notation $\left(\frac{a_1}{b_1}, \frac{a_2}{b_2}, \ldots, \frac{a_n}{b_n}\right)$ to denote the Montesinos tangle obtained by the tangle sum of the rational (possibly integral) tangles $\left(\frac{a_1}{b_1}\right), \left(\frac{a_2}{b_2}\right), \ldots, \left(\frac{a_n}{b_n}\right)$.

Since the sum of a rational tangle and an integral tangle yields a rational tangle [**14**], a Montesinos (nonrational) tangle must have at least two nonintegral summands. The numerator closure of a Montesinos tangle is called a *Montesinos knot or link*. The model discussed in Section 7.4 predicts that all knots and links

arising from site-specific recombination fall within a small subfamily of Montesinos knots and links.

Note that $N(A_1 + A_2 + \cdots + A_n)$ is isotopic to the numerator closure of the sum of the same tangles re-ordered by a cyclic permutation and/or a reversal of order. (See Chapter 12 in [**12**] for a primer on Montesinos tangles and knots).

8.2. Double branch covers. Given a tangle T, then we will use $\widetilde{T}$ to mean the double cover of B^3, branched over t. In general, we will write dbc(K) to denote the three-manifold that is the double cover of S^3 branched over the set K. We now turn our attention to the (compact, connected and orientable) three-manifolds that arise as double branch covers of tangles or 4-plats.

If P is a rational tangle, then $\widetilde{P}$ is a solid torus, which we will denote by V_P. (See Figure 14.)

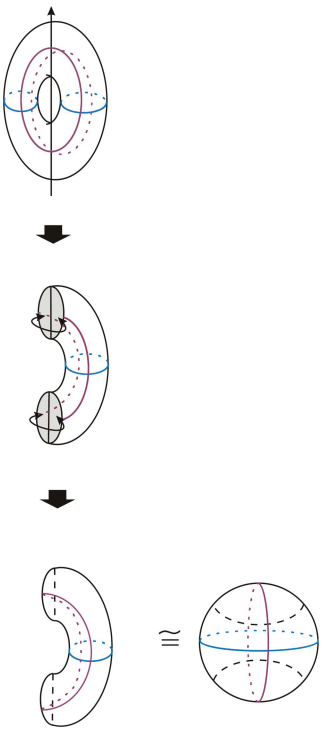

FIGURE 14. Double branch cover of a 3-ball branched over a rational tangle is a solid torus.

Schubert showed that dbc$(b(p,q))$ is a special type of well-understood three-manifold, the lens space $L(p,q)$. Two 4-plats $b(p,q)$ and $b(p',q')$ are equivalent if and only if their corresponding double branch covers, the lens spaces $L(p,q)$ and $L(p',q')$, are homeomorphic [**81**], so $b(p,q) = b(p',q')$ if and only if $p = p'$ and $q^{\pm 1} \equiv \pm q' \mod p$. (See Rolfsen [**79**], for a good introduction to lens spaces.)

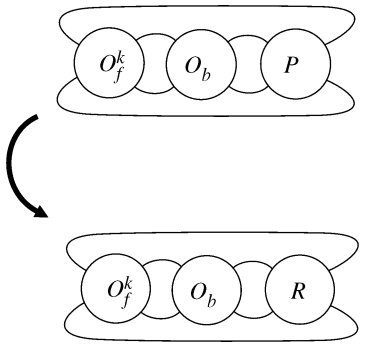

FIGURE 15. Tangle model of recombination: Substrate DNA is the numerator closure of the 3 tangles O_f^k, O_c and P. Recombination is modelled as replacing tangle P with tangle R.

8.3. An illustration of the tangle model.

8.3.1. *Tangle model for a tyrosine recombinase.* We can now describe in full detail a generalization, introduced in [**89**], of the original tangle model of Ernst and Sumners [**34**]. We illustrate this model with a generic member of the tyrosine family of recombinases, the protein *Flp* (pronounced 'flip'). Flp has served as the paradigm for site-specific recombination, and there are a number of proteins (including Cre, and λ Int acting on LR sites) whose products are, topologically speaking, identical to those of Flp.

