Math. Proc. Camb. Phil. Soc. (2004), **136**, 565
DOI: 10.1017/S0305004103007266 *Printed in the United Kingdom*

Tangle analysis of Gin site-specific recombination

By MARIEL VAZQUEZ

Department of Mathematics, University of California, Berkeley, CA 94720-3840, *U.S.A.*
e-mail: mariel@math.berkeley.edu

and DE WITT SUMNERS

Department of Mathematics, Florida State University, Tallahassee, FL 32306-4510, *U.S.A.*
e-mail: sumners@math.fsu.edu

(*Received* 21 *August* 2002; *revised* 18 *October* 2002)

Abstract

We use the tangle model to study the action of the site-specific recombinase Gin, an enzyme that can introduce topological changes on circular DNA molecules. Gin and its bound DNA are modelled as a 2-string tangle which undergoes changes during recombination, thereby changing the topology of the DNA substrate. We show that the tangles involved in the analysis are all rational tangles. This technique allows us to prove that, under the model's assumptions, there is a unique topological description of the enzymatic action. The Gin system is one of the few to date where tangle analysis can be carried out systematically and rigorously, yielding a single, biologically reasonable solution.

1. *Preview and motivation*

Site-specific recombination alters the genome of an organism by moving, inserting or inverting DNA segments. When acting on circular DNA molecules, site-specific recombinases can change the topology and geometry of the molecules [**25**, **31**]. The tangle model provides mathematical tools to analyze such changes [**26**, **27**, **28**]. The detailed tangle formalism and an application of the model to the Tn3 resolvase system were first presented in [**10**].

Here we analyze the enzymatic action of Gin, a site-specific recombinase encoded by bacteriophage Mu (a virus that infects bacteria). Gin inverts a DNA segment within the phage genome to extend the phage's range of infection. The phage genome contains two recombination sites that Gin recognizes, cleaves and rearranges to complete one round of recombination (Figure 1). Gin recombination is processive, *i.e.* it can carry out more than one recombination round while binding only once to its substrate. We study the case where, prior to recombination, the phage genome is a single circular DNA molecule, effectively the unknot in S^3. Recombination can then be analyzed by using tangles to model an enzyme such as Gin together with the DNA bound to the enzyme (see: [**29**]). Tangles, defined formally below, are

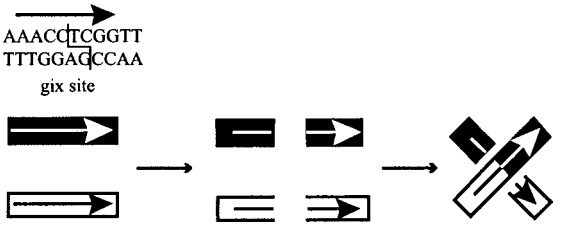

Fig. 1. *Local action of site-specific recombination*
The figure shows the two *gix* recombination sites undergoing recombination. The recombination sites are short (34 bp long) segments of DNA; in this representation the primary Crick-Watson twists are omitted. The base pair sequence of the *gix* sites is illustrated in the upper left corner. The sequence is not palindromic and therefore an orientation can be assigned to it; the figure shows one of the two possible orientations for each site. The "cut" through the nucleotide sequence indicates the cleavage region of the sites.

mathematical representations of the local interrelations between the enzyme and the DNA. Changes in the tangles represent local rearrangements, which can alter the topological structure of the circular DNA molecule, detected by changes of knot and link types. Tangle analysis seeks to characterize the local enzymatic action, given the observed knot types of DNA substrate and products after one or more recombination rounds.

In tangle analyses key simplifications usually occur if the tangles involved are known to be of a special form, *i.e.* to be rational tangles [**10**]. Previous applications of the tangle model to site-specific recombination have often been limited by the fact that tangle rationality had to be assumed in certain arguments (e.g. [**6**]). As in [**10**], we are able to prove rationality at the key points, by using double branched cyclic covering spaces of tangles and by using known results [**15**] on Dehn surgeries on knots. We find that solving tangle equations based on reasonable mathematical and biological assumptions systematically gives the same Gin recombinational mechanism as that suggested by direct biological reasoning, and, in addition, shows the uniqueness of the mechanism.

In this paper we will proceed as follows: first introduce the basic mathematical definitions of the tangle model (Section 1); then describe the biology of the Gin site-specific recombination system (Section 2); then specify the way the tangle model is applied to the analysis of site-specific recombination, in particular to the Gin system (Section 3); then give the mathematical results obtained in the Gin system analysis (Section 4); and finally outline some further challenges (Section 5), and conclusions (Section 6).

1. *Basic definitions for the tangle model and some previous mathematical results*

In this section we state some definitions needed to do tangle analysis and some previous results relevant to the present application of the model. Objects here defined live in the locally-flat PL or smooth category. A 2-string tangle is a pair (B, t) where B is a fixed oriented topological 3-ball in euclidean 3-space, and t is a pair of non-oriented mutually disjoint spanning arcs properly embedded in B [**3**, **22**]. Two 2-string tangles $B_1 = (B, t_1)$ and $B_2 = (B, t_2)$ defined in the same ball B, with identical endpoints for the spanning arcs are equivalent if there is an ambient isotopy fixed on the boundary of B mapping t_1 onto t_2. In that case we simply write $B_1 = B_2$.

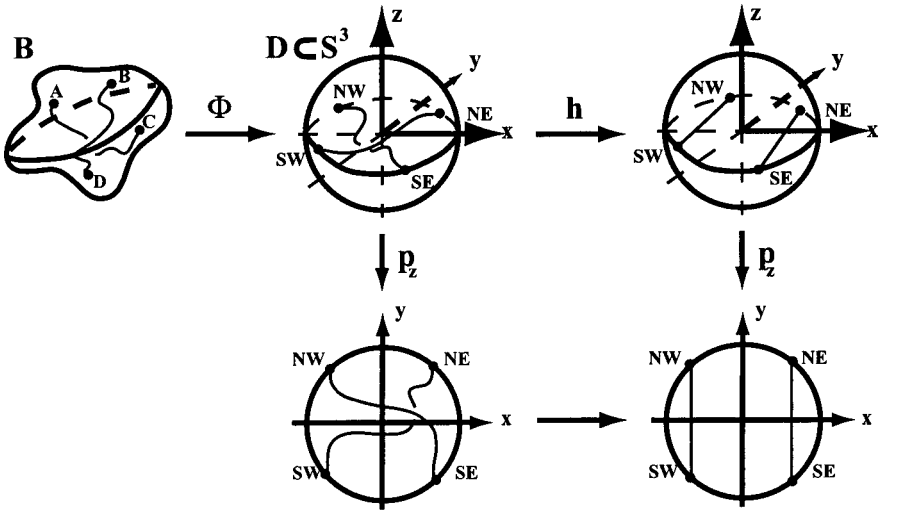

Fig. 2. *Definition*
The figure illustrates the definitions of 2-string tangle, rational tangle and tangle diagram. B is a topological 3-ball in S^3, D is the unit ball in S^3 and t two arcs properly embedded in D. The endpoints of the arcs t are fixed on the boundary of the ball at the four preferred equatorial points $\{NE, NW, SE, SW\}$. Φ is an orientation preserving homeomorphism of pairs mapping B together with its two arcs onto (D, t). h is an orientation preserving homeomorphism of pairs that is the identity on ∂D, and maps (D, t) onto (D, t_0), where $(D, t_0)$is the tangle on the upper right, also called the infinity tangle. p_z is the orthogonal projection onto the XY axis, that maps (D, t) and (D, t_0) onto their corresponding tangle diagrams.

