DNA Topology Review

Garrett Jones and Candice Reneé Price

Abstract DNA holds the instructions for an organism's development, reproduction, and, ultimately, death. It encodes much of the information a cell needs to survive and reproduce. It is important for inheritance and coding for proteins, and contains the genetic instruction guide for life and its processes. But also, DNA of an organism has a complex and interesting topology. For information retrieval and cell viability, some geometric and topological features of DNA must be introduced, and others quickly removed. Proteins perform these amazing feats of topology at the molecular level; thus, the description and quantization of these protein actions require the language and computational machinery of topology. The use of tangle algebra to model the biological processes that give rise to knotting in DNA provides an excellent example of the application of topological algebra to biology. The tangle algebra approach to knotting in DNA began with the study of the site-specific recombinase $Tn3$ resolvase. This chapter is a summary of some basic knot theory and biology. We then describe the tangle model developed by Ernst and Sumners using the $Tn3$ resolvase as an example. We conclude with applications of the tangle model to other biological problems.

Subject Classification 2010 57M25

1 Introduction

The DNA of any organism has a complex and interesting topology. One can take the view of it as two very long strands; as closed curves that are intertwined millions of times, perhaps linked to other closed curves or tied into knots; and,

G. Jones
Department of Mathematical Sciences, University of Wisconsin-Stevens Point, Stevens Point, WI, USA

C. R. Price (✉)
Department of Mathematics, University of San Diego, San Diego, CA, USA
e-mail: cprice@sandiego.edu

subjected to supercoiling in order to convert it into a compact form for information storage. For information retrieval and cell viability, some geometric and topological features must be introduced, and others quickly removed. Some proteins preserve the topology by passing one strand of DNA through another via a protein-bridged transient break in the DNA. This protein action plays a crucial role in cell metabolism, transcription, and replication. Other proteins break the DNA and recombine the ends by exchanging them to help regulate the expression of specific genes, mediate viral insertion into and deletion from the host genome, mediate transposition and repair of DNA, and generate antibody and genetic diversity. These proteins are performing important and incredible feats of topology at the molecular level; thus, the description and quantization of these protein actions requires the language and computational machinery of topology.

The topological approach to enzymology is an indirect method in which the descriptive and analytical powers of topology are employed in an effort to infer the structure of active protein–DNA complexes *in vitro* and *in vivo*. In the topological approach to enzymology experimental protocol, molecular biologists react circular DNA substrate with protein and capture protein signature in the form of changes in the geometry (supercoiling) and topology (knotting and linking) of the circular substrate. The mathematical problem is then to deduce protein mechanism and synaptic complex structure from these observations. The mathematics of topological objects, such as knots and tangles, are then used to solve these problems.

This chapter will discuss the background information of DNA topology by providing the definitions needed from knots and tangles. It will describe the tangle model: a developed sets of experimentally observable topological parameters with which to describe and compute protein mechanism and the structure of the active protein–DNA complex. Because, one of the important unsolved problems in biology is the three-dimensional structure of proteins, DNA, and active protein–DNA complexes in solution (in the cell), and the relationship between structure and function, this model utilizes the mathematics of knots and tangle to provide some solutions. It is the 3-dimensional shape in solution which is biologically important, but difficult to determine. The chapter will conclude with a brief discussion of some of the results utilizing the tangle model.

2 Knots and Links

Although knots have been used since the dawn of humanity, the mathematical study of knots is just under 300 years old. Not only has knot theory grown theoretically in that time, the fields of physics, chemistry, and molecular biology have provided many applications of mathematical knots.

A **knot** is defined as a closed, nonintersecting curve in $\mathbb{R}^3$. Formally, it is the embedding of a circle in three dimensions (Fig. 1). Intuitively, a knot can be simply thought of as a loop of rope with no end and no beginning.

A **link** (defined as a **catenane** by biologists) is a finite union of knots properly embedded in three-dimensional space. Each of these knots, which may be trivial,
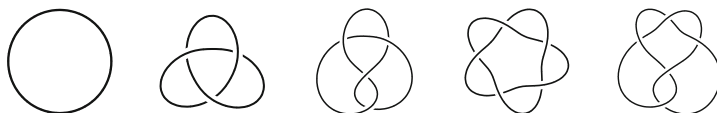
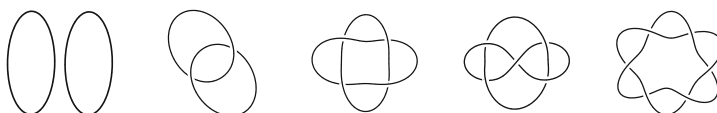**Fig. 1** Examples of simple alternating knots

Fig. 2 Examples of simple 2 component links

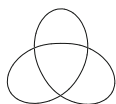**Fig. 3** Ambiguous and problematic intersections not allowed in knot diagrams


Fig. 4 A polygonal projection and a smooth projection of the knot with 3 crossings has eight possible knot diagrams, two are shown here

is known as a **component** of the link. We can view a knot as a 1-component link. From here, when discussing a property of the class of links of 1 or more component, we will use the terminology "link." When discussing a 1-component link property only, we will refer to the object as a "knot" (Fig. 2).

A **link projection** is the two-dimensional image of the three-dimensional link projected onto a plane. At each double point in the projection (a crossing involving only two line segments), it is not clear which portion of the link crosses over and which crosses under. To show this, gaps are left in the projection. At a crossing, the strand of the knot at the top of the crossing, represented by a solid line segment, is called the **overcrossing**. The strand that is at the bottom of the crossing is called the **undercrossing**, represented by a broken line segment.

It is known that problematic intersections (see Fig. 3) can be avoided so that all intersections correspond to double points. A link projection drawn with these criteria is called a **link diagram**. Knots and links are studied through their diagrams. Links that have diagrams that can be drawn using a finite number of polygonal circuits (i.e., closed paths) in three-dimensional space are called **tame** (Fig. 4). All other links are known as **wild** (Fig. 5). Most applications of knot theory concern only tame links, so we will only focus on this class of links.
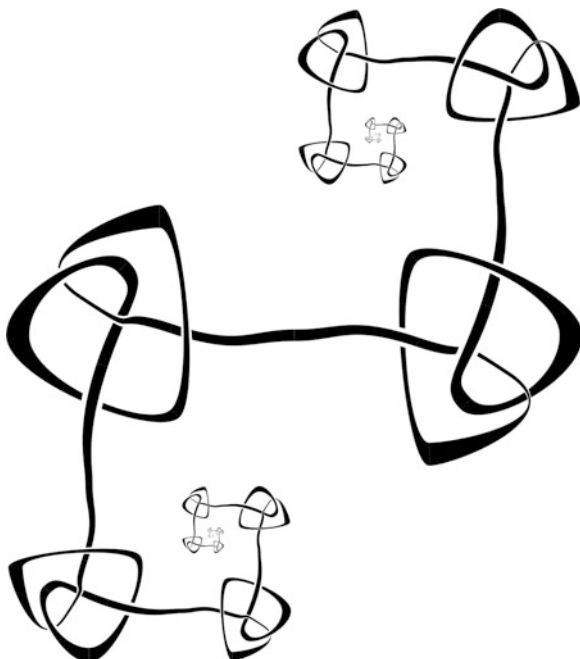
We say two links, K_1 and K_2, are **equivalent** if there is an ambient isotopy
between them. An **ambient isotopy** can be described as a continuous deformation
from one link diagram (K_1) to the other (K_2). It allows us to stretch, bend, and
twist the link however we would like; we just cannot cut it. Mathematically, two
links, K_1 and K_2, are **ambient isotopic** if there is an isotopy $h : \mathbb{R}^3 \times [0, 1] \to \mathbb{R}^3$
such that $h(s, i) = h_i(s)$ is a homeomorphism for all i where $h_0(K_1) = K_1$ and
$h_1(K_1) = K_2$ [12]. If two knots are equivalent, we refer to these knots as knots
of the same **knot type**, K, where K is the equivalence class under this equivalence
relation.

In 1926, Kurt Reidemeister proved that if we have two distinct diagrams of K,
we can go from one diagram to the other using **Reidemeister moves**, as described
in Theorem 2.1.

