A calculus for rational tangles: applications to DNA recombination

By C. ERNST

*Department of Mathematics, Western Kentucky University, Bowling Green,
Ky. 42101, U.S.A.*

and D. W. SUMNERS

*Department of Mathematics, Florida State University, Tallahassee,
Fla. 32306, U.S.A.*

1. *Introduction*

There exist naturally occurring enzymes (topoisomerases and recombinases),
which, in order to mediate the vital life processes of replication, transcription, and
recombination, manipulate cellular DNA in topologically interesting and non-trivial
ways [24, 30]. These enzyme actions include promoting the coiling up (supercoiling)
of DNA molecules, passing one strand of DNA through another via a transient
enzyme-bridged break in one of the strands (a move performed by topoisomerase),
and breaking a pair of strands and recombining them to different ends (a move
performed by recombinase). An interesting development for topology has been the
emergence of a new experimental protocol, the topological approach to enzymology
[30], which directly exploits knot theory in an effort to understand enzyme action.
In this protocol, one reacts artificial circular DNA substrate with purified enzyme *in
vitro* (in the laboratory); the enzyme acts on the circular DNA, causing changes in
both the euclidean geometry (supercoiling) of the molecules and in the topology
(knotting and linking) of the molecules. These enzyme-caused changes are
experimental observables, using gel electrophoresis to fractionate the reaction
products, and *rec A* enhanced electron microscopy [15] to visualize directly and to
determine unambiguously the DNA knots and links which result as products of an
enzyme reaction. This experimental technique calls for the building of knot-theoretic
models for enzyme action, in which one wishes mathematically to extract information
about enzyme mechanism from the observed changes in the DNA molecules.

This paper deals with the mathematics which arises in a topological model for
enzyme mechanism [25, 26, 27]. The mechanism of many enzymes involves local
(near the enzyme) interaction of two DNA strands. The mathematics which can be
used to model this 2-strand interaction is that of the 2-string tangle. When bound to
a circular DNA molecule, the enzyme naturally separates the DNA molecule into two
complementary tangles. Enzyme action on circular DNA can be viewed as tangle
surgery, in which the action of the enzyme is to delete one of these tangles, replacing
it by another. One regards these tangles as enzyme mechanism variables, and
experimental results pose equations relating these variables. One wishes to solve
these equations. In general, solving tangle equations is a difficult task. The job is
greatly simplified by the realization that most known DNA reaction products lie in

a well-understood class, that of 4-plats (2-bridge knots and links) [7]. Moreover, a great deal can be said about the factorization of 4-plats into tangle summands. The summands of interest are rational tangles [8]. Rational tangles are formed from the trivial tangle by twisting pairs of strands about one another, and look a great deal like DNA electron micrographs, in which pairs of DNA strands wind about each other. In the analysis of certain DNA experiments, one can use recent results on Dehn Surgery on 3-manifolds [9] to prove that the solutions to the tangle equations which arise must be rational tangles. Once the solutions are known to be rational tangles, the analysis becomes a matter of calculating the rational solutions, in which the rational tangle calculus is employed.

In Section 2, we describe the rational tangle calculus, and develop methods for solving equations involving rational tangles and 4-plats. In Section 3, we discuss the general problem of proving that solutions to certain tangle equations must be rational tangles. Section 4 discusses site-specific recombination, and Section 5 describes the tangle model for site-specific recombination. Sections 6 and 7 use the tangle model to analyse experimental results for the site-specific recombinant enzymes $Tn3$ resolvase and phage λ integrase.

A more complete account of applications of the rational tangle calculus to molecular biology will appear elsewhere [27].

2. Rational tangle calculus

The following discussion takes place in the smooth category. Unless otherwise specified, all ambient 3-manifolds will come equipped with an orientation, and 1-submanifolds will be unoriented. All homeomorphisms of ambient spaces will be assumed to be orientation-preserving. A 2-string tangle (or just tangle) is a pair (B, t), where B is a 3-ball and t is a pair of (unoriented) arcs properly embedded in B [8, 16]. We separate tangles into three types [16]. A tangle is rational if there exists a homeomorphism of pairs from (B, t) to the trivial tangle $(D^2 \times I, \{x, y\} \times I)$, where D^2 is the unit 2-ball in $\mathbb{R}^2$ and $\{x, y\}$ are points interior to D^2. A tangle is locally knotted if there exists a local knot in one of the strands; that is, there exists a 2-sphere in B meeting t transversely in 2 points, and such that the 3-ball it bounds in B meets t in a knotted spanning arc. A tangle is prime if it is neither rational nor locally knotted. Since rational and prime tangles contain no local knots, we say that they are locally unknotted. Figure 1 shows tangle diagrams of the three types.