There are three different types of tangles: rational, locally knotted and prime. A tangle (B, t) is *rational* if there is a homeomorphism of pairs, $h\colon (B, t) \to (D, t_0)$, where D is the unit ball in Euclidean 3-space, centered at the origin, oriented with the right-hand coordinate sytem as shown in Figure 2, and t is a pair of straight arcs in the equatorial (XY) plane, parallel to the Y axis, that intersect the x-axis at $-\sqrt{(2)}/2$ and $\sqrt{(2)}/2$. All tangle homeomorphisms are here required to be orientation preserving on the 3-ball B. Call (D, t_0) the "infinity" tangle, and $\{NE, NW, SE, SW\}$ the points where the arcs intersect the boundary of D (Figure 2). Note that in a rational tangle, if the end-points of one of the arcs of t is NE, then the other end-point of that same arc can be any of the three remaining equatorial points, and is not forced to be SE as suggested by the infinity tangle.

A tangle (B, t) is *locally knotted* if there is a 2-sphere S in B that intersects either of the two arcs in t transversely in two points, and such that the 3-ball bounded by S in B intersects t in a knotted arc with end-points on S. A tangle is *prime* if it is neither rational nor locally knotted. A tangle that is not locally knotted is referred to as *locally unknotted*; locally unknotted tangles are either prime or rational. Note that by definition any tangle is of one, and only one, of the three defined types.

We assign to each tangle (B, t) an orientation preserving homeomorphism of pairs, $\Phi\colon (B, t) \to (D, t_\Phi)$ that maps B onto the unit 3-ball D, and the endpoints of t onto NE, NW, SE and SW (Figure 2). In order to compare tangles in different 3-balls, from now on we consider all tangles as defined on the unit ball D (via Φ), with arcs anchored in the four equatorial points. This allows us to analyze tangles through their tangle diagrams (obtained by projecting the arcs onto the equatorial disk).

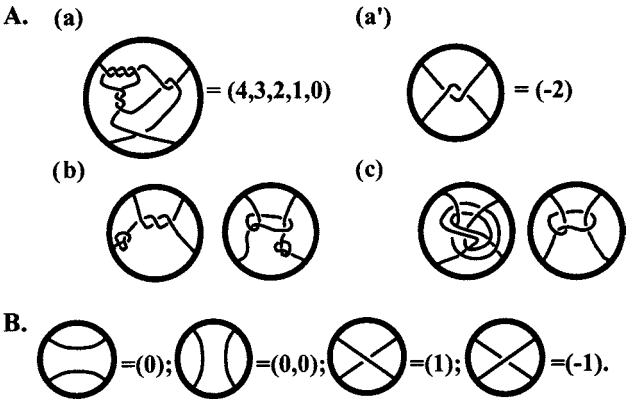

Fig. 3. *Tangle diagrams*
(A) The three different kinds of tangles: panels (a) and (a') show rational tangles; (b) locally knotted tangles; (c) prime tangles. The tangle in (a') is called an integral tangle. **(B)** Shown are the four trivial tangles. These tangles are rational, they are treated separately because their associated Conway symbols do not satisfy the convention. Note that the Conway symbol associated to the infinity tangle is $(0, 0)$.

Given the tangle (D, t_1), another tangle (D, t_2) equivalent to (D, t_1) can always be found such that the projection p_z of the arcs onto the equatorial disk is regular. A *tangle diagram* is the image of (D, t_2) under the projection p_z. Two tangle diagrams represent equivalent tangles if they differ by a finite sequence of Reidemeister moves in the interior of the equatorial disk (see [**1**]). Figure 3 shows tangle diagrams for the three different types of tangles: rational (Figures 3Aa, 3Aa', 3B); locally knotted (Figure 3Ab); prime (Figure 3Ac).

The classification of rational tangles is crucial for the tangle analysis of site-specific recombination. To each equivalence class of rational tangles corresponds a classifying vector, called the *Conway Symbol*. The Conway symbol, an integer entry vector $(a_1, a_2, \ldots, a_m)$, satisfies the following conditions: $|a_1| > 1$; all entries are non-zero, except possibly for a_m; and all entries have the same sign. Four exceptional tangles are excluded from this convention, they can be visualized in Figure 3, panel B, together with their standard vectors. The classification of rational tangles states that there exists a one-to-one correspondence between equivalence classes of rational tangles and the extended rational numbers $q/p \in Q \cup \{\infty\}$ with $p \in N \cup \{0\}, q \in Z$ and $(p, q) = 1$ [**3**, **4**, **14**]. Figure 3 panels A(a) and A(a') show rational tangles and their classifying vectors. The unique extended rational number $q/p \in Q \cup \{\infty\}$ associated to the Conway symbol $(a_1, a_2, \ldots, a_m)$ is obtained by a continued fraction calculation as follows:

$$\frac{q}{p} = a_m + \cfrac{1}{a_{m-1} + \cfrac{1}{a_{m-2} + \cfrac{1}{\cdots + \cfrac{1}{a_1}}}}$$

Several useful operations can be defined between tangles. Tangle addition is defined in Figure 4, panel A. Note that the sum of two rational tangles is not necessarily

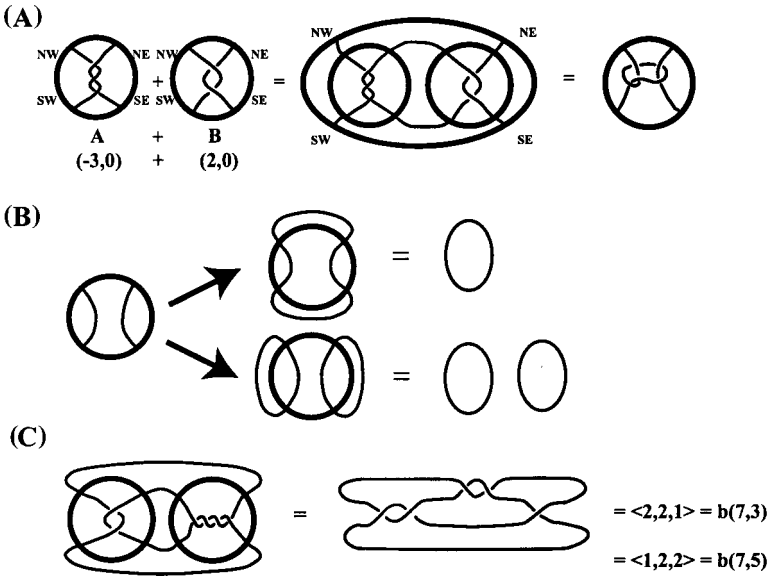

Fig. 4. *Operations on tangles*
(A) Tangle Addition: in this example the two intervening tangles are rational but their sum is prime. **(B)** The Numerator and Denominator operations produce knots and links. **(C)** The numerator of the sum of two rational tangles is a 4-plat. Every 4-plat can be drawn as a closed braid in 4 strands, with one untangled strand, as illustrated here. To this diagram can be associated two vectors with integer entries of the form $\prec c_1, c_2, ..., c_{2n+1} \succ$ and $\prec c_{2n+1}, c_2, ..., c_1 \succ$. If the crossing sign convention is as in the figure, the entries in the vector are positive, and they classify it. To each vector can be associated a rational number $\frac{\beta}{\alpha}$, as explained in the text; the resulting $b(\alpha, \beta)$ are the Conway Symbols for the 4-plat.

a rational tangle. Figure 4, panel B, is used to define N the *numerator* and D the *denominator* operations, that produce knots or 2-component links from tangles.