Theorem 2.1 (Reidemeister [33]) *Two link diagrams K_1 and K_2 are equivalent
if and only if they can be obtained from one another by a finite sequence of planar
isotopies and the three moves: twist, poke, and slide (Fig. 6).*

Given two knots, K_1 and K_2, a knot $K_3 = K_1 \# K_2$ can be constructed as seen in
Fig. 7. This knot is known as the **connected sum** of K_1 and K_2. A knot that cannot
be constructed in this manner using nontrivial knots is called **prime**. All prime knots
will be referred to as they are given in Rolfsen's table of prime knots [34].

Links can be split into two groups: alternating and nonalternating. A link is called
alternating if it has a diagram in which, when traveling around each component of

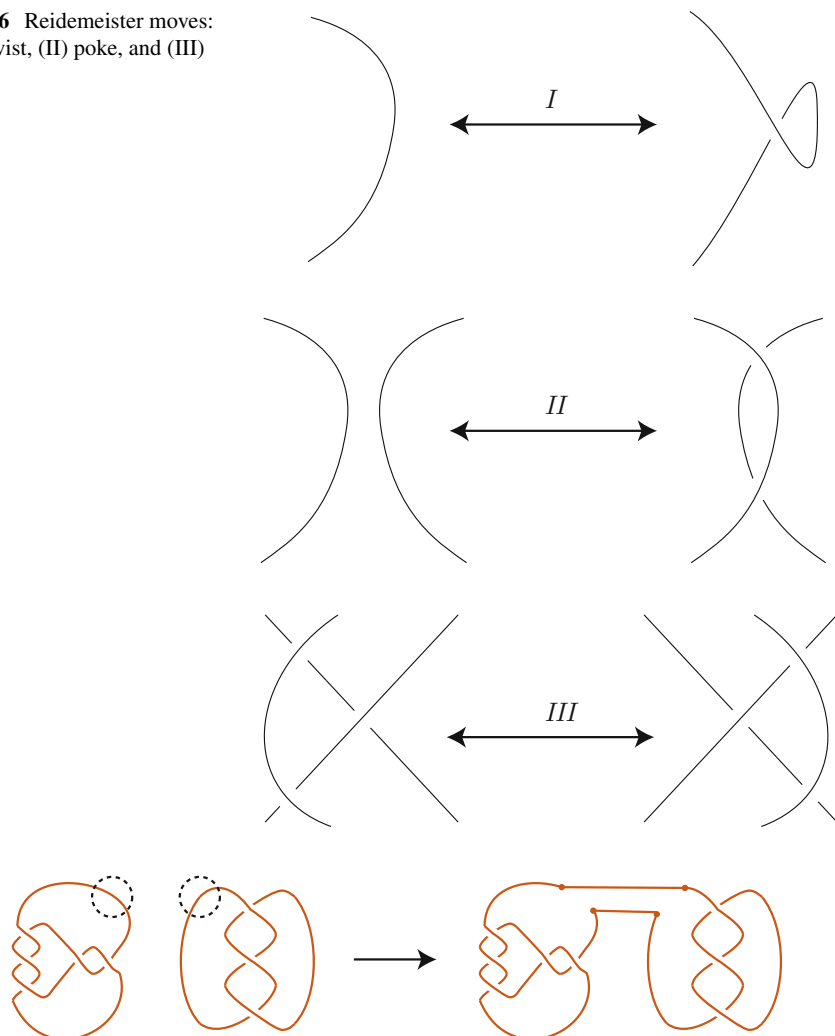**Fig. 6** Reidemeister moves: (I) twist, (II) poke, and (III) slide

Fig. 7 The connected sum of knots 5_2 and 3_1

the link, one alternates between overcrossing and undercrossings. A **nonalternating link** is one that is not alternating (i.e., every diagram has at least two overcrossings or two undercrossing in a row when traveling around the link).

An **oriented link** is a link for which each component has been given an orientation. An oriented link is **invertible** if it can be deformed to be the same link diagram with the opposite orientation [1]. The **mirror image** of a link, $\overline{L}$, is obtained by changing every overcrossing in the link to an undercrossing and vice versa. If L is equivalent to its mirror image, then we call L **amphicheiral** (or **achiral**). If L is not equivalent to $\overline{L}$, then it is **chiral**. Although not all links are achiral, most tables
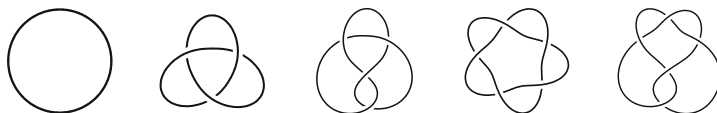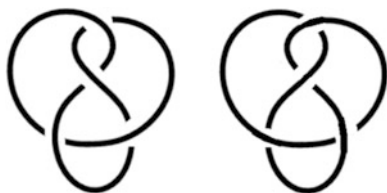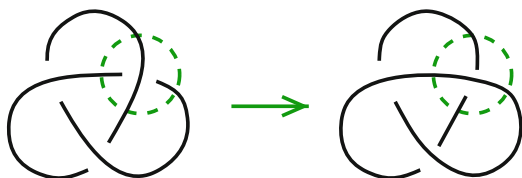
Fig. 8 4_1 knot and its mirror image

Fig. 9 Examples of minimum regular diagrams of the first five knots

Fig. 10 Example of a topological change: crossing change. This example shows the unknotting number of the knot 3_1

do not distinguish between a link and its mirror image. One example of a knot that is amphicheiral is the 4_1 knot (Fig. 8).

While Reidemeister moves are helpful to see if two links are equivalent, they are not as useful when showing that two links are not equivalent. Link invariants are utilized to show in-equivalence between two link diagrams. A **link invariant** is a specific quality of a knot or link type that does not change its value under ambient isotopy. Thus, if two links are equivalent, then their invariants are equal. Unfortunately, for a majority of invariants, the other direction is not usually true: equal invariant values for two link diagrams do not imply equivalent links.

One example of a link invariant is the minimum crossing number. The **minimum crossing number** is the minimum number of crossings over all knot diagrams of the knot type (Fig. 9).

Some invariants keep count of the number of topological changes made to a link diagram. Looking at a knot diagram, exchange locally overcrossings and undercrossings. This type of alteration may change the knot type. The **unknotting number** is the least number of crossing changes in a diagram of a knot to get to the trivial knot, minimized over all diagrams (Fig. 10).

The **linking number** is a link invariant for links of two or more components. It is calculated using the **crossing sign convention** (Fig. 11). The linking number is calculated by taking the sum of the crossing signs of each crossing between the different components of the link and dividing by two.

While the previous invariants give numerical quantities, other invariants can associate a polynomial to a knot type: the Alexander polynomial, the Jones

