JMB

ELSEVIER

Tangle Analysis of Xer Recombination Reveals only Three Solutions, all Consistent with a Single Three-dimensional Topological Pathway

Mariel Vazquez[1]*, Sean D. Colloms[2] and De Witt Sumners[3]

[1]*Department of Mathematics University of California at Berkeley, Berkeley CA 94720-3840, USA*

[2]*Division of Molecular Genetics Institute of Biomedical and Life Sciences, University of Glasgow Glasgow G11 6NU, Scotland UK*

[3]*Department of Mathematics Florida State University Tallahassee FL 32306, USA*

*Corresponding author

The product of Xer recombination at directly repeated *psi* sites on a circular unknotted DNA molecule is a right-hand four-noded catenane. Here, we use tangle equations to analyze the topological changes associated with Xer recombination at *psi*. This mathematical method allows computation of all possible topological pathways consistent with the experimental data. We give a rigorous mathematical proof that, under reasonable biological assumptions, there are only three solutions to the tangle equations. One of the solutions corresponds to a synaptic complex with antiparallel alignment of recombination core sites, the other two correspond to parallel alignment of cores. We show that all three solutions can be unified into a single three-dimensional model for Xer recombination. Thus the three distinct mathematical solutions do not necessarily represent distinct three-dimensional pathways, and in this case the three distinct tangle solutions are different planar projections of the same three-dimensional configuration.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: site-specific recombination; Xer recombination; topological mechanism; tangle equations; DNA knots

Introduction

Site-specific recombinases catalyze the exchange of genetic material between specific sites on a DNA molecule. This recombination mediates a number of biologically important processes including: integration and excision of viral DNA into and out of its host genome; inversion of DNA segments to regulate gene expression; resolution of multimeric DNA molecules to allow proper segregation at cell division; and plasmid copy number regulation.[1,2] Recombination reactions can be divided into two stages, synapsis and strand-exchange. In the first stage, the two recombination sites are brought together by the recombinase and accessory proteins to form a specific synaptic complex. In the second stage, the DNA strands are cleaved, exchanged and religated within this synapse, to form recombinant products. When the DNA substrate is circular, topological changes associated with recombination can be observed and quantified. These topological changes can be used to infer topological details of the synaptic complex and the strand-exchange mechanism,[3,4] and can be analyzed mathematically using the tangle method (reviewed in the next section).[5]

Site-specific recombinases are divided into two families, based on sequence similarity and reaction mechanism: the tyrosine recombinases, which include Flp, Cre, Int and XerC/XerD and recombine through a Holliday junction (HJ) intermediate;[6,7] and the serine recombinases, which include Tn*3* and γδ resolvases, as well as DNA invertases such as Gin, and recombine *via* double-strand cleaved intermediates.[8]

Many serine recombinases display topological selectivity, i.e. they distinguish between sites in different orientations, and between inter- and intramolecular sites. These systems also display topological specificity, i.e. the topology of the substrate uniquely determines the topology of the reaction product. Topological specificity implies that recombination occurs in a synapse with a defined topology, and that strand-exchange occurs by a fixed mechanism.

Abbreviations used: HJ, Holliday junction; 4-cat, four-noded torus catenane.

E-mail address of the corresponding author: mariel@math.berkeley.edu

Tangle analysis is a mathematically rigorous method used to analyze site-specific recombination reactions.[5] In tangle analysis, the experimentally determined topologies of the recombination substrates and product(s) are translated into a mathematical system of tangle equations. The solutions to the tangle equations correspond to all possible values for the synaptic complex and strand-exchange topologies consistent with the experimental results.

Tangle equations have been used successfully to analyze the topological changes occurring during recombination catalyzed by the serine recombinases Gin and Tn3 resolvase.[5,9,10] Serine recombinases catalyze processive recombination, giving a series of products, apparently by repeated application of the same recombination process without dissociation of the synaptic complex. Processivity and topological specificity make serine recombinases particularly appropriate for a systematic mathematical analysis.[9–12] Tangle analysis for tyrosine recombinases has generally proved more difficult. In most cases there is no topological selectivity, specificity or processive recombination (e.g. Int[13]). The experimental results therefore usually consist of only a single pair of substrate and product topologies, which do not provide enough information to deduce the topological pathway of recombination. Further mathematical and biological assumptions are then needed to produce a limited number of solutions to the tangle equations.[5,13,14]

Here, we provide a rigorous mathematical analysis, based on the tangle method, of the Xer site-specific recombination data reported by Colloms et al.[15] Our analysis builds on the work of Darcy,[16] and extends it by proposing a more detailed analysis of the system's underlying biology, which allows for a drastic reduction in the number of solutions. Furthermore, we give a three-dimensional interpretation of the solutions.

Xer recombination is catalyzed by a pair of tyrosine recombinases (XerC and XerD) and functions to keep multicopy plasmids (ColE1 and pSC101) in a monomeric state, thus ensuring stable plasmid inheritance.[17] In addition to XerC and XerD, recombination at plasmid resolution sites (e.g. psi from pSC101 and cer from ColE1) requires accessory proteins and sequences. The accessory proteins bind to accessory sequences to one side of the recombination core sites, to form a complex in which the two sites are interwrapped in a right-hand fashion.[18] Unlike recombination by most other tyrosine recombinases, Xer recombination at psi and cer displays topological selectivity and specificity,[15,19] making it an attractive target for tangle analysis, despite the absence of processive recombination.

Colloms et al.[15] showed that Xer recombination at psi and cer on unknotted substrates is strictly intramolecular, is efficient only at directly repeated sites (Figure 1(a)), and gives products with a specific topology: four-noded torus catenane (4-cat) with antiparallel sites (Figure 1(b)). This is
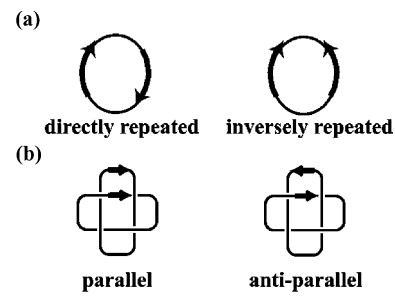

Figure 1. Site orientations. Recombination sites are usually short non-palindromic nucleotide sequences and, as such, can be assigned an orientation. (a) Two identical sites in a closed circle can be in direct repeat (left) or in inverted repeat (right). (b) Two identical sites in different components of a DNA torus catenane can be either parallel or antiparallel. The diagrams show a right-hand four-noded torus catenane (4-cat) with one recombination site on each component, and illustrate parallel and anti-parallel orientation of sites.

consistent with a requirement for a synapse with a fixed local topology/geometry, and a fixed strand-exchange mechanism. Here, we focus on recombination at psi, since the products of recombination at cer contain a HJ.[20,21] We propose a few reasonable biological assumptions and, using tangles, we systematically find all possible topologies for the synapse and the strand-exchange mechanism, proving mathematically that under our assumptions there are only three solutions to the Xer tangle equations. Two of these solutions correspond to the two topological models (Figure 2) proposed by Colloms et al.,[15] and have core sites aligned in parallel in the synaptic complex. The third solution has antiparallel core sites in the synaptic complex (and corresponds to a model shown by Bath et al.[19] and by Bregu et al.[22]). We then produce a 3D model
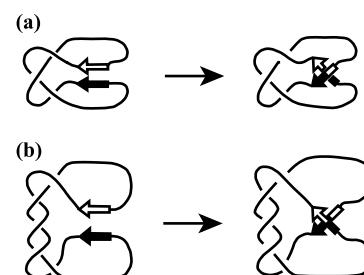

