

USING KNOT THEORY TO MODEL AND ANALYZE DNA REPLICATION
AND RECOMBINATION

by

Megan K. Farrell, B.S.

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Mathematics
December 2018

Committee Members:

David Snyder, Chair

Eugene Curtin

Anton Dochtermann

COPYRIGHT

by

Megan K. Farrell

2018

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

This work is protected by the Copyright Laws of the United States (Public Law 94-553, section 107). Consistent with fair use as defined in the Copyright Laws, brief quotations from this material are allowed with proper acknowledgment. Use of this material for financial gain without the author's express written permission is not allowed.

Duplication Permission

As the copyright holder of this work I, Megan K. Farrell, refuse permission to copy in excess of the "Fair Use" exemption without my written permission.

DEDICATION

In memory of my mother, who always supported my pursuit of mathematics.

ACKNOWLEDGMENTS

Thanks to my advisor, Dr. David Snyder, for all the assistance and encouragement he gave me during my mathematical studies.

Thanks to my mentor, Dr. Ray Treinen for his support and guidance through my academic career, and for motivating me to continue my mathematical studies.

Thanks to Dr. Eugene Curtin for his excellent teaching skills that piqued the interest of an adolescent mathematician.

Lastly, thanks to Dr. Anton Dochtermann for his assistance on this committee and for his wonderful instruction in my graduate academics.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS.....	v
LIST OF FIGURES.....	vii
CHAPTER	
I. INTRODUCTION	1
A Brief History of Knot Theory	1
About DNA.....	3
DNA as a Topological Ribbon	5
II. KNOT THEORY BACKGROUND	9
Reidemeister Moves and Isotopy.....	10
Tangles and 4-plats.....	13
Knot Polynomials	20
III. SITE-SPECIFIC RECOMBINATION.....	24
The Tangle Model.....	27
IV. TOPOISOMERASES.....	36
Type I Topoisomerases.....	37
Type II Topoisomerases	39
V. SIGNIFICANCE.....	42
VI. FUTURE DIRECTIONS FOR RESEARCH	43
BIBLIOGRAPHY	44

LIST OF FIGURES

	Figure	Page
1	Circular DNA Supercoiling	4
2	Positive (left) and Negative (right) Crossings.	6
3	Reidemeister Moves R1, R2, and R3 from top to bottom.	10
4	Achiral Figure Eight Knot.	11
5	From Left to Right, a Trivial, Rational, and Locally knotted tangle.	14
6	Left: Sum of two tangles. Right: Numerator closure.	17
7	Examples of Braids.	18
8	From left to right: L_+ , L_- , and L_0	20
9	Trefoil knot oriented clockwise	21
10	Crossing choice	21
11	L_+ crossing change	22
12	L_0 crossing change - Hopf Link.	22
13	From left to right, the Whitehead link and the 6_2 knot	30
14	Type I Topoisomerases Manipulation	38
15	Type II Topoisomerases manipulation on circular DNA.	40
16	Unknotting the 10_1 knot	41

I. INTRODUCTION

A Brief History of Knot Theory

We will begin this paper with a brief history of knot theory. From Celtic knots, sailor's knots, and even from Alexander the Great who cut the Gordian knot in 333 B.C., throughout history knots have played an essential role in human culture. However, the mathematical study of knots did not begin until 1883 when Johann Carl Frederick Gauss showed that the linking number of two knots could be found by the computation of the "Gauss Integral," or Linking Integral. See [15] for further detail. Applications of knot theory were first seen in the nineteenth century with William Thompson, also referred to as Lord Kelvin, and Peter Tait. After seeing Tait's colliding smoke ring experiments, Thompson conceived vortex atom theory, where the first attempt of applied knot theory was seen. Thompson believed atoms in the ether were knotted and linked vortex rings, and even though this was incorrect, it opened the door for knot theory to be applied to physics, chemistry, and DNA studies. Tait, agreeing with Thompson's theory, began to construct a table of elements to support Thompson's work. The table constructed by Tait was the first table of mathematical knots, some of which is still used to this day [16]. Tait's knot table contained knots with up to seven crossings, but this still left Tait unsatisfied – he wanted knots with even more crossings. Others joined Tait on his quest. When Tait had finally compiled a table of knots up to 11 crossings, totaling 1581 knots, Tait abandoned this mission. See [17] for Tait's original knot table, which was organized by crossing number.

We see interest in knot theory arise again in 1923 when J. W. Alexander [1] discovered what we now know as the Alexander Polynomial, which Alexander was able to prove was an invariant of the knot (that is isotopic knots will have

equivalent Alexander polynomials). This was the only known knot polynomial for over 50 years until, in 1984, Vaughan Jones [12] discovered yet another polynomial, later named (in Jones' honor) the Jones Polynomial. Jones' discovery inspired other researchers to look for additional knot polynomials. One such polynomial, the so-called HOMFLY polynomial, was discovered by a couple different research groups simultaneously (the name HOMFLY encodes the first letter of the last names of each of these mathematicians).

Applied knot theory came in to play after the discovery of the helical structure of DNA by James Watson and Francis Crick in the 1950s. This discovery opened the door for Liu and Davis, who discovered knotted DNA in the lab in 1981 [2]. In 1990, C. Ernst and De Witt Sumners [7] then applied a tangle model to site-specific recombination; this is where the study of modeling DNA recombination using knot theory indeed began. Modeling site-specific recombination using the tangle model will be discussed in more detail later; this will include discussing how the topology of DNA is modified via the manipulation of DNA strands by topoisomerase (I and II) enzymes. However, first, we will provide the necessary biological and mathematical background.

About DNA

DNA are long, thin molecules found in the nucleus of the cell that encode biological traits and are a mechanism for reproduction. DNA contains instructions in the form of a sequence of letters to construct other cellular components and is referred to as the blueprint of life. For us mere humans, our DNA sequence contains approximately 3 billion letters, but more complicated organisms can have even lengthier DNA sequences [8]. The letters in the DNA sequence are A (adenine), T (thymine), C (cytosine), and G (guanine) and they are attached to the sugar and phosphate molecules constructing “backbones” of the double helix. *Supercoiling* is the axis of DNA being twisted in space, or the over- and under-winding of the DNA strand. (See figure 1 for circular DNA supercoils.) Supercoiling is DNA’s process to make itself compact enough to fit into the nucleus of the cell while staying in order and not breaking. One can imagine an old, overused phone cord to visualize what supercoiled DNA would look like. When DNA is twisted, the strands become more tightly wound, which causes a great deal of strain on the strand. “DNA knots and links occur during replication (DNA copying) and recombination (DNA rearranging)” [2]. DNA must go through supercoiling to be condensed into the nucleus of a cell and thus must be acted upon by enzymes which control replication and other processes to reduce the torsional strain on the strand.

Supercoiling is useful in DNA recombination because it brings together sites of DNA that are otherwise not close to each other when the strand is elongated. To replicate though, DNA cannot be knotted, which is the result of a supercoiled strand. The enzyme dedicated to the unknotting of DNA by modifying the DNA topology is the *topoisomerase* enzyme family. Topoisomerases are enzymes that participate in the over- or under-winding of DNA. These enzymes are crucial to the

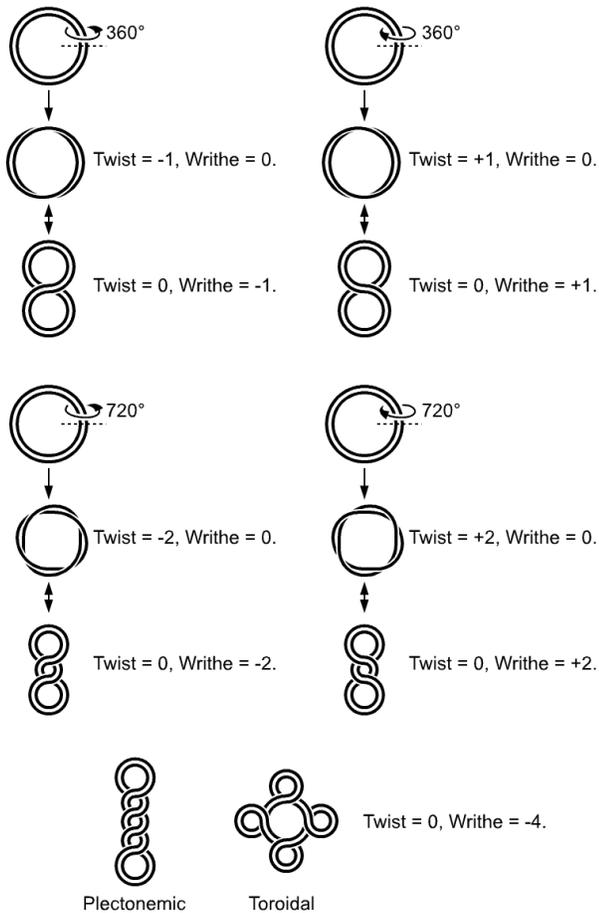


Figure 1: Circular DNA Supercoiling

reproduction of DNA because DNA must be unknotted in order to replicate itself. Essentially, topoisomerases cut a strand of DNA, allow another segment of DNA to pass through the cut, then reseal the broken segment. The topoisomerases families will be discussed in section IV of this paper.

