The Tangle Model for Site-specific Recombination: A Computer Interface and the TnpI-IRS Recombination System

Wenjing ZHENG, Christine GALLOY, Bernard HALLET and Mariel VAZQUEZ

ABSTRACT

The tangle method was first introduced by Ernst and Sumners in 1990. Since then the field has seen many advances which have been reviewed in [18,24] and more recently in a soon to appear review by Darcy. Rather than providing an in-depth review we here focus on the tools developed and used by our group. In particular we introduce TangleSolve, a computer implementation of the tangle method. We also present a novel visualization approach where we propose that solutions be classified modulo rigid spatial motions. We illustrate our recent work with the TnpI-IRS site-specific recombination system.

Mathematical Subject Classification 2000: 92-02, 92-04, 57M25, 57M10
Keywords: DNA topology, tangles, site-specific recombination, DNA knots

1. Introduction

Site-specific recombination is a biological process that introduces changes in the genetic material of an organism by insertion or excision of a foreign DNA into or out of the host genome, or by inversion or transposition of a DNA segment within a DNA molecule. The enzymes responsible for site-specific recombination are called *site-specific recombinases*. The short segment of specific sequence that a *recombination complex* (which may consist of one or more site-specific recombinases) recognizes is called a *recombination site*. Site-specific recombination is a two-step process. First, the enzymes recognize a pair of recombination sites (on the same or different DNA molecules), which are then aligned through manipulation of the global topology of the DNA molecule(s), and bound by the enzymes. This stage is known as *synapsis*, the enzyme and its bounded DNA are called the *synaptosome*, and the entire DNA molecule (bounded or unbounded) together with the enzyme is known as the *synaptic complex*. In the second step, the recombination sites are cleaved and rejoined with the ends exchanged (fig. 1).

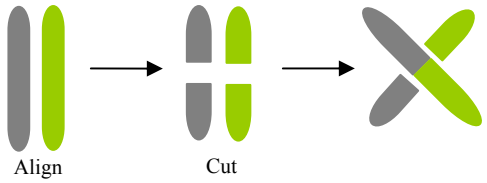


Figure 1: Local action of recombination

The DNA molecules prior to recombination form the *substrate,* and the DNA molecules after recombination form the *product*. We will mainly deal with circular substrates with two copies of the recombination site for a given enzyme. Typically the sites are not palindromic and therefore an orientation can be assigned to each site. Each site orientation induces an orientation of the whole DNA molecule. If these orientations agree, the sites are *directly repeated* (fig. 2-a). Otherwise, they are *inversely repeated* (fig. 2-b).

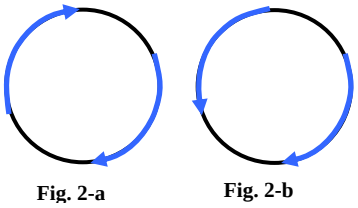


Figure 2: **a)** Directly repeated sites; **b)** inversely repeated sites.

Our goal is to use knot theory and low-dimensional topology to understand the changes in the global conformation of the DNA molecule(s) introduced by the enzyme during the first stages of recombination, and to unveil topological changes inside the synaptosome during recombination. To this end, circular substrates are often used because they best reflect the topological changes introduced on the substrate. The tangle method uses these changes to understand the action of the enzyme on the synaptic complex and the strand-exchange mechanism [24]. This method has been successfully used in the analysis of different site-specific recombination reactions [4,5,8,13,25,26].

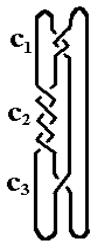
We first introduce the tangle method proposed by Ernst and Sumners in [10] (section 2). In section 3 we present TangleSolve, a computer implementation of the tangle method. In section 4 we use the method to analyze the experimental results on the TnpI-IRS recombination system of Tn4430, a transposon from *Bacillus thuringiensis* ([27]).

2. The Tangle Method

2.1 4-plats and tangles

A *4-plat* is a knot or link that can be obtained by braiding 4 strings and capping off the ends as illustrated in figure 3 ([6]). 4-plats admit a classification which is related to that of lens spaces. To each 4-plat can be associated a classifying vector $<c_1, c_2, \ldots, c_{2k+1}>$ (*Conway vector)*, such that $c_i>0$ for all i. Two 4-plats $<c_1, c_2, \ldots, c_{2k+1}>$ and $<d_1, d_2, \ldots, d_{2m+1}>$ are equivalent if and only if

$$\langle c_1, c_2, \ldots, c_{2k+1}\rangle = \langle d_1, d_2, \ldots, d_{2m+1}\rangle \quad \text{or} \quad \langle c_1, c_2, \ldots, c_{2k+1}\rangle = \langle d_{2m+1}, \ldots, d_2, d_1\rangle .$$



$$<c_1, c_2, c_3> = <2, 4, 1>$$

$$= <c_3, c_2, c_1> = <1, 4, 2>$$

Figure 3. Canonical form of 4-plats. If we start the braiding from top, the resulting vector is <2, 4, 1>. If we start from bottom, the resulting vector is <1, 4, 2>.

From the Conway vector, we can obtain a classifying rational number for each 4-plat through a continued fraction:

$$\frac{\beta}{\alpha} = \cfrac{1}{c_1 + \cfrac{1}{c_2 + \ldots \cfrac{1}{c_{2k+1}}}}, \text{ where } 0 < \beta < \alpha.$$

Hence, we can also denote a 4-plat by *b(α, β)*, its *Conway Symbol*. Furthermore, by the Classification Theorem of 4-plats ([20]), *b(α, β)* and *b(α',β')* are equivalent non-oriented links if and only if $\alpha = \alpha'$ and $\beta^{\pm 1} \equiv \beta' \pmod{\alpha}$. A simple calculation shows that the extended rational numbers associated to $<c_1, c_2, \ldots, c_{2k+1}>$ and $<c_{2k+1}, \ldots, c_2, c_1>$ are related by this congruence.