Figure 2. Topology of Xer recombination at psi. The two topological pathways proposed for Xer recombination at psi sites (black and white arrows) are shown.[15] The recombination products of both pathways are topologically identical: both are 4cats with antiparallel recombination sites. (a) The accessory sequences are wrapped around one another so as to trap three negative nodes. Upon formation of this synapse, XerC/D bind the two cores and catalyze strand-exchange to introduce an additional negative node. (b) Five (−) nodes are trapped in the synaptic complex and one (+) node is introduced by recombination.

for Xer recombination that fully accounts for the topology of the reaction on unknotted substrates. This single 3D model admits different planar projections, which correspond to the three different solutions to the tangle equations. We therefore demonstrate that distinct solutions to the tangle equations do not necessarily correspond to distinct 3D pathways, thus turning a previous limitation of the tangle method into a strength.

Tangle Method

The topology of recombination reactions can be

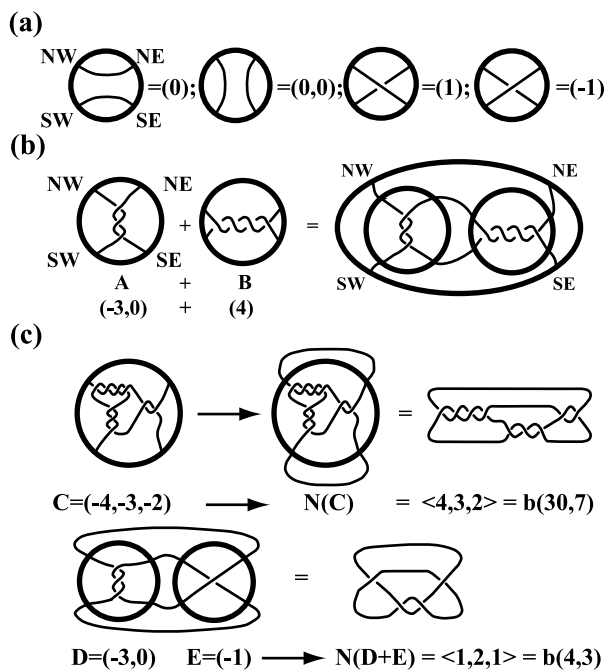

(a)

(b)

(c)

Figure 3. Tangles, tangle addition, the numerator operator and 4-plats. (a) The four trivial tangles are shown together with their canonical vectors (0), (0,0), (1) and ($-$1). Trivial tangles are the simplest examples of rational tangles. Intuitively, rational tangles are those that can be "unwound" to look like trivial tangles. (b) Two rational tangles, A and B, their canonical vectors, and their sum $A + B$. The tangle $B = (4)$ is an integral tangle. Integral tangles are formed as a horizontal row of twists and characterized by a canonical vector (k), where the integer k is the number of nodes. (c) The numerator operation converts a tangle, or a sum of tangles, into a knot or catenane (e.g. C into the catenane $N(C)$; $D + E$ into the catenane $N(D + E)$). If D and E are rational, as in this example, then $N(D + E)$ is a 4-plat knot or catenane.[25] The 4-plats are knots and catenanes formed by intertwining of four strings and joining the ends. Any 4-plat knot or catenane can be deformed "smoothly" to look as shown in the diagrams to the right of this panel, and can be represented by a canonical vector $\langle c_1, c_2, \ldots, c_{2k} + 1 \rangle$ with an odd number of positive integer entries and by a corresponding Conway symbol $b(\alpha, \beta)$. The canonical vector describes the sequence of nodes defining the 4-plat. Note that the 4-plat $\langle 1, 2, 1 \rangle$ shown in the bottom diagram is the 4cat product of Xer recombination on unknotted substrates, with Conway symbol $b(4,3)$.

studied using knots and tangles. The mathematics of knots and tangles are reviewed by Murasugi[23] and by Rolfsen,[24] and their use in the study of DNA recombination is reviewed by Sumners *et al.*[5] Tangles are used to model small portions of knots or catenanes containing two segments of inter-wrapped DNA that are bound to a recombination protein (illustrated in Figure 3(a) and (b)). Mathematically, a tangle is defined as a ball with two strands and a fixed direction p. Each tangle is studied through its tangle diagram, a 2D projection in the direction p. The most biologically relevant tangles in biology are the rational tangles. Rational tangles, relevant tangle operations, and the family of 4-plat knots and catenanes are illustrated in Figure 3, and are further discussed in Appendix I. Rational tangles and 4-plats have been classified, allowing them to be represented by integer entry vectors and rational numbers.[25–27]

In the tangle method, a recombination event is modeled by a system of two tangle equations (Figure 4):

$$N(O + P) = K_1 \qquad (1)$$

$$N(O + R) = K_2 \qquad (2)$$

where O, P and R are unknown tangles, and K_1 and K_2 are substrate and product of recombination (respectively). P is the parental tangle and contains only the DNA that is changed during recombination. The strand-exchange reaction is modeled by replacing P with the recombinant tangle R. The outside tangle O contains all topologically relevant DNA that is not changed during a recombination event. It is assumed that the strand-exchange mechanism (as represented by replacing P with R, while leaving O unchanged) is constant for a given recombinase.

The O tangle can be decomposed into two tangles: $O = O_f + O_b$. The outside bound tangle, O_b, contains all DNA wrapped by the recombinase and accessory proteins that is not changed by strand-exchange. O_f, the outside free tangle, contains any other topologically interesting DNA that is not bound by proteins. Distinguishing between O_f and O_b is sometimes necessary. For example, the action of a recombinase that displays topological selectivity on two different substrates[19] is modeled by two systems of tangle equations where different values of O_f are used to give different substrate topologies, while O_b, P and R are assumed to remain constant.[5]

A recombination reaction on a single substrate (K_1) that gives a single product topology (K_2) is modeled by a single system of two simultaneous tangle equations (1) and (2). K_1 and K_2 are determined experimentally (e.g. by gel electrophoresis and electron microscopy) and are often 4-plats (knots and catenanes formed by intertwining of four strings and joining of the ends; illustrated in Figure 3); O, P and R are the unknowns. It is possible to make simplifying assumptions for P. If both O and R are shown to
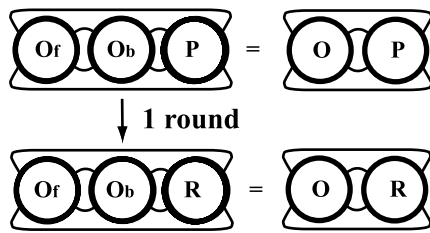
Figure 4. Tangle equations for site-specific recombination. One round of site-specific recombination is represented by a system of two tangle equations:

$$N(O_f + O_b + P) = N(O + P) = \text{substrate}$$

$$N(O_f + O_b + R) = N(O + R) = \text{product}$$

where O_f, O_b, O, P, and R are all tangles. P represents the parental tangle, containing the recombinase-bound cores of the recombination sites, where strand exchange takes place. R is the tangle obtained from P after recombination. O, the outside tangle, encloses all DNA not changed by strand-exchange. In the tangle method, O can be written as the sum of tangles, $O_f + O_b$. The tangle O_b corresponds to the DNA bound to the enzymatic complex, but not contained in P; O_f encloses the unbound DNA.