We model each of the substrates and products as the numerator closure of the sum of three tangles. Each tangle arc represents a segment of double-stranded DNA. In the tangle model pioneered by Ernst and Sumners [**34**], the cutting and joining of DNA is assumed to be completely localized: two of the tangles are unchanged by the action of the protein. In the substrate, the first tangle, P (Parental), represents the two short crossover sites that Flp recognizes and to which it chemically binds and then cuts, rearranges and re-seals. This action can be thought of as removing P and replacing it with a new tangle, R (Recombinant), in the product. The second tangle, O_c, represents the part of the DNA that is physically constrained, but unchanged, by the protein (O stands for Outside and c for constrained). The last tangle, O_f^k, represents the part of the DNA that is free (hence the subscript f) from protein binding constraints. O_f^k can vary depending on the amount of DNA supercoiling present at the time Flp acts. The superscript k indexes these different possibilities. (See Figure 15.)

In terms of tangles, this implies that the substrate and products can be modelled as:

$$N(O_f^k + O_c + P) = \text{substrate (before recombination)},$$
$$N(O_f^k + O_c + R) = \text{product (after recombination)}$$

where $k \in \{0, 1, 2, 3\}$ represents the degree of supercoiling in the substrate. O_f^k varies as k varies, so we obtain different products, as described below.

When Flp acts on DNA it yields a variety of torus knots (for inverted repeats) or links (for direct repeats), depending on O_f^k.

More precisely, when Flp acts on a DNA molecule with inverted sites, experiments have shown that the resulting DNA can be an unknot (with a different DNA sequence), or a knot with up to 11 crossings [**43**]. Crisona *et al.* have obtained images (using electron microscopy) of the simplest products, and have shown that they are the torus knots $b(1,1)$ (the unknot), primarily positive $b(3,1)$ and exclusively positive $b(5,1)$ [**23**]. This experimental evidence indicates that Flp begins with an unknotted DNA substrate with inverted repeats, $b(1,1)$ and converts it via tangle surgery into a torus knot $b(\pm(2k+1),1)$, where $k \in \{0,1,2,3\}$. See Figure 16.

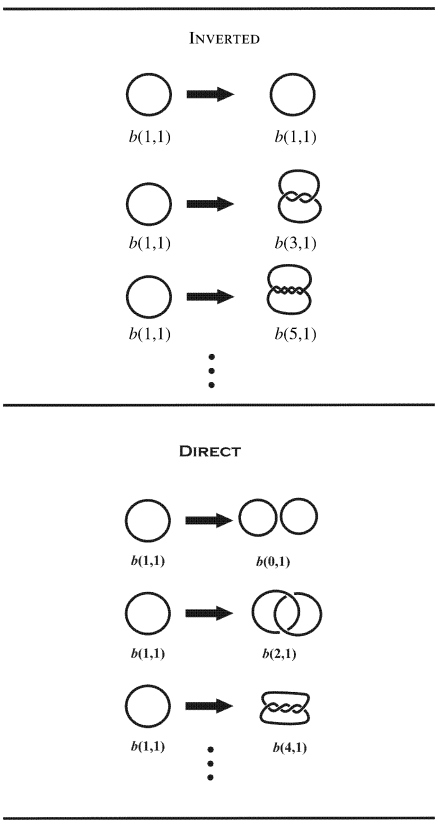

FIGURE 16. Substrate and products of Flp recombination.

We thus model the action of Flp on DNA with inverted repeats as:

Before: $N(O_f^k + O_c + P) = b(1,1) =$ unknot, for $k \in \{0,1,2,3\}$

After: $N(O_f^0 + O_c + R) = b(1,1) =$ unknot,

$N(O_f^k + O_c + R) = b(\pm(2k+1),1) =$ torus knot for $k \in \{0,1,2,3\}$.

8.3.2. *Tangle model for a serine recombinase.* Recall that, in contrast to a generic tyrosine recombinase, proteins in the serine family of recombinases, such as Tn3, require a fixed number of supercoils before they begin cutting and rejoining DNA. Once this requirement is met, they rearrange the DNA, occasionally multiple

times, before releasing it. The corresponding tangle equations: from substrate $N(O_f + O_c + P) = K_0$ to products $N(O_f + O_c + nR) = K_n$ are:

$$\left.\begin{array}{l} N(O_f + O_c + P) = K_0 \\ N(O_f + O_c + R) = K_1 \\ N(O_f + O_c + R + R) = K_2 \\ \qquad\vdots \\ N(O_f + O_c + \underbrace{R + \cdots + R}_{n}) = K_n \end{array}\right\} \begin{array}{l} \text{for the substrate,} \\ \\ \\ \\ \text{for the products.} \end{array}$$

These equations were first solved (*i.e.*, all constituent tangles have been characterized, given the substrate and product 4-plats) by Ernst and Sumners [**34**]. In particular, they found that $O = O_f + O_c = (-3, 0)$, $R = (1)$ and if P is rational, then $P = (2n, 3, 0)$, for $n \in \mathbb{N}$.