Fig. 11 Given an orientation, we can assign negative and positive crossings

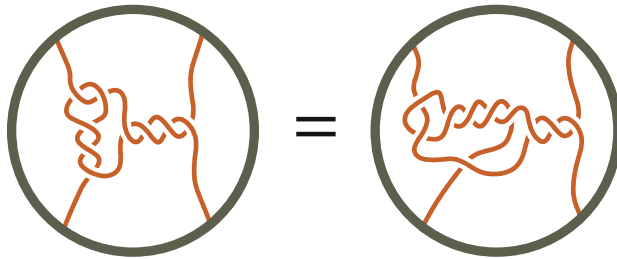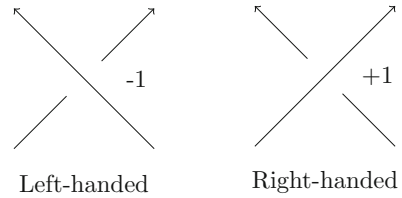

-1   +1

Left-handed   Right-handed

Fig. 12 Equivalent tangles

polynomial, and the HOMFLY-PT polynomial [3, 19, 25] or associate to a knot diagram even more complicated algebraic structures like chain complexes of abelian groups: Khovanov Homology and Knot Floer Homology [26, 30].
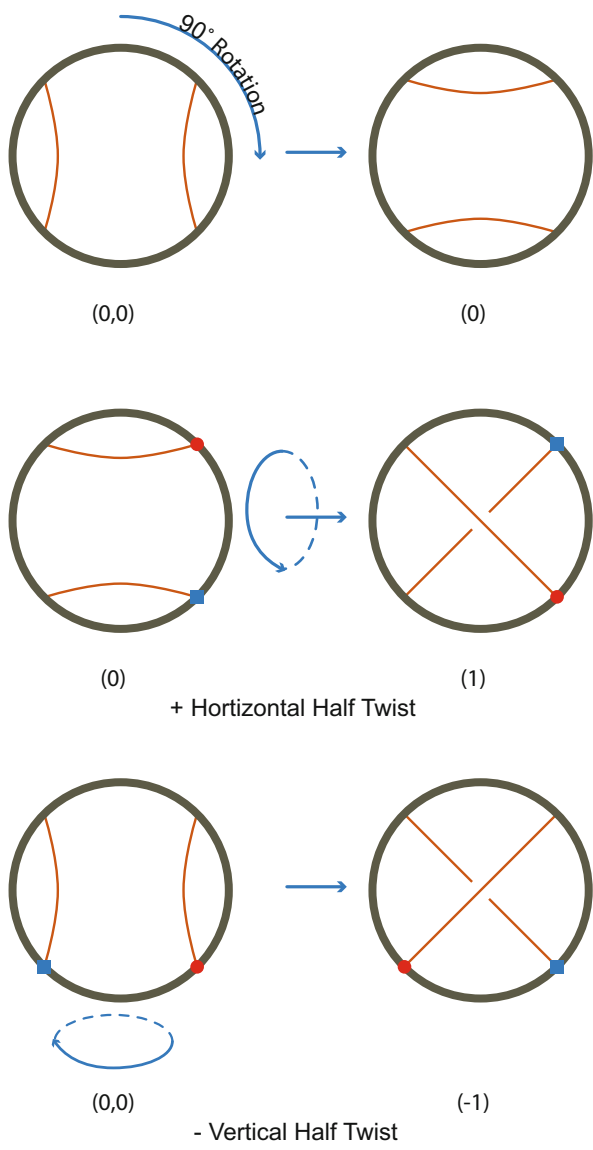
3 Tangles

An *n*-**string tangle** is defined as a pair (B, t) of a 3-dimensional ball B and a collection of disjoint, simple, properly embedded arcs, denoted t. An n-string tangle is formed by placing $2n$ points on the boundary of B and attaching n nonintersecting curves inside B such that $\partial B \cap t = \partial t$. We consider tangles $T_1 = (B, t_1)$ and $T_2 = (B, t_2)$ to be **equivalent** if there is an ambient isotopy of one tangle to the other keeping the boundary of the ball fixed (Fig. 12).

This work will focus on 2-**string tangles**. As part of the definition, we consider a 2-string tangle to be a pair (B, t) and a homeomorphism sending (B, t) to the unit ball in $\mathbb{R}^3$. We send the four endpoints of the arcs to the four equatorial points NW, NE, SE, and SW in the yz-plane described in $\mathbb{R}^3$ as the points:

$$NE : \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \; NW : \left(0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$
$$SE : \left(0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \; SW : \left(0, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right).$$

The simplest tangles are the zero tangle, denoted (0); the (∞)-tangle, denoted $(0, 0)$; the positive one tangle, denoted (1); and, the negative one tangle, denoted (-1) (Fig. 13).

Fig. 13 Simplest 2-string tangles: the (∞)-tangle, $(0, 0)$, is a 90° rotation of the zero tangle, (0). The positive one tangle, (1) is shown here as a positive horizontal half twist added to (0). The negative one tangle, (-1) is shown as a negative vertical half twist added to $(0, 0)$

(0,0)   (0)

(0)   (1)

+ Hortizontal Half Twist

(0,0)   (-1)

- Vertical Half Twist

We can take the **sum** of two tangles, T_1 and T_2, creating a new tangle, $T_1 + T_2$ (Fig. 14). Another tangle operation is the **numerator closure**, which connects the northern endpoints with the shortest arc on the exterior of B and similarly the southern endpoints, resulting in a knot or link denoted $N(T)$. We can also perform this operation on a sum of tangles (Fig. 15).

This work will focus on 2-string tangle rational tangles. A 2-string tangle is **rational** if it is ambient isotopic to the zero tangle, allowing the boundary of the
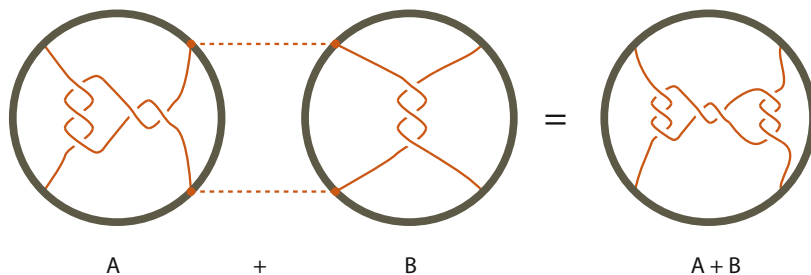
Fig. 14 Sum of two tangles

Fig. 15 Numerator closure of a tangle T and the numerator closure of the sum of two tangles T_1 and T_2 giving links $N(T)$ and $N(T_1 + T_2)$, respectively
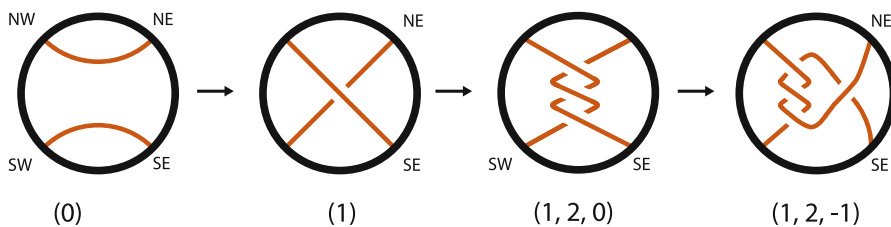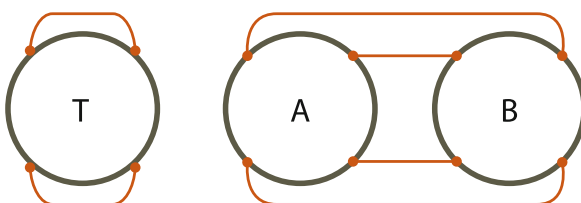

Fig. 16 Creating a rational tangle with Conway vector $(1, 2, -1)$

3-ball to move. A rational tangle diagram is created by starting with the zero tangle and interchanging the NE and SE boundary points a finite number of times creating **horizontal half twists**. Then, continue construction by interchanging the SW and SE boundary points a finite number of times creating **vertical half twists**. Continue in this manner, alternating between adding vertical and horizontal twists (Fig. 16).

John Conway associated to each rational 2-string tangle an extended rational number, $\frac{m}{n} \in \mathbb{Q} \cup \{\infty\}$, stating that there exists a 1–1 correspondence [11]. This number can be calculated using the **Conway vector**, denoted $(a_1, a_2, \ldots, a_i)$ where we choose i to be odd. This finite sequence of integers represents the sequence of moves performed on the zero tangle to produce a rational tangle. (Note: One can start with the (∞) tangle by rotating the zero tangle by $90°$.) Each integer represents the number of half twists given to the tangle, alternating between horizontal and vertical, ending with horizontal twists. The sign of the crossing follows that of Fig. 13.

If a tangle T is denoted $T(a_1, a_2, \ldots, a_i)$, then its extended rational number is calculated as:

$$\frac{m}{n} = a_i + \cfrac{1}{a_{i-1} + \cfrac{1}{a_{i-2} + \cfrac{1}{a_{i-3} + \cdots + \cfrac{1}{a_1}}}}$$

The numerator closure of a rational tangle, $\frac{m}{n}$, is referred to as a **2-bridge knot/link** denoted $N\left(\frac{m}{n}\right)$ or $< a_1, a_2, \ldots a_i >$. These links are also referred to as 4-**plats** and **rational knots/links**.

4 Biology Background

A crucial advancement in molecular biology was made when the structure of DNA was determined by James Watson and Francis Crick in 1953. Its structure revealed how DNA can be replicated and provided clues about how a molecule of DNA might encode directions for producing proteins [2].

Nucleic acids consist of a chain of linked units called **nucleotides**. Each nucleotide contains a deoxyribose, a sugar ring made of five carbon atoms which are numbered as seen in Fig. 17. This sugar ring then forms bonds to a single phosphate group between the third and fifth carbon atoms of adjacent sugar rings (Fig. 18). The backbone of a DNA strand is made from alternating phosphate groups and sugar rings. The four bases found in DNA are Adenine (A), Thymine (T), Cytosine (C), and Guanine (G). The shapes and chemical structure of these bases allow hydrogen bonds to form efficiently between A and T and between G and C. These bonds, along with base stacking interactions, hold the DNA strands together [2]. Each base is attached to the first carbon atom in the sugar ring to complete the nucleotide (Fig. 18).