---

be rational tangles, or sums of rational tangles, all solutions can be found in a mathematically rigorous fashion;[10,28] the computations involved are simple but tedious (shown for Xer recombination in Appendix II). The java applet TangleSolve[29] is available to do this calculation. To prove O and R rational, sophisticated low-dimensional topology tools may be required;[9,10] the goal cannot always be achieved, in which case extra assumptions are called for.[5,13,14,30] Finally, decomposing O into O_f and O_b unambiguously, and proving that there is no other solution (e.g. other prime or locally knotted) can be more difficult or even impossible.[16,28]

Results and Discussion

In Xer recombination *P* and *R* are not locally knotted, and *O* is rational

When Xer acts on an unknotted substrate ($K_1 = \langle 1 \rangle = b(1,1)$) the product is the 4-cat[15] ($K_2 = \langle 1,2,1 \rangle = b(4,3)$) and the corresponding system of tangle equations is:

$$N(O + P) = b(1, 1), \text{ the unknot,} \quad N(O + R)$$
$$= b(4, 3), \text{ the } 4 - \text{cat} \tag{*}$$

The implicit assumptions of constancy of enzymatic action and a fixed synapse topology (see Tangle Method, above) are supported by the experimentally observed topological specificity of Xer recombination.[15,19]

Theorem 1 (Xer recombination). *In system (*),* P *and* R *are either prime or rational, and* O *is rational.*

This result was implied without proof by Darcy;[16] a detailed proof is presented in Appendix I.

Mathematical and biological assumptions

Darcy computed three infinite families of rational solutions to the Xer tangle equations, and proposed experiments that could help narrow this list down.[16] Here, we show biological and geometrical arguments that can be combined with the existing experimental data to obtain more detailed assumptions on P and R. Our assumptions lead to three unique solutions to the system (*).

P *is chosen to be (0)*

We assume that P is a ball containing only the core regions of the recombination sites and the XerC/D recombinases. Any non-trivial DNA topology implicit in the synaptic complex is inside O. Given that the binding sites for XerC/XerD are very short DNA segments (~ 28 bp), it seems unlikely that they will contain excessive tangling. Therefore, the tangle P can be assumed to be a trivial tangle (Figure 3(a)). We chose $P = (0)$ to be consistent with previous discussions on site-alignment.[5] Any of the other three trivial tangles (Figure 3(a)) would be a valid choice.

Recombination sites are not coplanar

The two recombination sites contained in $P = (0)$ are said to be in parallel alignment if the arrows point in the same direction in the tangle diagram, and in antiparallel alignment otherwise. The concept of parallel or antiparallel alignment refers to local geometrical properties of the recombination sites considered as essentially straight lines in the tangle diagram, where two sites that do not cross over one another are either parallel or antiparallel. However, the tangle diagram is a projection of a 3D tangle in the direction p (Figure 6; and see the definition of tangle in Tangle Method, above). Parallel or antiparallel sites in the tangle diagram do not correspond to well-defined properties in the 3D tangle, unless the two sites are strictly coplanar. When the sites are not strictly coplanar (including the "slightly off plane" or "pseudo-planar" cases[31]), two projections of the same 3D object can always be found such that in one projection the sites appear in parallel alignment, and in the other projection the sites appear in antiparallel alignment.[5] Figure 6 illustrates how different planar projections of a solid ball containing two essentially straight strings may give rise to very different tangle diagrams. Taking different projections of the ball with two strings corresponds to choosing different directions p in the definition of the tangle (see Tangle Method, above), and thus changes the tangle. We assume here that the recombination mechanism starts and finishes with the recombination cores non-coplanar, implying that they can be considered as parallel, or
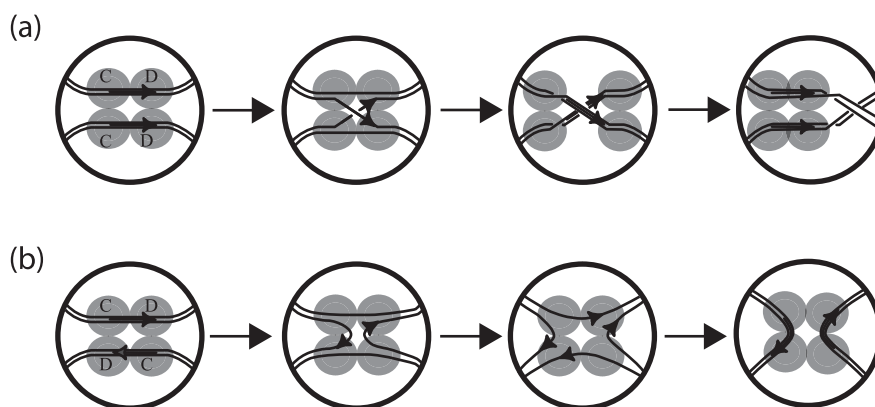
Figure 5. R is assumed to be integral or $(0,0)$. Tyrosine recombinases catalyse recombination *via* a HJ. (a) If in the tangle method one assumes $P=(0)$ with parallel sites, then if recombination occurs *via* an HJ, R is one of the rational tangles $(+1)$ (shown) or (-1) (not shown). In the particular case of Xer recombination, XerC and XerD act cooperatively at the recombination sites. After one round of recombination, a crossing is trapped within the recombinase tetramer. Further rounds of strand-exchange would require dissociation of the Xer synaptic complex and would result in a tangle $R=(k)$ integral. (b) If $P=(0)$ with antiparallel sites, one round of recombination (*via* an HJ) results in $R=(0,0)$. Visualizing further processive rounds is harder in this case. In Results we show that, for Xer recombination, all tangle solutions in the parallel sites case correspond to $R=(\pm 1)$ as predicted by the HJ. We thus assume that in the antiparallel sites case, R is simply $(0,0)$.

anti-parallel, depending on the direction in which they are observed.

