Journal of Knot Theory and Its Ramifications

BIOLOGICAL DISTANCES ON DNA KNOTS AND LINKS:
Applications to XER recombination

ISABEL K. DARCY
Mathematics Department
University of Texas at Austin
Austin, TX 78712

ABSTRACT

The mathematics of tangles has been very useful in studying recombinases which act processively and which require DNA to be in a certain configuration in order for the enzyme to act. Electron micrographs of the enzyme-DNA complex show the enzyme as a blob with DNA looping out of it. The configuration of the DNA within the blob cannot be determined from the electron micrographs. However, mathematics can in some cases determine the configuration of the DNA within the enzyme blob as well as the enzyme action.

In this paper, several theorems used to analyze recombinase experiments are summarized. In particular Xer recombinase, an enzyme which does not act processively is analyzed. Unfortunately, for enzymes which do not act processively, infinitely many possibilities exist. Several experiments are proposed to reduce this number and to emphasize both the usefulness and limitations of tangle analysis. Although the local action cannot be mathematically determined without more biological assumptions, it is possible to determine the topology of the synaptic complex through additional biological experiments.

Keywords: tangles, recombination, 4-plats, xer

Tangle calculus has been successfully used to study certain enzymes called recombinases [1]. In sections 1 and 2, the necessary biology and mathematics are summarized. In sections 3, 4, and 5, current techniques to analyze recombinase action are presented. In section 6, these techniques are applied to the enzyme Xer recombinase. Section 7 emphasizes some shortfalls in the mathematical analysis of the local action of an enzyme which does not act processively and suggests focusing on the entire synaptic complex instead. Section 8 analyzes four different synaptic complex models for Xer and proposes experiment which may determine the topology of the synaptic complex.

1. A Recombinase Primer

Recombinases are enzymes which cut two segments of DNA and interchange the ends and can thus change the topology of closed circular DNA. Some recombinases, called site-specific recombinases, bind to two specific sequences which are identical or nearly identical. If the two sequences appear on the same strand in the same direction (for example, –CTTGA——CTTGA–), then the sequences are oriented as direct repeats. If the two sequences appear on complementary strands in opposite directions (for example, –CTTGA—–___AGTTC_), then the sequences are oriented as inverted repeats. In the figures below, the arrows denote the orientation of these sequences. Recombination allows viruses to integrate into and excise out of host genomes and can be involved in the regulation of transcription by turning genes on or off by inverting segments of DNA. Note that deletions and fusions occur with direct repeats, resulting in a change in the number of components, and inversions occur with inverted repeats with no change in the number of components.

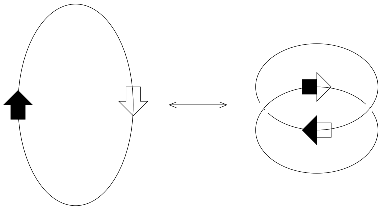


Fig. 1. Direct repeats recombination: deletion, fusion.

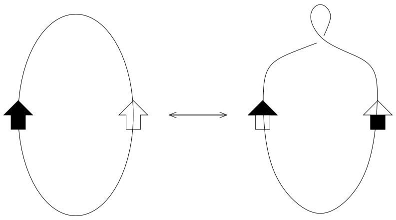


Fig. 2. Inverted repeats recombination: inversion.

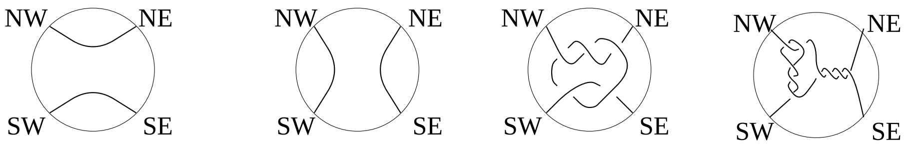
Since they break and rejoin DNA strands, recombinases produce knots and links when acting on closed circular DNA. Two main techniques, gel electrophoresis and electron microscopy, can be used to identify these knots and links. Gel electrophoresis does not give as much information as electron microscopy but is much simpler. In gel electrophoresis, an agarose gel is used to separate some knots and links by crossing number and sometimes even by knot type [2][3]. Agarose is a sugar polymer which forms a matrix through which the negatively charged DNA can travel when a positive electrode is put at the bottom of the gel. For DNA molecules with the same number of base pairs, gel velocity is determined by the average geometric conformation of each molecule as it migrates through the obstruction field imposed by the gel. Knotted DNA travels faster than unknotted DNA because a knot is more compact than an unknot when tied in a molecule with the same number of base pairs. Similarly, for low crossing number knots, gel velocity is determined by the crossing number of the knot. In some circumstances knots of the same crossing number can also be separated by gel electrophoresis. For example, torus knots have been separated from twist knots [3][4]. Thus the five crossing torus knots can also be distinguished from the five crossing twist knots via gel electrophoresis. However, no one has been able to separate out all of the six crossing knots or all of the seven crossing knots, etc.

Electron microscopy is another method used to identify knotted DNA. The DNA is coated with the protein RecA in order to thicken the strand to better identify crossings. This RecA coated DNA is spread out on a slide, and electron micrographs (EMs) are taken. If the crossings can be determined, then the knot can be identified. However, this method is time consuming, requires equipment and expertise to which some labs may not have easy access, and can be misleading. Certain knots spread out better and thus their EMs are easier to identify. And if an EM cannot be identified, it is ignored. Due to the difficulties with EM, biologists do not always fully identify the knotted and linked products, relying instead on information such as crossing number obtained through the much easier and faster technique of gel electrophoresis. Fortunately some mathematical analysis of these experiments only require knowledge of the crossing number.

Electron micrographs of recombinases bound to DNA show an enzyme “blob” with 2-3 loops of DNA sticking out of this blob. Biologists would like to know what is happening within this enzyme blob, but since this cannot be determine from the pictures, the mathematics of tangles have been used instead to model the enzyme action [5].

2. A 2-string Tangle Primer

For a more in depth introduction to 2-string tangles see chapter 9 in [6]. For a more in depth introduction to rational 2-string tangles see [7].



(0)
0-tangle

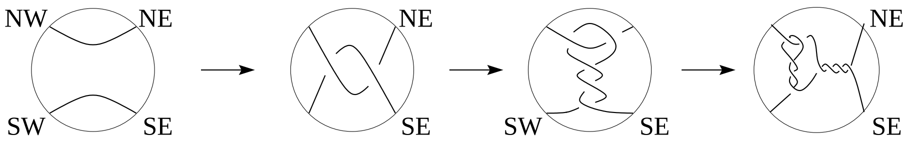
(0,0)
infinity tangle

a non-rational
tangle

(2, 3, 4)
30/7-tangle

Fig. 3. Some tangles.

A 2-string tangle is a pair (B, t) where B is a 3-dimensional ball and t is a pair of arcs properly embedded in B. The enzyme blob can be thought of as the 3-dimensional ball and the DNA strands within the enzyme blob can be thought of as arcs embedded in the 3-ball. A tangle is rational if it can be formed from either the zero tangle or the infinity tangle by alternating between rotating the NE and SE endpoints of the tangle and rotating the SW and SE endpoints of the tangle. Observe that the tangle (2, 3, 4) is obtained from the zero tangle by rotating NE and SE 2×180^o, followed by rotating SW and SE 3×180^o, and then rotating NE and SE 4×180^o (figure 4). By convention, the tangle corresponding to $(c_1, ..., c_n)$ always ends with horizontal crossings. Thus the tangle with two horizontal crossings followed by three vertical crossings in figure 4 is the (2, 3, 0) tangle since it ends with zero horizontal crossings.



(0) 0-tangle

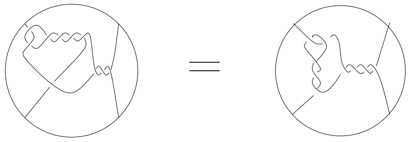
(2) 2-tangle

(2, 3, 0) 2/7-tangle

(2, 3, 4) 30/7-tangle

Fig. 4. Drawing the (2,3,4) tangle

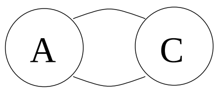
The tangles are called rational because the rational number corresponding to its continued fraction can be used to identify the tangle. For example (2, 3, 4) corresponds to the $\frac{30}{7}$ tangle since $4 + \frac{1}{3+\frac{1}{2}} = \frac{30}{7}$. Two tangles are said to be equivalent if one can be deformed (via an ambient isotopy) to the other keeping the boundary of the 3-ball fixed. Two rational tangles are equivalent if and only if the rational numbers corresponding to their continued fraction are the same. For example the tangle (-2, -4, 1, 3) is the same as the tangle (2,3,4) since both tangles correspond to $\frac{30}{7}$. A rational tangle is called integral if its rational number is an integer and thus consists solely of horizontal crossings.



(-1, -1, -4, 1, 3)

(2, 3, 4)

Fig. 5. The $\frac{30}{7}$-tangle



A + C

Fig. 6. Adding tangles

Tangles can be added as shown in figure 6. A knot or link can also be formed from a tangle or a sum of tangles by taking its numerator closure as shown in figure 7 or its denominator closure as shown in figure 8. Both the numerator closure and the denominator closure of a rational tangle give a particular type of knot or link called a 4-plat (or 2-bridge knot/link). All DNA knots and links that have been identified so far are 4-plats or a composite of 4-plats. This is actually not as surprising as it may at first seem since most small crossing knots and links are 4-plats. In fact all knots of less than eight crossings are 4-plats or composites of 4-plats. For a list of 4-plats knots up to 10 crossings and 4-plat links up to 9 crossings see http://www.utdallas.edu/∼darcy. The 4-plat corresponding to $N(a/b)$ is $S(a,-b)$. Four-plats are prime knots when a is odd and are prime links when a is even. To avoid additional terminology, we will avoid using the normal 4-plat notation and will instead use $N(a/b)$, rational tangle notation.