Note, in particular, that since O is always the vertical tangle $(-3, 0)$, then the free part of DNA, O_f, does not vary. This single, fixed O_f is what made the serine recombinase tangle equations more tractable than the generic integrases, whose equations involve a family of tangles O_f^k, indexed by k.

8.4. Strategy. Given the set of tangle equations above, whose products (4-plats) are known, the goal is to determine the constituent tangles (up to equivalence as discussed above). The interplay of tangles and 4-plats with their corresponding double branch covers is the key to many of the results in tangle calculus. For instance, if C and D are tangles, and D is a rational tangle, then the dbc$(C + D)$ is obtained by gluing $\widetilde{C}$ and $\widetilde{D} = V_D$ along annuli that are the lifts of their corresponding gluing disks. If D is integral, then the gluing annulus is boundary reducible, and dbc$(C + D) \simeq \widetilde{C}$ [**9**].

The sum and subsequent numerator closure of two tangles C and D induces a gluing of the boundaries of their respective double branch covers $\widetilde{C}$ and $\widetilde{D}$. If $N(C+D)$ yields a 4-plat $b(p, q)$, then $\widetilde{C} \cup_h \widetilde{D}$ must be the lens space $L(p, q)$, where h is the map that takes $\mu_{\partial\widetilde{C}}$ to $p\lambda_{\partial\widetilde{D}} + q\mu_{\partial\widetilde{D}}$. In particular, when C and D are both rational, $\widetilde{C} = V_C$ and $\widetilde{D} = V_D$ are solid tori, and they form a Heegaard splitting $V_C \cup_h V_D$ of $L(p, q)$.

Replacing tangle P in $N(O + P)$ by tangle R to obtain $N(O + R)$ is called *tangle surgery*. If P and R are rational tangles, then tangle surgery corresponds to replacing V_P with V_R in the double branch cover, and thus corresponds to different Dehn fillings of $\widetilde{O^k}$. In the special case of $N(O^k + P) = b(1, 1)$ (the unknot) where dbc$(b(1, 1)) = S^3$, the tangle surgery corresponds to Dehn surgery on the knot complement $(\widetilde{O^k})$ in S^3. Note by the preceding section the knot is trivial iff O^k is rational.

Ernst and Sumners proved that, given the above equations for a generic integrase, P and R are rational for both direct and inverted repeats [**34**]. More recently in [**9**], Buck and Verjovsky give a different proof for the rationality of R for direct and inverted repeats. So the tangle surgery of replacing P with R corresponds to Dehn surgery in the double branch covers.

Thus the strategy is to use restrictions on the type of Dehn surgeries of $S^3 = \mathrm{dbc}(b(1,1))$ that yield lens spaces. This in turn restricts the possible tangle surgeries, which in turn restricts the exact tangle solutions.

In [**9, 10**], Buck and Verjovsky consider the tangle model for a generic member of the tyrosine recombinase family without any *a priori* assumptions on the constituent tangles. They gave a complete classification of all possible solutions for O_f^k, O_c, P and R to the systems of equations arising from both the direct and inverted cases.

The proofs utilize a fair amount of three-manifold machinery. In particular, they utilize the cyclic surgery theorem, recent work of Kronheimer, Mrowka, Ozsváth and Szabó [**57**], and work of Ernst [**31, 32**].

The results in [**9, 10**] helped illuminate some of the biochemical steps of site-specific recombination. In particular, the elucidation of the tangle O_c helped determine that Flp aligned the sites in an antiparallel alignment [**42**].

9. Conclusion and future directions

There has been dramatic progress in understanding site-specific recombination through topological probing. For a generic tyrosine recombinase, the tangle solutions are completely classified [**9, 10**]. For a number of serine recombinases, the tangle solutions are also known, *e.g.* [**34**].

While the tangle model has been successful in illuminating some of the steps involved, there are many site-specific recombinases whose precise product knot or link type is still uncharacterized and so the tangle model cannot be harnessed. A powerful new strategy combines a predictive model, such as in [**8, 7**], together with the tangle model to probe the recombinase mechanism further.

