Introductory explanation to the tangle method to Gin site-specific recombination, both wild-type and mutant.

Abstract

Based on previous results, we will examine how Gin site-specific recombination may be modeled as a system of tangle equations, and that these tangles are rational tangles. Furthermore, if we make certain assumptions, then there is a unique topological description which leads to a unique biologically reasonable solution. Also, it has been shown that by introduction of mutant Gin, a wider range of products are found.

Table of Contents

1. Introduction – A background of Gin site specific recombination; and also mutant Gin.

2. Tangles – Mathematical definitions including classifications, 4-plats, operations, and theorems.

3. The Tangle method as it pertains to Gin site specific recombination.

4. Results of tangle model applied to mutant Gin site specific recombination.

5. Conclusion and Results.

1. Introduction.

The theory of knots and tangles taken from the mathematical area of Topology have some rather interesting implications to molecular biology, and more specifically to that of site-specific recombination via the enzyme Gin, which, represented as a mathematical tangle, enacts topological alterations in circular DNA molecules.

Both topoisomerases and recombinases are enzymes which act as devices in the processes of recombination, transcription, and replication of DNA, and these enzymes acting upon the DNA causes topological changes [3]. A recombinase is defined as an enzyme which induces recombination [3]. The method to be described is used via the recombinase Gin in the reproduction of the bacteriophage Mu.
A bacteriophage is a virus that infects a host bacterium, and uses the bacterium’s own machinery in order to reproduce itself. When the resources of the bacterium are depleted, the bacterium is destroyed, and the new phages are released, infecting new bacteria [9]. Mu is called “temperate” because it uses either of two methods for reproduction. In One pathway, called the “lytic pathway,” the DNA of the virus upon infecting a bacterium, is replicated in the cytoplasm of the host. The second pathway, called the “lysogenic pathway,” is when the phage’s DNA is actually integrated into the genome of the host bacterium [1].

As shall be seen in section 3, the tangle model is used to describe site-specific recombination in a mathematically cohesive method. It will be shown that the action of Gin on the circularized DNA molecules of Mu induces topological changes in the molecule, and actually changes the genetic code of the G-segment of the bacteriophage, which is used in encoding a essential part of the phage’s anatomy.

2. **Tangles.**

We begin by defining the three families of tangles. A tangle is defined as an ordered pair \((B,t)\), where \(B\) is topologically equivalent to a 3-dimensional ball, and \(t\) is a pair of spanning arcs that are mutually disjoint and non-oriented (embedded in \(B\)) [See Figure 1 for a visual representation]. We have three classifications of tangles; the rational, locally knotted, and the prime tangles [1].
A rational tangle is intuitively defined by taking $p_z(B)$, where $p_z$ is the projection to the unit circle on the XY coordinate plane. If we allow our four points, NW, NE, SW, and SE to move freely along the unit circle, and we are able through these free movements to “untie” the arcs into a tangle such that our arcs are non-crossing, then we have a rational tangle [1,2].

For the locally knotted tangle, this action of separating the intertwined strands cannot be done. However, it is more specific than just this. Visually, the locally knotted tangle has just what its name implies, a little knot that is tied with one of the arcs. This is in contrast with when a knot is tied between the two arcs, this does not imply locally knotted. The locally knotted classification appears to only pertain to a knot that is tied in an individual arc (that is, the arc is self-intersecting). If a knot is not locally knotted, then it is locally unknotted [2].

The third classification is a prime tangle. A prime tangle is neither rational nor locally knotted. Notice that we may still have a prime tangle with which the two arcs are tied in a knot with each other, so long as one of the arcs is not tied in a knot with itself (belonging to the locally knotted classification).
Figure 2: Found on (568) A. (a),(a') are rational tangles, where (a') is called an integral tangle; (b) locally knotted; and (c) prime. B. The four trivial tangles with their associated Conway symbols (do not satisfy convention)[all are rational].