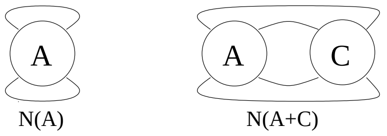


Fig. 7. Numerator closure

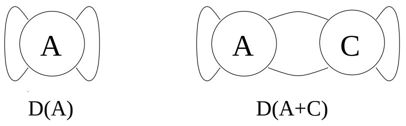


Fig. 8. Denominator closure

Useful facts about tangles:

(1) Both $N(a/b)$ and $D(a/b)$ are 4-plats. The knot/link $N(a/b)$ is the 4-plat $S(a,-b)$. The knot/link $D(a/b)$ is the 4-plat $S(b,a)$.

(2) The tangle corresponding to $\frac{a_1}{b_1}$ is the same as the tangle corresponding to $\frac{a_2}{b_2}$ if and only if $\frac{a_1}{b_1} = \frac{a_2}{b_2}$.

(3) The 4-plats $N(a_1/b_1)$ and $N(a_2/b_2)$, $a_i \geq 0$, are are the same if and only if $a_1 = a_2$ and $b_1 b_2^{\pm 1} \cong 1 \pmod{a_1}$.

(4) $N(\frac{a_1}{b_1} + \frac{a_2}{b_2}) = N(\frac{a_1 b_2 + a_2 b_1}{a_1 b_2' + a_2' b_1})$ where $a_2' b_2 - a_2 b_2' = 1$ [5].

(5) $\frac{a_1}{b_1} + \frac{a_2}{b_2} \neq$ a rational tangle unless either $b_1 = \pm 1$ or $b_2 = \pm 1$.

(6) $\frac{a}{b} + t = \frac{a+bt}{b}$

(7) $D(A+C) = D(A) \# D(C)$.

(8) $N(\frac{a_1}{b_1} + ... + \frac{a_n}{b_n})$ is a Montesinos knot which is not a 4-plat unless at most two of the tangles $\frac{a_i}{b_i}$ are non-integral.

(9) N(A + C) = 4-plat implies at least one of A or C is rational or locally knotted [8][9].

(10) If A is a prime tangle and C is locally unknotted and not the ∞-tangle, then A + C is a prime tangle [9][10].

The tangle, C, is locally unknotted if every 3-ball which intersects the two arcs of C in exactly two points contains an unknotted spanning arc. A tangle is locally knotted if it is not locally unknotted. A tangle is prime if it is neither rational nor locally knotted.

WARNING: The sum of two rational tangles is a rational tangle if and only if at least one of the two rational tangles is integral.

Even though the numerator closure of the sum of two rational tangles is the numerator closure of a rational tangle by (4), the sum of two rational tangles is not a rational tangle unless one of the tangles is integral (5), (6). The denominator closure of the sum of two tangles, A and C, is $D(A)\#D(C)$ which is a composite knot unless one of $D(A)$ or $D(C)$ is the unknot. But $D(a/b) = S(b,a) =$ unknot if and only if $b = \pm 1$, i.e., a/b is an integral tangle. Since 4-plats are prime and the denominator closure of a rational tangle is a 4-plat and thus prime, the sum of two rational tangles is not a rational tangle unless one of the tangles is integral. In this case the tangle $\frac{a}{b} + t$ is the rational tangle $\frac{a+bt}{b}$ (6). The sum of rational tangles where at least two of the tangles is non-integral is called a Montesinos tangle.

3. Tangle Calculus and Processive Recombination

Tn3 Resolvase was the first recombinase to be studied using the mathematics of tangles. The mathematical model describing Tn3 Resolvase action is shown in figure 9 [5]. In general, the substrate DNA is written as the numerator closure of the sum of the two tangles U and P. The tangle U stands for the outside tangle, which is outside of the enzyme action and remains unchanged during the reaction (U is used instead of O to avoid confusion with the 0 tangle = zero tangle). The recombinase action can be modeled by replacing the tangle P with the tangle R. We thus have the following two equations with 3 unknowns U, P, and R:

$$N(U+P) = \textit{substrate}, \quad N(U+R) = \textit{product}$$

In processive recombination the enzyme remains bound to the DNA but acts multiple times. Processive recombination has been modeled by adding additional copies of R thus obtaining the following additional equations:

$$N(U+R+R) = \textit{2nd processive product}$$

$$N(U+R+R+R) = \textit{3rd processive product}$$

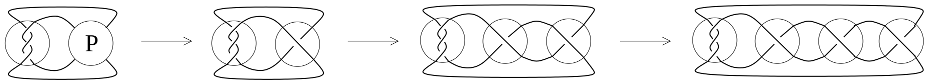
$$\cdot$$

$$\cdot$$

$$\cdot$$

$$N(U+R+\ldots+R) = \textit{nth processive product}$$

One can then solve these equations for U and R, giving new biological information. For example, Ernst and Sumners determined that Tn3 Resolvase uses the following configuration [5].



N(U + P) = unknot N(U + R) = N(2/1) N(U + R + R) = N(5/2) N(U + R + R + R) = N(8/3)

Fig. 9. Tn3 Resolvase recombination

Note that the tangle P only appears in one equation $N(U + P)$ = unknot. Thus it is not possible to find a unique solution for P. If P is rational, solving the equation, $N(\frac{-1}{3} + P)$ using tangle fact (4) gives $P = \frac{x}{1+3x} = (x, 3, 0)$ [5]. However, P need not be rational (figure 10). See section 6 for a more in depth discussion of non-unique solutions.

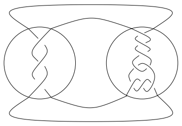


Fig. 10. $N(U + P) =$ unknot

Ernst and Sumners have proven that if the first three products of processive recombination are known, then the solutions for U and R are unique (assuming P = 0-tangle) using the following theorems:

Theorem 1 [**5**] *Let U and R be tangles such that $N(U + iR)$ =4-plat for some $i \geq 2$, and $N(U + jR) \neq N(U + iR)$ for some j. Then R is a rational tangle. If $i \geq 3$, then R is an integral tangle.*

Proof [10].

If R were locally knotted, then $N(U + iR), i \geq 2$ would be composite. Since 4-plats are prime, R cannot be locally knotted. Suppose R is a prime tangle. By tangle fact (9) $U + (i - 1)R$ is rational or locally knotted. By tangle fact (10), R prime implies $(i - 1)R$ prime and U must be the ∞-tangle or locally knotted.

Note that U cannot be the ∞-tangle. If U were the infinity tangle, then $N(U + iR) = D(iR) = D(R)\#...\#D(R)$. Since 4-plats are prime, $D(R) =$ unknot. But $N(U + iR) = D(iR) =$ unknot $= D(jR) = N(U + jR)$, a contradiction. Thus if U is locally unknotted, R must be rational.

If $i \geq 3$, then R does not have parity ∞ since 4-plats have at most two components. If $i \geq 3$, U is locally unknotted, and R is not integral, then if U is not integral, $U + R$ and $(i - 1)R$ are prime by tangle facts (5) and (10). But $N(U + iR)$ =4-plat would then contradict tangle fact (9). If U is integral and if R is rational, then $N(U + iR) =$ 4-plat, $i \geq 3$, if and only if R is integral by tangle fact (8). Thus if U is locally unknotted and $i \geq 3$, R must be integral.

Suppose U is locally knotted. Then if U' is the tangle formed from U by removing the local knot, then $N(U' + iR)$ =unknot, since 4-plats are prime. $N(U + jR) \neq N(U + iR)$ implies that $N(U' + jR) \neq N(U' + iR)$. Since the unknot is a 4-plat and U' is locally unknotted, R is rational if $i \geq 2$ and integral if $i \geq 3$ $\square$.

Theorem 2 [11] *If $N(U+P)$ =4-plat and $N(U+R)$ =4-plat where $P=\frac{a_1}{b_1}$, $R=\frac{a_2}{b_2}$, $a_1b_2-a_2b_1 \neq \pm 1$, then U is either a rational tangle or ambient isotopic to a sum of two rational tangles.*

Proof. If $N(U+P)$ =4-plat and $N(U+R)$ =4-plat where $P=\frac{a_1}{b_1}$, $R=\frac{a_2}{b_2}$, $a_1b_2-a_2b_1 \neq \pm 1$, then the cyclic surgery theorem implies that the double branched cover of the tangle U is a Seifert fibered space. In [11], Ernst proved that this means that U is ambient isotopic to a Montesinos tangle □.

[11] actually stated that U is a Montesinos tangle, but the following counterexample (figure 11) to corollary 10 in [11] shows that when gluing U back to P, additional twists may be necessary. In fact in this case there is no rational or Montesinos tangle solution to the system of equations $N(U+0)=N(5/-1)$ and $N(U+2)=N(7/-3)$. Fortunately by pushing any extra crossings into P and R, using flypes [7], and the euler bracket function [12], it can be shown that given the three equations, $N(U+iR)=K_i$ for $1 \leq i \leq 3$ where the K_i are 4-plats, then U is either rational or the sum of two rational tangles.

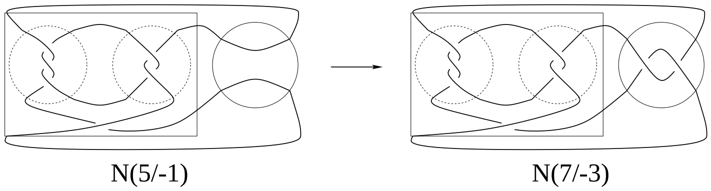


Fig. 11. $N(U+0)=N(5/-1)$ and $N(U+2)=N(7/-3)$

Theorem 3 [10] *Let U and R be tangles such that $N(U+iR)=K_i$ 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. Then there is at most one solution for U and U is either rational or the sum of two rational tangles.*

Theorem 4 [10] *Let U and R be tangles such that U is either rational or a sum of two rational tangles, and $R=(n)$. Moreover, suppose that $N(U+iR)=K_i$ for $0 \leq i \leq 3$, where the K_i are 4-plats with crossing number c_i. Then $|n|$ is determined as follows:*

(i) If $c_0 < c_3 > c_2 > c_1$, then $|n| = c_3 - c_2$.