The bonds between the sugars and the phosphate group give a direction to DNA strands. The asymmetric ends of the strands are called the 5′ (**five prime**) and 3′

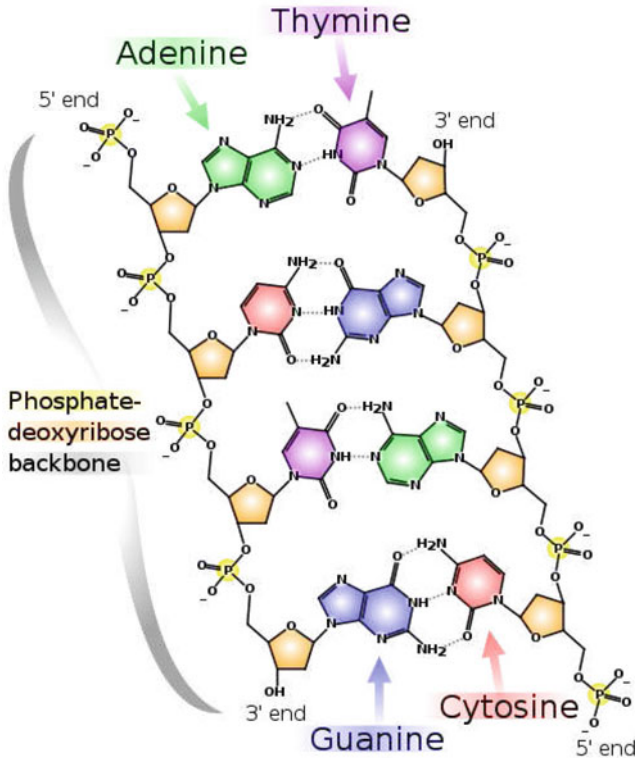**Fig. 17** Sugar ring made of five carbon atoms. Courtesy of [45]
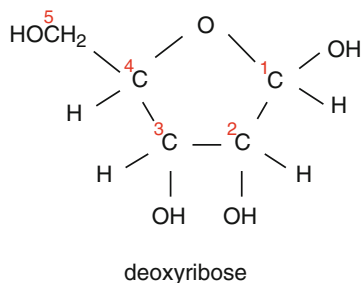
deoxyribose

Fig. 18 Deoxyribonucleic acid. Using the direction convention given to DNA strands, we read this sequence as $ACTG$, or equivalently $CAGT$. Courtesy of [46]

(**three prime**) ends, with the $5'$ end having a phosphate group attached to the fifth carbon atom of the sugar ring and the $3'$ end with a terminal hydroxyl group attached to the third carbon atom of the sugar ring (Fig. 18). The direction of the DNA strands is read from $5'$ to $3'$. In a double helix, the direction of one strand is opposite to the direction of the other strand: the strands are **antiparallel** [2].

Besides the standard linear form, a molecule of DNA can take the form of a ring known as **circular DNA**. One way to model circular DNA mathematically is as an annulus, **R**, an object that is topologically equivalent to $S^1 \times [-1, 1]$. The axis of **R** is $S^1 \times \{0\}$. With this model, we can choose an orientation for the axis of **R** and use the same orientation on $\partial\mathbf{R}$; thus, the axis and boundary curves of **R** have a parallel orientation. Note that this is a different convention than the biology/chemistry orientation. We use geometric invariants twist and writhe, denoted **Tw** and **Wr**, to describe the structure of the circular DNA molecule. **Writhe** can be determined by viewing the axis of **R** as a spatial curve and is measured as the average value of the sum of the positive and negative crossings of the axis of **R** with itself, averaged over all projections [28]. The sign convention for a crossing is
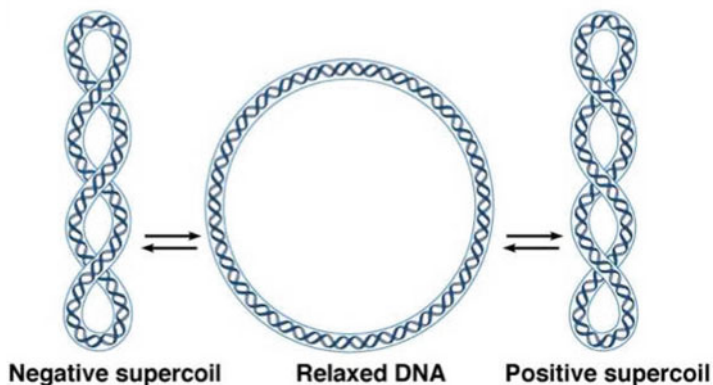
Fig. 19 Cartoon of negative, relaxed, and positive supercoiled DNA. Reproduced with permission from [22]

given in (Fig. 11). **Twist** is defined as the amount that one of the boundary curves of **R** twists around the axis of **R** [4].

One relationship between **Tw** and **Wr** is expressed in the following law:

LAW 4.1 (CONSERVATION LAW [20])

$$\mathbf{Lk(R) = Tw(R) + Wr(R)}$$

where **Lk(R)** is the linking number of the oriented link formed by the two boundary curves of **R** with a parallel orientation.

We say that a DNA molecule is **supercoiled** when **Wr**$\neq$ 0 (Fig. 19). Native circular DNA appears negatively supercoiled under an electron microscope, i.e., **Wr** < 0 (Fig. 20) [4].

Recall that the structure of DNA is a double-stranded helix, where the four bases are paired and stored in the center of this helix. While this structure provides stability for storing the genetic code, Watson and Crick noted that the two strands of DNA would need to be untwisted in order to access the information stored for transcription and replication [2]. They foresaw that there should be some mechanism to overcome this problem.

4.1 Transcription and Replication

DNA can be viewed as two very long strands; as closed curves that are intertwined and perhaps linked to other closed curves or tied into knots, and supercoiled. Thus, the main topologically interesting forms that circular DNA can take: supercoiled,
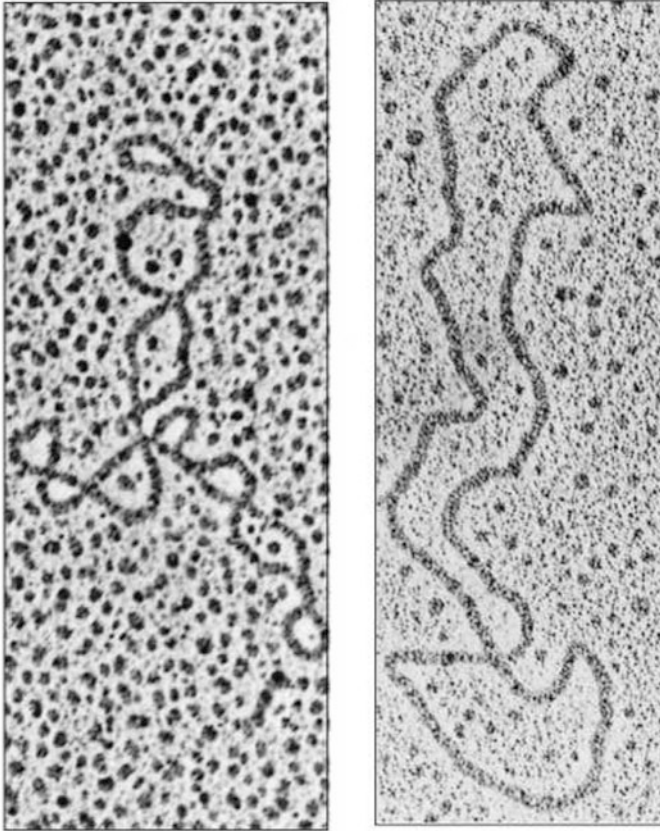
Fig. 20 Two examples of supercoiled DNA seen through an electron microscope. Reproduced with permission from [22]