An explanation of the Conway symbols will now be given. Take, for example the tangle diagram from Figure 2 above, part A (a). Notice that the associated Conway symbol is (4,3,2,1,0). The rightmost number must represent the rightmost horizontal set of crossings, and the only time a rational tangle can assume a value of zero in its Conway symbol is when the rightmost set of crossings is not horizontal, as in the case of our example. In this case, the first instance of crossings is a vertical crossing, and so we have a value of 1 following the 0 from right to left. We then proceed to the next instance of horizontal crossings, as the Conway numbers will always alternate between horizontal and vertical crossings. In this case, 2 crossings can be seen in the second horizontal position. Then, coming to the next vertical crossings, we have a value of 3. Finally, in the leftmost set of crossings is horizontal, and there are four crossings in this position, and so the first entry in the Conway vector is 4.

We see that a rational number can be calculated from the Conway number of an associated tangle. Let A be a rational tangle such that A is represented by its associated Conway number A=(a1,a2,…,am), where each ai is an integer, |a1|>1, ai≠0 for each i=1,2,…,m-1. Also, for each pair of elements, ai•aj>0 (In other words, the elements of A are either all positive, or all negative). A tangle which meets each of these requirements is defined to be of the canonical form. Furthermore, if we let A be any rational tangle, (not necessarily of canonical form), then it must necessarily be homeomorphic to a tangle of the form of A. As such, we can define our equivalence classes as the set of all tangles of the canonical form. Not coincidentally, A’ will have the same rational number obtained from its associated
Conway symbol by way of a continued fraction as does \( A \), provided that they belong to the same equivalence class. From the Conway symbol \((a_1, a_2, \ldots, a_m)\), the rational number is obtained as follows [1]:

\[
\frac{p}{q} = a_m + \frac{1}{a_{m-1} + \frac{1}{a_{m-2} + \cdots + \frac{1}{a_1}}}.
\]

Returning to our example in the previous paragraph of \((4,3,2,1,0)\), we have:

\[
0 + \frac{1}{1 + \frac{1}{2 + \frac{1}{3 + \frac{1}{4}}} = \frac{30}{43}}
\]

In regards to the Conway symbols associated with the rational tangles; perhaps part of the reason that a rational tangle is aptly named is that there is an important relationship between its Conway symbol and that of the extended rational numbers. Two tangles are equivalent \([A_1 = A_2]\) if they have the same rational number. We see that rational tangles form an equivalence class, and are in one-to-one correspondence to the set of rational numbers plus infinity (Rational Tangle Classification Theorem) [3,10].

Now, addition of tangles \( A \) and \( B \) is defined by “gluing” the NE of tangle \( A \) to the NW of tangle \( B \), and the SE of \( A \) to the SW of \( B \). Note that the addition of rational tangles need not be rational. Other operations on tangles are illustrated below, and are denoted by \( N(A) \) and \( D(A) \) where \( N \) is the operation of “numerator” and \( D \) is the operation of “denominator.” The numerator is obtained by gluing the NE to the NW and the SE to the SW of one diagram. Similarly, the denominator is obtained by gluing the NE to the SE. These operations can act on the sum of two tangles, and are used convert tangles into knots and links within a closed-loop representation, as will be seen in section 3.

A knot is a smooth unoriented embedding of the circle within \( S^3 \). Two knots are equivalent if there exists a homeomorphism \( h:S^3 \rightarrow S^3 \) [11]. These homeomorphisms have the same effect as any
manipulation of a knotted string where we are not allowed to pass the string through itself, nor are we allowed to cut the string. Also, a link is a disjoint union of knots in $S^3$. Once again, knots are equal if they are homeomorphic [11].

4-Plats are used in the addition of tangles, and are obtained by taking the numerator of a rational tangle, or of the sum of two rational tangles. Any 4-plat can be represented using a 4-braid where the strings are interwoven into each other. As an example, we have the following 4-plat:

$$= <2,1,1>.$$  The vector representation for 4-plats is similar to that of tangles, however, there are several differences. Our canonical format is to have all positive entries, as a 4-plat with negative values for the crossings can be deformed homeomorphically into a 4-plat with all positive crossings. Also, instead of alternating between horizontal and vertical crossings, in the case of a 4-plat, we alternate sets of crossings between bottom and top rows, and it is not required that the rightmost set of crossings represent horizontal crossings.