New avenues in related directions include looking at 3-string tangles, for proteins that use additional sequences in addition to the crossover sites [**26**], and developing a 3-dimensional version of the tangle model [**94**].

By now, the reader is hopefully convinced that DNA Topology has deep biological consequences, *e.g.* for replication and recombination. In particular, DNA knots and links can be dangerous (*e.g.* linked daughter chromosomes) or helpful (*e.g.* as probes for understanding recombination).

We have discussed that it is not (yet) easily possible to understand fully DNA topology via purely biological techniques. For example, electrophoretic migration does not yield the precise knot type, and electron microscopy becomes increasingly difficult as the number of crossings increases. Additionally, the mechanism of type II topoisomerases, whose primary function is to change DNA knot or link type, is still not precisely understood. Thus, mathematicians can help predict DNA topology, for example by restricting the possibilities for recombination products as described in Section 7.4. This in turn can illuminate the biological processes involved.

There are a variety of future research trajectories for both mathematicians and biologists at this rich interface. In addition to the trajectories mentioned earlier, we note that we are only beginning to understand how other DNA-rearranging proteins, such as transposases, form DNA knots and links. Hopefully this article has stimulated the reader to explore—and contribute to!—some of these areas.

References

[1] C. Adams, *Introduction to knot theory*, Applications of Knot Theory: Proceedings of Symposia in Applied Mathematics (2008).

[2] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman, *Gapped blast and psi-blast: a new generation of protein database search programs.*, Nucleic Acids Res **25** (1997), 3389–3402.

[3] J.M. Berger, *Type II topoisomerases*, Curr. Opin. Struct. Biol. **8** (1999), 26–32.

[4] T. Biswas, H. Aihara, M. Radman-Livaja, D. Filman, A. Landy, and T. Ellenberger, *A structural basis for allosteric control of DNA recombination by λ integrase*, Nature **435** (2005), 1059–1066.

[5] S. Bleiler, *Knots prime on many strings*, Trans. Amer. Math. Soc. **282** (1984), 385–401.

[6] T.C. Boles, J.H. White, and N.R. Cozzarelli, *Structure of plectonemically supercoiled DNA*, J. Mol. Biol. **213** (1990), no. 4, 931–51.

[7] D. Buck and E. Flapan, *A topological characterization of knots and links arising from site-specific recombination*, J. Mol. Biol. **374** (2007), 1186–1199.

[8] ———, *A topological characterization of knots and links arising from site-specific recombination*, J. Phys. A **40** (2007), 12377–12395.

[9] D. Buck and C. Verjovsky Marcotte, *Tangle solutions for a family of DNA-rearranging proteins*, Math. Proc. Cambridge Philos. Soc. **139** (2005), no. 1, 59–80.

[10] ———, *Classification of tangle solutions for integrases, a protein family that changes DNA topology*, Journal of Knot Theory and its Ramifications **16** (2007).

[11] G. Buck and E.L. Zechiedrich, *DNA disentangling by type-2 topoisomerases*, J. Mol. Biol. **340** (2004).

[12] G. Burde and H. Zieschang, *Knots*, De Gruyter Studies in Mathematics, Berlin, 2003.

[13] Călugăreanu, *Sur las classes d'isotopie des noeuds tridimensionnels et leurs invariants*, Czech Math. J. **11** (1961), 588–625.

[14] Quach thi Câm Vân, *On a theorem on partially summing tangles by Lickorish*, Math. Proc. Cambridge Philos. Soc. **93** (1983), no. 1, 63–66.

[15] Jason Cantarella, Robert B. Kusner, and John M. Sullivan, *On the minimum ropelength of knots and links*, Invent. Math. **150** (2002), no. 2, 257–286. MR1933586 (2003h:58014)

[16] S.D. Colloms, J. Bath, and D.J. Sherratt, *Topological selectivity in Xer site-specific recombination*, Cell **88** (1997), no. 6, 855–64.

[17] N.R. Cozzarelli, T. Christian-Boles, and James H. White, *Primer on the topology and geometry of DNA supercoiling*, DNA Topology and its Biological Effects (1990), 139–184.

[18] N.R. Cozzarelli, M.A. Krasnow, S.P. Gerrard, and J.H. White, *A topological treatment of recombination and topoisomerases*, Cold Spring Harb. Symp. Quant. Biol. **49** (1984), 383–400.