A *2-string tangle* is an ordered pair *(B,t)*, where *B* is a fixed oriented 3-ball in S^3 and *t* is a pair of non-oriented disjoint arcs properly embedded in *B*, whose endpoints lie on the bounding sphere. For each *(B,t)* there is an orientation

preserving homeomorphism

$$\Phi : (B,t) \to (D,t_\Phi)$$

that maps *B* onto the unit 3-ball *D*, and *t* onto t_Φ (two straight arcs in D connecting the preferred equatorial points NE with SE, and NW with SW). The endpoints of *t* map to the 4 special equatorial points {NW, NE, SE, SW}. Notice that in order to compare tangles defined in different 3-balls, we shall define a 2-string tangle more precisely as the triple *(B, t, Φ)*. This way we may consider, without loss of generality, all tangles as defined on the unit 3-ball D with strings attached to the 4 special equatorial points. Two tangles (D, t_1) and (D, t_2) are equivalent if there is an ambient isotopy that takes t_1 to t_2 while fixing the endpoints.

A *tangle diagram* is a planar representation obtained by considering a regular projection of the 3D tangle onto a plane (usually the equatorial plane) and by keeping track of under-and over crossing information (figure 4). Two tangle diagrams represent the same tangle if they are equivalent up to a finite sequence of Reidemeister moves.

There are three different types of tangles: rational, locally knotted and prime.

Fig. 4-a **Fig. 4-b** **Fig. 4-c** **Fig. 4-d** **Fig. 4-e**

Figure 4: Planar projections of tangles.
a) infinity tangle; **b)** trivial tangle; **c)** a rational tangle; **d)** a locally knotted tangle; **e)** a prime tangle.

A tangle is *rational* if the strings can be continuously deformed into the boundary 2-sphere of the 3-ball (fig. 4-a, 4-b, 4-c). Rational tangles are of special importance in tangle analysis of site-specific recombination, and we will elaborate on them later in this section. A tangle *(B, t)* is *locally knotted* if there is a 2-sphere *S* inside the 3-ball B that intersects one of the two arcs of *t* transversely in two points, and such that the 3-ball bounded by *S* holds *t* as a knotted arc with endpoints on *S* (as in fig. 4-d). If a tangle is neither rational nor locally knotted, then it is called a *prime tangle* (fig. 4-e). Any 2-string tangle must belong to exactly one of the above three categories.

John Conway established that there is a bijection between equivalence classes of rational tangles and the extended rational number ([7]). Conway's

classification proof relies on properties of the double branched cyclic covering spaces of the tangles. A simpler proof which uses bracket polynomials was given in [15].

Each equivalence class of rational tangles can be represented by its *Conway symbol*, a classifying vector of integer entries $(a_1, \ldots, a_n)$ such that: $|a_1| > 1$, all entries are nonzero except possibly for a_n, and all entries have the same sign. This scheme applies to all but four exceptional rational tangles, as in figure 5.

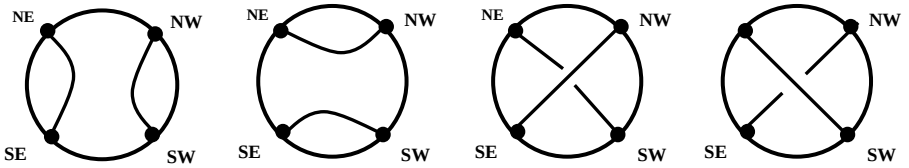


Fig. 5-a Fig. 5-b Fig. 5-c Fig. 5-d

Figure 5: The four exceptional rational tangles.
a) infinity tangle (0,0); **b)** trivial tangle (0); **c)** tangle (+1); **d)** tangle (-1)**.**

Again, starting with the Conway vector $(a_1, \ldots, a_n)$ and using a continued fraction we can obtain an unique extended rational number that identifies the tangle:

$$\frac{q}{p} = a_n + \cfrac{1}{a_{n-1} + \cfrac{1}{a_{n-2} + \cfrac{1}{\ldots a_2 + \cfrac{1}{a_1}}}},$$

where $\frac{q}{p} \in \mathrm{Q} \cup \left\{\frac{1}{0}\right\}$, $q \in N \cup \{0\}$, $q \in \mathrm{Z}$ and $gcd(p, q) = 1$.

Several operations can be defined on all types of tangles: tangle addition (figure 6-a), the numerator closure (figure 6-b) and the denominator closure (figure 6-c). We can also combine these operations by first adding two tangles and then taking their numerator or denominator (figure 6-d). The numerator and denominator operations convert tangles into knots and links. In particular, the numerator operation relates rational tangles with the family of 4-plats. If A is a rational tangle, then the result of taking its numerator is a 4-plat $N(A)$. And if A and B are rational tangles, then $N(A + B)$ is a 4-plat ([6]), even though $A + B$ need not be rational.