That is, zeroes do not exist in our vector, unless you have the $<0>$ knot:

If $<c_1,c_2,...,c(2n+1)>$ (there are always an odd number of entries) is the vector representation of a 4-plat $K$, then $K$ is characterized by the rational number defined as:

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

We denote $K$ by $b(\alpha,\beta)$. Additionally, if $\alpha = 0$ and $\beta = 1$, then $K$ is the unlink, and also if $\alpha = 1 = \beta$, then $K$ is the unknot. Otherwise, $\alpha > \beta > 0$ [1].

In the article *A calculus for rational tangles: applications to DNA recombination*, by Ernst and Sumners [3], we find the following important propositions:
Lemma 2.1: Let $X = u/v$ and $Y = x/y$ be two rational tangles. Then $N(X+A) = b(\alpha, \beta)$ is a 4-plat, and
$\alpha = |xv+yu|$. Furthermore, $\beta$ is determined as follows:

(a) If $\alpha = 0$, then $\beta = 1$;
(b) If $\alpha = 1$, then $\beta = 1$;
(c) If $\alpha > 1$, let $\sigma = \text{sign}(vx+yu)$. Then $\beta$ is uniquely determined by $\alpha > \beta > 0$ and $\beta = \sigma(vy'+ux') \mod \alpha$,
where $x'$ and $y'$ are integers such that $xx'-yy'=1$.

Example: Let $A$ and $B$ be tangles with the rational numbers $A = 2$ and $B = 23/17$. According to our lemma, when we take the tangle sum of $A+B$, and allow this to resolve into a 4-plat, we have $u = 2$, $v = 1$, $x = 23$, and $y = 17$, giving us that $\alpha = |1*23+17*2| = 57$. $\sigma = 1$, since each of our $u,v,x,$ and $y$ are all positive. Also, we have that $23x'-1+17y'$ yields the integer solution $x'=3$, $y'=4$, giving us that $\beta = (1*4+2*3) \mod 57 = 10$, hence $N(A+B) = b(57,10)$ [8].

Theorem 2.1: Let $A$ and $B$ be two locally unknotted tangles and $K = N(A+B)$ be a 4-plat. If $K$ exists, then $A$ or $B$ is rational.

These postulates are used to prove our results in the later sections, when we apply these topological concepts and theorems to site-specific recombination. On a final note, a given tangle will have one of three parity classifications (0), (1), and (0,0). We have the following table for determining parity:

![Parity Table]

(0) (1) (0,0)

3. Gin.

We now discuss the aforementioned theory as it pertains to the reproductive functions of the bacteriophage Mu. In particular, we will be looking at the lytic pathway of Mu’s ability to reproduce itself by essentially hijacking a host bacterium, and using the Gin enzymes already present in the host. This enzyme-DNA complex can be modeled using tangles [8].

The Bacteriophage begins by landing on the outer edge of the bacterium, detecting with its tail fibers whether or not the bacterium is a suitable host. Then, the DNA of the bacteriophage is harpooned from the capsid of the phage down the tail, and into the bacterium. While inside of the capsid of Mu, the DNA is linear, and very heavily supercoiled, however once inside the host, the DNA immediately becomes circular to protect against destructive enzymes present in the host. These circular DNA strands
DNA Topology
Author: Andrew Dynneson
Professor: Mariel Vazquez
Partner: Katrina Wono

contain encoding information for the construction of new bacteriophages. The DNA will be replicated by using the bacterium’s resources instead of its own [6].

The G-segment of the genome encodes the tail fibers. Gin is called an “invertase” because it causes an inversion at a G-segment. Gin inverts this segment by cutting and recombining at two locations on the genome. These two specific locations are called the gix sites. This is where we have the term “site-specific,” as it pertains to site-specific recombination. When the G-segment is inverted, this leads to some of the resulting offspring to be able to infect different species of bacteria then that of their parent. The effect of site-specific recombination via the enzyme Gin is that it extends the host range of the bacteriophage [1,4]. Each of the two recombination sites are 34 base-pairs long, and are denoted gix L and gix R. Gin causes a double-stranded break at each site, then conducts either a 180 degree rotation, or a full 360 degree rotation, depending on whether the code at the gix sites are directly or inversely repeated. Then Gin reconnects the strand. Given that the code at the gix site is not palindromic¹, we assign an orientation² to the letters of one gix site arbitrarily [1]. An orientation for the other site will arise naturally, as the code at both gix sites are identical, they may just be read in a different direction. If our orientations are both in the same direction, either clockwise or anti-clockwise about the circular DNA, then our gix sites are said to be in “direct repeat.” Similarly, if one orientation goes clockwise, while the other is anti-clockwise, then our gix sites are said to be in “inverted repeat” [8].