DNA as a Topological Ribbon

We can visualize DNA as a topological ribbon, by imagining the two DNA backbones as the edges of the ribbon and the axis of DNA would be the vertical middle of the ribbon. If one takes the ends of the DNA strand and attaches them together, one has a circular band. One may wonder if a Möbius strip would occur with a circular DNA strand. Due to DNA matching specific sequences together, a Möbius strip will not be formed. Instead, the circular DNA forms closed loops and a $(2, m)$ -torus knot or link [2, 8]. Topoisomers are DNA molecules that differ only in linking number [2]. To maintain the equilibrium population of topoisomers of large DNA rings, two families of topoisomerases have evolved. The primary function of these is to inter-convert topoisomers. The two families of topoisomerases are Type I, which cleaves onto one backbone to change the linking number by one step, and Type II, which cleaves onto both backbones and change the linking number in steps of two. Next, we define the linking number and discuss how to examine DNA as a topological ribbon.

The spatial complexity of DNA can be represented using the equation (discovered by White in 1969 [10]): $Lk = Tw + Wr$, or (in words) the linking number equals the twist plus the writhe. Tw represents twist, how the DNA is winded through space, and Wr represents writhe, how the axis itself is contorted through space. The next question one may ask is how exactly is the twist and writhe computed. Intuitively, the *twist* is how many times the strand is, well, twisted around itself, similar to how many times the tie on a bag of bread is twisted, and the *writhe* is a measure of how many times the strand loops around itself. One can imagine taking a knot or link diagram in \mathbb{R}^3 and projecting into \mathbb{R}^2 . Then we could see the writhe where 2 points in the diagram intersect in \mathbb{R}^2 .

We define the *linking number* by taking an oriented link diagram D and to each crossing in D assign either a +1 or a -1, depending on the orientation of the two strands forming the crossing (See figure 2). Then,

$$lk(D) = \frac{1}{2} \sum_{crossings} \pm 1$$

The *writhe* of an oriented diagram is the sum of the signs of all its crossings and is denoted $w(D)$.



Figure 2: Positive (left) and Negative (right) Crossings

When it comes to DNA, we think of the twist as a way to measure how the two backbones of the DNA strand are wrapped around its central axis. If the linking number is positive, we say the DNA strand is *positively supercoiled*. Similarly, a negative linking number would imply the DNA strand is *negatively supercoiled*. The linking number cannot be altered without breaking a DNA strand since the linking number is a topological invariant. Later, we will discuss how enzymes change the topology of DNA by changing the linking number, but first we shall examine the topology of DNA.

Scientists examine DNA as a topological ribbon through a process called *electron microscopy*. Electron microscopy is a method used to visualize DNA in which the DNA molecule is coated so it thickens and stiffens, thus allowing the precise knot or link type of the DNA strand to be determined. Thus, the scientists can view the molecule and then determine the sign of each axis crossing which

leads to classifying the knot or link type of the DNA molecule. This process can be laborious especially with extremely long or highly knotted DNA strands. The other issue with this method is there are only a few (dozens) published electron microscopy images plus not all labs develop the expertise required to perform this task [2]. Since it is difficult to determine the topology of the DNA by a direct approach, the indirect approach, called the topological approach to enzymology, is implemented using cloning techniques. Genetically engineered circular substrate molecules are used for the topological enzymology experiments to make the topological changes easier to detect.

Agarose gel electrophoresis is an additional technique used to separate DNA knot and link types. Gel electrophoresis is used after the reaction has taken place to separate the reaction product into knots and links having different types. “First, the DNA products are put at the top of an agarose gel. Then a positive charge is put at the bottom of the gel which attracts the negatively charged DNA. The smaller (or more knotted up) the DNA product, the faster it will travel. After gel electrophoresis, the DNA knots and links are directly observed using electron microscopy” [7]. Since electron microscopy can be both difficult and time-consuming, it is helpful to use topological methods of characterizing knotted and linked DNA. “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). Under standard conditions, knots of greater MCN migrate more rapidly than those with lesser MCN” [2]. The issue is when one DNA strand moves faster than another: there is no way to tell by how many more crossings the MCN will be larger.

Another issue is the sheer amount of knots that exist – 1,701,936 knots with a minimum crossing number of 16 or less. So even if we did know the MCN

of the knot, we would have quite a few guesses as to which type of knot it is. The next step in trying to determine the MCN of DNA knots would be analyzing the DNA strand in 2-dimensional gel electrophoresis, which allows scientists to better separate prime knots with the same MCN. Unlike gel electrophoresis, unfortunately, 2-dimensional gel electrophoresis cannot distinguish relative migration of DNA knots with the same MCN. Thus, both scientists and mathematicians alike are interested in new methods for determining the DNA knot or link type.

II. KNOT THEORY BACKGROUND

Knot theory is applied to the study of DNA replication processes. A mathematical knot is an embedding of a simple closed curve into 3-dimensional Euclidean space. An unknot is a closed loop with no crossings. The minimum number of crossing changes needed for a knot to become the unknot is referred to as the *unknotting number* of the knot. We define a *link* as the union of a finite number of disjoint knots in three-dimensional space. In particular, a knot can also be considered to be a one-component link. Knot theory is a sub-field of a larger branch of mathematics called *topology*. Topology is the study of spaces and spatial properties that are preserved under continuous transformation. In topology, a circle, the boundary of a square, the boundary of a triangle, and the boundary of any regular polygon are all topologically equivalent, as they can be continuously transformed one into the other by a bijection (a one-to-one and onto function). Imagine a bowl constructed out of clay, then the the clay is pressed down and flattened, into a plate. This indicates that a bowl would be topologically equivalent to a plate. Topological equivalence allows a geometric object to be twisted or stretched, but breaking or cutting the object is not allowed. For knot diagrams, a knot is not allowed to pass through itself, as allowing strands to pass through one another would result in crossing changes and thus change the knot type. The interested reader is referred to [10] for a more mathematical exposition of these ideas.

Reidemeister Moves and Isotopy

Reidemeister's Theorem states that two links can be continuously deformed into each other if and only if any diagram of one can be transformed into a diagram of the other by a sequence of Reidemeister moves. Reidemeister moves are changes to knot or link diagrams shown in figure 3. The rest of the knot or link remains unchanged by these manipulations and is also not shown at the crossings changed by the Reidemeister moves. After any finite sequence of Reidemeister moves, the initial and manipulated diagrams are *isotopic*. In other words, knot diagrams D_1 and D_2 are isotopic if D_1 can be manipulated by finitely many Reidemeister moves to transform to D_2 . Two knot diagrams are *regular isotopic* if one can be transformed into the other without the use of the $R1$ move. Isotopy is also an equivalence relation in the collection of knot diagrams.

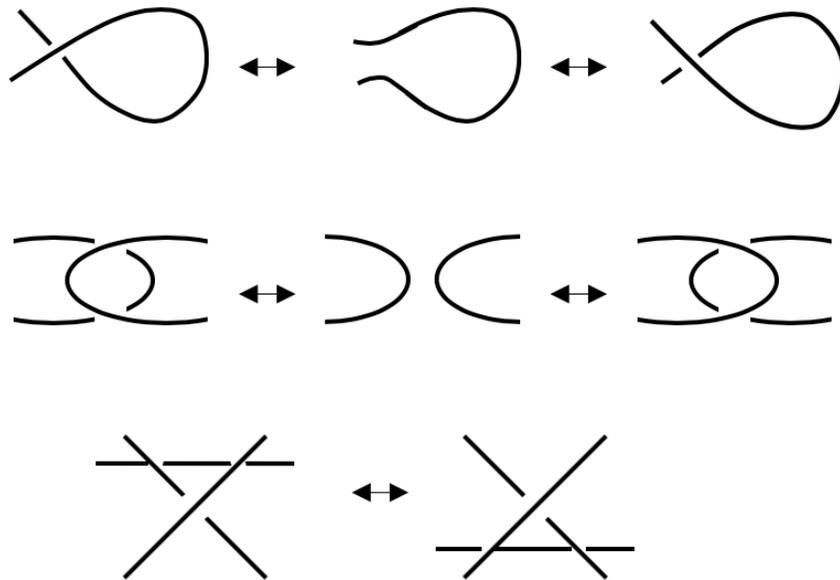


Figure 3: Reidemeister Moves R1, R2, and R3 from top to bottom

A knot is called *achiral*, or *amphichiral*, if it is isotopic to its mirror image.