Assumptions on R

Like other tyrosine recombinases, XerC/XerD recombine through an HJ intermediate.[20,21,32] The proposed mechanism for this recombination (Figure 5) suggests that if $P=(0)$ with parallel sites, then R is one of the tangles $(+1)$ or (-1).[33] Since we have assumed that the sites are non-coplanar, one can go from the $P=(0)$ parallel case, to the $P=(0)$ antiparallel case simply by rotating in 3-space (Figure 6, I.B and I.C). The same projection transforms $R=(\pm 1)$ into $R=(0,0)$ antiparallel (Figure 6, II.B and II.C), which is consistent with the results predicted by the HJ mechanism of recombination (Figure 5).[33]
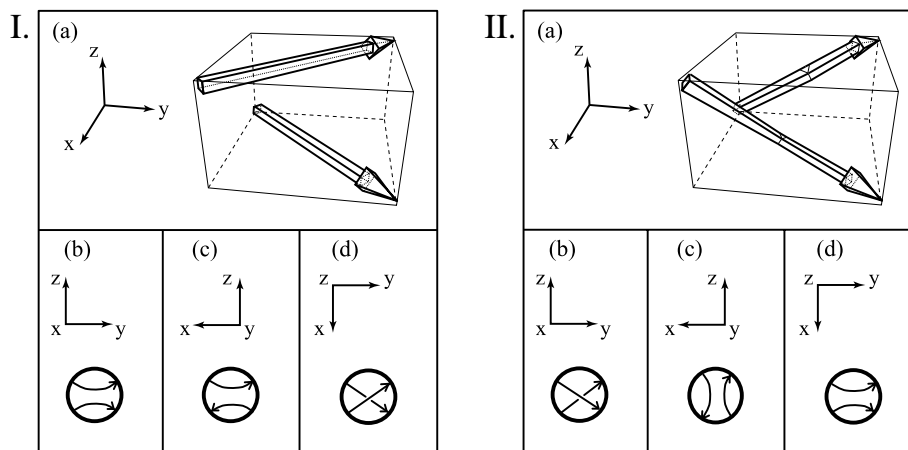

Figure 6. The topology of strand exchange is projection-dependent. Here, we use a trivial tangle before (I) and after (II) recombination to show how one can go from parallel to antiparallel alignment simply by changing projections. The arrows represent the core regions of the recombination sites, the two recombinase-bound DNA segments that undergo strand-exchange. A tangle is defined as a ball, two strings, and a fixed direction p (used to produce the tangle diagram). Tangles are analyzed through their tangle diagrams (specified by p); therefore, changing the direction p is equivalent to changing the tangle. In order to ease visualization of the spatial objects, we embed the core sites in a cube (instead of a bounding ball). The bottom of the Figure illustrates how two non-coplanar strands inside the cube can project to three different tangles, both before (left) and after (right) recombination. (I) Before recombination we show: (b) down the x-axis, $P=(0)$ parallel; (c) down the y-axis, $P=(0)$ antiparallel; (d) down the z-axis, $P=(-1)$. (II) After recombination we show: (b) down the x-axis, $R=(+1)$; (c) down the y-axis, $R=(0,0)$ antiparallel; (d) down the z-axis, $R=(0)$ parallel.

When $P = (0)$ parallel, multiple rounds of strand-exchange, with a reset step between each round, would give apparently processive recombination and an integral R tangle.[34] In Theorem 2, below, we assume that R is integral, to ensure we have covered all reasonable possibilities.

Three solutions for Xer recombination are found by tangle analysis

Theorem 2 (Xer recombination). *Let* O, P *and* R *be tangles that satisfy the Xer system of equations*:

N(O+P)=b(1,1), *the unknot with sites in direct repeat*
N(O+R)=b(4,3), *the 4-cat with antiparallel sites.*

If P *is assumed to be (0) with parallel or anti-parallel sites, and* R *is assumed to be integral or (0,0), respectively, the only solutions to the system are {O=(−3,0), R=(−1)}; {O=(−5,0), R=(+1)}; {O=(−4,0), R=(0,0)}.*

Proof. The calculations are presented in detail in Appendix II. The solutions can be obtained and visualized using TangleSolve†.[29] For $P = (0)$ with parallel sites and R integral the system has four solutions (Figure 7(a)): {O=(−3,0), R=(−1)}; {O=(−5,0), R=(+1)}; {O=(−1), R=(+5)}; {O=(+1), R=(+3)}. The last two solutions (shown in Figure 7(a)c and (a)d), produce a parallel 4-cat (Figure 1(b)), which contradicts the experimental results,[15] where the product was the antiparallel 4-cat. Thus, the tangle analysis proves that if P is (0) parallel and R is integral, there are only two solutions (Figure 7(a)a and (a)b) to the system of equations for Xer recombination on unknotted substrates with directly repeated sites. These solutions correspond to the two topological models proposed by Colloms *et al.* (and see Figure 2).[15] The solutions are consistent with the assumption of non-processive recombination that predicted $R = (\pm 1)$. This leads to the corresponding assumption in the $P = (0)$ anti-parallel case, namely $R = (0,0)$. Then, as shown in detail in Appendix II, a unique solution {O=(−4,0), R=(0,0)} is obtained (Figure 7(b)). □

Summing up, there are only three solutions to the tangle equations arising from the data reported by Colloms *et al.*[15] The synaptic complex fixes three, four or five negative nodes prior to recombination, and strand-exchange corresponds to changing tangle $P = (0)$ into tangle $R = (−1)$, (0,0) or (+1), respectively. The tangle method and our biologically based assumptions ensure that these are the only topological mechanisms to explain the experimental data.

A 3D unification of the solutions

We here show how the three solutions produced by the tangle method can be interpreted as different
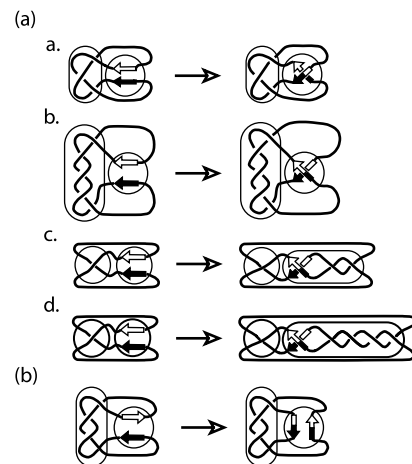
Figure 7. Solutions to the tangle equations. All solutions to the tangle equations obtained when the substrate of Xer recombination is unknotted and the product is the 4-cat. (a) The four solutions obtained when assuming $P = (0)$ with parallel sites and R integral. Two of these solutions (c and d) yield catenanes with parallel sites, and are therefore not consistent with the experimental data.[15] (b) The unique solution obtained if $P = (0)$ with anti-parallel sites and $R = (0,0)$.

projections of a single 3D model for the synaptic complex. Figure 8 presents a molecular model for the Xer synaptic complex at *cer* or *psi*. This model and the animated cartoon‡ are used to visualize a new interpretation for the three solutions {O=(−3,0), R=(−1)}, {O=(−5,0), R=(+1)} and {O=(−4,0), R=(0,0)}. In the molecular model, the Cre/*loxP* crystal structure is used to model the XerC/D recombinase core complex.[34,35] The accessory sequences are wrapped three times around a sandwich of two PepA hexamers and one ArgR hexamer, approximately as suggested by Sträter *et al.*,[36] and in agreement with the result from the tangle method that O fixes three, four or five nodes.