If A is a rational tangle then $N(A)$ is a 4-plat. It is known that 4-plats, 2-bridge and rational are different names for the same family of knots and links [**3**]. The numerator operation relates the family of rational tangles with that of 4-plats. If A or B are any rational tangles (including the trivial tangles), then $N(A+B)$ is a 4-plat as illustrated in Figure 4. 4-plats are defined informally in the figure as closed braids in 4 strands. Each 4-plat K is characterized by a vector of positive integers $\prec c_1, c_2, \ldots, c_{2n+1} \succ$ (Figure 4C), and by a rational number

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

In this case, the knot as link K is denoted by $b(\alpha, \beta)$. By convention, if $\alpha = 1 = \beta$ then K is the unknot; if $\alpha = 0$ and $\beta = 1$ then K is the unlink of two unknotted components; in all other cases $0 < \beta < \alpha$. The Classification theorem of 4-plats (reviewed in [**3**]) states that $b(\alpha, \beta)$ and $b(\alpha', \beta')$ are equivalent as non-oriented links if, and only if, $\alpha = \alpha'$ and $\beta^{\pm 1} \equiv \beta' (\mathrm{mod} \alpha)$.

The classifications of rational tangles and of 4-plats permit one to find solutions to equations of the form $N(A+B) = K$ where A and B are rational tangles. Ernst and Sumners [**10**] proved the following lemma that translates the topological formulation

of $N(X+A)=K$ into a numerical problem. This result is key to solving systems of rational tangle equations such as those arising from site-specific recombination reactions, as will be shown in Sections 3 and 4.

LEMMA 1·1 ([**10**]). *Given two rational tangles $X=(u/v)$ and $A=(x/y)$ then $N(X+A)=b(\alpha,\beta)$ is a 4-plat, where $\alpha=|xv+yu|$, and β is determined as follows*:

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

In general, from site-specific recombination events arise one or more systems of tangle equations of the form:

$$\begin{cases} N(O+P)=K_0 \\ N(O+R)=K_1 \\ N(O+2R)=K_2 \\ \quad\cdots \end{cases}$$

where K_0 is the substrate, K_1 is the product of one round of recombination, K_2 is the product of two rounds and so on (see [**8**, **29**]). Substrate and products of recombination are known 4-plats $(K_0,K_1,K_2,\ldots)$, but the intervening tangles O, P and R are unknown. In order to solve the system of equations one needs to show that $O+P$, $O+R$, $O+2R$ etc., are rational tangles, or sums of rational tangles. Determining when the intervening tangles are rational, or sums of rational tangles is not easy. Ernst and Sumners have extensively studied this problem [**10**–**13**]. The next few concepts and results will be used in Section 4 to prove that the tangles O and R arising in the system corresponding to Gin recombination are rational.

A 3-manifold M is *irreducible* if any 2-sphere embedded in M bounds a 3-ball in M. M has *incompressible boundary* if any curve that is essential in the boundary of M is also essential in M (*i.e.* no essential curve in the boundary bounds a disk in M). For example the solid torus $T=D^2\times S^1$ is irreducible and has compressible boundary. The following result gives a characterization of 2-string tangles in terms of their double branched cyclic covering spaces. This result will be used to detect tangle rationality.

THEOREM 1·2 ([**22**]). *Let $X=(D,t)$ be a tangle and X' its double branched cyclic covering, then*:

(i) *X is rational $\Leftrightarrow$ X' is a solid torus*;
(ii) *X is prime $\Leftrightarrow$ X' is irreducible and has incompressible boundary*;
(iii) *X is locally knotted $\Leftrightarrow$ X' is not irreducible.*

Given tangles A and B, and a 4-plat K such that $N(A+B)=K$, a few things can be said about the nature of A and B. Since all 4-plats different from the unknot or the unlink of two unknotted components are prime, if A and B are tangles such that $N(A+B)=K$ a 4-plat, then at most one of A or B is locally unknotted. This follows from the observation that if each of A and B contain a local knot, then $N(A+B)=K$ cannot be prime (or unknotted, or the unlink of two unknotted components).

THEOREM 1·3 ([**2**, **10**, **22**]). *Let A and B be locally unknotted tangles and K be a 4-plat such that $N(A+B)=K$, then A or B is rational.*

Stronger statements like the following can be made for systems of equations.

THEOREM 1·4 ([**10**]). *Suppose that X is a tangle, and that there exist tangles A_i for $1 \leqslant i \leqslant 3$, with A_2 and A_3 locally unknotted, such that the following three equations hold*:

(i) $N(X + A_1) = b(1, 1)$;

(ii) $N(X + A_2) = b(\alpha, \beta)$ *with* $\alpha > 1$;

(iii) $N(X + A_3) = b(\alpha', \beta')$ *with* $\alpha' > 1$.

If $|\alpha - \alpha'| > 1$, *then the 2-fold branched cyclic covering of X, denoted by X', is a torus knot complement. If in addition $\alpha = 2$, 3 or 4, then X' is a solid torus.*

The proof of this result makes use of the Cyclic Surgery Theorem [**7**] and of known results of Dehn surgery on torus knots [**23**].

Finally we state a result that will be used in the proof of Theorem 4·2.

THEOREM 1·5 ([**13**, **22**]). *If R is a prime tangle, and O is a locally unknotted tangle such that $O \neq (0, 0)$, then $O + R$ is a prime tangle.*

These results will be used in Section 4 to prove the theorems for Gin action. In the next section we introduce the Gin recombination system.

2. *Gin, the site-specific recombination system of bacteriophage Mu*

Bacteriophages are viruses that infect bacteria (see [**17**] for more information about viruses). There are two possible pathways of viral infection: the *lytic pathway* and the *lysogenic pathway*. During a lytic pathway, the viral DNA replicates in the bacterial cytoplasm, its genes are expressed and the virus is reproduced. The new daughter viruses lyse their host (*i.e.* kill the cell). In the lysogenic pathway the phage DNA becomes integrated into the host genome. Upon replication of the bacterial chromosome the phage DNA is also replicated and passed to daughter cells after division; no new viruses are created, and no cells are lysed. In some cases, the viral DNA excises from the host chromosome and initiates replication, phage reproduction and lysis of the bacterium. Viruses that can follow either pathway are called *temperate*.