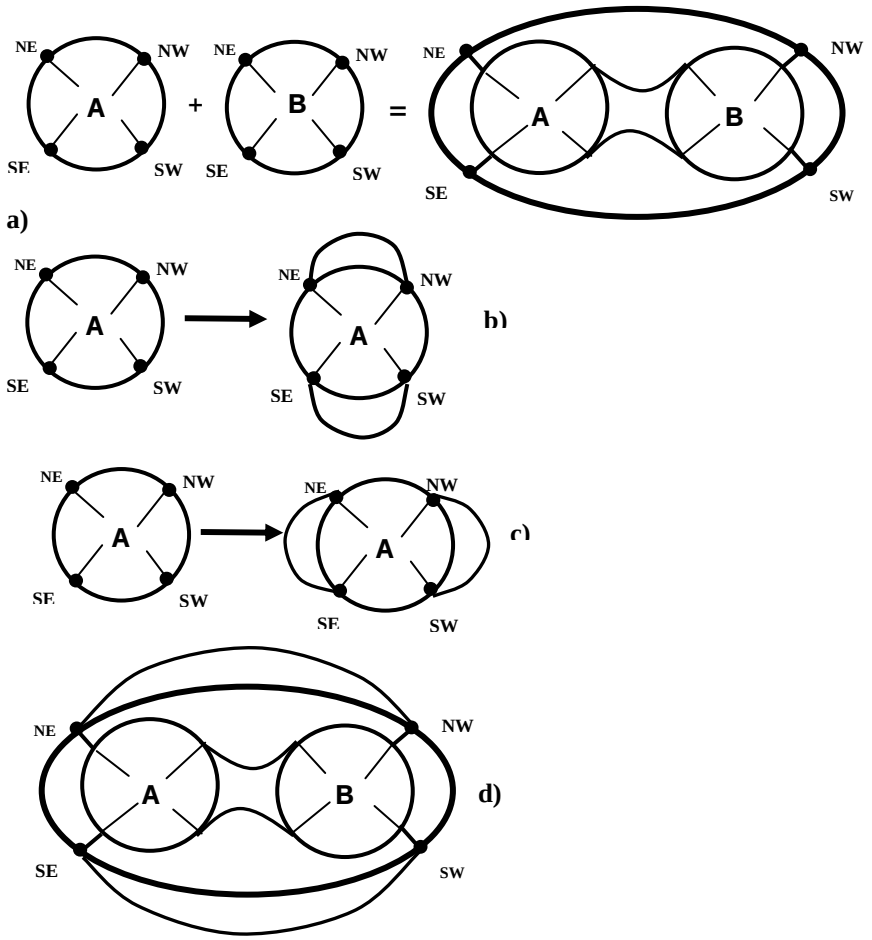


Figure 6: Arithmetic operations on tangles.
a) Tangle addition; **b)** Numerator closure; **c)** Denominator closure;
d) Numerator closure of a sum of tangles.

2.2. Tangle Method

The tangle method was proposed by Ernst and Sumners in [10], and is reviewed in [18,24]. The method was motivated by the idea that one can model the synaptosome consisting of the enzyme and a pair of recombination sites as a 2-string tangle, and that most of the recombination products are 4-plats. This method uses tangles to model the changes in topology of the synaptic complex before and after recombination, based on the known substrate and product topologies, and on a few justified assumptions. The first biological assumption is that the enzymatic mechanism is constant and independent of the geometry and topology of the substrate population, and that changes in the DNA molecule due to the enzymatic action only take place inside the synaptosome. This means that for any type of substrate, the

pre-recombination synaptosome is always the same, and difference in substrate type is only reflected in the DNA topology outside of the synaptosome. Also, the DNA outside the synaptosome is to remain unchanged throughout the recombination event. The second assumption is that the 2-string tangle that models the synaptosome can be divided into the sum of two tangles P and O_b. The P tangle represents the region inside the synaptosome that contains the recombination sites, and thus whose topology can be modified by the recombination event. The O_b tangle represents the remaining region inside the synaptosome that stays constant during recombination. The local enzymatic action of breaking and rejoining the ends of the sites is modeled as tangle surgery in which a P tangle is replaced by an R tangle. Under these assumptions, we have that:

Pre-recombination synaptosome = $O_b + P$
Post-recombination synaptosome = $O_b + R$

By the biological assumptions mentioned above, for a given enzyme both the pre- and post-recombination synaptosomes are always the same, regardless of the substrate's knot type, so the tangles $\{O_b, P, R\}$ are enzyme-specific constants. The topology of the substrate's knot type is reflected in the DNA outside of the synaptosome; let a tangle O_f denote this portion, which is unchanged throughout the recombination. Therefore, knowing the knot types of the substrate and product for a single recombination event gives rise to two tangle equations in 4 variables $\{O_f, O_b, P, R\}$:

$N(O_f + O_b + P)$ = substrate
$N(O_f + O_b + R)$ = product

The tangles on the left hand side of the equations represent the DNA in the synaptic complex, and the knots on the right hand side represent DNA free of enzyme. The first equation is before recombination, and the second is after. The strategy is to compute the $\{O_f, O_b, P, R\}$ from the substrate and product knot types. But since O_f and O_b represent regions of DNA unchanged during recombination, let $O = O_f + O_b$ model these unchanged regions. Thus, we can instead consider the following system of equations, and solve for $\{O, P, R\}$:

$N(O + P)$ = substrate
$N(O + R)$ = product

Figure 7 shows an example of tangle analysis of the mechanism of Tn3 resolvase [10, 21]. It illustrates Tn3 acting on two very different types of

substrates: the unknot (left) and the right-handed 4-crossing torus link (right). Based on the assumption of constant enzyme mechanism, the tangles $\{O_b, P, R\}$ are constants for both types of substrates. These tangles reflect the global action (binding mechanism) and the local action (tangle surgery changing P into R) of the enzyme. Only O_f differs in the two events to reflect the differences in substrate types.

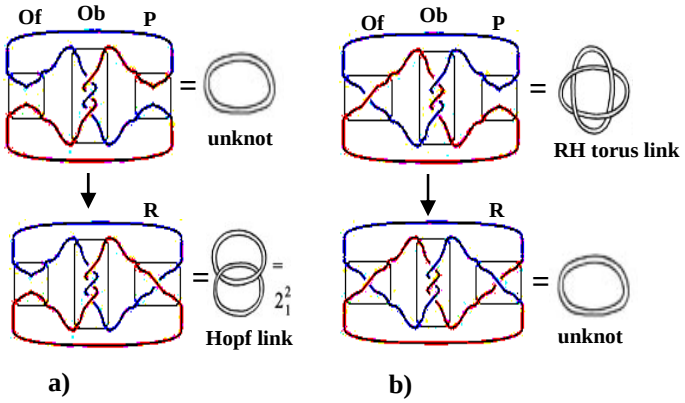


Figure 7: Action of Tn3 Resolvase on two different substrate types.

a) substrate is the unknot and product the hopf link.

b) substrate is the right-handed torus link and product the unknot**.**

Note that $\{O_b, P, R\}$ are constant, while O_f reflects the difference in substrates. The enzyme action is modeled as replacing tangle P with R.

3. TangleSolve

Using the tangle method each site-specific recombination reaction, where a substrate of specific topology is converted into a product of specific topology, can be rewritten as a system of two tangle equations. Certain enzymes perform two or more rounds of recombination in a single reaction (*processive recombination*). In that case, if the substrate and products of recombination are all 4-plats, then the processive recombination event is translated into a system of two or more tangle equations, as follows:

$$N(O+P)=b(a_0,b_0)$$
$$N(O+R)= b(a_1,b_1)$$
$$\vdots$$
$$\vdots$$
$$N(O+nR)= b(a_n,b_n)$$

Here $b(a_0,b_0)$ is the substrate, and $b(a_i,b_i)$, for $i=1,\ldots,n$ are the 4-plat products.