knotted, linked, or a combination of these. DNA is kept as compact as possible when in the nucleus, and these three states help or hinder this cause. However, when transcription or replication occur, DNA must be accessible [41]. **Ribonucleic acid (RNA)** is a nucleic acid made up of a chain of nucleotides (Fig. 21). There are three main differences between RNA and DNA: (a) RNA contains the sugar ribose, while DNA contains a different sugar, deoxyribose; (b) RNA contains the base uracil (U) in place of the base thymine (T), which is present in DNA; and, (c) RNA molecules are single stranded, but have interesting tertiary structure. **Transcription** is the process of creating a complementary RNA copy of a sequence of DNA. Transcription begins with the unwinding of a small portion of the DNA double helix to expose the bases of each DNA strand. The two strands are then pulled apart creating an opening known as the **transcription bubble**. During this process,
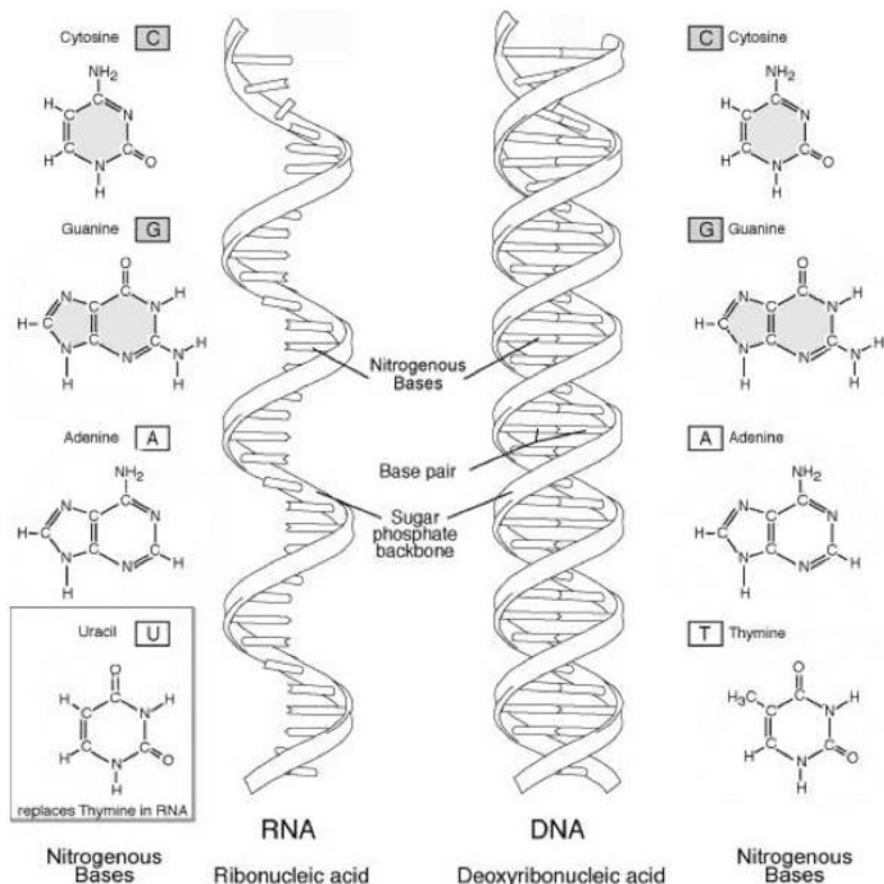
Fig. 21 Like DNA, **ribonucleic acid (RNA)** is a nucleic acid made up of a long chain of nucleotides. Courtesy of [47]

DNA ahead of the transcription bubble becomes positively supercoiled, while DNA behind the transcription bubble becomes negatively supercoiled (Fig. 22).

DNA replication is the process that starts with one DNA molecule and produces two identical copies of that molecule. During replication, the DNA molecule begins to unwind at a specific location and starts the synthesis of the new strands at this location, forming **replication forks** (Fig. 23, left). The DNA ahead of the replication fork becomes positively supercoiled, while DNA behind the replication fork becomes entangled, creating pre-catenanes, a state where the DNA molecules are beginning to form linked DNA molecules (Fig. 23, center). A topological problem occurs at the end of replication, when daughter chromosomes must be fully disentangled before mitosis occurs (Fig. 23, right) [41]. Topoisomerases play an essential role in resolving this problem.
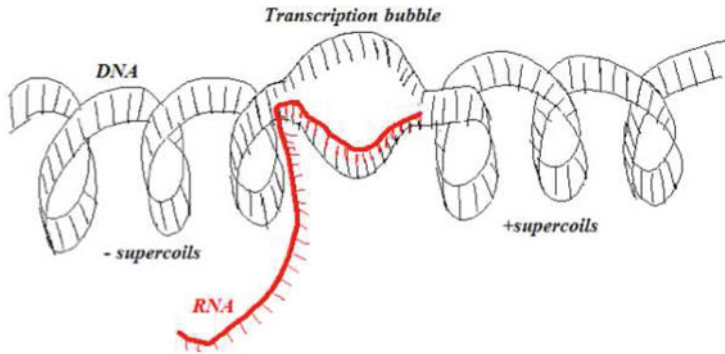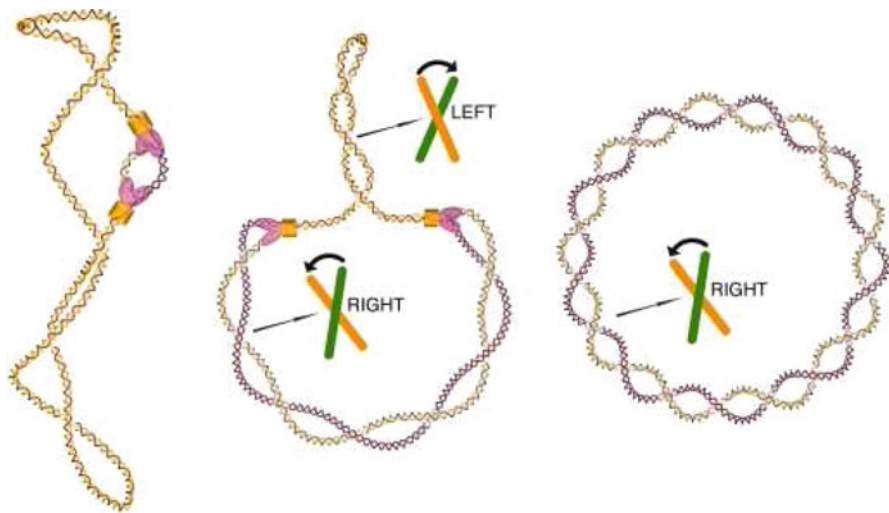
Fig. 22 Transcription-driven supercoils

Fig. 23 Topological changes to DNA during replication of circular DNA. The process of replication begins with negatively supercoiled DNA. The replication forks are shown in purple and gold. Partially replicated DNA molecule: the replicated portions of the DNA are interwound with positive (right-handed) crossings, creating a pre-catenane, while the remaining unreplicated DNA is still negatively (left-handed) supercoiled. Completely replicated DNA shown as a DNA catenane with positive (right-handed) crossings. Used with permission from [48]

4.2 Topoisomerase

Topoisomerases are proteins that are involved in the packing of DNA in the nucleus and in the unknotting and unlinking of DNA links that can result from replication and other biological processes. These proteins bind to either single- or double-stranded DNA and cut the phosphate backbone of the DNA. A **type I topoisomerase**
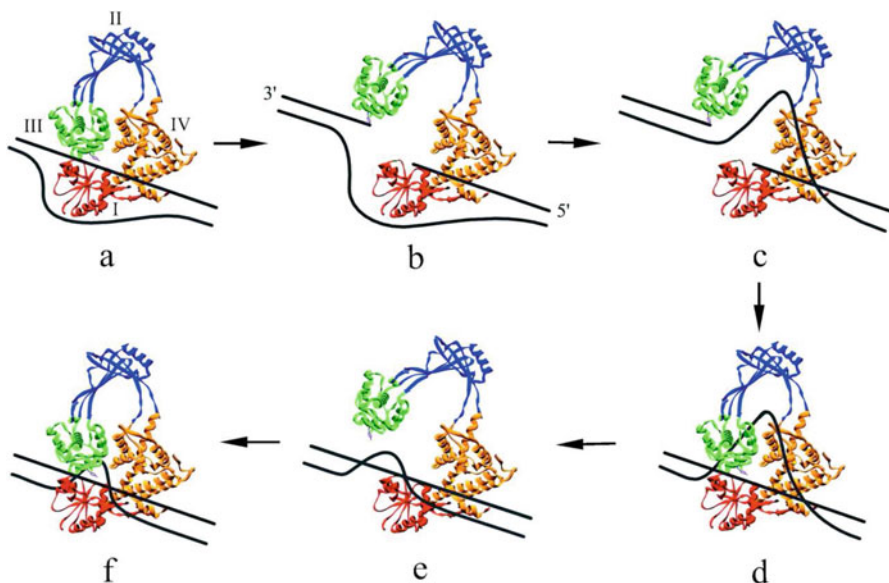
Fig. 24 Schematic of topoisomerase I action. Used with permission from [10]

cuts one strand of a DNA double helix allowing for the reduction or the introduction of stress (Fig. 24). Such stress is introduced or needed when the DNA strand is supercoiled or uncoiled during replication or transcription. **Type II topoisomerase** cuts both phosphate backbones of one DNA double helix, passes another DNA double helix through it, and then reseals the cut strands (Fig. 25). This action does not change the chemical composition and connectivity of DNA, but potentially changes its topology.