Bacteriophage Mu is a temperate bacteriophage with a wide range of infection, e.g. *Escherichia coli* K-12 strains, *Escherichia coli C* strains, *Citrobacter freundii*, *Shigella sonnei*, *Serratia marcescens*, *Enterobacter cloacae* and several *Erwinia* species. The phage genome consists of 37-42 kilobases (kb) of double-stranded DNA (*i.e.* Crick–Watson DNA double-helix, abbreviated by dsDNA) containing the 3.0 kb invertible G-segment. G-segment inversion extends the range of bacterial hosts that bacteriophage Mu can infect, and is mediated by the phage encoded site-specific recombinase Gin. Locally, Gin binds the DNA substrate at each of two 34 base pairs (bp) long recombination sites, called *gix L* and *gix R*. Gin mediates a double-strand break at each site, rotates the broken ends, and reconnects (recombines) the DNA (Figure 1). dsDNA is composed of two sugar-phosphate backbones, each sugar is bound to one of 4 bases (nucleotides A, G, C, T), which carry the genetic information. The nucleotide sequence defined by each of the the *gix* sites can be represented as a specific string in 34 letters. Since the string is not palindromic, one can assign a local orientation to each site (Figure 1). In the cell, the G-segment lies almost completely in one of the domains between the *gix* sites (Figure 5). Since the substrate DNA molecule is

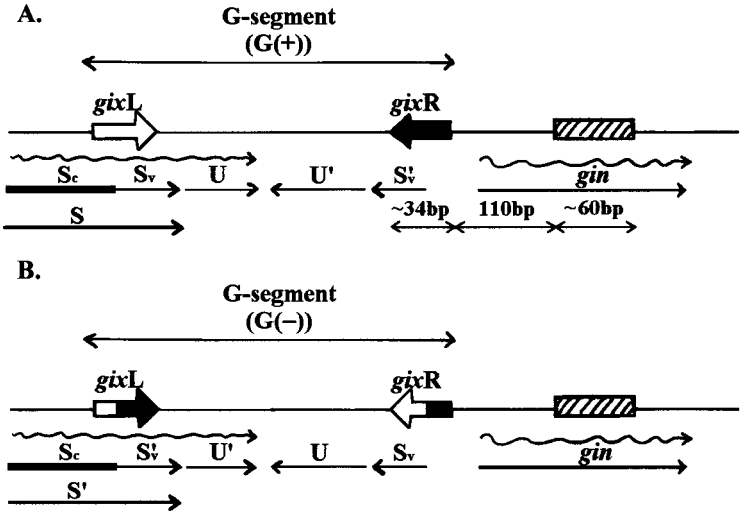

Fig. 5. *G-segment*
(A) The figure shows the region from the host genome that contains the G-segment prior to recombination. The G-segment is in the $G(+)$ orientation. S_c, S_v, and U are genes that encode for the tail-fiber proteins that allow the phage to infect certain strains of bacteria. The hatched region corresponds to the recombinational enhancer sequence. **(B)** After inversion the G-segment is in the $G(-)$ orientation, thus inducing transcription of $S' = S_c + S'_v$ and U' genes. Expression of S' and U' genes allows infection of different strains of bacteria.

circular, the orientation of each site induces an orientation on the circle. If the orientations induced by the two sites are the same, then the sites are *directly repeated*, otherwise *inversely repeated*. In the cell, when acting on unknotted substrates, Gin recombination is efficient only on inversely repeated sites, where a single round of recombination reverses (inverts) the orientation of the G-segment, switching from G(+) to G(−) (Figure 5). Nonetheless, when subjected to certain experimental conditions, somewhat efficient recombination has also been achieved on substrates with directly repeated sites [**19**], although no G-segment inversion occurs in this case.

Gin recombination requires three accessory factors: negatively supercoiled DNA; the host-encoded protein Fis (factor for inversion stimulation); and an enhancer sequence. The recombinational enhancer sequence (hatched region in Figure 5) is 60 bp long and is located approximately 110 bp to the right of *gix* R. The enhancer contains two binding sites for Fis. Both Gin and Fis must bind to the DNA molecule, and the enhancer must be present for full recombinational activity. Prior to recombination, Gin binds the *gix* sites, and Fis binds the enhancer, thus forming the *synaptic complex*.

At the sequence level, G-segment inversion results in genomic changes as illustrated in Figure 5. When observed in a molecule-wide context, Gin recombination on circular DNA substrates can result in topological changes. Such changes have been characterized experimentally [**18**, **19**]. In those experiments, all substrates were genetically identical, circular, unknotted DNA molecules with two *gix* sites. Two sets of experiments were done, one with directly repeated recombination sites, the other with inversely repeated sites. Substrates were reacted with purified enzyme harvested from living cells. Reaction products were fractionated by gel electrophoresis. Roughly speaking, DNA knots and links with same, small, crossing number migrate in the gel with same velocity [**24**], resulting in an array of discrete bands in the gel.

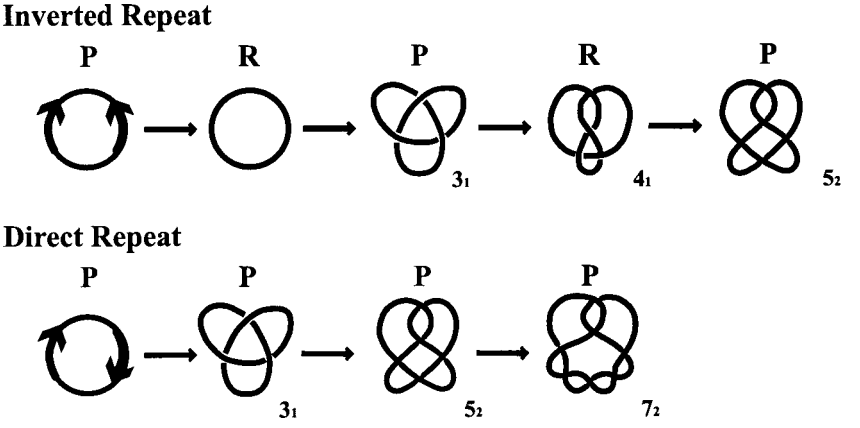

Fig. 6. *Gin products*
Products of Gin recombination on unknotted substrates with inversely repeated (upper row) or directly repeated (bottom row) *gix* sites. $G(+)$ and $G(-)$ denote orientation of the G-segment.

Gin recombination is *processive*, *i.e.* multiple recombination events can occur at a single binding encounter between enzyme and DNA substrate. Several topologically different products were observed as expected for a processive recombination event. It was shown in [**19**] that in the case of inversely repeated *gix* sites, products of recombination migrated as knots with 0, 3, 4 and 5 crossings; and products had 3, 5, and 7 crossings for directly repeated sites. Frequency distributions of the product knot and link types were established, and G-segment orientations of the products (G(+) or G(−)) were determined. Some gel bands were isolated, samples were taken from each band, and the molecules in the samples were purified. The resulting DNA molecules were treated with RecA protein and observed under the electron microscope. Electron micrographs (EM) were used to determine the knot type of each product. From these analyses it was concluded that, when acting on inversely repeated sites, Gin produces, in the following order, the unknot in G(–) orientation, the (–)trefoil in G(+), the figure-8 knot in G(–) and the (–)5-twist knot in G(+). Products of recombination for directly repeated sites were, in order, the (–)trefoil, the (–)5-twist knot and the (–)7-twist knot, all in the G(+) orientation (Figure 6). In [**19**] a model was proposed to account for the experimental data. We here present a mathematically rigorous analysis of Gin recombination. Using the tangle model under biologically reasonable assumptions, we show that the model proposed in [**19**] is the unique topological mechanism for the enzymatic action.