Systems of tangle equations corresponding to processive and non-processive recombination have been studied extensively, and in many cases all possible solutions have been characterized ([4,5,8,10-13,24-26]). Computing the solutions is usually not mathematically challenging but can be very tedious. In Saka and Vazquez ([19]) we introduced TangleSolve, a user-friendly computer implementation of the tangle method. TangleSolve is a java stand-alone program and web-based applet which offers a user-friendly interface for analyzing and visualizing recombination mechanisms. TangleSolve can be accessed from http://bio.math.berkeley.edu/TangleSolve/.

TopoICE-R is another computer implementation of the tangle model which is available through KnotPlot [9]. TangleSolve and TopoICE-R have complementary features and can be used by both mathematicians, biologists and computational biologists interested in the tangle method.

TangleSolve is illustrated in figure 8. There are three panes: input, selection and display. In the first pane the user can select between three tabs: Tangle Diagram, Processive and non-processive.

The first tab illustrates a recombination event where the user specifies the tangles *O* and *R* involved in the equations. *O* is presented as a rational tangle, or a sum of rational tangles. *O* and *R* are inserted in Conway form. TangleSolve illustrates *N(O)* and *N(O+R)* by showing both the tangles, and the resulting 4-plat, and gives their corresponding Conway vectors for the 4-plats.

For the second and third tabs the user inputs substrate and product(s) of recombination, and TangleSolve computes all solutions (that are rational or sums of rational tangles) to the corresponding system(s) of tangle equations. Input forms accepted include the Conway vector (e.g. *<1,2,1>*), Conway notation (e.g. *b(4,1)*), number of crossings (e.g. 4), or knot diagram which can be selected from the embedded knot table (provided by D. Rolfsen). TangleSolve assumes *P=(0)* and *R=(k)* for any integer *k*. This assumption is based on the fact that we chose the *P* tangle to be small enough to contain only the core regions of the recombination sites, and therefore *P* is trivial, and Likewise, *R* is expected to be very simple, to just reflect the local action on the recombination sites. If furthermore the sites inside *P* are assumed to be in parallel alignment then the only plausible recombinant tangles are integral. In the case where *P=(0)* and *R=(±1)*, TangleSolve has a feature that will provide equivalent solutions where *P=(0,0)* and *R=(0)*, obtained by rotating the tangles in space (see figure 8, top right window).

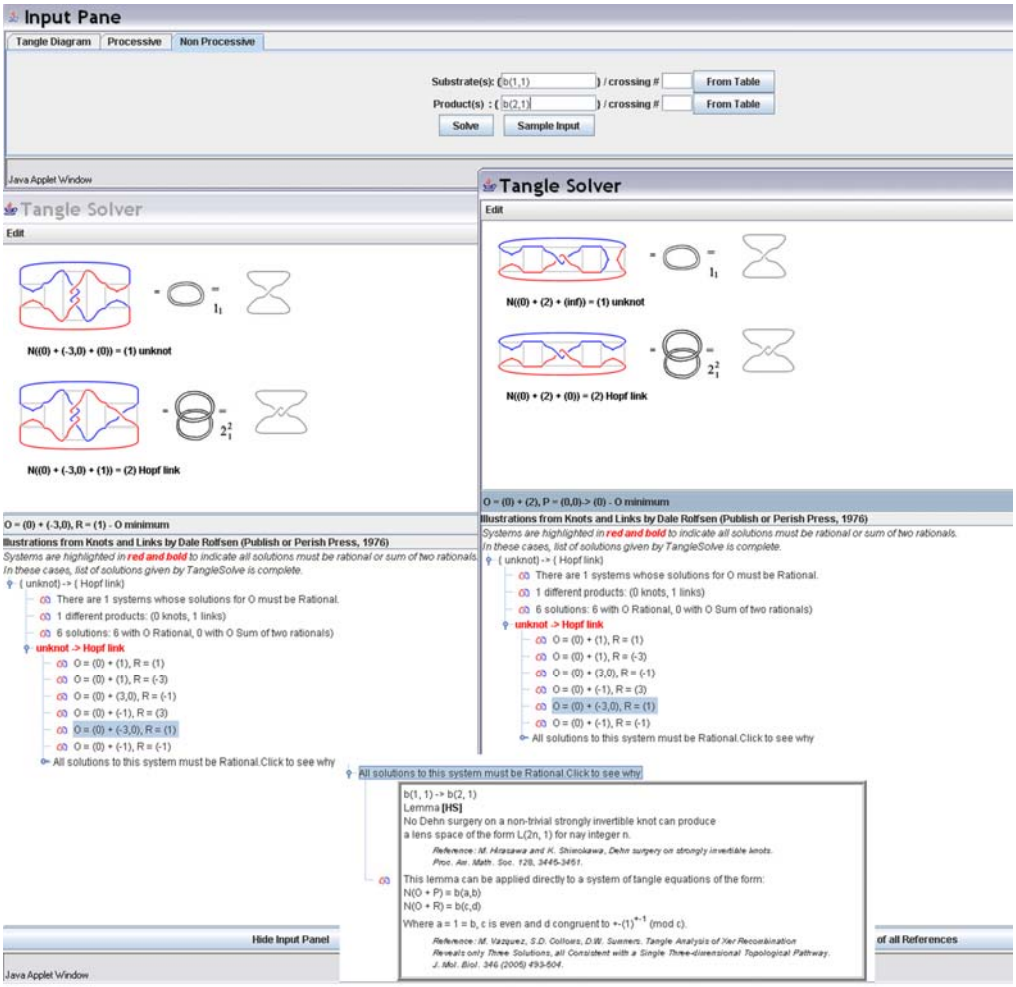


Figure 8: Snapshot of TangleSolve. We here illustrate the software TangleSolve (http://bio.math.berkeley.edu/TangleSolve) used to solve TnpI-IRS system of tangle equations.

The second tab corresponds to processive recombination. The user can input substrate and products of recombination. The products can be listed in order of appearance, if the order is known, and otherwise as an unordered list. In the unordered case the software will separately output the number of solutions to the systems of tangle equations involving 1, 2,…, *n* products. This is useful as in some cases the observed products may come from 2 or more separate reactions, and may not be accounted for by a single recombination event.