In order to compare tangles, we need to think of them as having 'the same' boundary. As in [6], we define the model 2-sphere S^2 in $\mathbb{R}^3$ to be the boundary of the unit 3-ball D^3 in $\mathbb{R}^3$, equipped with 4 distinguished equatorial points $P = \{$NE, SE, SW, NW$\}$. We require that every tangle comes equipped with a boundary parametrization, that is, a homeomorphism $\Phi : (\partial B, \partial t) \to (S^2, P)$. So a tangle is a triple $B = (B, t, \Phi)$. Two tangles (B, t, Φ) and (B', t', Φ') are isomorphic if there is a homeomorphism $H : (B, t) \to (B', t')$ such that $\Phi = \Phi'H$ on ∂B. If X and Y are isomorphic tangles, we write $X = Y$. Let p be the projection of D^3 into the equatorial plane, and choose a homeomorphism $\Psi : B \to D^3$ such that Ψ extends Φ and such that the image of the arcs t under $p\Psi$ is a regular projection in the interior of D^2. A tangle diagram is the image of (B, t) under $p\Psi$. Two tangle diagrams represent isomorphic
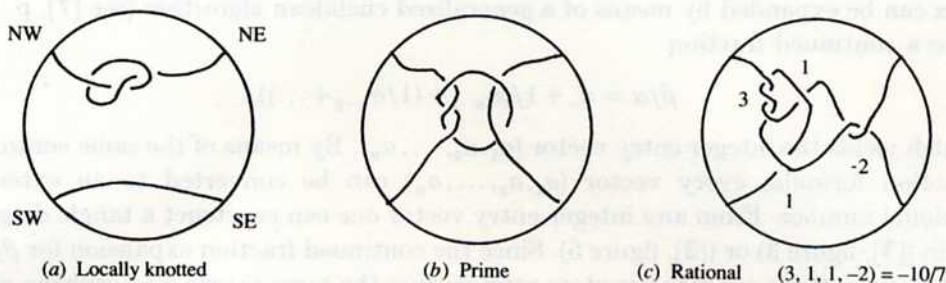
(*a*) Locally knotted   (*b*) Prime   (*c*) Rational $(3, 1, 1, -2) = -10/7$

Fig. 1. The three types of tangles.

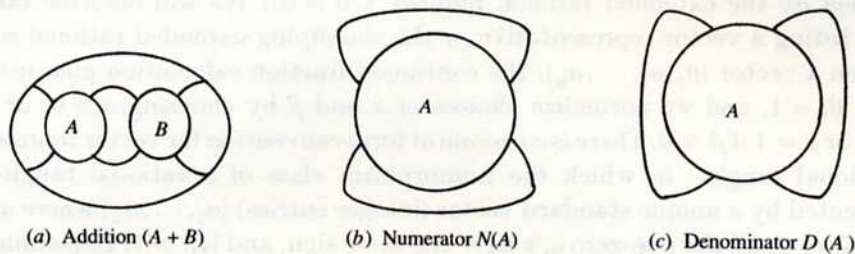(*a*) Addition $(A + B)$ (*b*) Numerator $N(A)$ (*c*) Denominator $D(A)$

Fig. 2. Tangle constructions.

tangles if and only if the diagrams are related by a finite sequence of Reidemeister moves in the interior of D^2.