(ii) If $c_0 = c_3$ and $c_1 = c_2$, then $|n| = c_3 - c_2 = c_0 - c_1$.

(ii) If $c_0 = c_3$ and $c_1 \neq c_2$ then $|n| = 1$.

(iii) If $c_0 \neq c_3$ and $c_2 \leq c_1$, then $|n| = c_0 - c_1$.

The tangle facts (3), (4), and (7) can be used to solve for the tangle U, or similarly, an algorithm given in [10] can be used.

Note that only the crossing numbers of the 4-plats were needed to determine $|R|$. This is very useful for analyzing biological experiments since gel electrophoresis cannot fully identify a knot but can identify its crossing number for small crossing knots.

4. Tangle Distances and Recombination

The above section gave a unique solution to tangle equations (with $P = 0$) resulting from processive recombination where U and R are constant. Tangle equations can be difficult to solve, however, if the enzyme (such as one of the recombinases, λ-Integrase, Flp, and mutant Gin) does not require the DNA to be in a specific configuration before the enzyme acts (meaning U is not constant) or if the recombinase (such as one of the recombinases, λ-Integrase, Xer, and possibly Flp) does not act processively. Distributive recombination adds additional complications. In distributive recombination, the enzyme releases the DNA between each round of recombination. Thus in the above equations, the enzyme first binds onto the substrate, $N(U_1 + P)$. One round of recombination is modeled as before by replacing P with R thus obtaining product 1, $N(U_1 + R)$. This product is then released by the enzyme. The enzyme may then rebind product 1. However, the DNA configuration may have changed since it was released and product 1 may now be written as $N(U_2 + P)$ where U_2 may be completely different from U_1. The second round of recombination is modeled again by replacing P with R thus obtaining product 2, $N(U_2 + R)$. Note that for all rounds of recombination, it is assumed that the enzyme is always replacing a P tangle with an R tangle. Thus if an enzyme sometimes replaces a P tangle with an R tangle and at other times replaces a P' tangle with an R' tangle where $P \neq P'$ and/or $R \neq R'$, the above model does not work for this enzyme.

Whether or not an enzyme requires a unique configuration, acts processively or distributively, or even if it uses more than one mechanism, tangle distances can be used to model enzyme action.

Definition The link L_2 is said to be obtained from the link L_1 by a *[P,R] move* if L_2 is obtained from L_1 by replacing a P tangle with an R tangle.

Definition The $[P, R]$-distance between L_1 and $L_2 = d_{[P,R]}(L_1, L_2) =$ the minimum number of $[P, R]$ moves required to change L_1 into L_2 where the minimum is taken over all diagrams of the knot.

In biological terms $d_{[P,R]}(L_1, L_2)$ is the minimum number of times an enzyme needs to act to convert L_1 into L_2 if the enzyme uses only $[P, R]$ moves. When P is the zero tangle, $[0, R]$ move will be shortened to R move and $d_{[0,R]}(L_1, L_2)$ will be shortened to $d_R(L_1, L_2)$

For example, $d_{+2}(0_1, 3_1) = 1$ (figure 12) and $d_{30/7}(0_1, 5_2) = 1$ (figure 13).

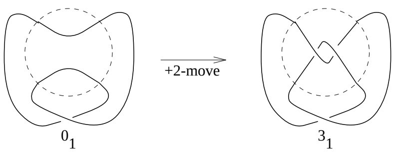


Fig. 12. $d_{+2}(0_1, 3_1) = 1$.

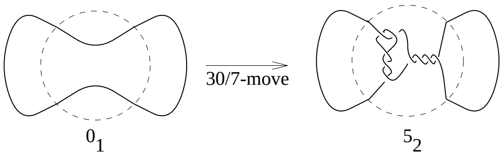
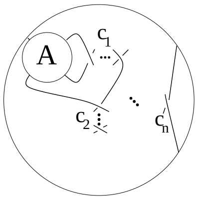


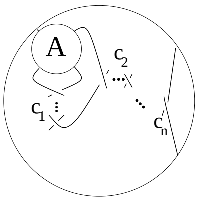
Fig. 13. $d_{30/7}(0_1, 5_2) = 1$.

Definition $A \circ B =$
where $B = (c_1, ..., c_n)$



n odd

or



n even

When P and R are rational, the following two theorems can be used to determine when the $[P, R]$ distance between two 4-plats is one. The first theorem assumes $P =$ 0-tangle and is a generalization of [13] which classified $d_2(N(a/b), 0_1) = 1$, of [14] which classified $d_2(N(a/b), 0_1^2) = 1$ and of [15], [16] and unpublished by John Berge which classified $d_2(N(a/b), N(z/v)) = 1$ and a modification of theorem 1 in [17].

Theorem 5. [17] *Let $R = t/w$-tangle, $(w, t) = 1$, and $ay - bx = 1$. Then the following are equivalent for $|t| \geq 2$. For $t = \pm 1$, (2) and (3) are equivalent and imply (1):*

(1) $d_R(N(a/b), N(z/v)) \leq 1$.

(2) If $w \not\cong \pm 1$ mod t, $N(z/v) = N((tb + (tk + w)a)/(-ty - (tk + w)x))$
or $N((-tx+(tk+w)a)/(-ty+(tk+w)b))$.
Else $w \cong \pm 1$ mod t and $N(z/v) = N((tp^2b + sa)/(-tp^2y - sx))$
or $N((-tp^2x + sa)/(-tp^2y + sb))$
where $s = tp(-q + pk) \pm 1$, $(p, q) = 1$, $p > 0$

(3) N(a/b) = N(U + 0) and N(z/v) = N(U + t/w) has the following solutions when $|t| \geq 2$: If $w \neq \pm 1$ mod t, then U must be rational and $U = \frac{a}{b+ka}$ or $\frac{a}{-x+ka}$. If $w = \pm 1$ mod t, then U must be ambient isotopic to a sum of at most two rational tangles and $U = (U_1 + U_2) \circ (h, 0)$ where $U_1 = \frac{-bj+a(d-kj)}{pb+a(pk-q)}$ or $\frac{xj+a(d-kj)}{-px+a(pk-q)}$ and $U_2 = j/p$, $pd - qj = 1$, and $h = \frac{-w\pm1}{t}$ if $\frac{-w\pm1}{t} \in Z$. If $t = \pm 1$, then the above list contains all solutions when U is ambient isotopic to a sum of rational tangles.

Proof.

$d_R(N(a/b), N(z/v)) \leq 1$ if and only if there exists a U such that $N(a/b) = N(U + 0)$ and $N(z/v) = N(U + t/w)$. By Ernst's theorem 2, U is either a rational tangle or ambient isotopic to the sum of two rational tangles. If U is rational, N(U + 0) = N(a/b) implies by tangle fact 4 that $U = \frac{a}{b+ka}$ or $\frac{a}{-x+ka}$ and $N(z/v) = N(U + t/w) = N((tb + (tk + w)a)/(-ty - (tk + w)x))$ or $N((-tx + (tk + w)a)/(-ty + (tk + w)b))$. If U is ambient isotopic to the sum of two rational tangles, $U_1 + U_2$, then since $N(U + 0)$ is a 4-plat, $U = (U_1 + U_2) \circ (h, 0)$. Solving $N((U_1 + \frac{j}{p}) \circ (h, 0) + 0) = N(a/b)$ implies $U_1 = \frac{-bj+a(d-kj)}{pb+a(pk-q)}$ or $\frac{xj+a(d-kj)}{-px+a(pk-q)}$ where $pd - qj = 1$. $N((U_1 + \frac{j}{p}) \circ (h, 0) + t/w) = N(U_1 + \frac{j}{p} + (t/w) \circ (h, 0)) = N(U_1 + \frac{j}{p} + \frac{t}{ht+w})$. If U_1 or U_2 is integral, $(U_1 + U_2) \circ (h, 0)$ is rational. If U_1 and U_2 are non-integral, $N(U_1 + U_2 + \frac{t}{ht+w})$ is a 4-plat if and only if $ht + w = \pm 1$, i.e. $w = \pm 1$ mod t in which case $h = \frac{-w\pm1}{t}$ if $\frac{-w\pm1}{t} \in Z$.

By tangle facts 4 and 6, if $s = tp(-q+pk) \pm 1$, $N(U_1 + U_2 + \pm t) = N((tp[pb+(pk-q)a] \pm a)/(-tp[py+(pk-q)x] \mp x)) = N((tp^2b+sa)/(-tp^2y-sx))$ or $N((tp[-px+(pk-q)a] \pm a)/(-tp[py-(pk-q)b] \pm b)) = N((-tp^2x+sa)/(-tp^2y+sb))$ $\square$.

A computer program which solves for U_1, U_2, t/w given a, b, z, v is available at http://www.utdallas.edu/∼darcy. Since U is known, the above theorem also gives a classification of oriented distance one oriented 4-plats when $t \neq \pm 1$ [18].

For example solving $N(U+0) = N(5/-1)$ and $N(U + \frac{t}{ti \pm 1}) = N(7/-3)$ gives $U = 1/2 + 1/3$, $t = -2$, and if $w = -2i+1$, then $h = -i$ or if $w = -2i-1$, then $h = -i-1$. In particular, $U = 1/3 + 1/2$, $t = -2$ and $U = (1/3 + 1/2) \circ (-1, 0)$, $t = 2$ (figure 11) are the only solutions when t/w is integral. Also note that since a 2-move is equivalent to a crossing change, this also implies that N(7/-3) can be obtained from N(5/-1) by changing one negative crossing to a positive crossing but not by changing one positive crossing to a negative crossing.