¹ A string is palindromic if the code is the same if read in either direction; for example: ATTCGAGCTTA.

² Note that this orientation is localized, that is, the global orientation may be in an opposite direction, without contradiction.
It can quickly be seen that in the case of directly repeated gix sites, one round of recombination would lead to opposite contradictory and mismatched orientations at the sites. As a result, the enzyme must continue to rotate and undergo an additional turn before reconnecting the strand. As a result, the recombinant DNA is identical at the G-segment as that of the parental DNA in this case. In the second case, for inversely repeated gix sites, one round of recombination leads to an opposite orientation at the G-segment, and thus the tail fibers have a different encoding (see figure 3 below) [1,4].

**Figure 3**
In either case, the DNA after recombination underwent a change in supercoiling (in Topology, this is a change in knot or link type), and if we represent this recombination as a sum of tangles, we see this works very nicely with the Tangle method. For example, we observe one such molecule as it undergoes recombination, and in the resulting diagram, we see that there has been a change in the topological structure of the DNA:

Figure 4: Changes in supercoiling due to recombination

We assume that the DNA–enzyme complex can be represented by tangles, and that the mechanism of recombination is constant and not dependent on the topology nor geometry of the substrate [1]. Also, let E denote the enzyme, Ob be the part of the DNA which is unchanged by the reaction but still bound by the enzyme E, Of be the part of the DNA that is not bound by the enzyme, and P be the site that is changed during the reaction (where P stands for “parental”). Also, denote the reaction by E=Ob+P. Then, let K0=N(Of+Ob+P) be our substrate molecule. We then assume that recombination acts in the way described in figure 4, where cuts are made in the DNA, and recombined at the gix sites. This is called “tangle surgery.” Then, in this tangle equation, since P is the only thing changed by the recombination
action, P is replaced by the tangle R (where R stands for “recombinant”). This represents one round of recombination. Then, when the enzyme has acted on the molecule, we have our product K1=N(OF+Ob+R). If we let O=Of+Ob, then we can continue to recombine at the gix sites, to get the products: k2=N(O+R+R), ... , kn=N(O+nR) [8].

It is important to note that the tangle method only involves rational tangles, and so we need not be concerned with other tangle types in the following methods [8]. If we have the resulting 4-plat formed during recombination, then we are able to deduce mathematically what our original conformations of the DNA are. In [1], we find the following theorem as it pertains to site-specific recombination:

Suppose tangles O, P, and R satisfy the following equations:

(a) \( N(O+P)=\langle 1 \rangle \) (the unknot)
(b) \( N(O+R)=\langle 1 \rangle \)
(c) \( N(O+R+R)=\langle 3,1 \rangle \) (the trefoil knot)

Then \((O,R)\in\{(\langle 2,0,1 \rangle,\langle 1 \rangle ); (\langle 4,1,1 \rangle,\langle 1 \rangle )\}\) [That is, two possible solutions]. If we have the additional condition that:

(d) \( N(O+R+R+R)=\langle 2,1,1 \rangle \) (figure-8 knot)

Then \((O,R)=(\langle 2,0,1 \rangle,\langle 1 \rangle )\) is the unique solution.

We can also extend this solution to predict another round of recombination: \( K4=\langle 2,2,1 \rangle \) (the 5-twist knot). The proof of these results makes use of lemma 2.1, theorem 2.1, theorems to detect tangle rationality, known results of Dehn surgery on torus knots, and others found in [1]. What has happened here is we have the resulting topological structures which can be detected experimentally. We can then deduce from these structures by use of the Tangle method what the recombination has started with originally. Notice in this case that after three rounds of recombination, we see that we began with two trapped negative supercoils in O.