Three different views of the molecular model can be seen by performing two rigid motions, m1 and m2, on the complex, and projecting onto the plane of the page. Equivalently, one can imagine an observer moving around the protein/DNA complex and stopping to look at it from three different spatial locations. We define the tangle P as a ball enclosing the two recombination cores prior to recombination. The endpoints (a, b, c, and d) of the strings lie on the sphere that bounds the ball, and when projected in the direction p (orthogonal to the page) they project to cardinal points (NW, NE, SW, and SE) on the tangle diagram. The first projection (Figure 8(a)) corresponds to $P = (0)$ with parallel sites, and a, b, c, and d map to NW, NE, SW, SE, respectively. The complement of P (i.e. the entanglement not contained in P) contains three crossings and corresponds to $O = (−3,0)$. Rotating around the
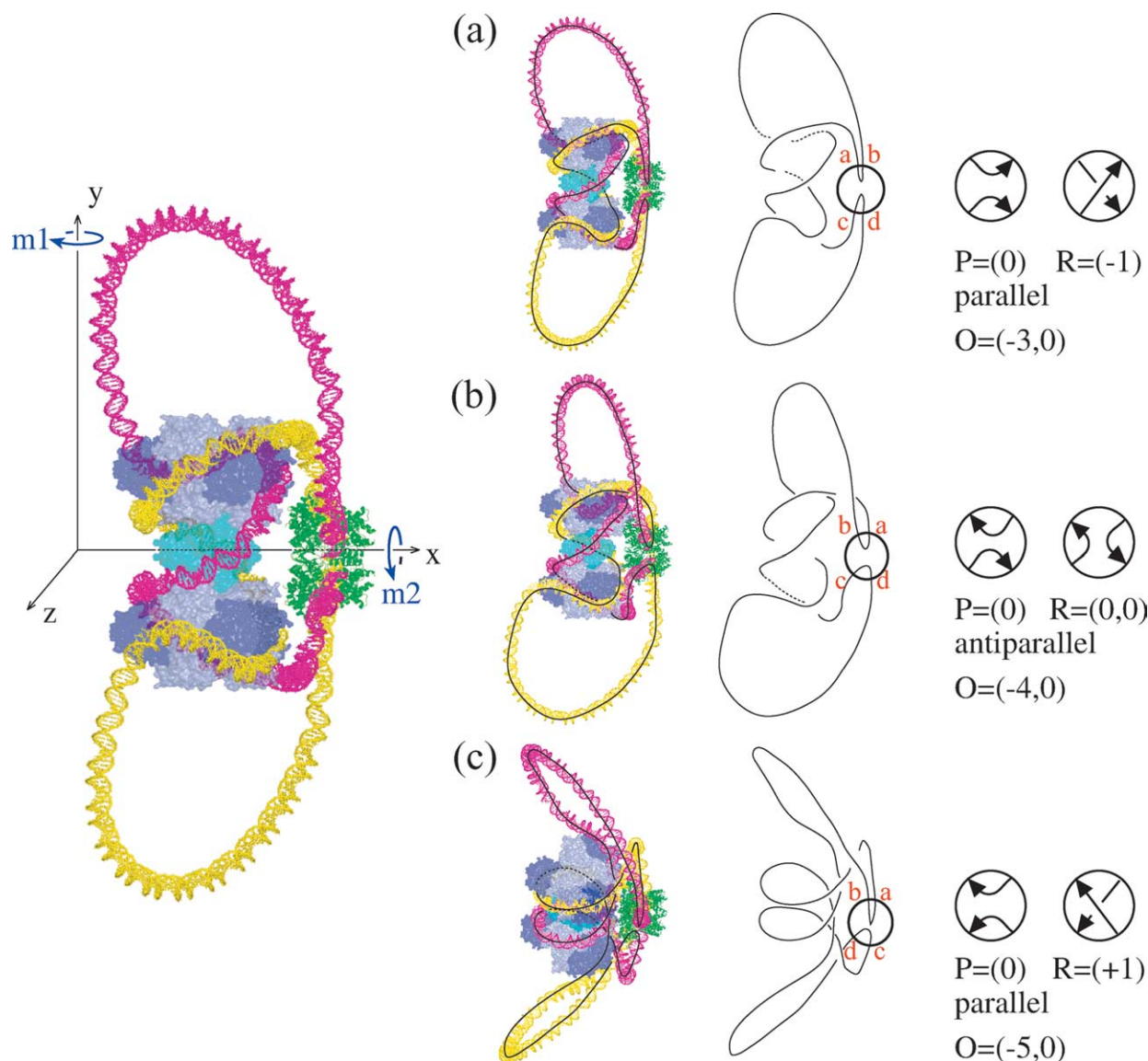
Figure 8. A possible 3D interpretation of the solutions to the Xer tangle equations. The left panel shows a synapse between two directly repeated Xer recombination sites in an unknotted circular DNA molecule. The Cre/*loxP* synapse (1CRX, shown as green ribbons) has been used to model the XerCD/recombination core site complex. The accessory sequences are interwrapped approximately three times around two hexamers of PepA (1GYT, represented as a transparent blue molecular surface) and one hexamer of ArgR (1XXB, shown as a transparent cyan surface), as suggested by Sträter *et al.*[36] The recombinases cleave and rejoin the DNA at the core sites yielding a 4-cat with antiparallel recombination sites. The DNA is colored so that after recombination, one product circle will be magenta and the other will be yellow. The recombinase core complex is slightly off-planar,[35] so that the cores appear either parallel or antiparallel in different projections. The right panel shows three projections of the same complex and their tracings, corresponding to the three different solutions to the tangle equations. The P tangle (circled) contains only the recombination cores. The four DNA arms (a, b, c, and d) are labeled as they enter the P tangle. The O tangle is the complement in 3-space of P. (a) The projection presented here is the same as in the left panel and corresponds to the solution $O=(-3,0)$, $P=(0)$ parallel and $R=(-1)$. (b) After a left-hand rotation around the y-axis (transformation m1), the magenta helix emanating from endpoint b crosses over the yellow helix from a. New tangles are defined with respect to this projection to obtain $O=(-4,0)$, $P=(0)$ antiparallel and $R=(0,0)$. (c) After a right-hand rotation around the x-axis (transformation m2), the magenta helix from c crosses over the yellow helix from d, thus yielding $O=(-5,0)$, $P=(0)$ parallel and $R=(+1)$. The projection in (c) contains additional crossings within the accessory sequences of each site. These crossings do not contribute to the O tangle. An animated cartoon of this molecular model can be found at http://bio.math.berkeley.edu/xeranim.

y-axis (transformation m1) gives $P=(0)$ but now the sites are antiparallel, and a, b, c, and d map to NE, NW, SW, and SE, respectively (Figure 8(b)). The complement of P is $O=(-4,0)$, which corresponds to the second tangle solution. Finally, rotating around the x-axis (transformation m2, Figure 8(c)), gives $P=(0)$ parallel (a, b, c, and d map to NE, NW, SE, and SW, respectively) and $O=(-5,0)$, which

corresponds to the third solution to the tangle equations. In each individual case it can be seen how recombination converts P into R: case 1, $P = (0)$ parallel, $R = (-1)$; case 2, $P = (0)$ antiparallel, $R = (0,0)$ antiparallel; case 3, $P = (0)$ parallel, $R = (+1)$. Thus, a single 3D structure can be interpreted as three different tangle solutions when viewed from different angles in space.