Fig. 14. Signed crossings.

However if $t = \pm 1$, U need not be ambient isotopic to a Montesinos tangle and thus the distance one classification is not yet complete. If $t = \pm 1$, then unless the classification determines that the distance is one (and thus there exists a U which is ambient isotopic to a Montesinos tangle), then it is unknown whether or not the distance is one unless other mathematical bounds apply or a projection is found in which such a [P,R]-move can be made. For example theorem 5 gives no information as to whether or not $d_1(6_2^*, 2_1^2) = 1$. However, the following example (figure 15) in which U is not ambient isotopic to a Montesinos tangle shows that $d_1(6_2^*, 2_1^2) = 1$.

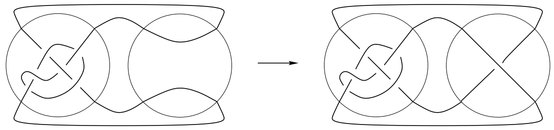


Fig. 15. $d_1(6_2^*, 2_1^2) = 1$.

Theorem 6. [17] *The link* L_2 *can be obtained from the link* L_1 *by replacing the* f_1/g_1*-tangle with the* f_2/g_2*-tangle if and only if* $d_{t/w}(L_1, L_2) = 1$, *where* $\frac{t}{w} = \frac{g_1f_2-g_2f_1}{h_1g_2-i_1f_2}$, h_1, i_1 *such that* $g_1h_1 - f_1i_1 = 1$. *In this case if* $\frac{t}{w} = (b_1, ..., b_k)$ *and* $\frac{f_2}{g_2} = (d_1, ..., d_m)$ *then there exists a continued fraction expansion of* $\frac{f_1}{g_1} = (c_1, ..., c_n)$, n *odd such that* $\frac{t}{w} = \frac{g_1f_2-g_2f_1}{h_1g_2-i_1f_2} = (d_1, ..., d_m - c_n, -c_{n-1}, ..., -c_1)$ *and* $\frac{f_2}{g_2} = \frac{th_1+wf_1}{ti_1+wg_1} = (b_1, ..., b_k + c_1, ..., c_n)$

These theorems can be used to determine all possible $[P, R]-$moves to change a particular 4-plat into another 4-plat where P and R are rational tangles except in the case when $P = \frac{f_1}{g_1}$, $R = \frac{f_2}{g_2}$ and $f_1 g_2 - f_2 g_1 \neq \pm 1$ [17]. Theorem 5 can be used to determine all possible R-moves. Theorem 6 states that solving the equations:

$$N(U + \frac{f_1}{g_1}) = substrate \qquad N(U + \frac{f_2}{g_2}) = product \tag{4.1}$$

is the equivalent to solving the equations:

$$N(U' + 0) = substrate \qquad N(U' + \frac{t}{w}) = product \tag{4.2}$$

where $\frac{t}{w} = \frac{g_1 f_2 - g_2 f_1}{h_1 g_2 - i_1 f_2} = (d_1, ..., d_m - c_n, -c_{n-1}, ..., -c_1)$, for some continued fraction expansion of $\frac{f_1}{g_1} = (c_1, ..., c_n), n$ odd and $\frac{f_2}{g_2} = (d_1, ..., d_m)$. See figure 16 below.

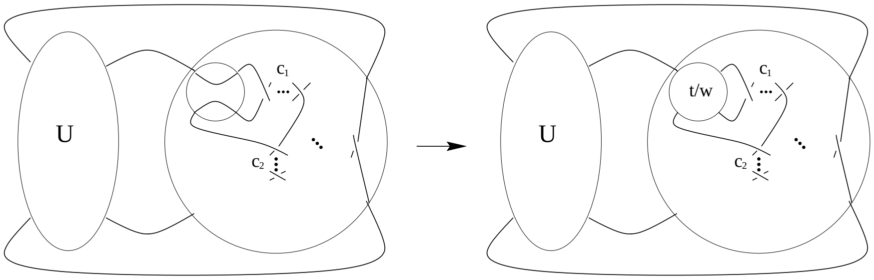


Fig. 16. $N(U + \frac{f_1}{g_1}) \rightarrow N(U + \frac{f_2}{g_2})$.

Since $A + B \circ (c_1, c_2, ..., c_n) = A \circ (c_n, c_{n-1}, ..., c_1) + B$ when n is odd, $N(U + \frac{f_1}{g_1}) = N(U + 0 \circ (c_1, ..., c_n)) = N(U \circ (c_n, ..., c_1) + 0)$ and $N(U + \frac{f_2}{g_2}) = N(U + \frac{t}{w} \circ (c_1, ..., c_n)) = N(U \circ (c_n, ..., c_1) + t/w)$ Thus if U is a solution to eq. (4.1), then $U \circ (c_n, ..., c_1)$ is a solution to eq. (4.2).

$N(U' + 0) = N(U' + (c_1, ..., c_n) \circ (-c_n, ..., -c_1)) = N(U' \circ (-c_1, ..., -c_n) + (c_1, ..., c_n)) = N(U' \circ (-c_1, ..., -c_n) + \frac{f_1}{g_1})$ and $N(U' + t/w) = N(U' + (d_1, ..., d_m) \circ (-c_n, ..., -c_1)) = N(U' \circ (-c_1, ..., -c_n) + (d_1, ..., d_m)) = N(U' \circ (-c_1, ..., -c_n) + \frac{f_2}{g_2})$ Thus if U' is a solution to eq. 4.2, then $U' \circ (-c_1, ..., -c_n)$ a solution to eq. 4.1.

Since $U' \circ (-c_1, ..., -c_n) \circ (c_n, ..., c_1) = U'$ and $U \circ (c_n, ..., c_1) \circ (-c_1, ..., -c_n) = U$, U is a solution to eq. 4.1 if and only if $U = U' \circ (-c_1, ..., -c_n)$ for some U' where U' is a solution to eq. 4.2.

Note that U is rational if and only if U' is rational and U is ambient isotopic to a sum of two rational tangles if and only if U' is ambient isotopic to a sum of two rational tangles. Thus U is ambient isotopic to a sum of at most two rational tangles when $f_1 g_2 - f_2 g_1 \neq \pm 1$ which corresponds to the case $t \neq \pm 1$.

Normally, when $f_1 g_2 - f_2 g_1 = \pm 1$, U does not need to be ambient isotopic to a sum of rational tangles. However, Hirasawa and Shimokawa have proved that U must be rational when the substrate is unknotted and the product is $N(2z/1)$ when $f_1 g_2 - f_2 g_1 = \pm 1$.

Theorem 7 [19] If $N(U+\frac{f_1}{g_1})=$ unknot and $N(U+\frac{f_2}{g_2})=N(2z/1)$ where $f_1g_2-f_2g_1=\pm1$, then U is rational.

Theorems 8 and 10 can be used when not all the knots involved are 4-plats and are a generalization of [20] and [21], respectively.

Theorem 8 [16] If $d_{t/w}(L,N(a/b))=1$, then for some integer n, and $s\cong w$ mod t, $|H_1(\widetilde{L})|\cong<m,L_1\mid tnL_1+sm,aL_1-bnm>$.

Corollary 9 If $d_{t/w}(N(a/b),K)=1$, then there exists an integer n such that $|H_1(M_K)|\equiv|aw+tbn^2|\pmod{ta}$.

Theorem 10 [22] If $d_t(K,K')=1$, then there exist $a\in H_1(M_K)$ and $a'\in H_1(M_{K'})$ such that $\lambda(a,a)\equiv n/|H_1(M_K)|$ and $\lambda(a',a')\equiv m/|H_1(M'_K)|\pmod 1$, where

$$m=n=\frac{\pm1}{t}(|H_1(M_K)|-|H_1(M'_K)|)$$

or

$$-m=n=\frac{\pm1}{t}(|H_1(M_K)|+|H_1(M'_K)|)$$

Note that theorem 10 also give information regarding $\frac{t}{kt\pm1}$ moves since these moves are equivalent to $\pm t$ moves.

5. Computer algorithm for solving tangle equations

To solve $N(U+P)=N(a/b)$ and $N(U+R)=N(c/d)$ where P and R are rational, theorem 5 can be used to find all rational solutions for R assuming P is the 0 tangle. If $R\neq1/j$, theorem 5 can also be used to find all solutions for U. Theorem 6 can then be used to find all rational solutions for R given any rational P. All solutions for U can then be found unless $R=1/j$ was a solution when $P=0$ which corresponds to the case $f_1g_2-f_2g_1=\pm1$ where $P=\frac{f_1}{g_1}$ and $R=\frac{f_2}{g_2}$

The classification of distance one 4-plats gives a formula which can be solved in order to determine which [0, t/w] moves are possible when $t\neq\pm1$. A computer program which outputs all possible moves when $t\neq\pm1$ is available at http://www.utdallas.edu/~darcy. The program is currently being updated to be more user-friendly. It will eventually allow the user to input the knot/link types of the substrate and all products in order to calculate all possible reaction pathways noting whether processive recombination, distributive recombination, and/or multiple products from one round of recombination are possible. Currently, one can input each pair of knots and then analyze the output by hand. However, these calculations are easily programmable and will be available soon. Although using Ernst and Sumners' theorem 4 is the fastest method to determine R in the case of processive recombination, the program allows for other possibilities in addition to processive recombination and will not require P to equal the zero tangle.