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

[20] M.M. Cox, *Mobile DNA*, Berg, D.E. and M.M. Howe (eds), ch. DNA inversion in the 2 μm plasmid of Saccharomyces cerevisiae, pp. 661–670, Am. Soc. Microbiology, Washington DC, 1989.

[21] N. Craig, R. Craigie, M. Gellert, and A. Lambowitz (eds.), *Mobile DNA*, ASM Press, Washington, DC, 2002.

[22] N.J. Crisona, R. Kanaar, T.N. Gonzalez, E.L. Zechiedrich, A. Klippel, and N.R. Cozzarelli, *Processive recombination by wild-type Gin and an enhancer-independent mutant. Insight into the mechanisms of recombination selectivity and strand exchange*, J. Mol. Biol. **243** (1994), no. 3, 437–57.

[23] N.J. Crisona, R.L. Weinberg, B.J. Peter, D.W. Sumners, and N.R. Cozzarelli, *The topological mechanism of phage lambda integrase*, J. Mol. Biol. **289** (1999), no. 4, 747–775.

[24] Isabel K. Darcy, *Biological distances on DNA knots and links: applications to XER recombination*, J. Knot Theory Ramifications **10** (2001), no. 2, 269–294, Knots in Hellas '98, Vol. 2 (Delphi).

[25] M.R. Dennis and J.H. Hannay, *Geometry of Călugăreanu's theorem*, Proc. Roy. Soc. A **461** (2005), 3245–3254.

[26] I.K. Darcy, J. Luecke, and M. Vazquez, *Tangle analysis of difference topology experiments: applications to a Mu protein-DNA complex*, arXiv:0710.4150v1 (2007).
[27] R.E. Depew and J.C. Wang, *Conformal fluctuations of the DNA helix*, Proc. Natl. Acad. Sci. USA **72** (1975), no. 11, 4275–4279.
[28] De Witt L. Sumners (ed.), *New scientific applications of geometry and topology*, American Mathematical Society, Providence, RI, 1992, Papers from the American Mathematical Society Short Course held in Baltimore, Maryland, January 6–7, 1992.
[29] K.C. Dong and J.M. Berger, *Structural basis for gate-DNA recognition and bending by type IIA topoisomerases*, Nature **450** (2007), 1201–1206.
[30] E. Ennifar, J.E.W. Meyer, F. Buchholz, A.F. Stewart, and D. Suck, *Crystal structure of a wild-type Cre recombinase-loxP synapse reveals a novel spacer conformation suggesting an alternative mechanism for DNA cleavage activation*, Nucleic Acids Res. **31** (2003), no. 18, 5449–5460.
[31] C. Ernst, *Tangle equations*, J. Knot Theory Ramifications **5** (1996), no. 2, 145–159.
[32] ———, *Tangle equations. II*, J. Knot Theory Ramifications **6** (1997), no. 1, 1–11.
[33] C. Ernst and D. W. Sumners, *The growth of the number of prime knots*, Math. Proc. Cambridge Philos. Soc. **102** (1987), no. 2, 303–315.
[34] ———, *A calculus for rational tangles: applications to DNA recombination*, Math. Proc. Cambridge Philos. Soc. **108** (1990), no. 3, 489–515.
[35] ———, *Solving tangle equations arising in a DNA recombination model*, Math. Proc. Cambridge Philos. Soc. **126** (1999), no. 1, 23–36.