The third tab deals with non-processive recombination. In this case only one substrate and one product are specified. If the exact knot types are not known the user may specify the crossing number, in which case the software will solve all systems consistent with the input.

TangleSolve computes only solutions that are rational or sum of rational tangles. Using tools from low-dimensional topology it can sometimes be proven that these are the only possible solutions to the given system. When one such result is available TangleSolve highlights the header for the solutions in red, and by clicking on it one can read the corresponding theorem and its source (see figure 8, bottom right window). This feature is updated in a regular basis.

4. Tn4430 from *Bacillus thuringiensis* and its site-specific TnpI-IRS recombination system

4.1. The TnpI-IRS recombination system

In vitro experiments on the site-specific TnpI-IRS recombination system encoded in transposon Tn4430 showed that recombination at directly repeated IRS sites on a circular unknotted DNA substrate yields products with specific topology ([27]). We use tangles to analyze the topological changes during recombination. The synaptic complex is modeled as a 2-string tangle and the recombination event as tangle surgery. We compute all solutions to the tangle equations arising from this model. Furthermore we propose a 3-dimensional model for the recombination system by integrating solutions to the tangle equations that are consistent with the biological assumptions of the tangle method and with existing knowledge of the TnpI protein.

Bacillus thuringiensis is a bacterium that produces specific toxins that are lethal to a variety of insect species, but inoffensive to most other organisms Therefore, *B. thuringiensis* and its toxin crystals are used in organic farming to protect crops from harmful moths and butterflies. Transposons are segments of DNA that can move to different regions (transposition) within the genome of a single cell. Tn4430 is a transposon from *B. thuringiensis* that uses a replicative mode of transposition. During this process, co-integrate intermediates between the donor and target replicons are generated ([17]). Tn4430 encodes the TnpI protein, a member of the tyrosine site-specific recombinase family that catalyzes the site-specific recombination reaction used to resolve these co-integrate intermediates. Each Internal Resolution Site (IRS) of Tn4430 contains palindromic recombination core sites (IR1, IR2), where cleavage by Tnpl takes place, and two additional accessory motifs (DR1, DR2). In [27],Vanhooff *et. al*. showed experimentally that the accessory motifs are not necessary to promote efficient recombination *in-vivo* or *in-vitro*, but *in-vivo* they contribute to prevent intermolecular recombination events by limiting recombination to sites within the same molecule. *In-vitro* recombination reactions on circular DNA substrates with

two directed repeats of the full IRS sites (containing both the core sites and the accessory motifs) yield products with specific topology: all products are two-node links (Hopf Links). Results from recombination reactions occurring on knotted DNA molecules containing inverted IRS sites suggest that only 1 negative supercoil is trapped by the accessory motifs during synapse assembly ([27]). Unifying all these experimental results, Vanhooff *et al.* proposed a model for the TnpI-IRS synaptic complex where the two negative supercoils are trapped during the synapsis, giving way to the formation of the Hopf link product ([27], figure 10**a)**).

We apply the tangle method to analyze the results of above experiment, where TnpI-IRS recombination acting on unknotted circular substrate with directed repeats yields the Hopf link product. The corresponding system of tangle equations is:

Substrate:	$N(O + P) = b(1,1)$	the unknot	*(*)*
Product :	$N(O + R) = b(2,1)$	the Hopf link	

Here the *O* tangle is assumed unchanged by the enzyme throughout the recombination event, the *P* tangle is restricted to the recombinations sites (*i.e.* the DNA that is changed by recombination), and the local enzymatic action is modeled by replacing *P* with *R*. The goal of the model is to calculate the tangles *{O, P, R}*.

4.2 Solutions to the system (*) are rational

Before computing the tangles *{O, P, R}*, we should have an understanding of what type of tangles *{O, P, R}* are. To illustrate this process we follow for TnpI-IRS a similar approach to thoat taken in the analysis of Xer recombination [8, 26]. First, we need the following results:

Theorem 1([16]): Let $O = (B, t)$ be a tangle and O' be its double branched cyclic cover, then:
(i) *O* is rational ⇔ O' is a solid torus
(ii) *O* is prime ⇔ *O*' is irreducible and has incompressible boundary
(iii) *O* is locally knotted ⇔ O' is not irreducible.

Theorem 2([6]): If K_1 is the unknot and $K_2 = b(\alpha, \beta)$ is a 4-plat, then the double-branched cyclic cover K_1 is the 3-sphere, and that of K_2 is the lens space $L(\alpha, \beta)$.

Theorem 3 ([3,10,16]): Let *A* and *B* be locally unknotted tangles and *K* be a 4-plat such that *N(A+B)=K*. Then at least one of *A* or *B* must be a rational tangle.

Definition: A knot *K* in S^3 is *strongly invertible* if there is an orientation-preserving involution of S^3 that preserves *K* as a set and reverses the orientation of *K*.

Theorem 4([2]): The lens space *L(2,1)* cannot be produced by Dehn Surgeries on nontrivial strongly invertible knots.

Theorem 6: In the system of tangle equations arising from TnpI-IRS recombination:

Substrate: $N(O + P) = b(1,1)$ the unknot
Product : $N(O + R) = b(2,1)$ the Hopf link

the tangles *P* and *R* are locally unknotted, and *O* is rational.