For example the products of Tn3 Resolvase acting on unknotted substrate containing direct repeats are $N(2/1),N(5/2),N(8/3)$. When the recombinase equations were originally solved, it was assumed that $N(2/1),N(5/2),N(8/3)$ were the products of the first, second, and third round of recombination, respectively (figure 9). Suppose one instead assumes processive recombination, but allowing the same product for two different rounds of recombination and without assuming the order of the products. Thus the processive recombination equations are $N(U+iR) = K_i$, $i = 1, ..., n$, $n \geq 3$ where $K_i \in \{N(1/0), N(2/1), N(5/2), N(8/3)\}$. The three knotted/linked products must appear at least once in these equations; however, the substrate, N(1/0) need not. Recall that R must be an integral tangle by theorem 1, and note that $d_{2R}(K_1, K_3) = 1$. Thus there must exist an even t such that $d_t(K_1, K_3) = 1$. By using the program to calculate all possible integral distances between $N(1/0), N(2/1), N(5, 2)$, and $N(8/3)$, it is easily determined that exactly three round of recombination occurred and that $K_2 = N(5/2)$ and $\{K_1, K_3\} = \{N(2/1), N(8/3)\}$ where $t = \pm 2$ and thus $R = \pm 1$. Since U must be either a rational tangle or the sum of two rational tangles, one obtains either the original solution shown in Fig. 9 or that $N(11/3 + P)$ = unknot, $N(11/3 + -1) = N(8/3)$, $N(11/3 + -2) = N(5/3)$, $N(11/3 + -3) = N(2/1)$. Since $N(2/1)$ is a more abundant product than $N(8/5)$, it was a biologically reasonable assumption to assume that $N(2/1)$ is the product of one and not three rounds of recombination. Also $U = -1/3$ is more likely than $U = 11/3$ when the substrate is negatively supercoiled DNA. But with the program, it is easy to solve recombinase equations with as few or as many assumptions as desired; although in some cases, the list of solutions may be extensive (see section 6).

Since determining possible R moves is fast, an option to enter the crossing number of the knot rather than the actual identity of the knot will also be available to analyze experiments for which only the crossing number of the product is known. For example, if a seven crossing knot is produced, it is easier and much faster to mathematically analyze all seven crossing equations rather than to biologically identify the knot.

6. Example: Xer recombination

Xer recombinase acting on unknotted substrate produces only one product, the link $N(4/1)$ [23]. Thus there are only two equations for which to solve for 3 unknowns resulting in an infinite number of solutions.

$$N(U_1 + P) = unknot \qquad\qquad N(U_1 + R) = N(4/1)\ . \tag{6.1}$$

This enzyme cannot act processively. Thus there is no experiment that can be done to reduce the infinite number of solutions to a finite number. However, what mathematics can do is list all possible solutions and propose experiments to reduce this list to a smaller number of infinite solutions which can then be analyzed for which solutions are the most biologically relevant using additional biological assumptions. For example, if P and R are restricted for biological reasons to have at most 4 crossings, then the list becomes finite. In this section, we will list all possible solutions given a few assumptions and propose experiments to reduce this list.

If P is the zero tangle and R is the rational tangle t/w, by theorem 5, R is $\frac{1}{j} = (j,0)$ or $\frac{+3}{1+3j} = (3,j,0)$ or $\frac{+5}{1+5j} = (5,j,0)$ or $\frac{-1+4k}{4+j(-1+4k)} = (-4,j,k,0)$ where j and k are any integers. By theorem 6, if P is the rational tangle $f_1/g_1 = (c_1, ..., c_n)$, n odd, and R is the rational tangle f_2/g_2, then $f_2/g_2 = \frac{th_1+wf_1}{ti_1+wg_1}$, and $g_1h_1 - f_1i_1 = 1$. Or equivalently, $f_2/g_2 = (j, c_1, ..., c_n)$ or $(3, j, c_1, ..., c_n)$ or $(5, j, c_1, ..., c_n)$ or $(-4, j, k, c_1, ..., c_n)$. However, P and R may not be rational.

For some recombinases one round of recombination will result in various products rather than a unique product. In this case one is solving the equations:

$$\begin{array}{ll}
N(U_1 + P) = \text{substrate} & N(U_1 + R) = \text{product 1} \\
N(U_2 + P) = \text{substrate} & N(U_2 + R) = \text{product 2} \\
 & \vdots \\
N(U_n + P) = \text{substrate} & N(U_n + R) = \text{product n}
\end{array}$$

If the substrate is unknotted DNA and at least two different products result from one round of recombination, then P is rational by lemma 11. Unfortunately Xer only produces a unique product and thus lemma 11 cannot be used to show that P is rational for the recombinase Xer.

Lemma 11 [5] If $N(U_i+P) =$ unknot, $i = 1, 2$ and $U_1 \neq U_2$, then P is rational.

In order to obtain more equations, Colloms et al [24] used different substrates. The enzyme λ-Integrase is another recombinase which produces $(p, 2)$ torus knots, $N(p/1)$, p odd, when acting on substrates containing inversely oriented sites and produces $(p, 2)$ torus links, $N(p/1)$, p even, when acting on substrate containing directly oriented sites. It was determined by gel electrophoresis that the majority of the $(p, 2)$ torus links contained $4, 6, 8, 10$,and 12 crossings. When these $(p, 2)$ torus links were used as substrates for Xer recombination, products resulted containing 7, 9, and 11 crossings. It has previously been shown that Xer will not act on the 4 crossing torus link [23]. In order to determine which of the other $(p, 2)$ torus links produced which products, the six and eight crossing torus links were purified from the mixture of torus links. Xer recombinase acting on the six crossing torus link produced a seven crossing product. When the eight crossing torus link was used as a substrate, a nine crossing product resulted.

$$N(U + P) \neq N(4/1) \tag{6.2}$$

$$N(U_2 + P) = N(6/1) \qquad N(U_2 + R) = \textit{seven crossing product} \tag{6.3}$$

$$N(U_3 + P) = N(8/1) \qquad N(U_3 + R) \ = \ \textit{nine crossing product} \tag{6.4}$$

Eq. (6.2) cannot reduce the number of mathematical solutions, but can be used for biological analysis. Eq. (6.4) involves a nine crossing knot. Unfortunately, theorem 5 only applies to 4-plats. Up to mirror images, there are 49 nine crossing prime knots of which 25 are 4-plats. Since it is unknown whether or not the nine crossing knot produced by Xer is a 4-plat, eq. (6.4) will not be used in this paper to reduce the number of mathematical solutions. There are other methods such as theorem 8 which may be of use, but these methods are not nearly as efficient as theorems 5 and 6. Since most small crossing knots are 4-plats and the number of knots with a particular crossing number increases exponentially, products with at most 7 crossings are strongly preferred.

Eq. (6.3) can be used to reduce the number of mathematical solutions. Recall that by eq. (6.1), if P is the zero tangle and R is the rational tangle t/w, R is $\frac{1}{j}, \frac{+3}{1+3j}, \frac{+5}{1+5j}$ or $\frac{-1+4k}{4+j(-1+4k)}$. Recall that theorem 5 cannot be used to eliminate $R = \frac{1}{j}$. All seven crossing 4-plats knots (which includes all 7 crossing prime knots) were checked using the computer program to determine if $R = \frac{+3}{1+3j}$, or $\frac{+5}{1+5j}$ satisfies eq. (6.3). If $K = N(21/8) = 7_7^*$, then $R = \frac{+3}{1+3j}$ is still a possible solution. A C program was written to determine what rational tangles $t/w, w \ncong \pm 1$ satisfy both equations. If $K = N(13/-10) = 7_3$, then $R = \frac{-37}{4-37j}$ is still a possible solution. If $K = N(15/4) = 7_4^*$, then $R = \frac{-9}{4-9j}$ is still a possible solution. If K is a seven crossing composite knot, then theorem 8 rules out everything except $R = \frac{1}{j}, \frac{+3}{1+3j}, \frac{-13}{4-13j}$.

Seven crossing 4-plat links were also checked, and it was determined that $R = \frac{+3}{1+3j}$ is the only rational tangle which satisfies both eq. (6.1) and (6.3) when $P = 0$ in which case K is the link $N(18/5) = 7_2^2$. However, a link is not the biologically expected product. The six crossing link was constructed with a recombination site on each component of the link. Thus if P and R are the same in eq. (6.1) and (6.3), the product must be a knot and not a link (see figure 1). Therefore, seven crossing non 4-plat links were not considered.

Colloms et al [23] were also able to determine the orientation of N(4/1), the product of Xer recombination on unknotted DNA substrate. Tn3 Resolvase is another site-specific recombinase. Tn3 Resolvase binds to specific DNA sequences called res sites. Tn3 Resolvase will not act on N(4/1) when the two res sites are oriented in a parallel fashion, but will act when the sites are in anti-parallel orientation (figure 17). A substrate that contained two res sites in addition to two sites for Xer was used to determine that the product of Xer recombination acting on unknotted substrate is N(4/1) with an anti-parallel orientation (Lk = -2). In Figure 18, the Xer binding sites are represented with open triangles and the res sites are represented by black arrows.