[36] R. Feil, *Conditional somatic mutagenesis in the mouse using site-specific recombinases*, Handbk. Exp. Pharmacol. **178** (2007), 3–28.
[37] M.D. Frank-Kamenetskii, *Unraveling DNA*, VCH Publishers, New York, NY, 1993.
[38] S.J. Froelich-Ammon and N. Osheroff, *Topoisomerases poisons: harnessing the dark side of enzyme mechanism*, J. Biol. Chem. **270** (1990), 21429–21432.
[39] F.B. Fuller, *The writhing number of a space curve*, Proc. Natl. Acad. Sci. USA **68** (1971), 815–819.
[40] J.R. Goldman and L. H. Kauffman, *Rational tangles*, Adv. Appl. Math. **18** (1997), 300–332.
[41] O. Gonzalez and R. de la Llave, *Existence of ideal knots*, J. Knot Theory Ramifications **12** (2003), no. 1, 123–133. MR1953628 (2003j:57010)
[42] I. Grainge, D. Buck, and M. Jayaram, *Geometry of site alignment during Int family recombination: Antiparallel synapsis by the Flp recombinase*, J. Mol. Bio. **298** (2000), 749–764.
[43] I. Grainge and M. Jayaram, *The integrase family of recombinase: organization and function of the active site*, Mol. Microbiol. **33** (1999), no. 3, 449–56.
[44] N. D. Grindley, K. L. Whiteson, and P. A. Rice, *Mechanisms of site-specific recombination*, Ann. Rev. Biochem. **75** (2006), 567–605.
[45] A. Y. Grosberg, A. Feigel, and Y. Rabin, *Flory-type theory of a knotted ring polymer*, Phys. Rev. E Sta.t Phys. Plasmas Fluids Relat. Interdiscip. Topics **54** (1996), no. 6, 6618–6622.
[46] F. Guo, D.N. Gopaul, and G.D. van Duyne, *Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse*, Nature **389** (1997), no. 40–46.
[47] K.A. Heichman, I.P. Moskowitz, and R.C. Johnson, *Configuration of DNA strands and mechanism of strand exchange in the Hin invertasome as revealed by analysis of recombinant knots*, Genes Dev **5** (1991), no. 9, 1622–1634.
[48] R. H. Hoess, A. Wierzbicki, and K. Abremski, *The role of the loxP spacer region in P1 site-specific recombination*, Nucleic Acids Res. **14** (1986), no. 5, 2287–300.
[49] D.S. Horowitz and J.C. Wang, *Torsional rigidity of DNA and length dependence of the free energy of DNA supercoiling*, J. Mol. Biol. **173** (1984), 75–91.
[50] Jim Hoste, Morwen Thistlethwaite, and Jeff Weeks, *The first 1,701,936 knots*, Math. Intelligencer **20** (1998), no. 4, 33–48.
[51] R. Kanaar, A. Klippel, E. Shekhtman, J. M. Dungan, R. Kahmann, and N. R. Cozzarelli, *Processive recombination by the phage Mu Gin system: implications for the mechanisms of DNA strand exchange, DNA site alignment, and enhancer action*, Cell **62** (1990), no. 2, 353–66.
[52] T. Kanenobu and H. Murakami, *Two-bridge knots with unknotting number one.*, Proc. Am. Math. Soc. **98** (1986), 499–502.
[53] V. Katritch, J. Bednar, D. Michoud, J. Dubochet, and A. Stasiak, *Electrophoretic mobility of DNA knots*, Nature **384** (1996), no. 6605, 122.