Otherwise, the knot is said to be *chiral*. An oriented knot is *invertible* if it is isotopic to itself with reversed orientation, called the reverse of the knot. For example, the figure eight knot is regular isotopic to its mirror image, so the figure eight knot is achiral. See figure 4 to observe the steps in transforming the figure eight knot into its mirror image. To break down the steps, first the knot is rotated 180° , which is called an $R0$ move. Next, the arc on the left can be imagined as being lifted up and over the knot, which is a series of $R2$ and $R3$ moves. The last two steps can be thought of both as a series of $R0$ smoothing moves.

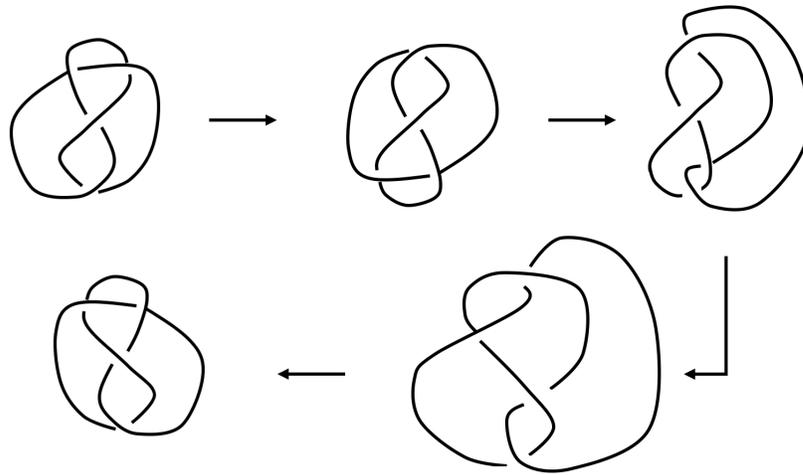


Figure 4: Achiral Figure Eight Knot

In \mathbb{R}^3 , we have the concept of *ambient-isotopy*, which is the idea that we are continuously moving the strings of the original knot to transform into a new knot. By definition, a knot K is *ambient-isotopic* to a knot L if by a continuous sequence of homeomorphisms of \mathbb{R}^3 from time $t=0$ to $t=1$ and if there exists a continuous function $H: \mathbb{R}^3 \times [0, 1] \rightarrow \mathbb{R}^3$ such that

- (1) $h_0 = H(-, 0)$ is the identity $\mathbb{R}^3 \rightarrow \mathbb{R}^3$
- (2) for all $t \in [0, 1]$, $h_t = H(-, t)$ is a homeomorphism $\mathbb{R}^3 \rightarrow \mathbb{R}^3$
- (3) if $h_1 = H(-, 1)$, then $h_1(K) = L$.

Isotopy of knot diagrams using Reidemeister moves is the topological analog of ambient-isotopy of knots and links in \mathbb{R}^3 . This brings us to the following theorem:

Theorem 1. *(See Gilbert and Porter [10]) Two knots are ambient-isotopic if and only if they possess isotopic diagrams.*

It should also be noted that one can show two tangles are equivalent by performing a finite sequence of Reidemeister moves which do not move the endpoints of the tangles.

Tangles and 4-plats

Conway developed the theory of tangles in 1967 [5], and then Ernst and Sumners built upon Conway's work in 1990 [7]. Knot theory and tangle theory are quite similar, with the exception that for tangles we focus on a location inside a ball and imagine the boundary of the ball fixed so we can only operate on the strings inside. We will first discuss some definitions of tangle theory before we how tangles are used to model DNA replication. A *3-ball*, or *three-dimensional ball*, is defined as $\{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 \leq 1\}$, which is also referred to as the unit ball. Next, we define an *n-tangle* to be the proper embedding of a disjoint union of n arcs into a 3-ball. The four endpoints of a tangle diagram are typically labeled as NE, NW, SW, SE. If we imagine projecting the 3-ball onto \mathbb{R}^2 , then we would have a unit disk where the y -axis represents north and south, and the x -axis represents east and west. The formal definition of a *tangle* (B, t) , where t consists of two unoriented arcs, is “a three-dimensional ball B , containing two disjoint arcs together with a finite number (possibly zero) of disjoint simple closed curves that are all represented by t , such that the intersection of t with the boundary of B is precisely the set of endpoints of the two arcs. Furthermore, there is a fixed orientation-preserving homeomorphism from B to the three-dimensional unit ball that takes the endpoints of the arcs to the specified points, $\{\text{NE}, \text{NW}, \text{SW}, \text{SE}\}$ ” [9]. We say that two tangles are *equivalent* if there exists an orientation-preserving homeomorphism between the two tangles, or, equivalently if we can transform one tangle into the other via a finite sequence of Reidemeister without moving the endpoints of the arcs.



Figure 5: From Left to Right, a Trivial, Rational, and Locally knotted tangle

A *trivial* tangle is essentially a ball with two strings that are not tangled or twisted. The 0 -tangle has two horizontal unknotted arcs and the ∞ -tangle has two unknotted vertical arcs, called *strands*. There are three types of tangles (See figure 5), not counting the trivial tangles, and these three tangles are mutually exclusive, no tangle can be in more than one family. A tangle is said to be *locally knotted* if it contains a knotted strand. A *rational* tangle is a 2-tangle that is homeomorphic to the trivial 2-tangle by a map of pairs consisting of the 3-ball and two arcs. Lastly, a tangle is *prime* if it is neither rational nor locally knotted. We are mainly interested in discussing rational tangles, which are also referred to as a family of tangles that can be transformed back into the trivial tangle by a twisting of the endpoints [9]. It should be noted that rational tangles are *not equivalent* to the trivial tangle because, for equivalence of tangles, the endpoints cannot be moved when transforming one into the other. Rational tangles are interesting and amenable to analysis because of the vertical and horizontal twists in them, which are similar to the electron micrographs of DNA strands [9].

Rational tangles also can be represented by vectors and continued fraction decomposition via their isotopy classes, which was discovered by Conway. For the vector representation, we will have a vector in the form $(\alpha_1, \alpha_2, \dots, \alpha_n)$ where each $\alpha_i \in \mathbb{Z}$ for all $1 \leq i \leq n$. Each α_i is an instruction on how to untwist the tangle to transform it into the 0-tangle, denoted (0) , or the 0-tangle rotated 180° , denoted $(0, 0)$, also referred to as the ∞ -tangle. Thus, all α_i for $1 \leq i \leq n - 1$ must be nonzero

to represent a rational tangle. The vector representation is used to instruct on how to transform from the rational tangle to the trivial tangle. Begin with α_n horizontal half-twists, then perform α_{n-1} vertical half-twists, then α_{n-2} horizontal half-twists, and so on repeating the alternating pattern. If α_n is zero, this indicates to begin untwisting the tangle using α_{n-1} vertical half-twists. The positive integers represent positive twists, which implies a positive slope of the overcrossing strand. Similarly, a negative integer corresponds to a negative slope for the overcrossing strand. The vector representation has a relation to the rational representation of the tangle. Let $p, q \in \mathbb{Z}$, $p, q \neq 0$, then each vector representation $(\alpha_1, \alpha_2, \dots, \alpha_n)$ that corresponds to a specific tangle has unique $\frac{p}{q}$ so that:

$$\frac{-p}{q} = \alpha_n - \frac{1}{\alpha_{n-1} - \frac{1}{\alpha_{n-2} - \frac{1}{\dots - \frac{1}{\alpha_2 - \frac{1}{\alpha_1}}}}}$$

If two fractions are equal, then the two rational tangles which they represent are of the same class, i.e. *isotopic*. If you have a vector representation that is only one integer, say (α_0) , we call that an *integer tangle* since its vector consists of only one integer. Integer tangles are made in a way similar to the way rational tangles are made. Begin with the trivial tangle and do a horizontal twist of the two endpoints NE and SE. Once one performs a vertical twist of the two endpoints NW and NE, a rational tangle could be obtained, and only horizontal twists are allowed for integer tangles. Thus the integer tangles are a subfamily of the rational tangles.

Now we discuss tangle operations. First, tangle addition is defined as, given

two tangles A and B , the gluing of NE of A to NW of B , and SE of A to SW of B [19]. Generally, the addition of two tangles will yield a new third tangle, with the exception of adding the trivial tangle. The trivial tangle acts as an identity for tangle addition in the sense that any tangle, A , summed with the trivial tangle will result in a tangle equivalent to the original tangle, A . The *Montesinos tangles* are a class of prime tangles obtained by the addition of rational tangles.