It is important to note that in the proposed 3D structure (Figure 8) there are only three plectonemic interwraps between the accessory sequences. Two of the nodes seen in the $O = (-5,0)$ projection, and one of the nodes seen in the $O = (-4,0)$ projection, come not from interwrapping of the accessory sequences, but are "alignment nodes" between the accessory sequences and the outside DNA. A different 3D structure, where O consists of five plectonemic interwraps between the two sets of accessory sequences, and a strand exchange mechanism that introduces one positive node, is also a valid solution, giving the observed topology on unknotted substrates.[15] In a simple representation analogous to that of Figure 8, this solution would correspond in the tangle equations to $P = (0)$ parallel, $O = (-5,0)$ and $R = (+1)$. However, this model with five plectonemic interwraps between accessory sequences does not predict the observed topology when torus knots or catenanes are used as substrates.[19] Therefore, two 3D solutions to the tangle equations that have identical values for O, P and R, and give the correct predictions for the unknotted substrates, are not equivalent, and give different predictions on catenated and knotted substrates.

Other site-specific recombination systems

Normal action of tyrosine recombinases produces, for each substrate–product pair, a system of two tangle equations. Such systems generally lead to infinitely many solutions, some of which cannot even be characterized.[13,16] Reducing the number of solutions to a small finite number often requires additional assumptions. Here, we have further extended the tangle method by showing that different tangle solutions may simply correspond to different projections of the same 3D object. This further reduces the number of solutions obtained by binning them in 3D equivalence classes, where two solutions are equivalent if they can be interconverted by rotations and translations in 3-space. We are currently trying to generalize this process to apply it to other tyrosine recombinases. The application to other enzymes with topological selectivity and specificity such as Cre is straightforward, as it mimics the present work. However, the general case opens a new and challenging mathematical problem, as well as non-trivial applications that could give relevant new insights into the mechanism of action of enzymes lacking both topological selectivity and specificity, such as Int.

Concluding Remarks

Our study builds on Darcy's results,[16] and goes beyond them by providing reasonable biological assumptions to reduce the number of solutions to the tangle equations from infinite to three, and by showing that the three solutions obtained are the only possible explanations to the experimental data and can be interpreted as three different views of the same 3D molecular model. We also provide, in two Appendices, the mathematical proof for the rationality of the tangle O and the detailed calculations of the tangle solutions. For the first time, for tyrosine recombinases, we are able to present a finite number of well-defined solutions with comparatively weak assumptions. The procedure of fusing all solutions obtained by tangle analysis can be extended to other tyrosine recombinases and can be applied to classify, and thus simplify, complicated solution sets (e.g. Int),[13] and to propose topological mechanisms for these enzymes.

In the case of Xer recombination, there could be other possible 3D models, for example some that would consist of five plectonemic interwraps between the two sets of accessory sequences. The topological information is not sufficient to elucidate a detailed biochemical mechanism of Xer recombination. At this stage, more information is needed about the actual 3D structure of the Xer/accessory proteins/DNA complex in solution. Crystallographic data, coupled with our results could lead to a better model.

Bath *et al*. used catenanes as substrates for Xer recombination.[19] Mathematical analysis of these data suggests two models.[30] Among these there is a preferred model where, if $P = (0)$ parallel, the accessory proteins fix three nodes $(O_b = (-3,0))$ before recombination, and recombination adds one negative node to the domain. In general, analyses of this type on knotted and catenated substrates can help nail down the geometry of the synaptic com-plex when the tangle method applied to a single experiment fails to provide enough information.

Appendices

This material gives basic definitions used here, as well as detailed proofs for Theorems 1 and 2 in Results and Discussion. Appendix I is directed to a more specialized audience familiar with the mathematical terminology, while Appendix II is intended for all audiences. Appendix I gives more formal definitions for the three types of tangles, states the classification theorem for rational tangles, and presents the proof that, for Xer recombination, P and R are locally unknotted tangles, and O is rational. Appendix II reviews in detail the elementary calculations used to find solutions to the tangle equations.

Here, we chose to use predominantly the term link instead of catenane (contrary to the main text).

Appendix I

In Xer Recombination, *P* and *R* are Locally Unknotted and *O* is Rational

A knot is a simple closed curve in R^3. A catenane (link) is composed of two or more such curves, which can be intertwined. Knots and links are studied through their knot diagrams, planar projections with strand breaks at crossings that encode their topology unambiguously. In a knot diagram, each crossing of one DNA segment over another is referred to as a node. Knots and links can be described by "crossing number", which is the minimum number of nodes in a planar projection (of the knot or catenane).

In this work a tangle, denoted by (B,t), refers to a "two-string tangle" defined as a fixed topological three-ball B in Euclidean 3-space, two non-oriented mutually disjoint spanning arcs t properly embedded in B, and a fixed direction (p) in 3-space. A projection in the direction p determines the tangle diagram for (B,t). There are three different types of tangles, locally knotted, rational and prime. A tangle (B,t) is locally knotted if there is a two-sphere S in B that intersects either of the two strands t transversely in two points, and such that the three-ball bounded by S in R^3 intersects that strand in a knotted arc (Figure 9(a)). Rational tangles are defined informally in Figure 3. Formally, a tangle is
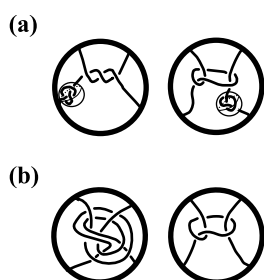
rational if there is an orientation preserving homeomorphism of pairs from (B,t) onto (D,t_0), where D is the unit ball in 3-space (oriented with the right-hand coordinate system) centered at the origin, and (D,t_0) is a trivial tangle (Figure 3(a)). The classification of rational tangles,[26,27] due to J. H. Conway, states that there exists a one-to-one correspondence between equivalence classes of rational tangles and the extended rational numbers (set of rational numbers together with ∞). To each extended rational number q/p is associated a unique Conway vector $(a_1,a_2,\ldots,a_n)$ by means of a continued fraction as follows:

$$q/p = a_n + 1/(a_{n-1} + 1/(a_{n-2} + 1/(\cdots + 1/a_1)\cdots)))$$

By convention, p and q are relatively prime, p is a positive integer or zero, and q can be any integer. The Conway vector must satisfy the conditions that $|a_1| > 1$, all entries have the same sign and are non-zero, except for the last one that can be zero. The tangle $A = (0,0)$ corresponds to $q/p = \infty$; if $p \neq 0$ and $q = 0$ then $q/p = 0$ corresponds to $A = (0)$. The Conway vector provides a way to construct any tangle from the trivial tangle $(0,0)$ thus showing that, intuitively, rational tangles are those that can be "unwound" by successive untwistings of pairs of adjacent endpoints.

Finally, prime tangles are those that are neither rational nor locally knotted. Two different prime tangles are illustrated in Figure 9(b).

Tangle addition and numerator were defined in Figure 3(b) and (c). Rational tangles are related to the family of 4-plat knots and links by the numerator operation: if A and B are rational tangles, then $N(A+B)$ is a 4-plat (Figure 3(c), bottom).[25] Each 4-plat can be characterized by a canonical vector $\langle c_1,c_2,\ldots,c_{2n+1}\rangle$, as illustrated by the examples in Figure 3(c). A rational number $\beta/\alpha = 1/(c_1 + 1/(c_2 + 1/(\cdots + 1/c_{2n+1})\cdots)))$, calculated as a continued fraction from the entries in the canonical vector, can also be used. The 4-plats are denoted either by their canonical vector, or by their Conway symbol $b(\alpha,\beta)$.

The following results are used in the tangle analysis of Xer recombination.