Proof:
The tangles *O* and *P* are summands of the unknot in the proposed system of tangle equations. If any of them had a local non-trivial knot *T* then the sum and numerator operation would preserve this local knot, and *N(O + P)* would result in either *T* or a composite knot *T#K*, for some knot *K*. This would in turn contradict the first equation *N(O + P)=b(1,1)*. So *O* and *P* are locally unknotted. Similarly, if *R* had a local knot, then it would be preserved by the sum and numerator operations, and this would contradicts that *N(O + R)=b(2,1)*, which has no local knots. Therefore the tangles *{O, P, R}* are locally unknotted, that is, each of them is either a prime tangle or a rational tangle. Now it remains to show that *O* is rational. For this, it suffices to prove that *O* is not a prime tangle.
Let *O'*, *P'*, and *R'* be the double-branched cyclic cover of *O*, *P*, and *R*, respectively. Suppose *O* is a prime tangle. Then, since *N(O + P)* is a 4-plat, and *O* and *P* are locally unknotted, it follows from theorem 3 that *P* must be rational. When lifting the tangles in the first equation to their double-branched cyclic covers, the tangle sum and numerator operation is translated into the union of *O'* and *P'* along their common boundary. So we have that:

$$O' \cup P' \cong S^3 \quad \ldots (1)$$

where S^3 is the double-branched cyclic cover of the unknot (theorem 2). Since we concluded earlier that *P* is rational, then by theorem 1, *P'* is a solid torus, and therefore the tubular neighborhood of a knot *K* (not necessarily trivial) in S^3.

From equation (1), we know that *O'* is the knot exterior of *K* in S^3. Next, by theorem 2, the double-branched cyclic cover of the Hopf link *b(2,1)* is the lens space *L(2,1)*. So, the second equation of the system lifts to:

$O' \cup R' \cong L(2,1)$ … (2).

Since *O'* is the knot exterior of *K* in S^3, it follows from equations (1) and (2) that *L(2,1)* is obtained by Dehn surgery on the knot *K*. However, the covering transformation *P'* is an orientation-preserving involution that maps *K* onto itself but with reversed orientation (recall that *P'* is the regular neighbourhood of *K*); thus *K* is a strongly invertible knot. In other words, *L(2,1)* is obtained by Dehn surgery on a strongly invertible knot *K*. Then by theorem 4, *K* must be the trivial knot. This implies *P'* is an unknotted solid torus and $O' = S^3 - P'$ is also an unknotted solid torus, then by theorem 1, *O* is a rational tangle. This conclusion contradicts the initial assumption that *O* is a prime tangle, and since *O* is locally unknotted, then *O* must be a rational tangle. □

We have shown that for the system of tangle equations:

Substrate:	$N(O + P) = b(1,1)$	the unknot
Product :	$N(O + R) = b(2,1)$	the Hopf link

The tangles *P* and *R* are locally unknotted, and *O* is rational.

Now, we will make some biologically reasonable assumptions on the tangles *P* and *R*. These assumptions will help reduce the list of possible solutions for our system.

Assume P is the trivial tangle

Recall that *P* is taken to be inside a ball containing only the core regions of the recombination sites (i.e. where the breakage and rejoining takes place). In Tn4430, the core recombination sites are of approximately 28bp ([27]). Such short segments of DNA are unlikely to cause tangling inside *P*. Therefore, *P* can be any of the four simplest tangles in figure 5. We choose *P = (0).* Any difference in the synaptosome tangle as result of this choice will be trapped in the resulting *O* tangle.

Alignment of the recombination sites

Recall that the non-palindromic sequence of the recombination sites induces a natural direction on each site which we represent with arrows. The two recombination sites on our parental tangle *P = (0)* are in parallel alignment if

both arrows point in the same direction in the tangle diagram, otherwise they are said to be in antiparallel alignment.

However, the concept of parallel and antiparallel alignment represent a local geometric property of the recombination sites and is well-defined only in the tangle diagram, which is a planar projection of the 3-dimensional tangle. Unless the two sites are strictly coplanar, we can always obtain for the same 3-D tangle a planar projection with parallel alignment of the sites and another planar projection with antiparallel sites, as illustrated in figure 9.

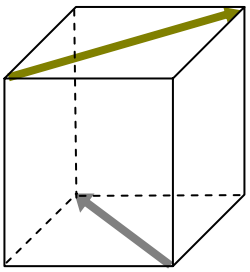


Figure 9: Recombination sites are not strictly co-planar, the relative orientation of the recombination sites depend on the projection taken. **a)** view from top; **b)** view from right; **c)** view from front.

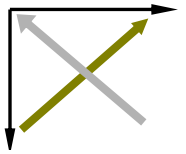


Fig. 9-a

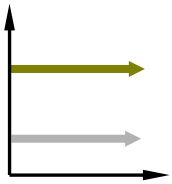


Fig. 9-b

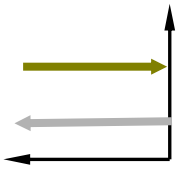


Fig. 9-c

To take into account the most general situations, we will assume that the recombination sites are not coplanar in 3-dimensional space, and hence they can be in parallel or antiparallel alignment in the tangle diagram based on the direction in which the projection is taken. Biological evidence suggests that enzymes in the tyrosine family present a pseudo-planar conformation at the synapse where the sites are presented in anti-parallel alignment. However co-planarity, in a strict mathematical sense, is unlikely to occur in nature (*cf.* argument and references in [26]).

Assumptions on R

As a member of the tyrosine site-specific recombinase family, TnpI catalyzes the recombination of the IRS sites through a Holiday junction intermediate ([21,27]). If in our tangle model, $P = (0)$ with parallel sites and the enzyme recombines via HJ then $R = (+1)$ *or* (-1) tangle, and if $P = (0)$ with antiparallel sites then $R = (0,0)$ ([27]).We will solve the equations for the more general case where R is an integral tangle $R=(k)$,or the infinity tangle.

4.3 The system of tangle equations for the TnpI-IRS system of recombination admits three biologically relevant solutions

From our previous discussions we proceed with the assumptions that $P = (0)$, $R = (k)$ for some integer k, and O is a rational tangle. We also know from the experimental results of Vanhooff et. al. in [27] that the unknotted substrate contains two directly repeated sites and that the resulting Hopf link is negative (due to the induced orientation). Now, we will find solutions $\{O, R\}$ that satisfy the system:

$N(O + P) = b(1,1)$	unknot with directly repeated sites
$N(O + R) = b(2,1)$	the (-)Hopf link

where $P = (0)$, $R = (k)$ for some integer k, and O is a rational tangle.

Solutions to the above system can be found using the following result:

Theorem 7([10]): Given two rational tangles $X = x/y$, A = u/v.
Then N(X + A) = b(a, b) is 4-plat, where $a = |xv + yu|$, and b is determined as follows:
1. If $a = 0$ or 1, then $b = 1$;
2. If $a > 1$, then $0 < b < a$, and $b \equiv s(vy'+ux')(mod\ a)$, where $s = sgn(xy + yu)$, and x', y' are integers such that $xx' - yy' = 1$.