3. *Applying the tangle model to site-specific recombination*

In Ernst and Sumners [**10**] the authors successfully introduced and used tangle analysis (reviewed in [**8**, **30**]) to elucidate the topological mechanism of the Tn3 resolvase.

The motivation for the tangle model [**29**] can be seen from electron micrographs of the Tn3 resolvase synaptic complex that show unkotted DNA molecules bound to site-specific recombinases as dark spots (enzymes) with two emanating arcs (DNA). This led to a model where the enzyme-DNA complex is a 2-string tangle E ("enzyme"). The complement $S^3 - E$, together with the two DNA arcs not bound

to the enzymatic complex, is another tangle O_f ("outside free" tangle). The synaptic complex, before recombination, is represented by the tangle equation $N(O_f + E) = b(1, 1)$. For simplicity the recombination sites are confined in the interior of a ball $P \subset E$, and therefore recombination ("cut-and paste") is restricted to P.

Certain assumptions, both biological and mathematical, have to be made before doing tangle analysis. The first assumption is that the enzyme/DNA complex can be represented as a sum of tangles $E = O_b + P$, where O_b ("outside bound") contains the DNA bound to the enzymatic complex, but that remains unchanged after recombination. The second assumption is that site-specific recombination acts by tangle surgery, where the tangle P is converted into a tangle R after one round of site-specific recombination. One round of recombination can then be modelled by a system of two tangle equations:

$$N(O_f + O_b + P) = K_0 = \text{ substrate}$$
$$N(O_f + O_b + R) = K_1 = \text{ product.}$$

The third assumption is that the mechanism of recombination is constant, independent of substrate geometry and topology. What this means is that if all substrates are topologically identical (all have the same knot or link type, K_0), then O_f, O_b, P and R do not change from one event to another. It also implies that if the same enzyme were to act on topologically different substrates, then all substrate differences would be acounted for in O_f, and the other tangles would remain unchanged from one experiment to the other (except perhaps for the site orientation that could affect R) thus reflecting enzyme binding (O_b). Finally, it is assumed that processive recombination acts by tangle addition, *i.e.* after n rounds of recombination, P is converted into $nR = R + R + \cdots + R$. For simplicity we call $O = O_f + O_b$ the "outside tangle", n rounds of recombination are then modelled as a system of n equations, where O, P and R are unknown

$$N(O + P) = K_0 = \text{ substrate}$$
$$N(O + R) = K_1 = \text{ first product}$$
$$\cdots$$
$$N(O + nR) = K_1 = n\text{th product.}$$

It has been observed, through electron micrographs, that when the substrates are unknotted, the strands emanating from the enzymatic complex are untangled, thus implying that O_f is a trivial tangle (*i.e.* one of (0), (1), (–1), (0.0)). Therefore, any topological feature of the tangle O mainly reflects what is happening upon DNA binding by the site-specific recombinase and its accessory proteins.

The work of Ernst and Sumners allows to find solutions for O and R that are rational or sums of rational tangles, when substrate and products of recombination are known 4-plats [**10**–**13**]. If only two equations are provided, it is sometimes impossible to find a unique solution pair (O, R) for the equations. If two or more products are observed, and are known to result from processive recombination, then a small set of solutions for tangles O and R can usually be obtained with only a few rounds of recombination. If a unique solution pair is achieved then results of further rounds can be predicted.

Showing that O and R are rational tangles is difficult when only two equations are available. If two or more products of recombination are known, there are results that make the task easier, as illustrated by Theorem 1·4 [**10**] that will be used in the analysis of Gin recombination. These results make use of the theory of Seifert fibered spaces, and of the Cyclic Surgery Theorem [**7**].

For the particular case of Gin recombination, the assumption of constant mechanism has been verified experimentally under standard conditions [**19**]. Hin is another site-specific recombinase whose action is believed to be identical to that of Gin. Electron-micrographs of the Hin recombinase enzymatic complex show the synaptic complex as a dark spot with three emanating, untangled, DNA loops [**16**]. Extending this to Gin and Fis, one can infer that the enzymes together with the bound DNA can be regarded as a 3-string tangle. However, 3-string tangles are complex objects and are difficult to classify. Recent work due to Emert and Ernst [**9**] and to Cabrera (personal communication) give some characterization of rational 3-string tangles. It remains difficult to manipulate these objects, and no tools are yet available to detect rationality of 3-string tangles. This represents an apparent limitation to the tangle analysis of Gin recombination. Nonetheless, there exists biological evidence that the accessory protein Fis does not play an active role in recombination, and that its participation in the synaptic complex is transitory. We assume that one of the three loops emerging from the 3-string tangle enzymatic complex remains completely fixed by Fis throughout the recombination event. If we then deform the ball that represents the enzymatic complex to include that loop in its interior, the resulting figure is one in which only two loops emanate from the ball, *i.e.* a 2-string tangle (or sum of 2-string tangles). Furthermore, from the electron-micrographs, O_f can be chosen to be (0). In this case $O = O_f + O_b = O_b$, and solving the equations for O and R will give all necessary information to reconstruct the initial conformation of the DNA inside the ball (tangle O), and the process of strand-exchange (tangle R).

4. *Results*: *tangle analysis of Gin recombination*

We first consider the case of Gin recombination on unknotted substrates with inversely repeated *gix* sites, and G-segment in the G(+) orientation (K_0). Products of recombination have been shown experimentally to be, in order, K_1 the unknot with G(–), K_2 the (–) trefoil (3_1) with G(+), K_3 the figure-8 knot (4_1) with G(–), and K_4 the (–)5-twist knot (5_2) with G(+) ([**10**], illustrated here in Figure 6). From these data results a system of five tangle equations: one substrate $N(O + P) = K_0$ (equation (i)); and four products $N(O + iR) = K_i$ (equations (ii), (iii), (iv), (v)). We will here show (Theorem 4·2) that four equations are sufficient to ensure a unique solution to the system, and accurately predict the product in the fifth equation. Before stating the theorem we give a very simple claim that, given the data, restricts the choices for the intervening tangles O, P and R. The claim uses the concept of parity.

The *parity* of a tangle [**29**] describes which pairs of points in $\{NE, NW, SE, SW\}$ are connected by each tangle arc. A tangle (A, t) has *parity* (0) (denoted by $A \approx (0)$), if the arcs of t connect NW to NE, and SW to SE; A has *parity* (1) if the arcs connect NW to SE and SW to NE; and, A has *parity* $(0, 0)$ if the arcs connect NW to SW and NE to SE.

Parity	$N(O+R)$	knot/link	$N(O+R+R)$	knot/link
$O \sim (1)$ $R \sim (1)$		link		knot
$O \sim (0)$ $R \sim (0)$		link		link
$O \sim (\infty)$ $R \sim (0)$		knot		knot
$O \sim (1)$ $R \sim (0)$		knot		knot
$O \sim (0)$ $R \sim (1)$		knot		link
$O \sim (\infty)$ $R \sim (1)$		knot		knot

Fig. 7. *Parity table*
(Note: in this table the tangle (0,0) is denoted by (∞).)