Given two tangles $\{A, B\}$, we define tangle addition as shown in Figure 2(*a*), and denote the result as $A + B$. Note that $A + B$ may contain a simple closed curve, in which case $A + B$ is not a 2-string tangle. Addition is an associative operation: $A + (B + C) = (A + B) + C$. The numerator and denominator constructions applied to tangle A are shown in Figure 2(*b*, *c*), and denoted $\{N(A), D(A)\}$. If A and B are tangles, we define $N(A + B)$ and $D(A + B)$ in an analogous manner. If A is a tangle, then each of $\{N(A), D(A)\}$ is either a knot or a link of 2 components. We note that the knot (link) $N(A + B)$ is topologically equivalent to that obtained by glueing A to B along their 'common' boundary (after relabelling the 4 endpoints in ∂B). We say that a tangle A has the parity of (0) if the string which enters at the NW position exits at the NE position. Likewise, A has the parity of (1) if the string which enters at the NW position exits at the SE position, and A has the parity of ∞ if the string which enters at the NW position exits at the SW position. The concept of parity was considered in [3], where it was called string attachment class. Finally, if A denotes the isomorphism class of a tangle, then $(-A)$ denotes the isomorphism class of the mirror image of A, obtained by reversing every crossover in any projection of A.

Rational tangles admit very nice classification scheme (see [8, 11]). They can be represented by rational numbers, by vectors with integer entries, or by 2×2 integer matrices of determinant $+1$.

RATIONAL TANGLE CLASSIFICATION THEOREM ([8]). *There exists a* $1-1$ *correspondence between isomorphism classes of rational tangles and the extended rational numbers* $\beta/\alpha \in \mathbb{Q} \cup \{1/0 = \infty\}$, *where* $\alpha \in N \cup \{0\}$, $\beta \in \mathbb{Z}$, *and* $\gcd(\alpha, \beta) = 1$.

A proof of this Theorem 2·1 can be found in [19] or [7], p. 196. Every rational number

β/α can be expanded by means of a generalized euclidean algorithm (see [7], p. 187) into a continued fraction

$$\beta/\alpha = a_n + 1/(a_{n-1} + (1/a_{n-2} + \ldots)),$$

which yields the integer entry vector $(a_1, a_2, \ldots, a_n)$. By means of the same continued fraction formula, every vector $(a_1, a_2, \ldots, a_n)$ can be converted to an extended rational number. From any integer-entry vector one can construct a tangle diagram as in ([3], figure 3) or ([2], figure 5). Since the continued fraction expansion for β/α is not unique, there are many vectors representing the same tangle isomorphism class. For example, the vectors (0) and $(0, 3, 0)$ represent the tangle classified by the rational number zero, and the vectors $(0, 0)$ and $(1, -1, 5)$ represent the tangle classified by the extended rational number $1/0 = \infty$. We will describe tangles by either listing a vector representative or the classifying extended rational number.

Given a vector $(a_1, a_2, \ldots, a_n)$, the continued fraction calculation guarantees that $\gcd(\alpha, \beta) = 1$, and we normalize choices of α and β by choosing $\alpha > 0$, or $\beta = 1$ if $\alpha = 0$, or $\alpha = 1$ if $\beta = 0$. There is a canonical form convention for vector representation of rational tangles, in which the isomorphism class of a rational tangle can be represented by a unique standard vector (integer entries) $(a_1, \ldots, a_n)$ where $a_i \neq 0$ for $1 \leqslant i \leqslant n-1$, all the non-zero a_i's have the same sign, and $|a_1| > 1$. Depending on the rational number β/α, the standard vector is of even or odd length, and the non-zero entries are either all positive or all negative. Figures $3(a)$ and $3(b)$ show canonical vectors and their diagrams; Figure $3(b)$ is the canonical form for the tangle in Figure $1(c)$. The above convention excludes the four exceptional tangles $\{(0), (\pm 1), (0, 0)\}$ which are shown in Figure $3(c)$. We will call the standard vector for a rational tangle the Conway symbol for the rational tangle, and take the four vectors above as the Conway symbols for the four exceptional tangles.