4. Mutant Gin.

Mutant Gin produces a much wider range of knotted products [4], including torus knots, which have never been seen in wild-type Gin recombination [1]. It has also been shown that, where wild-type Gin is dependent on the protein FIS, mutant Gin is FIS independent. Furthermore, mutant Gin does not require negative supercoiling to initiate the strand exchange [4].

Also, by changing the length of the G-segment, we have some different results in supercoiling. For example, if we have that the G-segment is between 1.4 and 2.4 kb in length, then mutant Gin can recombine using a plethora of synaptic complexes, which is the number of trapped supercoil (which differs from that of the wild-type). The number of trapped negative supercoils is at least 1 to 12. If we restrict the G-segment to 0.3 kb, then we also restrict the number of functional synaptic complexes,
thereby restricting the number of possible torus knots produced by mutant Gin. Restricting the G-segment to 0.1 kb significantly diminishes the products. This is so restrictive that it has been shown that only one or two negative supercoils can be accommodated between the gix sites. The -1 complex is restricted, but the -2 complex can undergo more than 9 rounds of recombination, and is not restricted.

5. Conclusion and Results.

Tangles are incredibly useful for developing a method of representing the site-specific recombination of the enzyme Gin and mutant Gin in the reproductive faculties of the bacteriophage Mu. We have explored the definitions and classifications of tangles and 4-plats, and their application to the supercoiling of Mu’s circularized DNA. We have also discussed the operations of addition and numerator, and its connection to recombination. Also, given the product of recombination, we are able to use this method to calculate the substrate. We have seen that the effect of site-specific recombination via the enzyme Gin is that the host range of Mu is extended. Finally, by introducing an altered form of the enzyme Gin, we see that there are vast differences between the resulting supercoiling induced by mutant Gin when it is compared to that of wild-type Gin.

Also, we have seen that in the case of wild-type Gin, only certain original conformations of a DNA molecule are conducive to site-specific recombination. This is offset somewhat by introduction of mutant Gin, and mutant Gin accepts a much wider range of original topological conformations.

The program TangleSolve, which is found on [2] is used to calculate the original conformations, as well as to predict further products, if recombination where to be repeated. Katrina Wono has organized the resulting data found using TangleSolve, seen below, where each successive round of recombination by the mutant Gin is pictured:

1 (-) supercoils trapped (antiparallel)
unknot → unknot → trefoil knot → +5 torus knot → +7 torus knot

![Diagram 1](image1)

2 (-) supercoils trapped (parallel)
unknot → unknot → trefoil knot → figure-8 knot → -5 twist knot → -6 twist knot → -7 twist knot → -8 twist knot → -9 twist knot

![Diagram 2](image2)
3 (-) supercoils trapped (antiparallel)
un knot → figure-8 knot → < 1, 2, 1, 1, 1 > → < 1, 4, 1, 1, 1 > → < 1, 1, 6, 1, 1 >

4 (-) supercoils trapped (parallel)
un knot → +trefoil knot → -5 twist knot → < 1, 2, 1, 1, 1 > → < 1, 2, 1, 2, 1 >

Author: Andrew Dynneson
Professor: Mariel Vazquez
Partner: Katrina Wono
5 (-) supercoils trapped (antiparallel)
unknot → -6 twist knot → ≪ 1, 2, 1, 3, 1 ≫ → ≪ 1, 4, 1, 3, 1, 1 ≫ → ≪ 1, 6, 1, 3, 1 ≫ → ≪ 1, 8, 1, 3, 1 ≫ → ≪ 1, 10, 1, 3, 1 ≫
6 (-) supercoils trapped (parallel)
unknot → +5 torus knot → -7 twist knot → (1, 4, 1, 1, 1) → (1, 4, 1, 2, 1)

8 (-) supercoils trapped (parallel)
unknot → +7 torus knot → -9 twist knot → (1, 1, 1, 6, 1) → (1, 2, 1, 6, 1) → (1, 3, 1, 6, 1) → (1, 4, 1, 6, 1)
\[ N((0) + (-8,0) + (0)) = (1,2,1,6,1) \text{ knot} \]

\[ N((0) + (-8,0) + (5)) = (1,3,1,6,1) \text{ knot} \]

\[ N((0) + (-8,0) + (6)) = (1,4,1,6,1) \text{ knot} \]
References