4.3 Recombinase

In various biological processes, there often is a need to integrate, excise, or invert portions of a DNA molecule. For example, gene expression is often regulated by the absence or presence of repressor or promoter sites. Inserting a promoter or repressor site can result in the expression, or lack of expression, respectively, of a particular gene. Another example is the insertion of viral DNA into its host cell. Insertion of the viral DNA into the host genome allows it to replicate and continue its life cycle. **Recombination** is a process involving the genetic exchange of DNA where DNA sequences are rearranged by proteins known as **recombinases** [2]. **Site-specific recombination** is an operation on DNA molecules where recombination proteins, **site-specific recombinases**, recognize short specific DNA sequences on the recombining DNA molecules. First, two sequences from the same or different
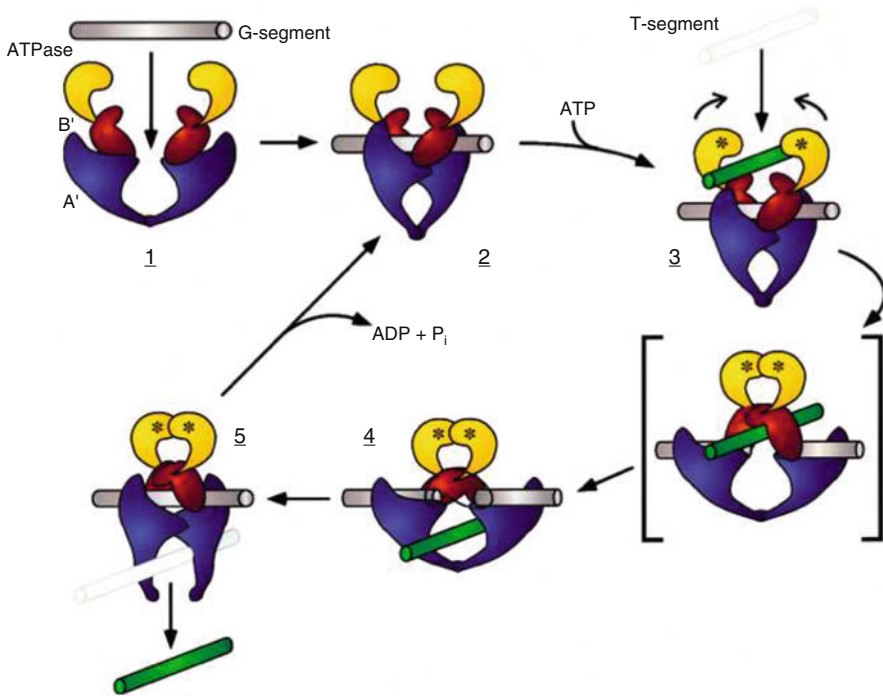
Fig. 25 Schematic of topoisomerase II action. Used with permission from [5]

DNA molecule are drawn together. The recombinase then introduces a break near a specific site, known as a **recombination site**, on the double-stranded DNA molecule. The protein then recombines the ends in some manner and seals the break (Fig. 26). We call this DNA-protein complex a **synaptic complex**. We will call the part of the synaptic complex that consists of only the protein together with the part of the substrate DNA bound to the protein, the local synaptic complex. After synapsis occurs, the recombinase then cleaves the DNA at the recombination sites and rejoins the ends by exchanging them. The specific way in which the exchange occurs is determined by the particular protein [21, 39, 43].

The DNA sequence of a recombination site can be used to give an orientation to this site. When two sites are oriented in the same direction, the sites are called **direct repeats** (Fig. 27). Recombinase action on direct repeats normally results in a change in the number of components, taking knots to links and links to knots or a link with a higher number of components (Fig. 27). If the two sites are oriented in opposite directions, the sites are called **inverted repeats** (Fig. 28). The action of a recombinase on inverted repeats normally results in no change in the number of components (Fig. 28).

There are two families of site-specific recombinases: **tyrosine recombinases** and **serine recombinases**. Tyrosine recombinases break and rejoin one pair of DNA
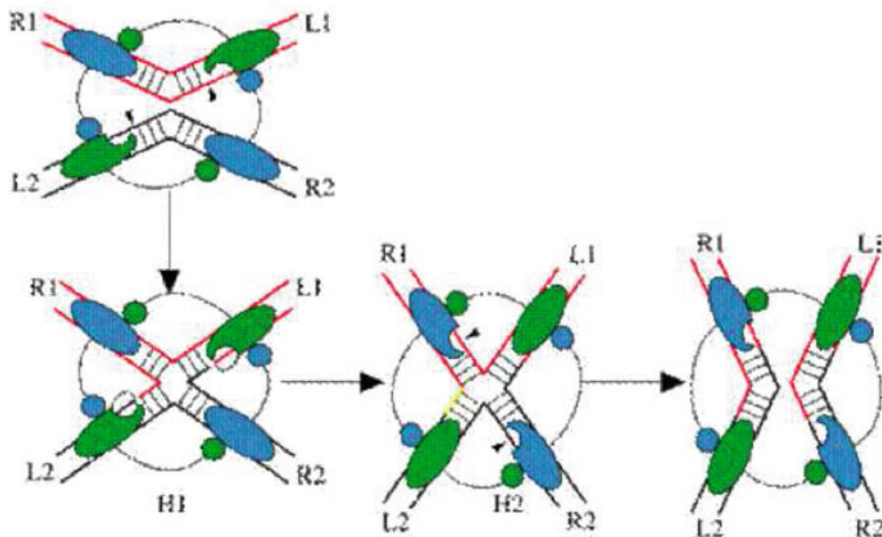
Fig. 26 An example of a site-specific recombinase mechanism where the protein makes breaks one strand of the double helix, recombines it, and then does the same with the other strand

Fig. 27 Recombinase action on direct repeats
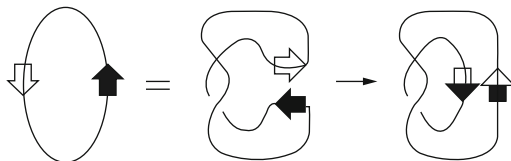
Fig. 28 Recombinase action on inverted repeats
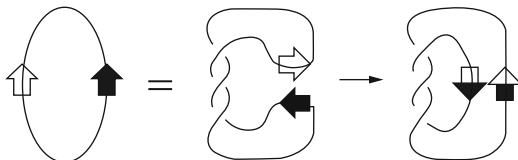
strands at a time (Figs. 26, 29). Serine recombinases introduce double-stranded breaks in DNA and then recombines them in some manner (Fig. 30) [36].

5 Tangle Model

DNA encodes much of the information a cell needs to survive and reproduce. If we were to unwind the chromosomes from one human cell and place the DNA strands end to end, it would span approximately 2 m [9]. All of this DNA is packed
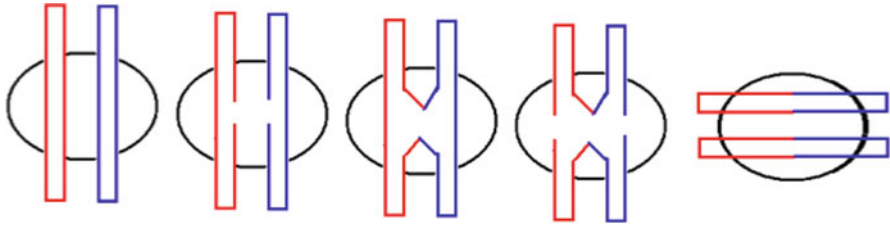
Fig. 29 Schematic of tyrosine recombinase action: single-stranded breaks. We model the tyrosine protein as a black ball, while the double-stranded DNA is modeled by red and blue rectangles

Fig. 30 Schematic of serine recombinase action: double-stranded breaks. We model the serine protein as a black ball, while the double-stranded DNA is modeled by red and blue rectangles
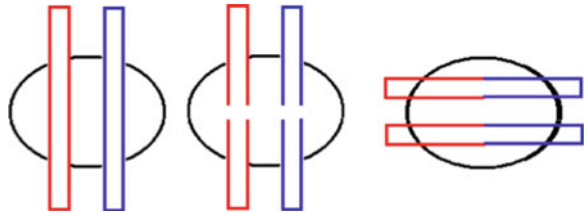
inside the nucleus of a cell whose diameter is measured on the scale of micrometers, that is one thousandth of a millimeter, 0.001 mm. The DNA must not only be arranged to sit inside such a small space, but it must also be organized so that the information it contains is accessible. Inside this complex environment, vital functions like transcription and replication must take place. It is no surprise, then, that various mechanisms have evolved over time to change the structure of the DNA. One mechanism is the action of proteins.