The system of tangle equations for the TnpI-IRS system is:

$N(x/y + 0) = b(1,1)$
$N(x/y + k) = b(2,1)$

The essence of the algorithm is to express the integers $\{1,1,2,1\}$ in the Conway symbol of each 4-plat in terms of the unknowns $\{x, y, k\}$ using theorem 7 and then solve for the unknowns from these algebraic expressions.

As the computations are very tedious, we will not present them here. The detailed calculations and visualization of the solutions can be done using TangleSolve ([19]) (section 3).

From theorem 6, the solutions for the above system must satisfy O rational tangle and P and R locally unknotted. Furthermore, using biological evidence we have justified the more rigid assumptions $P = (0)$ and R integral. We obtain all possible solutions to the system using TangleSolve, and select the

ones that are consistent with the experimental results in [27].

The solutions to the tangle equations are:
a) $O = (-1), R = (-1)$
b) $O = (-3,0), R = (1)$
c) $O = (-2,0), R = (0,0)$
d) $O = (1), R = (-3)$
e) $O = (1), R = (1)$
f) $O = (3,0), R = (-1)$
g) $O = (2,0), R = (0,0)$
h) $O = (-1), R = (3)$
Solutions a,b,d,e,f,h, are illustrated in figure 8.

Theorem 8: Let *O, P, R* be tangles that satisfy the recombination system of TnpI-IRS:
$N(O + P) = b(1,1)$ unknot with directly repeated sites
$N(O + R) = b(2,1)$ the (-)Hopf link
If we assume $P = (0)$ and $R = (0, 0)$ *or* (k) for some integer k, then the only solutions to this system are: $\{O = (-1), R = (-1)\}$, $\{O = (-3,0), R = (1)\}$ and $\{O = (-2,0), R = (0,0)\}$.

Proof:
For the solutions (a)-(d), $N(O + R)$ gives the (-) Hopf link. While the solutions (e)-(h) produce the (+) Hopf link, which is not consistent with the experimental results in [27]. Therefore we consider only solutions (a)-(d). And since tyrosine recombinases recombine through a HJ junction, (d) is unlikely as we would expect R=(0,0) or R=(±1). Therefore, we conclude that solutions (a), (b), (c) are the only solutions consistent with the tangle method and experimental results.

4.4. A 3-D model for TnpI-IRS recombination system

In theorem 8 above, we have produced 3 solutions for the TnpI-IRS recombination system that are consistent with the experimental results from [27]. Notice that the proposed solutions $O = (-1), (-2,0)$ and $(-3,0)$ differ by at most 2 negative supercoils from one another. In the 3-dimensional frameworks, this small difference between the tangles suggests that the solutions may all be different projections of one single 3-dimensional tangle, and that the differences in the number of supercoils in the planar projections depend on how tight the supercoils are in space. Likewise the solutions for R change with the projection (fig. 10). We follow the approach used in our

analysis of Xer recombination ([26]). An animation of the 3-dimensional model proposed in [26] for Xer recombination can be accessed from http://math.berkeley.edu/~mariel/xeranim.html

The proposed biological model, and corresponding 3D model for TnpI-IRS recombination system are illustrated in figure 10. In the 3D model (fig 10b), depending on the direction of the view, one, two or three (-) supercoils are fixed by the binding mechanism to bring the two recombination sites together. However, note that in this model only one (-) supercoil is trapped between the accessory sequences as suggested by Vanhooff *et al.* ([27], fig 10a). The additional nodes in the solutions *O = (-2,0)* and *O = (-3,0)* are results of the alignment of the strings from the perspectives these projections were taken.

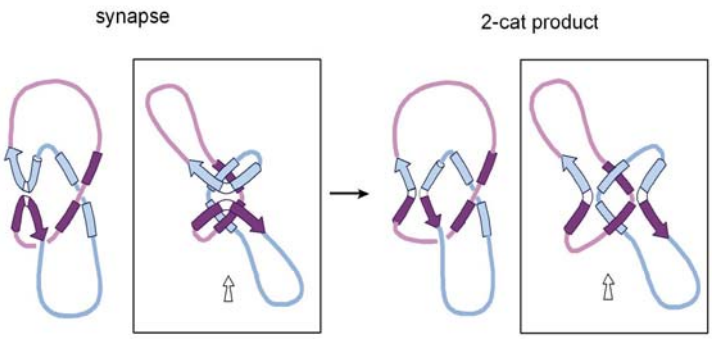


Fig. 11-a

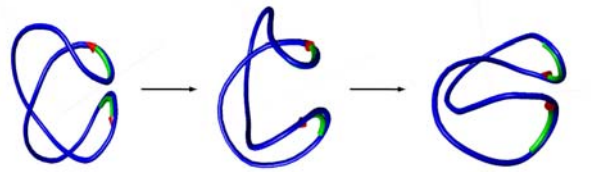


Fig. 11-b

Figure 10: Models for TnpI. **a)** Idealized model of the TnpI synapse designed from experimental data. **b)** Three different projections of a 3D model of the TnpI synapse. This 3D model is consistent with the tangle analysis, and with the model designed by the experimentalists (fig. a). Furthermore, it presents the three tangle solutions as three projections of the same 3D synapse. If tangles O and P were superimposed on the figure we could see that projections of the same 3D tangle taken from different directions reveal tangle diagrams with different number of supercoils if the tangle is not tightly twisted.

5. Concluding Remarks

We have used here the tangle model to analyze the binding mechanism of TnpI-IRS recombination system based on experimental results by Vanhooff et. al. [27]. Under the a few biological assumptions, we computed solutions to the tangle system arising from this modeling and we integrated these 2-D solutions into a 3-D model for the synaptic complex. The procedure of unifying solutions obtained by tangle analysis was first proposed for Xer recombination by Vazquez et .al. [26]. We must remark, however, that this approach is reasonable only if the solutions to be integrated are relatively similar. That is, they differ by only a small number of supercoils, as result of viewing the loosely twisted molecule from different directions, or one solution is simply the horizontal projection of the other. In the case of Xer ([26]) and TnpI-IRS recombination, this method is particularly effective because we know that the tangle O is a rational tangle and can compute all solutions using the tangle calculus. Hence we are certain that the synaptic complex should not have local knots or prime tangle projections and that its main topological feature are supercoils because all its possible planar projections are rational tangles.