CLAIM 4·1 (Inversely Repeated Sites). *The tangles involved in equations* (*i*), (*ii*) *and* (*iii*) *of Theorem* 4·2 *arising from Gin recombination on inversely repeated sites satisfy the following properties*:

$$0 \approx (0,0), R \approx (1) \text{ and } P \approx (0).$$

Proof. R cannot have parity $(0,0)$ since $N(0+2R)$ would have at least two components, but $N(0+2R)$ is a knot (only one component). All observed products of recombination are knots. Figure 7 applied to equations (*i*), (*ii*) and (*iii*), shows that the only "parity" possibilities for O and R are the following:

$$0 \approx (0,0) \text{ and } R \approx (0);$$
$$0 \approx (1) \text{ and } R \approx (0);$$
$$0 \approx (0,0) \text{ and } R \approx (1).$$

In the tangle model, O remains unchanged throughout the recombination event. It can readily be seen that if R and R + R had the same parity, then N(O + R) and N(O + R + R) would have the same G-segment orientation, thus contradicting the experimental data. Therefore R cannot have parity (0), and the only possibility is:

$$0 \approx (0,0) \text{ and } R \approx (1).$$

By a similar argument, P and R have different parities, so $P \approx (0,0)$ or $P \approx (0)$. But if $P \approx (0,0)$ and $O \approx (0,0)$ then $N(O+P)$ is a link, contradicting the substrate equation; therefore $P \approx (0)$.

THEOREM 4·2 (Inversely Repeated Sites). *Suppose that the tangles* $O, P,$ *and* R *satisfy the following equations*:

(i) $N(O+P) = \prec 1 \succ$
(ii) $N(O+R) = \prec 1 \succ$
(iii) $N(O+R+R) = \prec 3 \succ$
then $(O,R) \in \{((-2,0),(1));\ ((4,1),(-1))\}$.

If furthermore

(iv) $N(0+R+R+R) = \prec 2,1,1 \succ$ = *figure-8 knot*
then $(O,R) = ((-2,0),(1))$ *is the unique solution pair.*

Proof. First O and R are shown to be rational, then solutions to the tangle equations are computed.

O, P and R are locally unknotted tangles because each of them is a summand of the unknot via the numerator construction. Therefore, each of O, P and R is either rational or prime.

Assume that R is prime. Note that if $O = (0,0)$ then

$$\begin{cases} N(O+R) = D(R), \ \forall R \\ N(O+R+R) = D(R+R), \ \forall R. \end{cases}$$

If R is prime then $D(R+R)$ is a composite knot, contradicting equation (iii). Therefore O cannot be (0, 0). But by Theorem 1·5, R prime, O locally unknotted and $O \neq (0,0)$, imply that $O+R$ is prime. It is known that a 4-plat cannot have two prime summands, so equation (iii) cannot be satisfied if R is a prime tangle. We therefore conclude that R is a rational tangle.

Suppose now that O is prime. Theorem 1·3 applied to equation (i) implies that P is a rational tangle. Thus by Theorem 1·2, P' and R' (2-fold branched cyclic covers of P and R) are solid tori. Furthermore,

$$\begin{cases} N(O+P) = \langle 1 \rangle \\ N(O+R) = \langle 1 \rangle \end{cases} \Rightarrow \begin{cases} [N(O+P)]' \cong [N(O+R)]' \cong S^3 \\ \text{and } \ \partial O' = \partial P' = \partial R' \cong T^2 \end{cases}$$

where T^2 denotes the two-dimensional torus. $O' \bigcup_{T^2} P' \cong S^3$ and P' a solid torus, imply that O' is the bounded complement of some knot K_P in S^3; that is $O' \cong (S^3 \setminus N(K_P))$ where $N(K_P)$ is the interior of a tubular neighborhood of K_P in S^3. Likewise $O' \bigcup_{T^2} R' \cong S^3$ and R' a solid torus implies that $O' \cong (S^3 \setminus N(K_R))$ for some knot K_R in S^3.

The knots K_P and K_R are non-trivial because we assumed O prime. But if O' is the complement of a non-trivial knot, then there is a *unique* way to Dehn fill $(S^3 \setminus O')$ in order to retrieve S^3. In other words, if S^3 is obtained by Dehn surgery in the complement of a knot, and the knot is non-trivial, then this Dehn filling is unique up to ambient isotopy [**15**]. From Claim 4·1 we know that $R \approx (1)$ and $P \approx (0)$, therefore the Dehn fillings determined by P and R are not equivalent (the gluing diffeomorphisms are not ambient isotopic). We conclude that O' is the complement in S^3 of the trivial knot, *i.e.* O' is a solid torus, and therefore O is rational.

We have just proved that both O and R are rational. At least one of O and R must be an integral tangle (defined in Figure 3), because otherwise $N(O+R+R)$ would be a Montesinos knot which is not a 4-plat [**3**] thus contradicting equation (iii). Moreover, O cannot be integral, because integral tangles have parity (0) or parity (1), and O has parity $(0,0)$. Therefore R is integral, so $R = (r)$, $O = u/v$, where u, r are integers, and v is a non-zero natural number. Observe that $r \neq 0$, otherwise $N(O+R) = N(O+2R)$.

Lemma 1·1 is used to solve the tangle equations. The first part of the lemma states that the numerator of the sum of two rational tangles, $A_1 = p/q$ and $A_2 = u/v$, is a 4-plat $b(\alpha, \beta)$ where $\alpha = |pv + qu|$. This applied to equations (ii), and (iii) leads to the

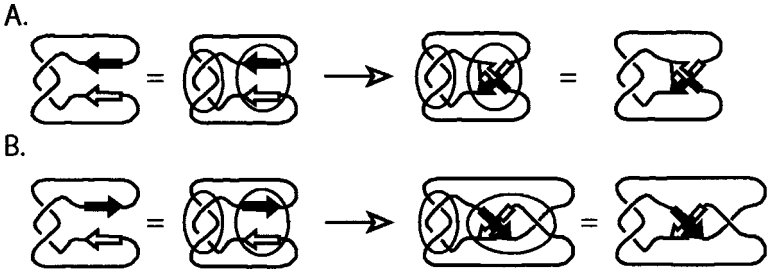

Fig. 8. *Solutions of the tangle equations*
Theorems 4·2 and 4·3 find unique solutions to the two systems of tangle equations posed by Gin recombination on inversely repeated (panel A), and directly repeated sites (panel B). The solutions suggest that two negative supercoils are bound by the accessory protein Fis, thus preparing the synaptic complex for recombination. This corresponds to a tangle $O = (-2, 0)$ that remains fixed throughout the recombination event. When acting on inversely repeated sites, the enzyme binds the two *gix* sites, and with each round of recombination it introduces one positive crossing in the domain ($R = (+1)$). When acting on direclty repeated sites, each round of recombination introduces two positive crossings ($R = (+2)$).

following system of equations, with integral unknowns u, r, v:

$$(1)\ldots. \begin{cases} |u + rv| = 1 \\ |u + 2rv| = 3 \end{cases}.$$

A set of ten solutions is obtained for the ordered pair $(u/v,\ r)$. Therefore there are ten possible solution pairs (O, R) of rational tangles; five different pairs $((-2, 0), (1))$, $((1), (-2))$, $((5), (-4))$, $((-2, -2), (2))$, $((4, 1), (-1))$, and their mirror images. Parity considerations lead us to discard any case where O is integral, and the case where $R = (2)$, leaving only two possible pairs and their mirror images. Furthermore only one from each of the two mirror images satisfies equation (iii) because the product knot is chiral (not equivalent to its mirror image). The complete set of tangle solutions (O, R) to the system of two tangle equations (ii) and (iii) is:

$$\{((-2, 0), (1)), ((4, 1), (-1))\}.$$

Of these two tangle solutions, only $(O, R) = ((-2, 0), (1))$ satisfies tangle equation (iv).