[54] V. Katritch, J. Bednar, D. Michoud, R. G. Scharein, J. Dubochet, and A. Stasiak, *Geometry and physics of knots*, Nature **384** (1996), 142–145.
[55] P. Kohn, *Two-bridge links with unlinking number one*, Proc. Am. Math. Soc. **113** (1991), 1135–1147.
[56] M. A. Krasnow, A. Stasiak, S. J. Spengler, F. Dean, T. Koller, and N. R. Cozzarelli, *Determination of the absolute handedness of knots and catenanes of DNA*, Nature **304** (1983), no. 5926, 559–60.
[57] P. Kronheimer, T. Mrowka, P. Ozsváth, and Z. Szabó, *Monopoles and lens space surgeries*, Ann. of Math. **165** (2007), 457–546.
[58] B. Laurie, V. Katritch, J. Dubochet, and A. Stasiak, *Geometry and physics of links*, Biophysical Journal **74** (1998).
[59] A.E. Leschziner and N.D.F. Grindley, *The architecture of the $\gamma\delta$ resolvase crossover site synaptic complex revealed by using constrained DNA substrates.*, Mol. Cell **12** (2003), 775–781.
[60] S. D. Levene and H. Tsen, *Analysis of DNA knots and catenanes by agarose-gel electrophoresis*, Methods Mol. Biol. **94** (1999), 75–85.
[61] C. Levine, H. Hiasa, and K.J. Marians, *DNA gyrase and topoisomerase IV: biochemical activities, physiological roles during chromosome replication and drug sensitivities.*, Biochem. Biophys. Acta **1400** (1998), 29–43.
[62] W. Li, S. Kamtekar, Y. Xiong, G. Sarkis, N.D.F. Grindley, and T.A. Steitz, *Structure of a synaptic resolvase tetramer covalently linked to two cleaved DNAs*, Science **309** (2005), 1210–1215.
[63] L.F. Liu and J.L. Davis, *Novel topologically knotted DNA from bacteriophage p4 capsids: studies with topoisomerases*, Nuc. Acids Res. **9** (1981), 3979–3989.
[64] A. Lobb, *A slice genus lower bound from $sl(n)$ khovanov-rozansky homology*, arXiv:math/0702393 (2007).
[65] A. Maxwell and A. D. Bates, *DNA topology*, Oxford University Press, 2005.
[66] S.K. Merickel and R.C. Johnson, *Topological analysis of Hin-catalysed DNA recombination in vivo and in vitro*, Mol. Microbiology **51** (2004), no. 4, 1143–1154.
[67] J.M. Montesinos, *Seifert manifolds that are ramified two-sheeted cyclic coverings*, Bol. Soc. Mat. Mexicana (2) **18** (1973), 1–32.
[68] ———, *Una familia infinita de nudos representados no separables*, Revisita Math. Hisp.-Amer. **33** (1973), no. IV, 32–35.
[69] ———, *Revêtements ramifiés de nœuds, espaces fibrés de seifert et scindements de heegaard*, Publicaciones del seminario Matematico de Galdeano **11** (1984), no. 3.
[70] M. Nollmann, J. He, O. Byron, and W.M. Stark, *Solution structure of the Tn3 resolvase-crossover site synaptic complex*, Mol. Cell **16** (2004), no. 1, 127–137.
[71] P. Ozsváth and Z. Szabó, *Knots with unknotting number one and heegaard floer homology*, Topology **44** (2005), 705–745.
[72] S. Pathania, M. Jayaram, and R.M. Harshey, *Path of DNA within the Mu transpososome. transposase interactions bridging two Mu ends and the enhancer trap five DNA supercoils*, Cell **109** (2002), no. 4, 425–36.
[73] William F. Pohl, *DNA and differential geometry*, Math. Intelligencer **3** (1980/81), no. 1, 20–27. MR617886 (83d:92120)
[74] I.R. Porteous, *Geometric differentiation for the intelligence of curves and surfaces*, Cambridge University Press, Cambridge, England, 2001.
[75] D.E. Pulleyblank, M. Shure, D. Tang, J. Vinograd, and H.-P. Vosberg, *Action of nicking-closing enzyme on supercoiled and nonsupercoiled closed circular DNA: Formation of a Boltzmann distribution of topological isomers*, Proc. Natl. Acad. Sci. USA **72** (1975), no. 11, 4280–4284.
[76] J.A. Rasmussen, *Khovanov homology and the slice genus*, arXiv:math/0402131v1 [math.GT] (2004).
[77] ———, *Khovanov-rozansky homology of two-bridge knots and links.*, Duke Math. J. **136** (2007), 551–583.
[78] P.A. Rice and T.A. Steitz, *Model for a DNA-mediated synaptic complex suggested by crystal packing of $\gamma\delta$ resolvase subunits*, EMBO J **13** (1994), no. 7, 1514–24.
[79] Dale Rolfsen, *Knots and links*, Publish or Perish Press, Berkeley, CA, 1971.