Acknowledgements

We are particularly grateful to Yuki Saka who created the first version of TangleSolve, and to Stefanus Jasin for his contributions to this project. We also thank K. Shimokawa for conversations relevant to this project, and to Dale Rolfsen for allowing us to use his knot table in TangleSolve. MV research was funded through NASA NSCOR04-0014-0017 and NIH-RIMI grant NMD000262. BH was funded through The Belgian FNRS and an ARC grant 01/06-268. CG received a FRIA fellowship.WZ was part of the Undergraduate Research Apprentice Program at UC Berkeley.

References

[1] C. Adams, The knot book, W. J. Freeman and Company (1994).

[2] S. Bleiler, and R. Litherland, Lens spaces and Dehn surgery, Proc. Am. Math. Soc. **107**(4), pp. 1127-1131 (1989).

[3] S. Bleiler, Knots prime on many strings, Trans. Am. Math. Soc. **282**(1), pp. 385-401 (1984).

[4] D. Buck, and C. Verjovsky Marcotte, Classification of tangle solutions for integrases, a protein family that changes DNA topology, eprint arXiv:math/0501173(2005).

[5] D. Buck, and C. Verjovsky Marcotte, Tangle solutions for a family of DNA-rearranging proteins, Math. Proc. Camb. Phil. Soc. **139**, pp. 59-80 (2005)
[6] G. Burde, and H. Zieschang, Knots, De Gruytier Studies in Mathematics 5, Walter de Gruyter (2003).
[7] J. H. Conway, An enumeration of knots and links, and some of their algebraic properties, Computational Problems in Abstract Algebra, pp. 329-358 (1970).
[8] I. Darcy, Biological distances on DNA knots and links: applications to XER recombination, J. Knot Theory Ramifications **10**(2), pp. 269-294 (2001).
[9] I. Darcy and R. G. Scharein, TopoICE-R: 3D visualization modeling the topology of DNA recombination, Bioinformatics **22**(14), pp. 1790-1791 (2006); doi:10.1093/bioinformatics/btl164.
[10] C. Ernst, and D. W. Sumners, A calculus for rational tangles: applications to DNA recombination, Math. Proc. Camb. Phil. Soc. **108**, pp. 489-515 (1990).
[11] C. Ernst, Tangle Equations, J. Knot Theory Ramif. **5**, pp. 145-159 (1996).
[12] C. Ernst, Tangle Equations II, J. Knot Theory Ramif. **6**, pp. 1-11 (1997).
[13] C. Ernst, and D. W. Sumners, Solving tangle equations arising in a DNA recombination model, Math. Proc. Camb. Phil. Soc. **126**, pp. 23-36 (1999).
[14] M. Hirasawa and K. Shimokawa, Dehn surgery on strongly invertible knots which yield lens spaces, Proc. Am. Math. Soc. **128**(11), pp. 3445-3451 (2000).
[15] J. Goldman and L. Kauffman, Rational tangles, Adv. in Appl. Math. **18**, pp. 300-332 (1997).
[16] W. Lickorish, Prime knots and tangles, Trans. Amer. Math. Society **267**, pp. 321-332 (1981).
[17] J. Mahillon and D. Lereclus, Structural and functional analysis of Tn4430: identification of an integrase-like protein involved in the co-integrate resolution process, The EMBO Journal 7(5), pp. 1515-1526 (1988).
[18] K. Murasugi, Knot theory and its applications (Translated by B. Kurpita), Birkhäuser (1996).
[19] Y. Saka and M. Vazquez, TangleSolve: topological analysis of site-specific recombination, Bioinformatics - Applications Note 18(7), pp.1011-1012 (2002).
[20] H. Schubert, Knoten mit zwei Brücken, Math. Zeit. **65**, pp. 133-170, (1956).

[21] W. Stark, D. Sherratt and M. Boocock, Site-specific recombination by Tn3 resolvase: topological changes in the forward and reverse reactions, Cell **58**(4), pp. 779-790 (1989).
[22] D.W. Sumners, Lifting the Curtain: Using topology to probe the hidden actions of enzymes, Notices of Amer. Math. Soc. **42**, pp. 528-537 (1995).
[23] D.W. Sumners, The role of knot theory in DNA research, In Geometry and Topology, pp. 297-318, Marcel Dekker (1987).
[24] D. W. Sumners, C. Ernst, N. Cozzarelli, and S. J. Splengler, Analysis of the mechanism of DNA recombination using tangles, Quarterly Review of Biophysics **28**(3), pp. 253-313 (1995).
[25] M. Vazquez and D. W. Sumners, Tangle analysis of Gin site-specific recombination., Math. Proc. Camb. Phil. Soc. **136**, pp. 565-582 (2004).
[26] M. Vazquez, S. D. Colloms, and D. W. Sumners, Tangle analysis of Xer recombination reveals only three solutions, all consistent with a single three-dimensional topological pathway, J. Mol. Biol. **346**, pp. 493-504 (2005).
[27] V. Vanhooff, C. Galloy, H. Agaisse, D. Lereclus, B. Revet, and B. Hallet, Self-Control in DNA site-specific recombination mediated by the tyrosine recombinase TnpI, Molecular Microbiology **60**(3), pp. 617-629, (2006).

M. VAZQUEZ, Mathematics Department, San Francisco State University, San Francisco, CA 94132 USA; and Mathematics Department, University of California at Berkeley, Berkeley, CA 94720, USA
mariel@math.sfsu.edu (to whom correspondence should be addressed)

W. ZHENG, Mathematics Department, University of California at Berkeley, Berkeley, CA 94720, USA
wenjingzy@gmail.com

C. GALLOY, Unité de Génétique, Université Catholique de Louvain, Belgium
christine.galloy@fundp.ac.be

B.HALLET, Unité de Génétique, Université Catholique de Louvain, Belgium
hallet@gene.ucl.ac.be