Next, the numerator closure and denominator closure tangle operations are defined. Given a tangle T , the numerator closure, $N(T)$, is formed by connecting the NW and NE endpoints and the SW and SE endpoints with an unknotted arc. (See figure 6) The denominator closure $D(T)$ of T would be connecting the NW and SW endpoints and then the NE and SE endpoints. Both the numerator and the denominator closure produce knots or links. With rational tangles, the resulting knot of a numerator closure is called a *4-plat* knot. Even though 4-plat knots can be formed by the numerator closure of rational tangles, and they can even be represented with vectors, all 4-plats are prime knots. Instead of using the $(\frac{p}{q})$ notation as we do for rational tangles, the notation for a 4-plat that is equal to the numerator closure of rational tangle $(\frac{p}{q})$ is denoted $b(p, q)$. Then $b(1, 1)$ would be a 4-plat representation of the unknot. Since the Montesinos tangles are rational, their numerator closure yields a knot referred to as the *Montesinos knot or link*. Since knots are one component links, we will refer to both types as *Montesinos links*. Buck and Flapan used a model to predict which knots and links arise as products of site-specific recombination, as discussed in the next section, and their model predicts such knots and links fall within the family of Montesinos knots and links [2].

A 4-plat (also known as two-bridge) is a knot made by braiding four strings and connecting the ends. By definition, bridge knots (or plats) are knots made by

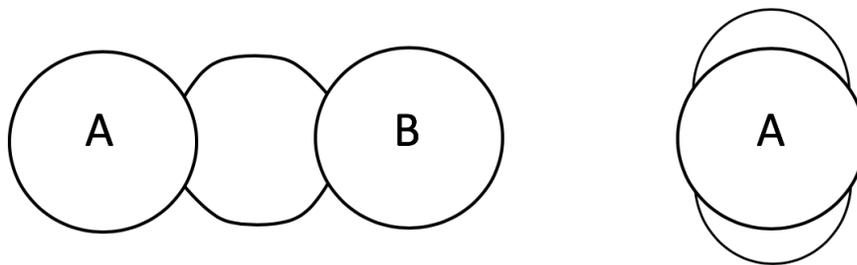


Figure 6: Left: Sum of two tangles. Right: Numerator closure

interlacing three or more strands and then attaching the ends together. The formal definition given in [9] states: a knot or link L is said to be a *two-bridge* knot or link if there exists a rational tangle A such that $L = N(A)$. With this definition, two-bridge knots and links are seen to be a subfamily of Montesinos links since the Montesinos links are the numerator closure of a sum of rational tangles. A (mathematical) *braid* is made by twisting n strands in such a way that they are between two horizontal bars and they begin, without loss of generality, at the top bar and intertwine then stop at the bottom bar (See figure 7). The *trivial n -braid* would have n parallel strands. Thus, a 4-plat is a form of a braid. Prime knots and certain links can also be classified as 4-plats (and hence, as braids), specifically prime knots with crossing number less than eight, and links with crossing number less than seven. “Since rational tangles and 4-plats are formed by twisting strings, we find them to be the perfect candidate for modeling DNA” [9]. Since DNA is supercoiled, the most common DNA knots and links are members of the 4-plats, as are most observed products of recombination experiments. A DNA knot is defined as the self-entanglement of a single DNA molecule; therefore, this excludes catenane structures that are formed by more than one chain.

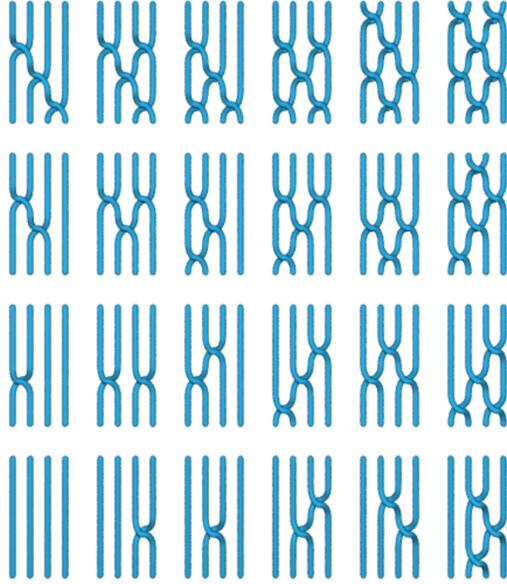


Figure 7: Examples of Braids

We also need the following definition:

Definition 1 (Twofold Branched Cover). *Let M and N be three-manifolds, and let $h : M \rightarrow M$ be an orientation-preserving homeomorphism with an order of two. Let $p : M \rightarrow N$ be defined such that $p(x) = p(y)$ if and only if either $x = y$ or $h(x) = y$. Suppose that p is a continuous onto open map. Let A denote the set of points $x \in M$ such that $h(x) = x$. If $B = p(A)$ is a one-manifold, then we say that M is a twofold branched cover of N branched over B [9].*

If $T = (B, t)$ is a tangle, then the twofold branched cover of the ball B over t is the twofold branched cover of T . For example, the twofold branch cover of the trivial tangle is a solid torus [9]. Then the twofold branch cover of any rational tangle will be the solid torus, because all rational tangles are homeomorphic to the trivial tangle. Thus we may infer that a tangle is rational if and only if its twofold branch cover is a solid torus [9]. If a tangle is neither rational nor trivial, its twofold branched cover B over t would have the boundary of the torus, but

would be some knot or link in S^3 .

Knot Polynomials

Knot polynomials are an excellent and efficient method for classifying knots. There are several types of polynomials. The first is the Alexander polynomial, which uses the minimum number of crossings, labeling of knot diagrams, and finding determinants [11]. The Alexander polynomial is also invariant under isotopy via Reidemeister moves. The approach used to define the Alexander polynomial is similar to the one used to develop the group presentation of knots constructed by Dehn. Another polynomial is the HOMFLY polynomial, described in more detail below. In general, each knot polynomial is an invariant under isotopy via Reidemeister moves.

Below we will compute the HOMFLY polynomial of the trefoil knot. The HOMFLY polynomial is defined inductively as follows:

$$P(\text{unknot}) = 1$$

$$lP(L_+) + l^{-1}P(L_-) + mP(L_0) = 0$$

where L_+, L_- , and L_0 are defined in figure 8

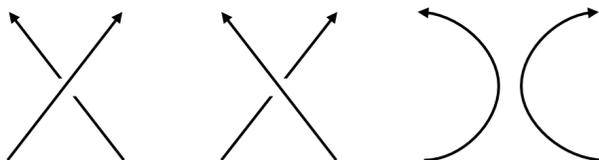


Figure 8: From left to right: L_+, L_- , and L_0

To compute the HOMFLY polynomial of the trefoil knot, we will begin with the oriented trefoil knot in figure 9.

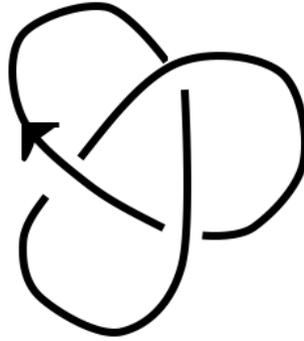


Figure 9: Trefoil knot oriented clockwise

Then we select a crossing to begin manipulation on the knot in order to determine its polynomial. Figure 10 shows which crossing we choose.

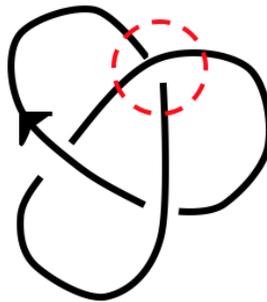


Figure 10: Crossing choice

Next we perform the crossing changes needed to obtain L_+ , L_- , and L_0 from figure 8. Since the crossing we chose is a negative crossing, the L_- is already in place. So, we will further refer to $l^{-1}P(L_-)$ as $l^{-1}P(\text{trefoil})$. To obtain L_+ , we perform a crossing change by switching the chosen crossing from an under-crossing to an over-crossing, as illustrated by figure 11.

By a sequence of Reidemeister moves, L_+ is isotopic to the unknot. Thus, we will further refer to $lP(L_+)$ as $lP(\text{unknot})$. Finally, the crossing change that creates L_0 will connect the arcs of the knot to delete the crossing. The result is the Hopf

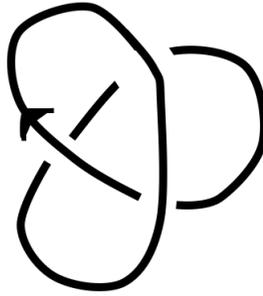


Figure 11: L_+ crossing change

link in figure 12.

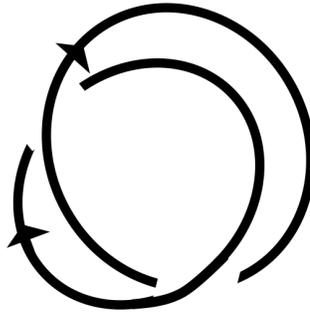


Figure 12: L_0 crossing change - Hopf Link

Now, we will refer to $mP(L_0)$ as $mP(\text{Hopf link})$. It should be noted here that it is assumed we already know the HOMFLY polynomial of the Hopf link. The reader may check for clarity, for a precise explanation see [10], but it will be given as $P(\text{Hopf link}) = l^3m^{-1} + lm^{-1} - lm$.