Fact 1. [25]If K_1 is the unknot and $K_2 = b(\alpha,\beta)$ is a 4-plat, then the double-branched cyclic cover of $N(O+P) = K_1$ is S^3, and that of $N(O+R) = K_2 = b(\alpha,\beta)$ is the lens space $L(\alpha,\beta)$.

Fact 2. [37]A tangle is rational if, and only if, its double-branched cyclic cover is a solid torus in S^3.

Lemma 1. [10,37,38]*If* A *and* B *are locally unknotted tangles and* b(α, β) *is a 4-plat such that* N(A+B)=b(α, β), *then at least one of* A *or* B *is rational*.

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.

Lemma 2. [39]*No Dehn surgery on a non-trivial strongly*

Figure 9. Other types of tangles: (a) locally knotted tangles; (b) prime tangles.

invertible knot can produce a lens space of the form
L(2p,1) *for any integer* p.

Following Isabel Darcy's suggestion, we use Lemma 2 to show that for Xer recombination O is rational.

Theorem 1 (Xer recombination). *The system arising from Xer recombination on unknotted substrates is*

$$N(O + P) = b(1, 1), \text{ the unknot}, \quad N(O + R)$$

$$= b(4, 3), \text{ the } 4 - \text{cat} \tag{*}$$

where O, P *and* R *are tangles. From this system,* P *and* R *are locally unknotted, and* O *is rational*.

Proof. Claim: *O, P* and *R* are locally unknotted tangles.

A local knot in O or P would result in at least one local knot in $N(O+P)$, implying that $N(O+P)$ is either a prime or a composite knot, and thus contradicting the first equation in the system. Similarly, a local knot in R would result in a local knot in $N(O+R)$, but $N(O+R)$ is the prime link (catenane) $b(4,3)$ and, as such, it has two unknotted components.

By the Claim, each of *O, P* and *R* is either prime or rational. We next prove that *O* is not prime, and is therefore rational.

Let O', P' and R' be the double-branched cyclic covers of *O, P* and *R*, respectively. Suppose that *O* is prime, then Lemma 1 applied to the first tangle equation $N(O+P)=b(1,1)$ implies that *P* is rational. By Fact 2, P' is a solid torus, i.e. P' can be seen as the regular neighborhood of some knot *K* in S^3. The first tangle equation lifts to $O' \cup P' \cong S^3 = (b(1,1))'$ where the union is taken along the common boundary of P' and O' (a 2D torus). From this equation, O' is the complement of *K* in S^3.

Likewise, from the second tangle equation $O' \cup R' \cong (b(4,3))'$ with $(b(4,3))' = L(4,3)$. But $O' = S^3 - K$, therefore $L(4,3)$ is obtained by Dehn surgery on the complement of the knot *K*. *K* is a strongly invertible knot (the covering transformation of the double-branched cover of $N(O+P)$ is an orientation preserving involution that maps *K* onto itself but reverses its orientation); then, by Lemma 2, *K* is a trivial knot, thus implying that O' is a solid torus. This contradicts the original assumption that *O* is prime. We conclude that *O* cannot be prime; and, since *O* is not locally knotted, *O* is rational.

Summarizing, Xer recombination on unknotted substrates can be written as the system (*) of two tangle equations with three unknowns *O, P* and *R*. Theorem 1 shows that *P* and *R* are locally unknotted tangles, and *O* is a rational tangle. Appropriate assumptions are made about *P* and *R*, namely $P = (0)$ and $R = (k)$ if the sites in *P* are parallel, or $R = (0,0)$ if the sites in *P* are anti-parallel. Under these conditions, all solutions to the system (*) can be obtained tangle analysis (see Appendix II), and thus all possible enzymatic mechanisms can be elucidated.

Appendix II

Computing Solutions to the Tangle Equations

Here we show, step-by-step, how to compute solutions to the Xer system of equations. This process involves direct application of the following result.

Lemma 3. [10]If *X* and *A* are two rational tangles, with classifying rational numbers u/v and x/y, respectively, then $N(X+A)=b(\alpha,\beta)$ is a 4-plat where $\alpha = |xv+yu|$, and β is determined as follows:

  (i) If $\alpha=0$ then $\beta=1$;
 (ii) If $\alpha=1$ then $\beta=1$;
(iii) If $\alpha>1$, then β is determined uniquely by $0 < \beta < \alpha$ and $\beta \equiv \sigma(vy'+ux') \pmod{\alpha}$, where $\sigma = \text{sign}(vx+yu)$ and y' and x' are integers such that $xx'-yy'=1$.

If $P=(0)$ with sites in parallel alignment and $R=(k)$ for some integer *k*, from the two equations in the system (*) result two equations for α:

$$1 = |0v + 1u| = |u|$$

$$4 = |kv + u|$$

By squaring both equations we obtain:

$$1 = u^2$$

$$16 = k^2v^2 + 2kuv + u^2$$

Subtracting the first from the second equation results in:

$$k^2v^2 + 2kuv - 15 = 0$$

We first solve the quadratic equation for *k* in terms of *u* and *v*, and then we substitute $u = \pm 1$ (since $u^2 = 1$) to express k in terms of *v*. By definition, *v* is positive, therefore there are eight possible solution pairs $(u/v, k)$: $(1/3,1)$, $(1,3)$, $(1,-5)$, $(1/5,-1)$ and their additive inverses (i.e. $(-1/3,-1)$ etc.).

Each entry in a pair is the classifying rational number for a rational tangle; the first entry corresponding to *O* and the second to *R*. We thus obtained four chiral pairs (i.e. mirror image pairs) of possible solutions (O,R) to the tangle equations. Only one of each chiral pair satisfies the system of tangle equations for Xer recombination since the 4-cat $b(4,3)$ is a chiral link (not equivalent to its mirror image). The resulting solutions to the tangle equations are $(-1/3,-1)$, $(-1/5,1)$, $(1,3)$ and $(-1,5)$, corresponding to the four tangle solutions shown in Figure 7(a). The two solutions shown in Figure 7(a)c and (a)d (corresponding to $(1,3)$ and $(-1,5)$) convert unknotted molecules with directly repeated sites to 4-cats with parallel sites, and are therefore not consistent with experimental results.

Now, if $P=(0)$ with antiparallel sites, then Figure 5 suggests that $R=(0,0)$, the infinity tangle, with classifying extended rational number $\infty = 1/0$. From Lemma 3, the corresponding equations are:

$$1 = |0v + 1u| = |u|$$

$$4 = |v + 0u| = |v|$$

thus leading to two possible solutions $u/v = 1/4$, and $u/v = -1/4$. From these, only $u/v = -1/4$ satisfies the system of tangle equations. This solution corresponds to $O = (-4,0)$, $P = (0)$ and $R = (0,0)$.