Understanding how a particular protein acts on DNA can be a difficult task. Proteins and their actions cannot be directly observed with the naked eye. Even with electron microscopy, there is not enough detail to see exactly how a particular protein binds to and acts on its substrate. We must rely on well-designed experiments to gain this knowledge. Additional use of mathematical models can help to further clarify the results obtained by experiments and this is exactly what was done to determine the action of a particular tyrosine recombinase called **Tn3 resolvase**.

In the 1990s, C. Ernst and D. Sumners developed the tangle calculus which was then successfully used to model the action of recombinases on circular DNA substrate [18]. In this model, the synaptic complex was represented by the numerator closure of a sum of 2-string tangles. A pair of 2-string tangle, O_b and P, represented the local synaptic complex, that is the protein and bound DNA. The parental tangle, P, contains the site where strand breakage and reunion takes place. The outside bound tangle, O_b, was the rest of the DNA in the local synaptic complex outside of the tangle P. Finally, another 2-string tangle, the outside free tangle, O_f, represented the DNA in the synaptic complex which is free and not bound to the protein. The action of the protein was then modeled as a tangle surgery, where the tangle, P, is replaced by a new tangle R. The knotted products which were observed
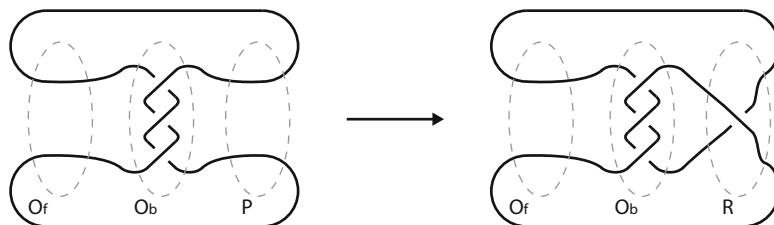
Fig. 31 A schematic of the tangle model. This particular example shows the first round of recombination for $Tn3$ resolvase

in experiments would allow for a system of tangle equations to be set up (see Fig. 31 for a visual):

$$N\left(O_f + O_b + P\right) = \text{substrate},\qquad(5.1)$$

$$N\left(O_f + O_b + R\right) = \text{product}.\qquad(5.2)$$

Several assumptions had to be made for this model to work [18, 37]. One assumption was that the local synaptic complex could be modeled with a 2-string tangle that subdivided into the sum of two tangles. It was assumed that the recombination takes place entirely inside the protein ball, while the substrate configuration outside the protein ball remains fixed. The protein mechanism in a single recombination event is assumed constant, and independent of the geometry and topology of the substrate. Also, it was assumed that **processive recombination**, consecutive reactions without releasing its substrate, could be modeled with tangle addition by adding the tangle R for each additional round of recombination:

$$N\left(O_f + O_b + P\right) = \text{substrate},\qquad(5.3)$$

$$N\left(O_f + O_b + R\right) = \text{1st round product},\qquad(5.4)$$

$$N\left(O_f + O_b + R + R\right) = \text{2nd round product},\qquad(5.5)$$

$$\vdots$$

$$N\left(O_f + O_b + \underbrace{R + R + \ldots + R}_{n}\right) = \text{nth round product}.\qquad(5.6)$$

Experiments with $Tn3$ resolvase acting on circular DNA substrate, which carried two copies of the recombination site, were carried out and the products of this reaction were observed. Resolvase typically mediates a single recombination event and releases the substrate. The principle product of the experiments were the Hopf link, $\langle 2 \rangle$, which was believed to be the result of this single recombination event. In about one in 20 encounters though, resolvase acts processively. Other products

observed were the Fig. 8 knot, $\langle 2, 1, 1 \rangle$, the result of two rounds of recombination, the Whitehead link, $\langle 1, 1, 1, 1, 1 \rangle$, the result of three rounds of recombination, and the 6_2 knot, $\langle 1, 2, 1, 1, 1 \rangle$, the result of four rounds of recombination [42, 44]. Using observations from electron micrographs of the synaptic complex, it was also assumed that O_f was the (0) tangle, thus we can reduce the tangle $O_f + O_b$ to one single tangle O. Using the information above, the following system of tangle equations could be set up:

$$N(O + P) = \langle 1 \rangle \text{ (the unknot)}, \tag{5.7}$$

$$N(O + R) = \langle 2 \rangle \text{ (the Hopf link)}, \tag{5.8}$$

$$N(O + R + R) = \langle 2, 1, 1 \rangle \text{ (the Fig. 8 knot)}, \tag{5.9}$$

$$N(O + R + R + R) = \langle 1, 1, 1, 1, 1 \rangle \text{ (the (+) Whitehead link)}. \tag{5.10}$$

Due to the amount of unknowns, it is not possible to explicitly solve for the tangle P; there are the infinite possibilities of a solution for any given O. However, there are biological and mathematical arguments to support the idea that $P = (0)$ [38]. Thus, with this assumption and using only the first three rounds of recombination and the tangle calculus, Ernst and Sumners were able to prove the following theorem about this system of equations:

Theorem 5.1 ([18]) *Suppose that tangles O, P, and R satisfy the following:*

$$N(O + P) = \langle 1 \rangle \text{ (the unknot)}, \tag{5.11}$$

$$N(O + R) = \langle 2 \rangle \text{ (the Hopf link)}, \tag{5.12}$$

$$N(O + R + R) = \langle 2, 1, 1 \rangle \text{ (the Fig. 8 knot)}, \tag{5.13}$$

$$N(O + R + R + R) = \langle 1, 1, 1, 1, 1 \rangle \text{ (the (+) Whitehead link)}. \tag{5.14}$$

Then, $\{O; R\} = \{(-3, 0), (1)\}$, and $N(O + R + R + R + R) = \langle 1, 2, 1, 1, 1 \rangle$.

Not only did this theorem show that the tangles O and R must equal $(-3, 0)$ and (1), respectively, but it also predicted that a fourth round of recombination would result in $\langle 1, 2, 1, 1, 1 \rangle$. This is exactly the product that was observed experimentally. This theorem can then be viewed as a mathematical proof that the synaptic complex structure as proposed by Wasserman et al. in [44] is the only possibility.

6 Further Applications of the Tangle Model

The use of tangle algebra to model the biological processes that give rise to knotting in DNA provides an excellent example of the application of topological algebra to biology. The tangle algebra approach to knotting in DNA began with the study of the site-specific recombinase $Tn3$ resolvase. It is assumed that this protein acts on

unknotted DNA processively, producing a series of products, thus providing ample information for systematic mathematical analysis [16]. The model arising from this assumption produced testable, and verified, predictions of knot products [44] and the tangle algebra approach made it possible to write down tangle equations that reflected the progressive repeat action of the protein [18].

A similar approach was taken to study the effect of many site-specific recombinase [6, 7, 13, 14, 29, 40]. Tangles continue to be used to describe the synaptic structure during recombination. One example is extending the tangle model to include 3-string tangles [15, 17, 23], and has also been used in [6, 8] to make predictions of the possible knots that may arise under different hypotheses about the substrate arrangement. Tangle algebras are also being used to study proteins that do not change the topological structure of DNA, but bind to it in interesting ways. This experimental technique has been the focus of many papers, yielding interesting results [24, 27, 31, 32].