Then, our equation becomes:

$$lP(\text{unknot}) + l^{-1}P(\text{trefoil}) + mP(\text{Hopf link}) = 0$$

Since $P(\text{unknot}) = 1$ and $P(\text{Hopf link}) = l^3m^{-1} + lm^{-1} - lm$, we have:

$$l(1) + l^{-1}P(\text{trefoil}) + m(l^3m^{-1} + lm^{-1} - lm) = 0$$

$$l^{-1}P(\text{trefoil}) = -l - m(l^3m^{-1} + lm^{-1} - lm)$$

$$l^{-1}P(\text{trefoil}) = -l - l^3 - l + lm^2$$

$$P(\text{trefoil}) = -2l^2 - l^4 + l^2m^2$$

This last polynomial is the HOMFLY polynomial of the trefoil knot.

The HOMFLY polynomial proves useful when one needs to obtain the polynomials of both the reverses and the mirror images of knots.

Proposition 1. *Let L be an oriented link, with reverse \bar{L} and mirror image L' .*

Then, $P(L)(l, m) = P(\bar{L})(l, m) = P(L')(l, m)$ [10].

III. SITE-SPECIFIC RECOMBINATION

DNA replicates (which will be discussed later in this paper) but also goes through a *recombination* process. Recombination facilitates genetic diversity. We are interested in the type of knots and links that occur after DNA recombination occurs. *Site-specific recombination* is defined as the creation of new genetic sequences out of pieces of existing genetic sequences, which has a consequence of topologically manipulating the DNA strands. Generally, these changes and genetic rearrangement achieved by the recombination introduces knots and links into the DNA strands, which is why we say the topology of the strand has been altered.

Site-specific recombination is a process by which a block of DNA is moved to another position on the molecule, or a block of viral DNA is integrated into a host genome. Site-specific recombination is mediated by an enzyme called *recombinase*, which reshuffles the genetic sequence and regulates genes. The process either deletes, inserts, or inverts the DNA segment [2, 18]. Once bound to the DNA, recombinase breaks the DNA on each side of the double helix and then exchanges the two segments to recombine them. The DNA molecule before recombination begins is called the *substrate*. The first stage of site-specific recombination is binding two recombinase molecules binding onto the closed, circular DNA. The binding sites are called *crossover sites* [3]. The second stage is the exchanging of the sites. The stage is followed by resealing the DNA segment. There may be only one exchange or there may be a sequence of multiple rounds of strand exchange, which is referred to as *processive recombination* [3]. Lastly, the rearranged DNA, referred to as the *product*, is released.

As with to topoisomerases, there are two families of the recombinase enzyme. Before recombination occurs, the *serine* enzyme traps a specific

number of supercoils by the binding of nonactive resolvase molecules. Similarly, *tyrosine* enzymes also trap many supercoils, but they differ by needing additional (accessory) proteins and enhancer sequences at the DNA recombination sites to assist with the unique complex that cleaves onto the DNA. Thus the serine and tyrosine enzymes differ in the mechanism used to manipulate DNA at the recombination sites via cutting and resealing the strand. The serine enzyme cleaves to four backbones of the DNA strand and can perform processive recombination. The tyrosine enzymes only cleave to two of the backbones, exchange them, reseal the broken segment, and then repeat this process on a different pair of backbones.

A single round of recombination by serine recombinase enzyme and a single round of recombination mediated by the tyrosine enzymes both result in multiple possibilities of DNA knots or links. In particular, we see after one round of recombination via Tn3 on an unknotted strand that the result is the Hopf link. Since site-specific recombination can result in supercoiled DNA transforming into links or knots, we are interested in categorizing such links and knots that arise after this process. “Several approaches have been developed to determine a particular DNA knot or catenane type, including utilizing the node number for knots, the Jones polynomial for catenanes, Schubert’s classification of 4-plats, and the HOMFLY polynomial [2].”

Let us now examine a circular DNA molecule. Due to the biological structure of DNA with its sequence of base pairs, the DNA strand has an orientation. If we examine the substrate at the recombination site, the substrate orientation is a factor to which knots and links arise after recombination. If the orientation of the substrate is the same as the DNA strand, we see *direct repeats* in the sequence. The result of recombination via direct repeats is a DNA *link* (biologically referred to as a catenane) of the two circular DNA strands. Otherwise,

the orientations are opposing, and we see *indirect repeats* in the site configuration. The result of recombination via indirect repeats is a DNA *knot* [7].

The model used to predict DNA knots and links resulting from recombination was developed by Dorothy Buck and Erica Flapan [2]. They approach this action topologically and generally, rather than focusing on the recombinase enzymatic action. Given a supercoiled strand of DNA, there can be single or multiple rounds of recombination. This model considered, from a topological standpoint, three types of knots and results of recombination. Buck and Flapan's model can predict, from previously uncharacterized data, knot and link types that may arise during recombination, determine the pathway of recombination, and predict any knot or link product that is of a new family must have arisen by distributive recombination. Next, we will examine the use of the tangle model on the Tn3 enzyme.

The Tangle Model

This section discusses the tangle model of recombination which allows us to use mathematical deduction to analyze the process. “The tangle model studies topological changes in DNA caused by the enzymes” [6]. Ernst and Sumners originally developed this model using Conway’s theory of tangles specifically for a recombinase called *Tn3 resolvase*, though their model can be used for other proteins as well. The purpose of the tangle model is to use reasonable biological assumptions about DNA during site-specific recombination and then use mathematics to derive a conclusion. “One of the goals of the tangle model is to compute the topology of the synaptosome (enzyme + bound DNA), before and after the enzymatic action” [6]. This model addresses the action inside the enzyme ball since that is where the DNA strands are cut and recombined. The observation of DNA using electron microscopy allows us to see the DNA strands winding about each other. We can use rational tangles to model protein-DNA complexes since the twisting of strands forms the complexes. A tangle is used to model DNA by assuming the recombinase enzyme is the 3-ball, and the strings are the recombination sites [18]. We cannot observe what exactly is happening inside the ball, so for example, the two sites could be more twisted than in our assumption, or there could be additional pieces of the substrate that intersect the ball. Our assumption is the most reasonable one, however. Also if precisely what was happening inside the ball was observable or detectable, we would not need to make and study this model. The other assumption made is that all of the enzymatic action takes place inside the ball and only the pieces of DNA inside the site of the tangle are affected. Finally, we assume the tangles are assumed to be rational.

To begin with, note that the enzyme-DNA complex can be represented as a sum of tangles. Each DNA substrate molecule can be visualized as a knot or link embedded in 3-space. Specifically, the DNA with the bounded enzyme(s) form a 2-string tangle, where a 2-string tangle is a 3-ball with two embedded strings [6]. Let E represent the enzyme, the part of the unchanged DNA bound to the enzyme, and P the site that changed during the reaction. Thus, the enzyme-DNA complex can be represented by $E = O_b + P$. Then we represent the tangle of DNA not bound to the enzyme by O_f . This leads us to one tangle equation, $N(O_f + O_b + P) = K_0$, which models the substrate molecule before recombination. The next assumption we must make is that recombination acts by tangle surgery, where the site tangle P is replaced by the recombinant tangle R after one round of recombination. By this assumption, P will be replaced by R after the recombination process, which leads us to two equations - the substrate molecule equation mentioned above and the product molecule equation for after recombination, $N(O_f + O_b + R) = K_1$. Assuming the mechanism of recombination is constant, independent of the substrate geometry and topology, means all of the substrate molecules are all of the same knot type. Then the tangles will not change from one event to another. The tangle O_f is the only tangle that would change if the substrate molecules were of different knot types. Now consider what happens after n rounds of recombination. We know after one round of recombination, P is replaced by R , then after n rounds, we have $nR = R+R+R+\dots+R$. Thus, the model for processive recombination is given by $N(O_f + O_b + nR) = K_n$ [9]. The reason we can assume we have this repeated addition of R is due to a biological assumption rather than a mathematical one. The biological assumption is that the enzyme acting via recombination performs the same act, even during multiple rounds of recombination events. Thus, this leads us to our

mathematical assumption that n rounds of recombination can be represented by nR , and also that processive recombination can be modeled by tangle addition.

Multiple rounds of strand exchange can occur before releasing the DNA; this process is known as *processive* recombination, which can be modeled by tangle addition [19]. DNA knots and links are formed as a result of a single recombination event or multiple rounds of processive recombination. When using the tangle model to analyze a specific enzyme, one first must prove rationality of the tangles in question, which requires deep results in topology, such as the Cyclic Surgery Theorem. Let us now discuss such theorems, specifically the results of the work done by Ernst and Sumners [9].