What can be said about tangle P? We showed that $P \approx (0)$ (Claim 4·1) and that P is locally unknotted (Theorem 4·2). However P appears only in equation (i), and a tangle equation in one rational unknown has infinitely many rational solutions [**10**]. The computation of the solutions (O, R) in Theorem 4·2 does not take into account equation (i), however the theorem shows that the system of four tangle equations has a unique solution pair (O,R). It is possible to argue, on biological grounds, that $P = (0)$ [**29**, **30**]. Note that assuming $P = (0)$ would lead to a unique solution to the system with only equations (i)–(iii).

In biological terms, when reacted on a substrate with *gix* sites in the inverted orientation, Gin is specific for the $(-2, 0)$ synaptic complex and each round of recombination adds one positive crossing to the substrate. The existence of a unique solution to the system of tangle equations arising from Gin recombination where only the substrate and three products are specified allows prediction of further products. For example the fourth product of recombination is predicted to be the $(-)5$-twist knot $\prec 2, 2, 1 \succ$ which was also observed experimentally [**19**].

In [**19**] it was found that, under certain conditions, Gin can also act on substrates with directly repeated *gix* sites. We will next show that the proposed mechanism for this reaction is very similar to the one seen in the previous case. The enzyme recognizes and fixes the two negative supercoils as before, emphasizing the specificity for the $(-2,0)$ synaptic complex. But the previous mechanism $R=(+1)$ would in this case result in a mismatch in the sequence, and the loose ends would not be able to religate, forcing the enzyme to act again. We show here that for directly repeated sites, two positive crossings are added after each single event of recombination.

THEOREM 4·3 (Directly Repeated Sites). *Suppose that the tangles O, P and R satisfy the following*:

(i) $N(O+P) = \prec 1 \succ$;
(ii) $N(O+R) = \prec 3 \succ$;
(iii) $N(O+R+R) = \prec 1,2,2 \succ = (-5)$ *twist*;
then $(O,R) \in \{((-2,0),(2));\ ((2,1,1,2),(-2))\}$.
If in addition
(iv) $N(0+R+R+R) = \prec 1,4,2 \succ = (-7)$ *twist*
then $(O,R) = ((-2,0),(2))$ *and*
(v) $N(0+4R) = (-9)$ *twist knot.*
More generally
(n+1) $N(0+nR) = -(2n+1)$ *twist knot.*

Proof. We proceed as in Theorem 4·2, by showing that O and R are rational before finding solutions to the tangle equations using Lemma 1·1.

If O or P are locally knotted, then N(O + P) is a non-trivial knot, contradicting equation (i). If R is locally knotted, then O + R is locally knotted. This means that $N((O+R)+R)$ is composite, contradicting equation (iii). Therefore O, P and R are locally unknotted.

If R is prime then, as in Theorem 4·2, O $\neq (0,0)$ is locally unknotted and by Theorem 1.5, O + R is prime, which contradicts Theorem 1·3. Therefore R must be rational.

If O is prime, then by Theorem 1·3 and equations (i), (ii) and (iii), P, R and R + R are rational. By Theorem 1·2, P', R' and $(R+R)'$ are solid tori.

$$b(1,1) = N(O+P) \Rightarrow S^3 \cong O' \cup_{T^2} P'$$

where $\partial O' = \partial P' = \partial R' \cong T^2$. Therefore O' is the bounded complement of some knot K in S^3, and K is non-trivial (since O is prime). Likewise,

$$b(3,1) = N(O+R) \Rightarrow L(3,1) \cong O' \cup_{T^2} R'$$

Then $L(3,1)$ is obtained by Dehn surgery on some non-trivial knot K. Furthermore $L(7,3)$ is obtained by another Dehn surgery on the same knot K. By Theorem 1·4, O' is is a solid torus (*i.e.* K is trivial). We conclude that O is a rational tangle.

Summarizing, O and R are rational, and P is locally unknotted. As in the proof of Theorem 4·2, we have that O is rational and R is integral. Solving the absolute value equations related to tangle equations (i), (ii) and (iii) leads to two mirror image solution pairs represented by $(O,R) = ((-2,0),(2))$ and $(O,R) = ((2,1,1,2),(-2))$. Only the first solution satisfies equation (iv), thus we have the unique solution $(O,R) = ((-2,0),(2))$ to the system of the four tangle equations (i)-(iv). With this we

can predict $N(O+4R) = 9_2$; in general $N(O+nR)$ is the negative twist knot $(2n+1)_2$. As before, P appears in only one tangle equation, and is therefore undetermined, but one can reasonably argue on biological grounds that $P = (0)$.

5. *Mutant Gin*

In the next few paragraphs we sketch some further steps that we have undertaken in the study of Gin and of a mutant of Gin that does not require accessory factors. Our results will be included in a future paper.

We have given a mathematical proof that Gin recombination is specific for the $(-2, 0)$ synaptic complex, and that upon recombination either one or two positive crossings (*i.e.* $R = (1)$ or $R = (2)$) are added to the domain, depending on the relative orientations of the sites. Furthermore, proper Gin recombination requires three accessory factors: negative supercoiling; an enhancer sequence; and the protein Fis. A few questions arise: how is Gin specific for the $(-2, 0)$ synaptic complex; what are the roles of the accessory factors? A mutant of Gin was created experimentally [**20**, **21**] to address these questions. Recombination by the mutant neither required Fis nor negative supercoiling. It is known that wild-type Gin mediates DNA inversion between two inversely repeated recombination sites, and that this action produces $(-n)$-twist knots exclusively. When reacted with inversely repeated sites [**5**], mutant Gin produced a much wider range of knot types, including torus knots. The spectrum of solutions was so complex that the order in which the products of recombination appeared could not be directly determined from the experiments. The first thing that one may ask is whether all the knots observed can be accommodated in a single recombination pathway, *i.e.* can a system of equations be stated and a solution pair (O, R) be found that generates the complete observed knot spectrum?

As a result, given the knot spectrum generated by mutant Gin, one has to find all possible systems of equations that accommodate the data. If there is no single system that accommodates all the data, then what is the smallest number of systems that will? Once the possible pathways are determined, the corresponding tangles O and R have to be proved rational, and solutions found. We observe that if O is rational and $P = (0)$, with $N(O + P) = \prec 1 \succ$, then $O = (-m, 0)$.

In addition computational methods can be used to find all possible systems of equations (reaction pathways) for a given set of products. In fact by assuming $R = (\pm 1)$ it can be proven that in mutant Gin recombination O cannot be constant, and therefore concluded that mutant Gin uses more than one synaptic complex of the form $O = (-m, 0)$.