[80] V.V. Rybenkov, C. Ullsperger, A.V. Vologodskii, and N.R. Cozzarelli, *Simplification of DNA topology below equilibrium values by type ii topoisomerases*, Science **277** (1997), no. 5326, 690–3.
[81] H. Schubert, *Über eine numerische Knoteninvariante*, Math. Z. (1954).
[82] N.C. Seeman, *Synthetic single-stranded DNA topology*, Applications of Knot Theory: Proceedings of Symposia in Applied Mathematics (2008).
[83] D. Shore and R.L. Baldwin, *Energetics of DNA twisting: I. Relation between twist and cyclization probability*, J. Mol. Biol. **170** (1983), 957–981.
[84] M. Shure and J. Vinograd, *The number of superhelical turns in native virion SV40 and minicol DNA determined by the band counting method*, Cell **8** (1976), 215–226.
[85] D.S. Silver, *Knot theory's odd origins*, American Scientist **94** (2006), 158–165.
[86] S.J. Spengler, A. Stasiak, and N.R. Cozzarelli, *The stereostructure of knots and catenanes produced by phage λ integrative recombination: implications for mechanism and DNA structure*, Cell **42** (1985), no. 1, 325–34.
[87] M.C. Smith and H.M. Thorpe, *Diversity in the serine recombinases*, Mol Microbiol **44** (2002), no. 2, 299–307.
[88] W.M. Stark, M.R. Boocock, and D.J. Sherratt, *Catalysis by site-specific recombinases*, Trends Genet **8** (1992), 432–439.
[89] D.W. Sumners, C. Ernst, S.J. Spengler, and N.R. Cozzarelli, *Analysis of the mechanism of DNA recombination using tangles*, Quarterly Review of Biophysics **28** (1995), no. 3, 253–313.
[90] A. Stasiak and A. Flamini, *Simulation of action of DNA topoisomerases to investigate boundaries and shapes of spaces of knots*, Biophys. J. **87** (2004), 2968–2975.
[91] O. Sundin and A. Varshavsky, *Terminal stages of SV40 DNA replication proceed via multiply intertwined catenated dimers*, Cell **21** (1980), no. 1, 103–14.
[92] S. Trigueros, J. Arsuaga, M.E. Vazquez, D.W. Sumners, and J. Roca, *Novel display of knotted DNA molecules by two-dimensional gel electrophoresis*, Nucleic Acids Res. **29** (2001), no. 13, E67–7.
[93] I. Torisu, *The determination of the pairs of two-bridge knots or links with gordian distance one*, Proc. Am. Math. Soc. **126** (1998), 1565–1571.
[94] M. Vazquez, S.D. Colloms, and D. Sumners, *Tangle analysis of Xer recombination reveals only three solutions, all consistent with a single three-dimensional topological pathway*, J. Mol. Biol. **346** (2005), no. 2, 493–504.
[95] M. Vazquez and D.W. Sumners, *Tangle analysis of Gin site-specific recombination*, Math. Proc. Cambridge Philos. Soc. **136** (2004), no. 3, 565–582.
[96] A.A. Vetcher, A.Y. Lushnikov, J. Navarra-Madsen, R.G. Scharein, Y.L. Lyubchenko, I.K. Darcy, and S.D. Levene, *DNA topology and geometry in Flp and Cre recombination*, J. Mol. Biol. **357** (2006), 1089–1104.
[97] A.V. Vologodskii, N.J. Crisona, B. Laurie, P. Pieranski, V. Katritch, J. Dubochet, and A. Stasiak, *Sedimentation and electrophoretic migration of DNA knots and catenanes*, J. Mol. Biol. **278** (1998), no. 1, 1–3.
[98] A.V. Vologodskii, *The topology and physics of circular DNA*, CRC Press, Boca Raton, FL, 1992.
[99] A.V. Vologodskii, W. Zhang, V.V. Rybenkov, A.A. Podtelezhnikov, D. Subramanian, J.D. Griffith, and N.R. Cozzarelli, *Mechanism of topology simplification by type ii DNA topoisomerases*, Proc. Natl. Acad. Sci. USA **98** (2001), 3045–3049.
[100] S.A. Wasserman and N.R. Cozzarelli, *Supercoiled DNA-directed knotting by T4 topoisomerase*, J. Biol. Chem. **266** (1991), no. 30, 20567–73.
[101] S.A. Wasserman, J.M. Dungan, and N.R. Cozzarelli, *Discovery of a predicted DNA knot substantiates a model for site-specific recombination*, Science **229** (1985), no. 4709, 171–4.
[102] J.D. Watson and F.H.C. Crick, *A structure for deoxyribose nucleic acid*, Nature **171** (1953), 737–738.
[103] Claude Weber, *Où en sont les mathématiques?*, ch. Questions de topologie en biologie moléculaire. Appendix by Andrzej Stasiak and Jacques Dubochet, pp. 328–347, Vuibert, 2002.
[104] J.H. White and N.R. Cozzarelli, *A simple topological method for describing stereoisomers of DNA catenanes and knots*, Proc. Natl. Acad. Sci. USA **81** (1984), no. 11, 3322–6.

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

[106] J.H. White, K.C. Millett, and N.R. Cozzarelli, *Description of the topological entanglement of DNA catenanes and knots by a powerful method involving strand passage and recombination*, J. Mol. Biol. **197** (1987), no. 3, 585–603.

[107] E.L. Zechiedrich and N.J. Crisona, *Methods in molecular biology: DNA topoisomerase protocols*, M. Bjornsti and N. Osheroff (eds), vol. 1, ch. Coating DNA with RecA protein to distinguish DNA path by electron microscopy, pp. 99–108, Humana Press, Totowa, NJ, 1989.

Department of Mathematics, and Centre for Integrative Systems Biology, Imperial College London, South Kensington, London SW7 2AZ, England