In 1990, Ernst and Sumners proved the correctness of the original model given by Wasserman about the action of TN3 resolvase [9]. The Tn3 enzyme works only through directly repeated recombination and acts upon negatively supercoiled DNA strands. Let T represent the *site tangle*, that is, the tangle before enzyme action. Let S denote the *substrate tangle*, or the tangling occurring outside of the enzyme. Let R represent the *recombination tangle*, *i.e.* the result of T after the recombination event has occurred. Then, we can view the original substrate as $N(S + T)$, and $N(S + R)$ as the substrate after recombination has occurred. Now we examine the main theorem proved by Ernst and Sumners [7].

Theorem 2. *Suppose that T, S, R are tangles satisfying the following equations:*

$$N(S + T) = \text{the unknot} \tag{1}$$

$$N(S + R) = \text{the Hopf link} \tag{2}$$

$$N(S + R + R) = \text{the figure eight knot} \tag{3}$$

$$N(S + R + R + R) = \text{the Whitehead link} \tag{4}$$

Then, S and R are both rational tangles, $S = (3, 0)$, $R = (1)$, and $N(S+R+R+R+R)$ is the 6_2 knot.

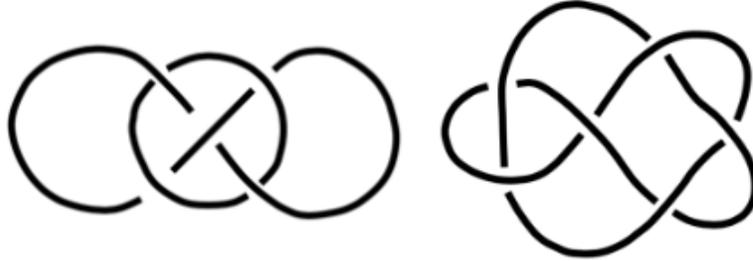


Figure 13: From left to right, the Whitehead link and the 6_2 knot

These assumptions in Theorem 1 are made because in 1985 biologists observed such knots and links resulting from the Tn3 resolvase acting on an unknotted substrate [8]. As mentioned earlier, biologists cannot directly observe what is happening during recombination, so this theorem offers a solution as to which knots S and R model, and how the resulting knots are produced. A significant part of this proof is merely showing that the two tangles S and R are both, in fact, rational tangles by using proof by contradiction [9]. This part of the proof is the longest portion; the interested reader should see [7, 9] for all details. The details we outline here are as given in [9]. First, we need a lemma, which we state without proof.

Lemma 1. *Let A and B be tangles. If $N(A + B)$ is a two-bridge link, then at least one of A and B is either a rational tangle or the connected sum of a rational tangle with $N(A + B)$. If $A + B$ is a rational tangle, then either both A and B are rational tangles or $N(A + B)$ is the unknot [9].*

The proof for Theorem 1 uses the Cyclic Surgery Theorem. To understand this theorem, we must first have a few definitions.

Definition 2. An n -dimensional manifold is a Hausdorff space M^n with countable base so that any $x \in M^n$ has a neighborhood homeomorphic to \mathbb{R}^n .

Definition 3. A chart of an n -dimensional manifold is an invertible map between an open subset of the manifold and an open subset of \mathbb{R}^n .

Definition 4. A Seifert manifold is a closed 3-manifold together with a decomposition into a disjoint union of circles, referred to as fibers, such that each fiber has a tubular neighborhood that forms a standard fibered torus.

Definition 5. Given a 3-manifold M with torus boundary components, we perform a Dehn filling by gluing in a solid torus by a homeomorphism of its boundary to one torus boundary component T of M . Dehn surgery on a knot involves removing a knotted solid torus from S^3 and then gluing the solid torus back along its boundary via a homeomorphism in such a way that the boundary of the torus may be attached differently to the boundary of the hole.

Definition 6. Let X and Y be topological spaces, and $f : X \rightarrow Y$ be a bijection. If both f and its inverse are continuous, then we call f a homeomorphism. If a homeomorphism between two manifolds, $f : M \rightarrow N$ maps charts in M to charts in N with the orientation preserved, then f is called an orientation-preserving homeomorphism.

Definition 7. Suppose M and N are three-dimensional manifolds. Let $h : M \rightarrow M$ be an orientation-preserving homeomorphism with an order of two. Let $p : M \rightarrow N$ be defined such that $p(x) = p(y)$ if and only if either $x = y$ or $h(x) = y$ and suppose p is continuous. Let $A = \{x \in M \mid h(x) = x\}$. If $B = p(A)$ is a one-manifold, then we say M is a twofold branched cover of N branched over B .

Definition 8. For any topological space X and a point $x_0 \in X$, the fundamental group of X based at x_0 is defined by a group denoted $\pi_1(X, x_0)$. The fundamental group gives us information about these topological spaces since it is a topological invariant. For details, see [10].

Theorem 3 (Cyclic Surgery Theorem). For a compact, connected, orientable, irreducible three-manifold M whose boundary is a torus T , if M is not a Seifert-fibered space and r, s are slopes on T such that their Dehn fillings have cyclic fundamental group, then the distance between r and s (the minimal number of times that two simple closed curves in T representing r and s must intersect) is at most 1.

We can restate this as follows: if two different Dehn surgeries on a knot both yield three-manifolds that have finite fundamental groups, and if the orders of such groups differ by more than one, then the knot on which we did Dehn surgery is either a trivial knot or a torus knot [9].

This brief introduction to the Cyclic Surgery Theorem is sufficient to understand the following proof.

Theorem 1 proof. Note that since the unknot, the unoriented Hopf link, and the figure eight knot are all achiral, there is an even number of tangle solution pairs for S, R , because for each solution to S, R , we must also consider the reverse of such solution, which gives us an overall even number of possibilities. If tangles A, B are rational tangles, tangle A is known, and $N(A + B)$ is a 4-plat, then there are infinitely many possibilities for the B . Luckily for us, biologists have observed the original tangle as $T = (0)$, which significantly simplifies proving S, R are rational tangles. Since, by assumption, $N(S+R)$ is the Hopf link, we can conclude that both S and R are locally unknotted since the Hopf link has two unknotted components.

If we assume both S and R are prime, the twofold branch cover of $N(S + R)$ would not be the solid torus, and thus implying $N(S + R)$ is not rational, which is a contradiction.

First, we will show R must be rational. By contradiction, assume R is not rational. If $S + R$ is rational, then, by Lemma 1, $N(S + R)$ would be the unknot. Since $N(S + R)$ is the Hopf link, $S + R$ cannot be rational. The figure eight knot is a two-bridge link and is $N(S + R + R)$ by the third assumption. Since R is not rational and $S + R$ is not rational, $S + R + R$ is not rational. By Lemma 1, $S + R$ or R must contain the figure eight knot since $S + R$ is not the connected sum of a rational tangle. Since $N(S + R)$ is the Hopf link, neither R or $S + R$ can contain the figure eight knot since the Hopf link's two components are unknotted. Thus R is rational.

Now we will show S must be rational. By contradiction, assume S is not rational. By the assumption, $N(S + T)$ is the unknot, which is trivially a two-bridge knot, and thus we can apply Lemma 1. Since S is not rational by the assumption, then T must either be rational or the connected sum of a rational tangle with the unknot, which is just a rational tangle. Thus, T is a rational tangle. Consider the twofold branch cover of the Hopf link, which has the fundamental group \mathbb{Z}_2 , and the twofold branch cover of the figure eight knot, which has the fundamental group \mathbb{Z}_5 [9].

By the Cyclic Surgery Theorem, the twofold branch cover of S must be either the complement of a torus knot or the complement of the trivial knot. The reason for this is because the fundamental groups are both finite and their orders differ by more than one, specifically three, so the knots must either be trivial or a torus knot. Due to work by Moser in 1971 [14], the twofold branched cover of S is a solid torus. Thus, S must be rational. This completes the proof that both S and R are rational tangles. Now, it nearly directly follows that the $N(S + R)$, $N(S + R + R)$,

and $N(S + R + R + R)$ must all be the Montesinos links by our previous definition of the Montesinos knots and links. Specifically, when $S = (3, 0)$ and $R = (1)$, it follows that $N(S + R + R + R + R)$ is the 6_2 knot.

□

It should be noted that the assumption that $T = (0)$ is not a mathematical assumption, rather a biological one. It is believed by biologists that before every recombination event it is a requirement for $T = (0)$ [9]. This leads us to the next theorem presented by Ernst and Sumners in 1990.

Theorem 4 (9). *There is no tangle S such that both $N[S + (0)]$ is the unknot and $N[S + (\pm 1)]$ is the Whitehead link.*

Ernst and Sumners then generalized their theorems with the goal of using the tangle model to model various enzymes, and not only Tn3, as we had originally assumed was the acting enzyme.