6. *Conclusions*

We have given mathematical proofs, based on the tangle model, that in Gin recombination the enzyme and accessory proteins bind the DNA substrates at the two recombination sites, fixing two crossings that correspond to $O = (-2, 0)$. For inversely repeated sites, each round of recombination adds a new crossing to the domain ($R = (+1)$). Similarly, for directly repeated sites $R = (+2)$. These models were previously proposed in an experimental context by Kanaar *et al.* [**19**] as possible mechanistic explanations to Gin data. Our work ensures that under the model's assumptions these solutions are unique. Since the reasoning here is substantially

independent of the arguments given in [**19**], the uniqueness proved for this model is rather convincing.

The Gin system is one of the few where the entire tangle analysis can be done rigorously, starting from simple and plausible assumptions, proceeding systematically without any *ad hoc* steps, and arriving at a unique, biologically reasonable solution. It is thus a potential paradigm for situations where the biologists have not been able to guess the mechanisms of enzymatic action by more intuitive reasoning. We presented a specific proposed application which involves some challenging steps, and that will be included in a future publication.

Acknowledgements. We are very grateful to Rainer Sachs and Javier Arsuaga for helpful comments. We would also like to thank Isabel Darcy for help in the early days, and Joe Gallian for valuable advice. M. Vazquez was mainly supported through a DGAPA-UNAM graduate fellowship. Additional funds were obtained from infrastructure grants from the NSF and the Burroughs Wellcome Fund to the Program in Mathematics and Molecular Biology, and NSF grant DMS-9971169 (MV).

REFERENCES

[**1**] C. C. Adams. *The knot book.* (W. H. Freeman and Company, 1994).
[**2**] S. A. Bleiler. Knots prime on many strings. *Trans. Amer. Math. Soc.* **282** (1984), 385–401.
[**3**] G. Burde and H. Zieschang. *Knots.* De Gruytier Studies in Mathematics. no. 5 (Walter de Gruytier & Co., 1985).
[**4**] J. H. Conway. An enumeration of knots and links, and some of their related properties. *In Computational Problems in Abstract Algebra.* Proc. Conf. Oxford 1967 (Pergamon Press, 1970), pp. 329–358.
[**5**] N. J. Crisona, R. Kanaar, T. N. Gonzalez, E. L. Zechiedrich, A. Klippel and N. R. Cozzarelli. Processive recombination by wild-type Gin and an enhancer-independent mutant insight into the mechanisms of recombination selectivity and strand exchange, *Journal of Molecular Biology* **243** (1994), 437–457.
[**6**] N. J. Crisona, R. L. Weinberg, B. J. Peter, D. W. Sumners and N. R. Cozzarelli. The topological mechanism of phage lambda integrase. *Journal of Molecular Biology* (4) **289(4)** (1999), 747–775.
[**7**] M. Culler, C. McA. Gordon, J. Luecke and P. B. Shalen. Dehn surgery on knots. *Ann. of Math.* (2)**125** (1987), 237–300.
[**8**] I. K. Darcy. Biological distances on DNA knots and links: applications to XER recombination. *In Knots in Hellas '98.* vol. 2 (*J. Knot Theory Ramifications*) **10** (2001), 269–294.
[**9**] J. Emert and C. Ernst. N-string tangles. *J. Knot Theory Ramifications* **9** (2000), 987–1004.
[**10**] C. Ernst and D. W. Sumners. A calculus for rational tangles: applications to DNA recombination. *Math. Proc. Camb. Phil. Soc.* (3) **108** (1990), 489–515.
[**11**] C. Ernst. Tangle equations. *J. Knot Theory Ramifications* (2) **5** (1996), 145–159.
[**12**] C. Ernst. Tangle equations II. *J. Knot Theory Ramifications* (1) **6** (1997), 1–11.
[**13**] C. Ernst and D. W. Sumners. Solving tangle equations arising in a DNA recombination model. *Math. Proc. Camb. Phil. Soc.* (1) **126** (1999), 23–36.
[**14**] J. R. Goldman and L. H. Kauffman. Rational tangles. *Adv. in Appl. Math.* **18** (1997), 300–332.
[**15**] C. M. Gordon and J. Luecke. Knots are determined by their complements. *Bull. Amer. Math. Soc. (N.S.)* (1) **20** (1989), 83–87.
[**16**] K. A. Heichman and R. C. Johnson. The Hin invertasome: protein-mediated joining of distant recombination sites at the enhancer. *Science* **249** (1990), 511–517.
[**17**] W. K. Joklik. *Virology* (Appleton-Century-Crofts, 1985).
[**18**] R. Kanaar. Site-specific DNA inversion in the genome of bacteriophage mu. A topological analysis. Ph.D. thesis. Rijksuniversiteit te Leiden (1988).
[**19**] R. Kanaar, A. Klippel, E. Shekhtman, R. Kahmann, J. M. Dungan and N. R. Cozzarelli. Processive recombination by the phage Mu Gin system: implications for the mechanisms of DNA strand exchange, DNA site alignment, and enhancer action. *Cell* **62** (1990), 353–366.
[**20**] A. Klippel, K. Cloppenberg and R. Kahmann. Isolation and characterization of unusual Gin mutants. *The EMBO Journal* (12) **7** (1988), 3983–3989.

[**21**] A. KLIPPEL, R. KANAAR, R. KAHMANN and N. R. COZZARELLI. Analysis of strand exchange and DNA binding of enhancer-independent Gin recombinase mutants. *The EMBO Journal* (3) **12** (1993), 1047–1057.

[**22**] W. B. R. LICKORISH. Prime knots and tangles. *Trans. Amer. Math. Society* (1) **267** (1981), 321–332.

[**23**] L. MOSER. Elementary surgery along a torus knot. *Pacific J. Math.* **38** (1971), 737–745.

[**24**] A. STASIAK, V. KATRITCH, J. BEDNAR, D. MICHOUD and J. DUBOCHET. Electrophoretic mobility of DNA knots. *Nature* **384** (1996), 122.

[**25**] W. M. STARK, M. R. BOOCOCK and D. J. SHERRATT. Site-specific recombination by Tn3 resolvase. *Trends Genet.* (9) **5** (1989), 304–309.

[**26**] D. W. SUMNERS. The role of knot theory in DNA research. *In Geometry and Topology* (Marcel Dekker, 1987), pp. 297–318.

[**27**] D. W. SUMNERS. Untangling DNA. *Math. Intelligencer* **12** (1990), 71–80.

[**28**] D. W. SUMNERS. Lifting the Curtain: Using topology to probe the hidden action of enzymes. *Notices Amer. Math. Soc.* **42** (1995), 528–537.

[**29**] D. W. SUMNERS, C. ERNST, N. COZZARELLI and S. J. SPENGLER. Mathematical analysis of the mechanisms of DNA recombination using tangles. *Quarterly Reviews of Biophysics* **28** (1995), 253–313.

[**30**] M. VAZQUEZ. Tangle analysis of site-specific recombination: Gin and Xer systems. PhD. thesis. Florida State University (2000).

[**31**] S. A. WASSERMAN and N. R. COZZARELLI. Biochemical topology: applications to DNA recombination and replication. *Science* (4753) **232** (1986), 951–960.