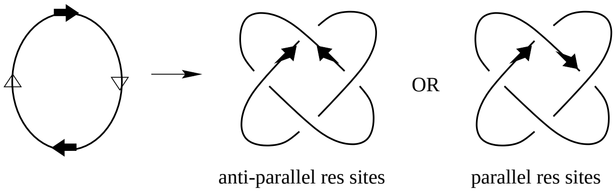


Fig. 17. Parallel versus anti-parallel

By theorem 5, $P = 0$ and $R = \frac{+3}{1+3j}, \frac{-37}{4-37j}, \frac{-13}{4-13j}$. results in N(4/1) with parallel sites (Lk = +2), whereas $P = 0$ and $R = \frac{-9}{4-9j}$ results in N(4/1) with anti-parallel sites. By Hirasawa and Shimokawa's theorem 7, U_1 is rational when $P = 0$ and $R = 1/j$. Thus N(4/1) is anti-parallel when $R = 1/j$. Thus by theorem 5, theorem 7, and tangle facts 4 and 8, if $P = 0$ and R is rational, then $U_1 = \frac{1}{-j}$ and $R = \frac{-9}{4-9j}$ or $U_1 = \frac{1}{-j-4}$ and $R = \frac{1}{j}$. If $P = f_1/g_1 = (c_1, ..., c_n)$ and if R is rational, then $R = f_2/g_2 = \frac{th_1+wf_1}{ti_1+wg_1} = \frac{-9h_1+(4-9j)f_1}{-9i_1+(4-9j)g_1}$ or $\frac{h_1+jf_1}{i_1+jg_1}$ and the solutions to eq. (6.1) assuming (6.3) also holds are of the form:

$$\begin{aligned} N((-j, -c_1, ..., -c_n) + (c_1, ..., c_n)) &= unknot \\ N((-j, -c_1, ..., -c_n) + (-4, -2, j, c_1, ..., c_n)) &= N(4/1) \end{aligned} \tag{6.5}$$

or

$$\begin{aligned} N((-j-4, -c_1, ..., -c_n) + (c_1, ..., c_n)) &= unknot \\ N((-j-4, -c_1, ..., -c_n) + (j, c_1, ..., c_n)) &= N(4/1) \end{aligned} \tag{6.6}$$

If U_2 is ambient isotopic to a sum of rational tangles, and R is rational, the solutions to eq. (6.3) assuming (6.1) also holds are of the form

$$\begin{aligned} N((6, -j, -c_1, ..., -c_n) + (c_1, ..., c_n)) &= N(6/1) \\ N((6, -j, -c_1, ..., -c_n) + (-4, -2, j, c_1, ..., c_n)) &= N(15/4) \end{aligned} \tag{6.7}$$

or

$$\begin{aligned} N((-1/3 + -1/3) \circ (-j-1, -c_1, ..., -c_n) + (c_1, ..., c_n)) &= N(6/1) \\ N((-1/3 + -1/3) \circ (-j-1, -c_1, ..., -c_n) + (j, c_1, ..., c_n)) &= N(15/4) \end{aligned} \tag{6.8}$$

or

$$\begin{aligned} N((6, k, -c_1, ..., -c_n) + (c_1, ..., c_n)) &= N(6/1) \\ N((6, k, -c_1, ..., -c_n) + (j, c_1, ..., c_n)) &= N(\frac{6(j+k)+1}{-6}) \end{aligned} \tag{6.9}$$

where the seven crossing product is $N(7/1)$ if $j+k = 1$ or $N(11/6)$ if $j+k = -2$.

If P and R are rational and if U_2 is ambient isotopic to a sum of rational tangles, the seven crossing product is $N(15/4)) = 7_4^*$, $N(7/1) = 7_1^*$, or $N(11/6)) = 7_2$. However, U_2 may not be ambient isotopic to a sum of rational tangles when $f_1g_2 - f_2g_1 = \pm 1$ which is the case in eq. (6.8) and (6.9).

Note that the two equations below not only imply that $d_{[P,R]}(0_1, 4_1^2) = 1$ and $d_{[P,R]}(6_1^2$, seven crossing product$) = 1$, but they also imply that $d_{[U_1,U_2]}(0_1, 6_1^2) = 1$ and $d_{[U_1,U_2]}(4_1^2$, seven crossing product$) = 1$:

$$N(U_1 + P) = unknot \qquad N(U_1 + R) = N(4/1) \tag{6.10}$$

$$N(U_2 + P) = N(6/1) \qquad N(U_2 + R) = \textit{seven crossing product} \tag{6.11}$$

Using theorem 5, if U_1 = 0-tangle and U_2 is rational, then $U_2 = 1/j$ or $U_2 = \frac{31}{-6+31j}$ is a possibility when the seven crossing product is $N(7/1) = 7_1^*$ and $U_2 = \frac{13}{-6+13j}$ is also a possibility when this product is $N(11/-9) = 7_2$.

If P or R are not rational, then U_1 and U_2 must be rational by tangle fact (9). But by theorem 5 and Hirasawa and Shimokawa's theorem 7, P is rational.

Although P must be rational, R may not be rational. Recall that if $U_1 = f_1/g_1$ and $U_2 = f_2/g_2$, $f_1g_2 - f_2g_1 \neq \pm 1$, then P and R are ambient isotopic to a sum of at most two rational tangles. Note, however, that if U_1 = 0-tangle = 0/1 and $U_2 = 1/j$, then 0j - 1(1) = -1 and thus R need not be rational or even ambient isotopic to a Montesinos tangle in this case.

Proposed experiment 1: Use any of the knots, $K = 4_1, 6_3, 7_5, 7_7, 8_1$, or any of the links $K = 5_1^2, 6_2^2, 6_3^2$, or their mirror images as substrate. Since $d_1(0_1, K) > 1$ by theorem 8, if the recombinase can act on any of these substrates, then P and R are either rational or ambient isotopic to the sum of two rational tangles. Depending on the resulting product, it may be possible to show that P and R are rational.

Drawback: Experiments are time consuming. First sufficient amounts of the knotted/linked substrate must be constructed. T4 topoisomerase acting on supercoiled DNA produces twist knots such as 4_1 and 8_1. The link 5_1^{2*} is a minor product of Tn3 resolvase recombination meaning that a very large reaction would be required in order to produce a sufficient amount to use as a substrate for Xer recombination. If the recombinase does not act on the constructed knot/link, no information regarding P and R would be gained. If the recombinase does act, then the infinite list of mathematically possible solutions can be shown to be complete; however, it may not be worth the time to rule out prime tangle solutions which may not be biologically reasonable.

7. P and R versus E and F

If a recombinase acts processively, mathematicians can easily solve for the tangles U and R, and the number of solutions will be finite. If the enzyme cannot act processively, mathematicians can list all rational tangle solutions for R assuming that P is a particular rational tangle, but this list will be infinite. It may also be possible to prove that P and R are rational if the enzyme will act on an appropriate knotted substrate (proposed experiment 1). However, no matter what biological experiment is done, the list of mathematically possible solutions for P and R will still be infinite for enzymes that do not act processively. Since crossings can be pushed in or out of the tangles P and R, many distances are equivalent. For example figure 18 shows that a t-move is equivalent to any t/w-move where $w \cong 1$ mod t. Experiments were proposed in this paper in part to point out the limitations of tangle analysis.

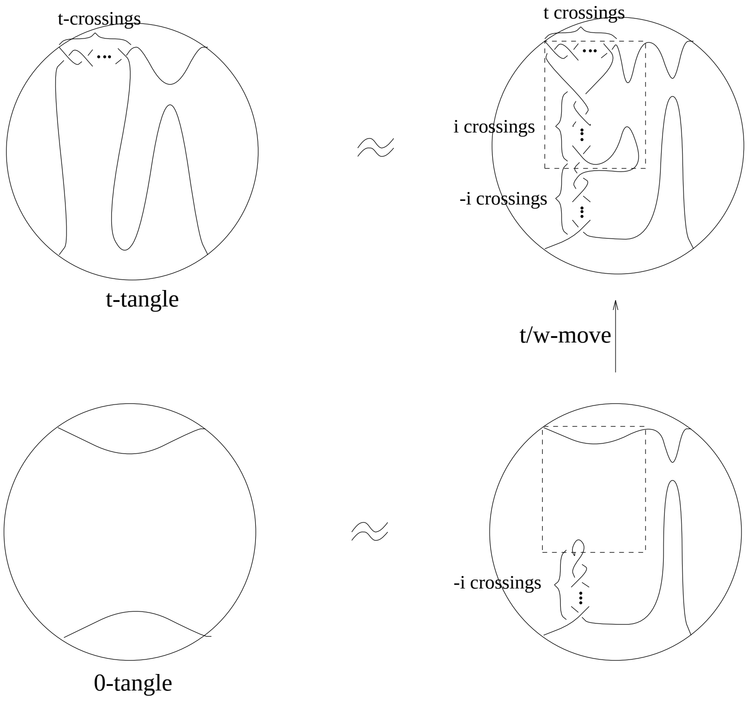


Fig. 18. Replacing the 0-tangle with the t-tangle via a t/w-move

Many biologists assume that for all recombinases, P and R have at most one crossing. However, [25] proposed that Xer may use a 1/2-move. This is mathematically possible. In fact, if P is the zero tangle, up to three crossings, $R = \infty, \pm 1, \pm 1/2, \pm 1/3$ are all possible solutions. However, if one only looks at the knot types of the products, then it is not possible to narrow this list further. Thus it is up to biologists to decide if the information gained is worth the experimental time.

The tangle model can still be very useful in analyzing Xer recombinase reactions. The full tangle model divides the tangle U into the tangles U_f and U_b where the tangle U_f is bound by the enzyme but outside of the enzyme action and the tangle U_f represents the free DNA not bound by the enzyme (figure 19). Because Xer produces a unique topological product for each substrate used, U_b is believed to remain constant for all reactions. The tangle model currently assumes that U_b is a two string tangle, but note that the more general model shown in figure 20 is also a possibility.

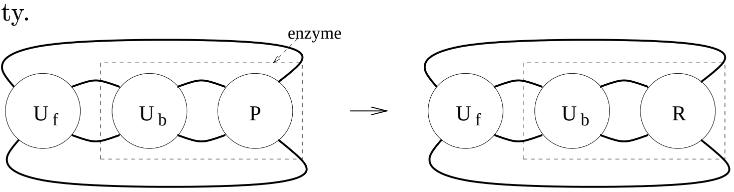


Fig. 19. Tangle model for recombination.

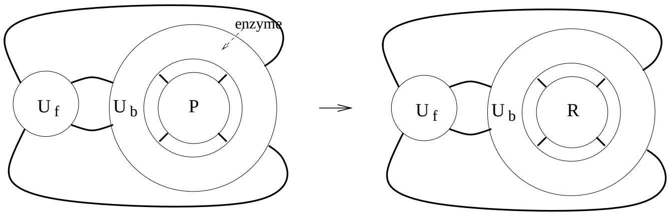


Fig. 20. Alternate tangle model for recombination.