Theorem 5. *Let L_0 be a two-bridge knot or link, and let L_1, L_2 , and L_3 be two-bridge knots or links that are not all the same. Suppose that for $0 \leq n \leq 3$, we have the equations $N(S+nR) = L_n$. Then there is at most one pair of tangles $\{S, R\}$ that satisfies all three of these equations. Furthermore, if such a pair of tangles exists, then S must be either a rational tangle or the sum of two rational tangles, R must be an integral tangle, and at least one of L_1, L_2 , and L_3 must be topologically chiral.*

It should be noted that Theorem 4 is, in fact, a generalization of Theorem 1.

This is clear by taking:

$$L_0 = \text{unknot}$$

$$L_1 = \text{Hopf link}$$

$$L_2 = \text{figure eight knot}$$

$$L_3 = \text{Whitehead link}$$

Ernst and Sumners also focused on results of recombination. As previously discussed, we see 4-plats as results of a system of tangle equations which are Montesinos links. Noting this, Ernst and Sumners proved the following theorem: [7]

Theorem 6. *Given the system of four simultaneous tangle equations $N(O + iR) = K_i$ for $0 \leq i \leq 3$ where the K_i are 4-plats and K_1, K_2, K_3 represent at least 2 different link or knot types, there is at most one simultaneous solution O, R , and this solution must be of the form R an integral tangle and O either a rational tangle or the sum of two rational tangles. Moreover, if there exists a solution, then at least one of the 4-plats K_i must be chiral.*

The proof presents methods to mathematically solve such complex tangle equations, but it also is useful in the biological analysis of enzyme reactions.

Buck and Flapan also took on the task of predicting which knots and links arise from site-specific recombination in general [3]. In their model, the beginning substrate is either the unknot, unlink, or a $T(2, m)$ torus knot. Using assumptions supported by biological evidence, they predicted that products arising from site-specific recombination must be in the family of Montesinos knots and links, specifically in the form $(\frac{1}{p}, \frac{1}{q}, \frac{r}{rs+1})$ where $p, q, r, s \in \mathbb{Z} [2, 3]$.

IV. TOPOISOMERASES

In this section, we discuss DNA replication. In order for DNA to replicate, it acts as a zipper and unzips down the central axis. A problem occurs as DNA unzips. Due to the supercoiling and knottedness of the DNA strand, if the DNA were to continue to unzip with no regulating enzyme, the torsional strain would become too great, and the strand would break. Imagine two shoelaces in a knot on top of the shoe. If we twist the shoe laces tightly, then pull the top two strings apart, eventually, we will not be able to pull the strings any tighter. These two shoe strings represent the two backbones of the DNA strand. Thus, we can see that the knottedness and twistedness of DNA obstruct replication, but without the assistance of enzymes, life would not be possible.

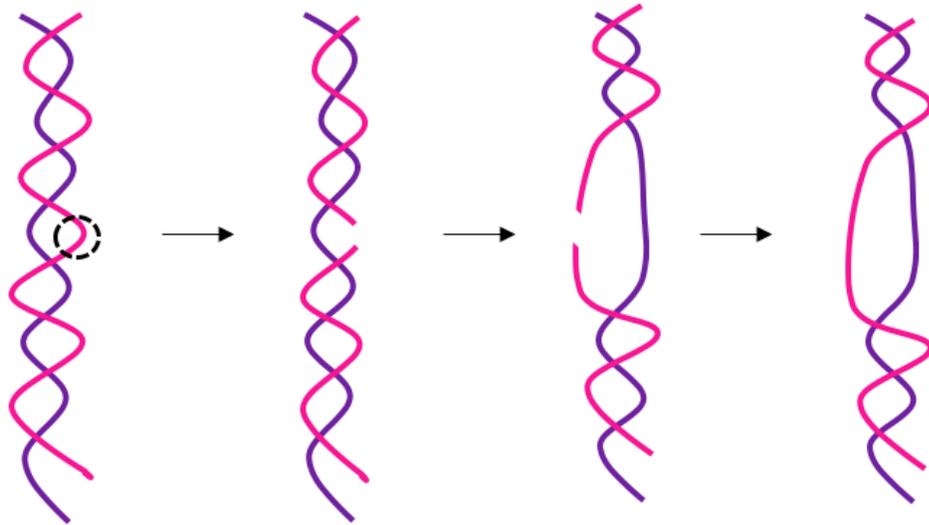
Since DNA cannot replicate itself while knotted, DNA must be topologically manipulated by enzymes for replication to occur. “The topoisomerases are a family of enzymes which maintain the natural structure of DNA molecules while changing their extrinsic topology, which is why they have the prefix *topo*” [8]. The topoisomerases enzymes are crucial to the survival of all DNA strands because they are responsible for unknotting, unlinking, and regulating the supercoiling of the DNA strand. The two families of topoisomerases are cleverly named type I and type II based on how many backbones they cleave onto and cut. Type I topoisomerases assist in alleviating excessive supercoiling so DNA can further unzip. Type II topoisomerases change the topology of the DNA strand by completely cutting through both backbones to allow a different part of the strand to pass through the opening, then resealing the separated pieces. Both type I and type II topoisomerases are critical for DNA replication to occur and thus are crucial for the life of cells to continue.

Type I Topoisomerases

Topoisomerase I solves the problem of excessive supercoiling during replication by cutting a backbone, allowing it to untwist around the other backbone, and then re-gluing the two ends. In this way, positive supercoiling is reduced, allowing the unzipping process (and thus replication) to continue. By regulating supercoiling during replication, Type I topoisomerases enable cells to grow and reproduce. However, not all cell growth and reproduction is a good thing. The growth of cancerous cells in bodies is often deadly. Topoisomerases I inhibitors have been used as cancer-fighting drugs because they prevent cancer cells from replicating. Unfortunately, such medications can also prevent healthy cells from replicating as well. “In particular, topoisomerases I inhibitors interfere with the growth of hair cells, which is why chemotherapy causes hair loss” [8].

The type I topoisomerases also play a hand in altering the DNA topology. The type I family has the sole function of regulating the amount of supercoiling of DNA by relaxing the supercoils. The type I topoisomerases convert a change in twist to a change in writhe by binding to the DNA molecule, making a break in one of the DNA backbones, and then passing the other backbone through before resealing and releasing the DNA [2] (See figure 14).

An example of when the type I topoisomerases naturally come into play is during DNA replication. When the DNA helix is unzipped, the molecule may be constrained where the unzipping fork is headed, so 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 great, 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].



n

Figure 14: Type I Topoisomerases Manipulation

Topoisomerases I change the topology of the DNA strand by untwisting one or more twists, resulting in a change of twist, Tw . In circular DNA, this also results in a change in writhe, Wr . Thus, the linking number of the DNA strand is altered in steps of one. The rules of topology clearly state no cutting or breaking is allowed, and type I topoisomerases break these rules. Thus, the DNA strand, D , before the topoisomerases I acts upon it is not topologically equivalent to the strand, D' , after this enzyme's action. In other words, there is no continuous function from strand D onto strand D' due to the action of this enzyme.

Type II Topoisomerases

The primary function of the type II topoisomerases are to change the DNA knot or link type, but these enzymes can also remove or add DNA supercoils. Type II topoisomerases simplify DNA topology by acting in a manner that preferentially unknots and unlinks DNA, a process known as *topological simplification*. It is quite interesting how type II topoisomerases effectively select which crossing to change in order to form a path to the unknot. Even when the DNA strand takes the form of a complex knot, this enzyme still recognizes the best method to unknot the strand. “At this point, we do not know yet which are the actual geometric parameters selected by topoisomerases to guide them to perform most efficient knots relaxation” [4]. We can examine a case of the type II topoisomerases unlinking two strands in bacterial genomic DNA, which has a circular structure. The type II topoisomerases occur naturally when the circular DNA is replicated (See figure 15). If a type II topoisomerase cannot unlink two circular molecules, then the bacterial cell cannot divide properly and essentially commits suicide [2]. Thus, the type II topoisomerases have been drug targets for infectious diseases and even cancer. Antibiotics inhibit the protein from doing its job of unlinking the circular molecules of the DNA to stop all replication and cause the cells to self-destruct, which then kills the bacterial infection. Thus understanding the mechanism of topoisomerases II has been an essential study for biologists and mathematicians.

What exactly is the mechanism of type II topoisomerases which allows them to change the DNA topology efficiently? The protein binds to a supercoil or a knot/link crossing and breaks the gate (G) segment to allow the transported (T) segment to pass through in a unidirectional manner [2]. After the G segment has



Figure 15: Type II Topoisomerases manipulation on circular DNA

been resealed, we can see the crossing has changed from an over-crossing to an under-crossing, or vice-versa. Thus, this enzyme has a crucial role in the dynamics of linked circular DNA strands, whereas the type I topoisomerases cannot effect unlinking and replicating.