References

1. C.C. Adams, *The Knot Book* (American Mathematical Society, Providence, 2004), An elementary introduction to the mathematical theory of knots, Revised reprint of the 1994 original. MR MR2079925 (2005b:57009)
2. B. Alberts, D. Bray, K. Hopkins, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Essential Cell Biology*, 2nd edn. (Garland Science/Taylor & Francis Group, New York, 2003)
3. J.W. Alexander, Topological invariants of knots and links. Trans. Am. Math. Soc. **30**(2), 275–306 (1928)
4. A.D. Bates, A. Maxwell, *Dna Topology* (Oxford University Press, Oxford, 2005)
5. J.M. Berger, S.J. Gamblin, S.C. Harrison, J.C. Wang, Structure and mechanism of DNA topoisomerase ii. Nature **379**(6562), 225–232 (1996)
6. D. Buck, E. Flapan, A topological characterization of knots and links arising from site-specific recombination. J. Phys. A **40**(41), 12377–12395 (2007). MR 2394909 (2010h:92064)
7. D. Buck, C.V. Marcotte, Tangle solutions for a family of DNA-rearranging proteins. Math. Proc. Camb. Philos. Soc. **139**(1), 59–80 (2005). MR 2155505 (2006j:57010)
8. D. Buck, K. Valencia, Characterization of knots and links arising from site-specific recombination on twist knots. J. Phys. A Math. Theor. **44**(4), 045002 (2011)
9. C.R. Calladine, H.R. Drew, B.F. Luisi, A.A. Travers, *Understanding DNA*, 3rd edn. (Elsevier Academic Press, Amsterdam, 2004)
10. J.J. Champoux, DNA topoisomerases: structure, function, and mechanism. Annu. Rev. Biochem. **70**, 369–413 (2001)
11. J.H. Conway, An enumeration of knots and links, and some of their algebraic properties, in *Computational Problems in Abstract Algebra (Proc. Conf., Oxford, 1967)* (Pergamon, Oxford, 1970), pp. 329–358. MR 0258014 (41 #2661)
12. P.R. Cromwell, *Knots and Links* (Cambridge University Press, Cambridge, 2004). MR MR2107964 (2005k:57011)
13. I.K. Darcy, Biological distances on DNA knots and links: applications to XER recombination. J. Knot Theory Ramifications **10**(2), 269–294 (2001), Knots in Hellas '98, Vol. 2 (Delphi). MR 1822492 (2002m:57008)
14. I. Darcy, J. Chang, N. Druivenga, C. McKinney, R. Medikonduri, S. Mills, J. Navarra-Madsen, A. Ponnusamy, J. Sweet, T. Thompson, Coloring the Mu transpososome. BMC Bioinf. **7**(1), 435 (2006)

15. I. Darcy, J. Luecke, M. Vazquez, Tangle analysis of difference topology experiments: applications to a mu protein-DNA complex. Algebr. Geom. Topol. **9**(4), 2247–2309 (2009)
16. P. Dröge, N.R Cozzarelli, Recombination of knotted substrates by tn3 resolvase. Proc. Natl. Acad. Sci. **86**(16), 6062–6066 (1989)
17. E. Eftekhary, *Heegaard floer homologies of pretzel knots*. arXiv preprint math/0311419.
18. C. Ernst, D. Sumners, A calculus for rational tangles: applications to DNA recombination. Math. Proc. Camb. Philos. Soc. **108**, 489–515 (1990)
19. P. Freyd, D. Yetter, J. Hoste, W.B.R. Lickorish, K. Millett, A. Ocneanu, A new polynomial invariant of knots and links. Bull. Am. Math. Soc. (N.S.) **12**(2), 239–246 (1985)
20. F.B. Fuller, Decomposition of the linking number of a closed ribbon: a problem from molecular biology. Proc. Natl. Acad. Sci. U. S. A. **75**(8), 3557–3561 (1978). MR 0490004 (58 #9367)
21. N.D.F. Grindley, K.L. Whiteson, P.A. Rice, Mechanisms of site-specific recombination. Annu. Rev. Biochem. **75**(1), 567–605 (2006), PMID: 16756503
22. J. Hardin, G.P. Bertoni, L.J. Kleinsmith, *Becker's World of the Cell*, 8th edn. (Benjamin Cummings, San Francisco, 2010)
23. H.C. Ibarra, D.A.L. Navarro, An algorithm based on 3-braids to solve tangle equations arising in the action of gin {DNA} invertase. Appl. Math. Comput. **216**(1), 95–106 (2010)
24. M. Jayaram, R.M. Harshey, The Mu transpososome through a topological lens. Crit. Rev. Biochem. Mol. Biol. **41**(6), 387–405 (2006)
25. V.F.R. Jones, A polynomial invariant for knots via von Neumann algebra. Bull. Am. Math. Soc. (N.S.) **12**, 103–111 (1985)
26. M. Khovanov, A categorification of the Jones polynomial. Duke Math. J. **101**(3), 359–426 (2000)
27. S. Kim, I.K. Darcy, *Topological Analysis of DNA-Protein Complexes*. Mathematics of DNA structure, function and interactions (Springer, Berlin, 2009), pp. 177–194
28. K. Murasugi, *Knot Theory and Its Applications* (Modern Birkhäuser Classics, Birkhäuser Boston Inc., Boston, 2008), Translated from the 1993 Japanese original by Bohdan Kurpita, Reprint of the 1996 translation [MR1391727]. MR 2347576
29. F.J. Olorunniji, D.E. Buck, S.D. Colloms, A.R. McEwan, M.C.M. Smith, W.M. Stark, S.J. Rosser, Gated rotation mechanism of site-specific recombination by *φc31 integrase*. Proc. Natl. Acad. Sci. **109**(48), 19661–19666 (2012)
30. P. Ozsváth, *Knot Floer Homology*. Advanced Summer School in Knot Theory, May 2009, International Center for Theoretical Physics
31. S. Pathania, M. Jayaram, R.M. Harshey, Path of DNA within the Mu transpososome. Cell **109**(4), 425–436 (2002)
32. C.R. Price, *A Biological Application for the Oriented Skein Relation* (ProQuest LLC, Ann Arbor, 2012), Thesis (Ph.D.)–The University of Iowa. MR 3078590
33. K. Reidemeister, *Knotentheorie* (Springer, Berlin, 1974), Reprint. MR MR0345089 (49 #9828)
34. D. Rolfsen, *Knots and Links*. Mathematics Lecture Series, vol. 7 (Publish or Perish Inc., Houston, 1990), Corrected reprint of the 1976 original. MR MR1277811 (95c:57018)
35. R.G. Scharein, *Interactive topological drawing*, Ph.D. thesis, Department of Computer Science, The University of British Columbia, 1998
36. M.C.M. Smith, H.M. Thorpe, Diversity in the serine recombinases. Mol. Microbiol. **44**(2), 299–307 (2002)
37. D.W. Sumners, Lifting the curtain: using topology to probe the hidden action of enzymes. Not. Am. Math. Soc. **42** (1995), 528–537.
38. D.W. Sumners, C. Ernst, S.J. Spengler, N.R. Cozzarelli, Analysis of the mechanism of DNA recombination using tangles. Q. Rev. Biophys. **28**, 253–313 (1995)
39. L.-P. Tan, G.Y.J. Chen, S.Q. Yao, *Expanding the scope of site-specific protein biotinylation strategies using small molecules*. Bioorg. Med. Chem. Lett. **14**(23), 5735–5738 (2004)
40. M. Vazquez, D.W. Sumners, Tangle analysis of Gin site-specific recombination. Math. Proc. Camb. Philos. Soc. **136**(3), 565–582 (2004). MR 2055047 (2005d:57013)
41. J.C. Wang, *Untangling the Double Helix* (Cold Spring Harbor Laboratory Press, 2009), DNA Entanglement and the Action of the DNA Topoisomerases

42. S.A. Wasserman, N.R. Cozzarelli, Determination of the stereostructure of the product of tn3 resolvase by a general method. Proc. Natl. Acad. Sci. U. S. A. **82**(4), 1079–1083 (1985). (English)
43. S.A. Wasserman, N.R. Cozzarelli, Biochemical topology: applications to DNA recombination and replication. Science **232**(4753), 951–960 (1986). (English)
44. S.A. Wasserman, J.M. Dungan, N.R. Cozzarelli, Discovery of a predicted DNA knot substantiates a model for site-specific recombination. Science **229**(4709), 171–174 (1985)
45. Wikipedia, *Deoxyribose—Wikipedia, the free encyclopedia*, 2012. Accessed 20 May 2012
46. Wikipedia, *DNA—Wikipedia, the free encyclopedia*, 2012. Accessed 20 May 2012
47. Wikipedia, *Nucleic acid—Wikipedia, the free encyclopedia*, 2012. Accessed 20 May 2012
48. G. Witz, A. Stasiak, DNA supercoiling and its role in DNA decatenation and unknotting. Nucl. Acids Res. **38**(7), 2119–2133 (2010)