Rather than combining U_b and U_f into the tangle U, it is also relevant to combine the tangles U_b and P into a tangle B and the tangles U_b and R into a tangle E with the resulting tangle equations $N(U_f + B)$ = substrate and $N(U_f + E)$ = product. There are two main differences between these equations and the equations involving U, P and R. First, it is possible to biologically determine U_f by taking an electron micrograph of the synaptic complex. Thus there are biological experiments which can reduce the infinite list of solutions for B and E to a finite or even unique topological answer (see proposed experiments 2 and 3 in section 8). Secondly, the tangles B and E represent the DNA bound by the enzyme and not just the local action which is represented by the tangles P and R. Since the tangles P and R represent the local action of breaking and rejoining two strands of DNA, these tangles are almost certainly 2-string tangles and thus U is also a 2-string tangle. There is no biological evidence that a third strand of DNA is directly involved in the local action. For some recombinases such as Tn3 resolvase, electron micrographs of the synaptic complex show that U_f is a 2-string tangle and thus E and B are also 2-string tangles. However, electron micrographs of the synaptic complex of the recombinase Gin show three strands of DNA looping out of the enzyme complex and therefore, U_f is a 3-string tangle. Currently the only mathematical method to handle this is to mathematically push one of the three strands into the enzyme complex and treat the tangles as though they were 2-string tangles.

8. Determining B and E.

Note that the tangle analysis done in section 6 also applies to equations (8) and (9) below assuming U_f is a 2-string tangle.

$$N(U_{f_1} + B) = \textit{unknot} \qquad N(U_{f_1} + E) = N(4/1) \tag{8.1}$$

$$N(U_{f_2} + B) = N(6/1) \qquad N(U_{f_2} + E) = \textit{seven crossing product} \tag{8.2}$$

Thus B must be rational. If R is rational and U_{f_2} is ambient isotopic to a sum of rational tangles, all the solutions to eq. (8.1) and (8.2) are listed in eq. (6.5) - (6.9). Since this list is rather extensive, in this section four different models will be analyzed. These models, drawn in figures 21-24, have been proposed to explain the topological product of Xer recombination on unknotted substrate with U_{f_1} = 0-tangle.

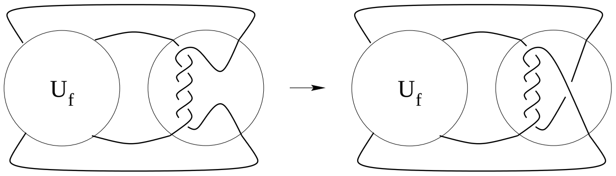


Fig. 21. $N(U_f + -1/5) \rightarrow N(U_f + 4/5)$

In figure 21, the -1/5 tangle is replaced by the 4/5 tangle. This is mathematically equivalent to replacing the zero tangle with the -25/4 tangle. However, by section 6, -25/4 is not a possible solution to equation (9). Thus Xer cannot use the model shown in figure 21.

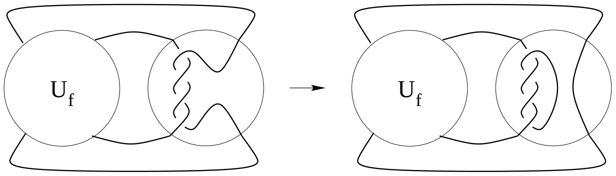


Fig. 22. $N(U_f + -1/4) \rightarrow N(U_f + R)$

In figure 22, the -1/4 tangle is replaced by a tangle containing the local link N(4/1). Thus any product according to this model would be of the form $N(4/1)\#K$ where K is some knot or link including possibly the unknot. If the substrate is a link, then U_f must have parity infinity. But if U_f has parity infinity, then K is a link and not a knot. However, there are no seven crossing links of the form $N(4/1)\#K$ where K is not a knot. Thus R cannot contain any local knots or links and Xer cannot use the model shown in figure 22.

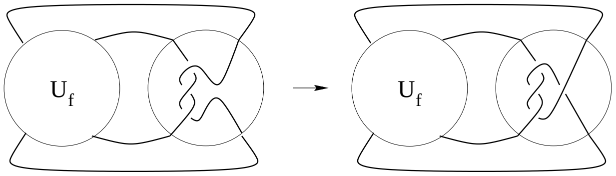


Fig. 23. $N(U_f + -1/3) \rightarrow N(U_f + -4/3)$

In figure 23, the -1/3 tangle is replaced by the -4/3 tangle. In this case if the product is a seven crossing knot, U_{f_2} must be a rational tangle. Suppose $U_{f_2} = \frac{x}{y}$, $N(\frac{x}{y} + \frac{1}{-3}) = N(6/1)$, and $N(\frac{x}{y} + \frac{4}{-3})$ = seven crossing product. Then $\frac{x}{y} = -1/3 =$ $(-3, 0)$ tangle and the seven crossing product is the knot 7_4 (figure 25).

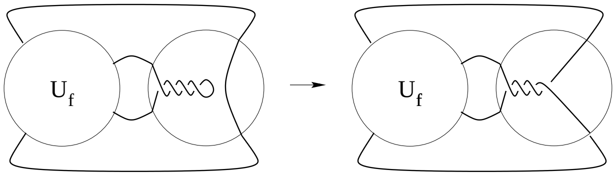


Fig. 24. $N(U_f + 1/0) \to N(U_f + 4)$

In figure 24, the infinity tangle is replace by the +4-tangle. If $N(U_{f_2} + \frac{1}{0}) = N(6/1)$ and $N(U_{f_2} + \frac{4}{1})$ = seven crossing product, then by eq. 6.8 and 6.9, $U_{f_2} = -31/6 = (-6, -5)$ and the seven crossing product is the knot 7_1^* (figure 26) or $U_{f_2} = -13/6 = (-6, -2)$ and the seven crossing product is the knot 7_2 (figure 27) or $U_{f_2} = (-1/3 + -1/3) \circ (2, 1, -1)$ and the seven crossing product is the knot 7_4^* (figure 28). Note, however, that U_{f_2} could be a prime tangle which is not ambient isotopic to a Montesinos tangle. In this case, the seven crossing product is not the knot 7_3 or its mirror image by theorem 8. Unfortunately, theorem 8 does not give any information regarding the other seven crossing knots.

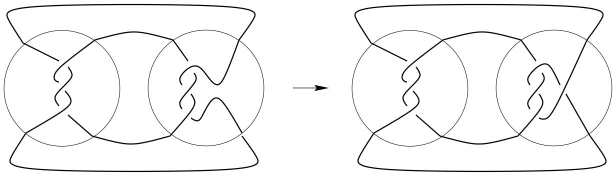


Fig. 25. $N(-1/3 + -1/3) \to N(-1/3 + -4/3) = 7_4^*$

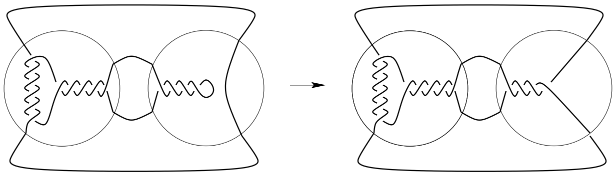


Fig. 26. $N(-31/6 + 1/0) \to N(-31/6 + 4) = 7_1^*$

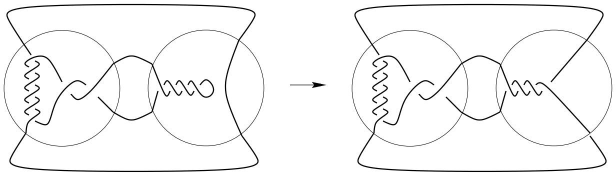


Fig. 27. $N(-13/6 + 1/0) \to N(-13/6 + 4) = 7_2$

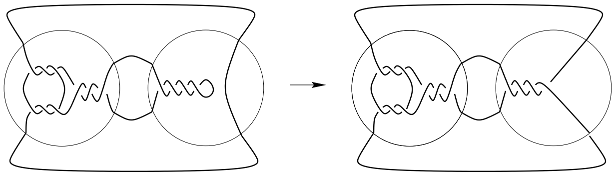


Fig. 28. $N((-1/3+-1/3)\circ(2,1,-1)+1/0)\rightarrow N((-1/3+-1/3)\circ(2,1,-1)+4)=7_4^*$

If U_{f_2} is ambient isotopic to a sum of rational tangles, then the only possible topological solutions assuming models 23 and 24 are drawn in figures 25 and 26-28, respectively. If the recombinase sites are used to orient the links N(4/1) and N(6/1), then the linking number of N(4/1) is -2 [23] and the linking number of N(6/1) is -3 [24]. Note that for the model shown in figures 24 and 26, the orientation of the recombinase sites when the substrate is unknotted is different than the orientation of the recombinase sites when the substrate is N(6/1), whereas the orientations agree for both substrates in figures 25, 27, and 28. If U_{f_2} is ambient isotopic to a sum of rational tangles and the orientation of the sites agree, then 7_1^* is not a possible product from $N(6/1)$ irregardless of the model proposed. The tangle U_f is also much simpler in model 25. Thus the figure 23 model is most likely biologically more reasonable than the figure 24 model for Xer recombination.

Note however, that the solutions drawn are topological solution. The geometrical solutions drawn in figure 29 are topologically equivalent to the solution drawn in figure 23 and thus are also possible models for Xer recombination (observe, though, that if the double helical nature of DNA is taken into consideration, ΔLk may be different in figures 23, 29 depending on assumptions regarding ΔTw). The top drawing in figure 29 is the same as figure 2b in [24].