Since the type II enzymes unknot the strand one crossing at a time, an essential topic of DNA research is studying what knots can be obtained by one single crossing change. One area of current knot theory research is classifying all knots that become unknotted after one single crossing change. The amount of crossing changes needed for the knot to become the unknot is referred to as the *unknotting number*. The simplest of a knot that becomes the unknot after one crossing change is the trefoil knot.

We define the *distance* between any two knots or links as the minimum number of crossing changes to transform K into L and is denoted $d(K, L)$. We may also see that $d(K, L)$ defines a metric on the set of knot types where $d(K, L)$ is less than or equal to the sum of the unknotting numbers of each K and L [2]. We call knots *adjacent* when $d(K, L) = 1$, which, from the biological standpoint, is equivalent to K being one type II topoisomerases move away from L . Due to the work of Rasmussen and Lobb, the lower bounds for knot adjacency are computable via a combinatorial approach, and this has been used to classify all knots K such that the unknotting number of K is one and K has nine or fewer crossings.

To give an example, we examine a 10-crossing knot that has an unknotting number of one. First, we select a crossing to change, see figure 16, from an under-crossing to an over-crossing. Next, we use the $R2$ Reidemeister move to separate the two loops. Lastly, the $R1$ Reidemeister move is used eight times to untwist the knot. The result is the unknot. Thus it is concluded the 10_1 knot has unknotting number one.

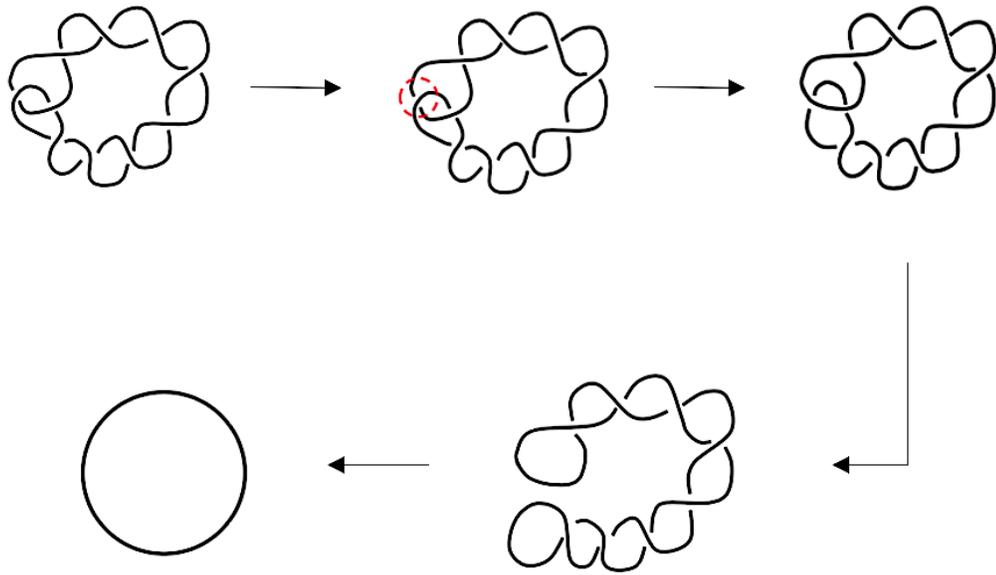


Figure 16: Unknotting the 10_1 knot

V. SIGNIFICANCE

Scientists are striving to comprehend the complexity of a single DNA strand's replication or recombination process, and knot theory is one tool being used to do so. The more scientists learn about knots, the more they learn about DNA. So we can see why interdisciplinary research between mathematical knots and molecular biology is becoming increasingly prevalent. The pharmaceutical industry makes use of knot theory too. Using knot theory to model DNA has many vital uses in modern-day pharmacology. Antibiotics, and even chemotherapy drugs, target type II topoisomerases with the goal of preventing diseased cells from replicating, which ideally stops the illness from spreading. Knowing the unknotting number is critical to drug development as well, because, with the unknotting number, one can accurately estimate how many steps it will take for a topoisomerase to unknot the DNA. This estimate leads to knowing how rapidly a disease will spread or progress through the body and how quickly the type II topoisomerase-inhibiting drugs will work to counteract the disease. "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 disease. As a result, these industries are now interested in site-specific recombinases as tools for precisely manipulating DNA." Lastly, understanding the specific knots involved in site-specific recombination may lead to the discovery of treatments for viral infections and genetic disorders [2].

VI. FUTURE DIRECTIONS FOR RESEARCH

The tangle model focuses mainly on specific types of recombinases enzymes, such as Tn3. There are many other enzymes which act on DNA by site-specific recombination that we have not modeled because their product knot or link has not been characterized [2]. Research on tangle operations on pairs of knots is a current area of investigation.

Since the topoisomerase enzymes are still not entirely understood by biologists, further mathematics research can assist in better predicting DNA topology through examining products of DNA recombination. Biologists are also curious as to which knot classifications assist in understanding enzymatic mechanisms. Moreover, biologists want to know if the operations on knots, links, and tangles suffices to simulate and model all of the different actions by enzymes on the DNA strand.

Another area of interest is examining how the changes made by the topoisomerase enzymes affect the DNA strand on a global level. Mathematicians are interested in how the topology of a strand changes through local crossing changes in the DNA knot or link.

BIBLIOGRAPHY

- [1] Alexander, J.W. (1928). “Topological invariants of knots and links.” *Trans. of the American Mathematical Society*. 30 (2): 275–306.
- [2] Buck, D., & Flapan, E. (2009). *Applications of knot theory: American Mathematical Society Short Course, January 4-5, 2008, San Diego, California*. Providence: American Mathematical Society.
- [3] Buck, D., & Flapan, E. (2007). “A topological characterization of knots and links arising from site-specific recombination.” *Journal of Physics A: Mathematical and Theoretical*, 40(41), 12377-12395.
- [4] Burnier, Y., Weber, C., Flammini, A., and Stasiak, A. “Local Selection Rules That Can Determine Specific Pathways of DNA Unknotting by Type II DNA Topoisomerases.” *Nucleic Acids Res.* 2007 Aug; 35(15): 5223–5231.
- [5] Conway, J. H. (1970). “An Enumeration of Knots and Links, and Some of Their Algebraic Properties.” In Leech, J. *Computational Problems in Abstract Algebra*. Oxford, England: Pergamon Press. pp. 329–358.
- [6] Darcy, I. K. (2008). Modeling protein–DNA complexes with tangles. *Computers & Mathematics with Applications*, 55(5), 924-937.
doi:10.1016/j.camwa.2006.12.099
- [7] Ernst, C., Sumners, D. (1990). “A calculus for rational tangles: Applications to DNA recombination.” *Mathematical Proceedings of the Cambridge Philosophical Society*, 108(3), 489-515. doi:10.1017/S0305004100069383
- [8] Flapan, E., & Averett, M. (2017). *Knots, molecules, and the universe: An introduction to topology*. Telangana: American Mathematical Society.

- [9] Flapan, E. (2009). When topology meets chemistry: A topological look at molecular chirality. Cambridge: Cambridge University Press.
- [10] Gilbert, N. D., and Timothy Porter. Knots and Surfaces. Oxford Science Publications, 1995.
- [11] Jackson, S, & Lim, N. “Molecular knots in biology and chemistry.” Journal of Physics: Condensed Matter, 27(35), Aug. 2015, <http://iopscience.iop.org/article/10.1088/0953-8984/27/35/354101>
- [12] Jones, V.F.R. (1985). “A polynomial invariant for knots via von Neumann algebra.” Bulletin of the American Mathematical Society (N.S.). 12: 103–111.
- [13] Kauffman, Louis H. (1987). “State models and the Jones polynomial”. Topology. 26 (3): 395–407.
- [14] Moser, Louise (1971). “Elementary surgery along a torus knot”, Pacific J. Math. 38: 737-745.
- [15] Ricca, R., & Nipoti, B. “Gauss’ Linking Number Revisited.” Journal of Knot Theory and Its Ramifications, vol. 20, no. 10, 2011, pp. 1325–1343. doi:10.1142/s0218216511009261.
- [16] Silver, D. “Scottish Physics and Knot Theory’s Odd Origins.” (n.d.). University of South Alabama, www.southalabama.edu/mathstat/personal_pages/silver/scottish.pdf.
- [17] Tait. (1885, July 20). On Knots. Retrieved from http://katlas.org/w/images/f/fe/Tait_-_On_knots_III.pdf
- [18] Tompkins, J. (2006) “Modeling DNA Using Knot Theory: An Introduction,” Rose-Hulman Undergraduate Mathematics Journal: Vol. 7 : Iss. 1 , Article 13

- [19] Vazquez, M. "Mathematical Background." (n.d.). Retrieved from <http://ewok.sfsu.edu/TangleSolve/MathematicalBackground.html>