All rational, and sum of rational, tangle solutions to a system of two or more tangle equations can be obtained using the software TangleSolve,[29] which is an implementation of Lemma 3 and other published results.[10,28,30,40,41]

References

1. Sadowski, P. D. (1993). Site-specific genetic recombination: hops, flips, and flops. *FASEB J.* **7**, 760–767.
2. Hallet, B. & Sherratt, D. J. (1997). Transposition and site-specific recombination adapting DNA cut-and-paste mechanism to a variety of genetic rearrangements. *FEMS Microbiol. Rev.*, 21.
3. Cozzarelli, N. R., Kraznow, M. A., Gerrard, S. P. & White, J. H. (1984). A topological treatment of recombination and topoisomerases. *Cold Spring Harbor Symp. Quant. Biol.* **49**, 383–400.
4. Stark, W. M. & Boocock, M. R. (1995). Topological selectivity in site-specific recombination. In *Mobile Genetic Elements* (Sherratt, D. J., ed.), pp. 101–129, IRL Press at Oxford University, Oxford.
5. Sumners, D. W., Ernst, C., Cozzarelli, N. R. & Spengler, S. J. (1995). Mathematical analysis of the mechanisms of DNA recombination using tangles. *Quart. Rev. Biophys.* **28**, 253–313.
6. Landy, A. (1999). Coming or going its another pretty picture for the lambda-Int family album. *Proc. Natl Acad. Sci. USA*, **96**, 7122–7124.
7. Nunes-Düby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T. & Landy, A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucl. Acids Res.* **26**, 391–406.
8. Stark, W. M., Boocock, M. R. & Sherratt, D. J. (1992). Catalysis by site-specific recombinases. *Trends Genet.* **8**, 432–439.
9. Vazquez, M. & Sumners, D. W. (2004). Tangle analysis of Gin site-specific recombination. *Math. Proc. Cambridge Phil. Soc.* **136**, 565–582.
10. Ernst, C. & Sumners, D. W. (1990). A calculus for rational tangles: applications to DNA recombination. *Math. Proc. Cambridge Phil. Soc.* **108**, 489–515.
11. McIlwraith, M. J., Boocock, M. R. & Stark, W. M. (1997). Tn3 resolvase catalyses multiple recombination events without intermediate rejoining of DNA ends. *J. Mol. Biol.* **266**, 108–121.
12. Wasserman, S. A., Dungan, J. M. & Cozzarelli, N. R. (1985). Discovery of a predicted DNA knot substantiates a model for site-specific recombination. *Science*, **229**, 171–174.
13. Crisona, N. J., Weinberg, R. L., Peter, B. J., Sumners, D. W. & Cozzarelli, N. R. (1999). The topological mechanism of phage lambda integrase. *J. Mol. Biol.* **289**, 747–775.
14. Grainge, I., Buck, D. & Jayaram, M. (2000). Geometry of site alignment during Int family recombination: antiparallel synapsis by the FLP recombinase. *J. Mol. Biol.* **298**, 749–764.
15. Colloms, S. D., Bath, J. & Sherratt, D. J. (1997). Topological selectivity in Xer site-specific recombination. *Cell*, **88**, 855–864.
16. Darcy, I. (2001). Biological distances on DNA knots and links: applications to Xer recombination. *J. Knot Theory Ramif.* **10**, 269–294.
17. Summers, D. K. & Sherratt, D. J. (1984). Multimerization of high copy number plasmids causes instability: ColE1 encodes a determinant essential for plasmid monomerization and stability. *Cell*, **36**, 1097–1103.
18. Alén, C., Sherratt, D. J. & Colloms, S. D. (1997). Direct interaction of aminopeptidase A with recombination site DNA in Xer site-specific recombination. *EMBO J.* **16**, 5188–5197.
19. Bath, J., Sherratt, D. J. & Colloms, S. D. (1999). Topology of Xer recombination on catenanes produced by lambda integrase. *J. Mol. Biol.* **289**, 873–883.
20. Colloms, S. D., McCulloch, R., Grant, K., Neilson, L. & Sherratt, D. J. (1996). Xer-mediated site-specific recombination *in vitro*. *EMBO J.* **15**, 1172–1181.
21. McCulloch, R., Coggins, W., Colloms, S. D. & Sherratt, D. J. (1994). Xer-mediated site-specific recombination at *cer* generates Holliday junctions *in vivo*. *EMBO J.* **13**, 1844–1855.
22. Bregu, M., Sherratt, D. J. & Colloms, S. D. (2002). Accessory factors determine the order of strand exchange in Xer recombination at *psi*. *EMBO J.* **21**, 3888–3897.
23. Murasugi, K. (1996). *Knot Theory, Its Applications* (Translated by B. Kurpita), Birkhauser, Boston, MA.
24. Rolfsen, D. (1976). *Knots Mathematics Lecture Series 7*, Publish or Perish, Berkeley, CA.
25. Burde, G. & Zieschang, H. (1985). Knots. In *de Gruytier Studies in Mathematics* (Gabriel, P., ed.), vol. 5, Walter de Gruyter, Berlin.
26. Conway, J. H. (1967). An enumeration of knots and links, and some of their algebraic properties. *Computational Problems in Abstract Algebra*, Pergamon, Oxford, UK pp. 329–358.
27. Goldman, J. R. & Kauffman, L. H. (1997). Rational tangles. *Advan. Appl. Math.* **18**, 300–332.
28. Ernst, C. & Sumners, D. W. (1999). Solving tangle equations arising in a DNA recombination model. *Math. Proc. Cambridge Phil. Soc.* **126**, 23–36.
29. Saka, Y. & Vazquez, M. (2002). TangleSolve: topological analysis of site-specific recombination. *Bioinformatics*, **18**, 1011–1012.
30. Vazquez, M. (2000). Tangle analysis of site-specific recombination: Gin and Xer systems. PhD dissertation in mathematics, Florida State University, Tallahassee, FL.
31. Van Duyne, G. D. (2001). A structural view of Cre-*loxP* site-specific recombination. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 87–104.
32. Arciszewska, L. K. & Sherratt, D. J. (1995). Xer site-specific recombination *in vitro*. *EMBO J.* **14**, 2112–2120.
33. Stark, W. M., Sherratt, D. J. & Boocock, M. R. (1989). Site-specific recombination by Tn3 resolvase: topological changes in the forward and reverse reactions. *Cell*, **58**, 779–790.
34. Gourlay, S. C. & Colloms, S. D. (2004). Control of Cre recombination by regulatory elements from Xer recombination systems. *Mol. Microbiol.* **52**, 53–65.
35. Gopaul, D. N., Guo, F. & Van Duyne, G. D. (1998). Structure of the Holliday junction intermediate in Cre-*loxP* site-specific recombination. *EMBO J.* **17**, 4175–4187.
36. Sträter, N., Sherratt, D. J. & Colloms, S. D. (1999).

X-ray structure of aminopeptidase A from *Escherichia coli* and a model for the nucleoprotein complex in Xer site-specific recombination. *EMBO J.* **18**, 4513–4522.

37. Lickorish, W. B. R. (1981). Prime knots and tangles. *Trans. Am. Math. Soc.* **267**, 321–332.

38. Bleiler, S. A. (1984). Knots prime on many strings. *Trans. Am. Math. Soc.* **282**, 385–401.

39. Hirasawa, M. & Shimokawa, K. (2000). Dehn surgeries on strongly invertible knots which yield lens spaces. *Proc. Am. Math. Soc.* **128**, 3445–3451.

40. Ernst, C. (1996). Tangle equations. *J. Knot Theory Ramif.* **5**, 145–159.

41. Ernst, C. (1997). Tangle equations II. *J. Knot Theory Ramif.* **6**, 1–11.