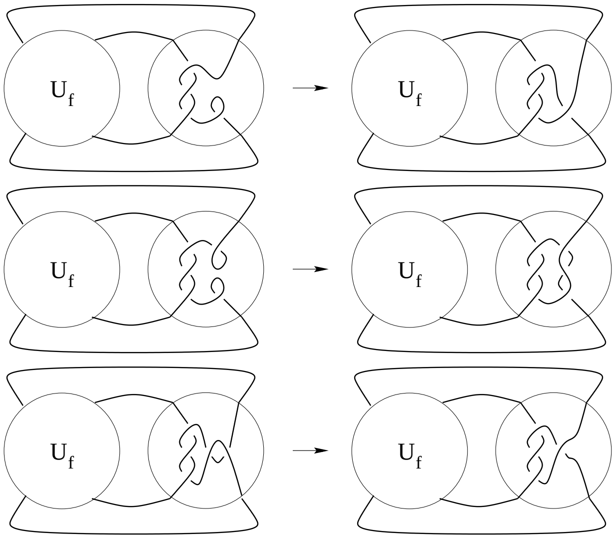


Fig. 29. Topologically equivalent to figure 23.

Proposed experiment 2. Take a picture of the synaptic complex using both the unknot and the link N(6/1) as substrate.

If U_{f_1} = 0-tangle, then by eq. (8.1), $B = 1/n$ tangle. If U_{f_2} = -1/3 tangle, then by equation (8.2), $B = \frac{-1+6k}{3+18k}$ for some k. If both equations hold with these U_{f_i}'s, then $B = \frac{-1}{3}$ is the unique solution.

If U_{f_1} = 0-tangle and if E is a rational tangle, then by equation 8, $E = \frac{4}{1+4k}$ tangle. Suppose U_{f_2} = -1/3 tangle. If both equations hold with these U_{f_i}'s, then $E = \frac{-4}{3}$ is the only rational tangle solution (figure 23). E cannot be a Montesinos tangle but can be ambient isotopic to a Montesinos tangle or some other prime tangle.

However, an electron micrograph showing U_{f_2} as a -1/3 tangle would look exactly like an electron micrograph showing U_f as a 3 tangle. Thus U_{f_2} = 3 tangle must also be considered. But by the solutions to eq. (6.10) and (6.11), if U_{f_1} = 0-tangle, then $U_{f_2} = 1/j$, $\frac{31}{-6+31j}$, or $\frac{13}{-6+13j}$. Thus $U_{f_2} \neq 3$ tangle.

Therefore, if we assume E is a sum of rational tangles and if is determined that U_{f_1} = 0-tangle and U_{f_2} = -1/3 tangle, then B must be the -1/3 tangle and E must be the -4/3 tangle.

Proposed experiment 3. Take a picture of the synaptic complex using both the unknot and a different knot or link as substrate.

Rather than redoing the Xer reaction using N(6/1) as substrate, more information may be gained by using another knot or link as substrate. If it is assumed that U_f is a rational tangle and that the model shown in figure 23 is correct, then it is easy to predict products of Xer recombination and to predict what U_f should look like. Rational tangles are often simpler than non-rational tangles, and in biology often the simpler model is the correct model.

If $U_f = \frac{x}{y}$ and the substrate is $N(a/b)$, then $N(\frac{x}{y} + \frac{-1}{3}) = N(a/b)$ and $\frac{x}{y} = \frac{v}{3v-a}$ where $bv^{\pm 1} = 1 \bmod a$. The product of recombination is $N(\frac{v}{3v-a} + \frac{-4}{3}) = N(\frac{-9v+4a}{-2v+a})$ The easiest knots/links to make in the lab are torus knots/links and negative twist knots.

If the torus knot/link N(a/1) is the substrate, then one simple U_f is $\frac{1}{3-a}$ with an expected product of $N(\frac{-9+4a}{-2+a})$. In particular if the five crossing torus knot, N(5/1) is the substrate of Xer recombination, then U_f is expected to be the $\frac{-1}{2}$ tangle and the product is expected to be a six crossing knot if the substrate contains inverted repeats. Note U_f cannot be the +2 tangle if U_{f_1} is the zero tangle. If both $U_f = \frac{-1}{2}$ and a six crossing knotted product is confirmed by biological experiments, then $B = -1/3$ and E is either a prime tangles which is not ambient isotopic to the sum of rational tangles or E must be the -4/3 tangle, a slightly stronger conclusion than that for the N(6/1) substrate.

If the twist knot, $N(\frac{a}{2})$, a odd with crossing number $\frac{a+3}{2}$ is the substrate, then one simple U_f is $\frac{2}{6-a}$ with an expected product of $N(\frac{4a-18}{a-4}) = (2, \frac{a-7}{2}, 1, 3)$ which is a link with one additional crossing if the crossing number of the negative twist knot is greater than four. If the substrate is the four crossing twist knot, an expected product is the two crossing link, N(2/1).

In particular, if the five crossing twist knot, N(7/2), is the substrate, then U_f is expected to be the -2 tangle with N(10/3), a six crossing link, as the expected product. If it is experimentally determined that U_f is the -2 tangle, then E must be either a rational tangle or ambient isotopic to a Montesinos tangle by theorems 5 and 6. However, an electron micrograph showing U_f as a -2 tangle, would look exactly like an electron micrograph showing U_f as a 1/2 tangle. Since it is mathematically unknown whether or not $d_{\frac{1}{2}}(0_1, 5_2) = 1$, the electron micrograph could be interpreted either way. However $d_{\frac{1}{2}}(N(4/1), N(10/3) \neq 1$. Thus if U_{f_1} is the zero tangle, U_f could not be the 1/2 tangle in the equations $N(U_{f_1} + B) = N(4/1)$ and $N(U_f + E) = N(10/3)$. The solutions in this case are B = -1/3 or 1/4 and E = -4/3 and both B and E must be rational.

Although it is not possible to mathematically determine P and R, these additional experiments could allow the exact determination of the topology of B and E. However, the experiments proposed are neither simple nor quick, but the list of possible models can be analyzed so that biologists can decide the relevance of an experiment.

Acknowledgments

The author would like to thank Jodi Schneider, Jonathan Bath, Sean Colloms, and David Sherratt for interesting discussions.

References

[1] D. W. Sumners, C. Ernst, S. Spengler, and N. Cozzarelli, *Analysis of the mechanisms of DNA recombination using tangles*, Quarterly Review of Biophysics 28 (1995) 253–313.

[2] A. Stasiak, V. Katritch, J. Bednar, D. Michoud, J. Dubochet, *Electrophoretic mobility of DNA knots*, Nature 384 (1996), no. 6605, 142–145.

[3] N. Crisona, R. Kanaar, T. Gonzales, E. Zechiedrich, A. Klippel, N. Cozzarelli, *Processive recombination by wild-type Gin and an enhancer-independent mutant*, J. Mol. Biol. 243 (1994), 437 - 457.

[4] S. Levene, H. Tsen, *Analysis of DNA knots and catenanes by agarose-gel electrophoresis*, preprint.

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

[6] K. Murasugi, *Knot Theory and Its Applications*, translated from the 1993 Japanese original by Bohdan Kurpita. Birkhuser Boston, Inc., Boston, MA, 1996

[7] J. R. Goldman, L. H. Kauffman, *Rational tangles*, Adv. in Appl. Math. 18 (1997), no. 3, 300–332.

[8] S. Bleiler, *Knots prime on many strings*, Trans. Amer. Math. Soc. 282 (1984), no. 1, 385–401.

[9] W. B. R. Lickorish, *Prime knots and tangles*, Trans. Amer. Math. Soc. 267 (1981), no. 1, 321–332.

[10] C. Ernst, D. W. Sumners, *Solving tangles equations arising in a DNA recombination model*, Math. Proc. Camb. Phil. Soc. 124 (1998).

[11] C. Ernst, *Tangle equations*, J. Knot Theory Ramifications 5 (1996), no. 2, 145–159.

[12] J. Roberts, *Elementary number theory, a problem oriented approach*, MIT Press, Cambridge, MA, 1977.

[13] T. Kanenobu, H. Murakami, *Two-bridge knots with unknotting number one*, Proc. Amer. Math. Soc. 98 (1986), 499–502.

[14] P. Kohn, *Two-bridge links with unlinking number one*, Proc. Amer. Math. Soc. 113 (1991), 1135–1147.

[15] I. Torisu, *The determination of the pairs of two-bridge knots or links with Gordian distance one*, Proc. Amer. Math. Soc., 126 (1998), no. 5, 1565–1571.

[16] I. Darcy, D. W. Sumners, *A strand passage metric for topoisomerase action,* in Knots '96, World Scientific, (1997), 267–278.

[17] I. Darcy, D. W. Sumners, *Rational tangle distances on knots and links*, Math. Proc. Cambridge Philos. Soc., 128 (2000), 497–510.

[18] I. Darcy, *Solving oriented tangle equations involving 4-plats*, preprint.

[19] M. Hirasawa, K, Shimokawa, *Dehn surgeries on strongly invertible knots which yield lens spaces*, Proc. Amer. Math. Soc., posted on May 18, 2000, PII S 0002-9939(00)05417-4 (to appear in print).

[20] P. Kohn, *Unlinking two component links*, Osaka J. Math. 30 (1991), 741–752.

[21] H. Murakami, *Some metrics on classical knots*, Math. Ann. 270 (1985), 35–45.

[22] I. Darcy, *Biological Metrics of DNA Knots and Catenanes*, Ph.D. thesis. Florida State University (1997).

[23] S. Colloms, J. Bath, D. Sherratt, *Topological selectivity in Xer site-specific recombination* Cell 88 (1997), 855–864.

[24] J. Bath, D. Sherratt, S. Colloms, *Topology of Xer recombination on catenanes produced by Lambda Integrase*, J. Mol. Biol. 289 (1999) 873-883.

[25] T. C. Hodgman, H. Griffiths, D. K. Summers, *Nucleoprotein architecture and ColE1 dimer resolution: a hypothesis*, Molecular Microbiology 29(2) (1998), 545-58.