A 4-plat (2-bridge) knot or 2-component link is one which admits a projection which consists of a braid on four strings, closed up as in Figure $4(a)$. In [1] it is shown that every 4-plat admits a projection consisting of a braid on four strings in which one string is free from crossings, as in Figures $4(b, c)$. There are classification schemes for 4-plats analogous to those for rational tangles. Given any integer-entry vector of odd length $\langle d_1, \ldots, d_{2k+1} \rangle$, we can construct a 4-plat as shown in Figure 4. Every 4-plat has a canonical form vector representation by an integer-entry vector $\langle c_1, \ldots, c_{2k+1} \rangle$ where $c_i > 0$ for all i. We call this vector a Conway symbol for the 4-plat. This convention excludes the 4-plat $\langle 0 \rangle$, the unlink of two unknotted components, so we take $\langle 0 \rangle$ as the Conway symbol in this case. Figure $4(b)$ shows a canonical form for the 4-plat knots of Figures $4(a, c)$. Two 4-plats represent the same knot (link) type if and only if they admit identical Conway symbols, or their Conway symbols become identical if one of them is reversed. Analogous to the rational tangle case, a classifying rational number for the knot (link) equivalence class of the 4-plat can be obtained from any vector representing the 4-plat via a continued fraction calculation: $\beta/\alpha = 1/(c_1 + (1/c_2 + \ldots))$. If one performs this continued fraction calculation on a Conway symbol for the 4-plat K $(K \neq \langle 0 \rangle, \langle 1 \rangle)$, one obtains $0 < \beta < \alpha$. Unless otherwise specified in the following, we will always choose to compute a classifying rational number for the 4-plat from a Conway symbol, and (following [7]) we write the 4-plat as $b(\alpha, \beta)$. The numbers α and β have a geometric interpretation in terms of a 2-bridge projection of a 4-plat (see [7], p. 183). With the crossover sign
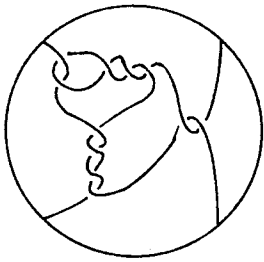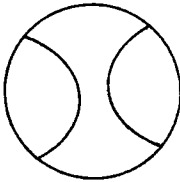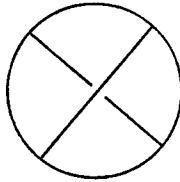
(a) Even (2, 3, 4, 2) = 67/30
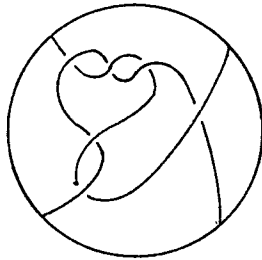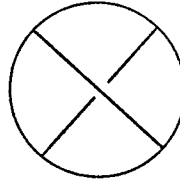

(b) Odd (–3, –2, –1) = –10/7
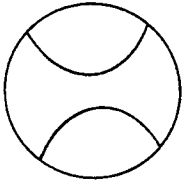
(0)   (0, 0)   (–1)   (+1)

(c) The exeptional tangles

Fig. 3. Canonical forms for rational tangles.

convention of Figure 4 (which agrees with the usual convention for representing generators of the braid group as in [7], and is opposite to that of [8, 25]), the 2-fold branched cyclic cover of $b(\alpha, \beta)$ is the lens space $L(\alpha, \beta)$. For example, the unknot $b(1, 1) = \langle 1 \rangle$ has S^3 as 2-fold branched cover, and the unlink of 2 unknotted components $b(0, 1) = \langle 0 \rangle$ has $S^1 \times S^2$ as 2-fold branched cover. 4-plats are classified by means of their 2-fold branched cyclic covers (see [7], p. 185):

4-PLAT CLASSIFICATION THEOREM ([21]). *Two* 4-*plats* $b(\alpha, \beta)$ *and* $b(\alpha', \beta')$ *are equivalent* (*as unoriented knots or links*) *if and only if* $\alpha = \alpha'$ *and* $\beta^{\pm 1} \equiv \beta' \pmod{\alpha}$.

4-plats and rational tangles are closely related via the numerator and denominator constructions. For example, for any integer x,

$$D((d_1, \ldots, d_{2k+1}, x)) = \langle d_1, \ldots, d_{2k+1} \rangle \quad \text{and} \quad N((d_1, \ldots, d_{2k+1}, x, 0)) = \langle -d_1, \ldots, -d_{2k+1} \rangle.$$

Given a rational number β/α with $0 < \beta/\alpha < 1$, the denominator construction applied to the tangle β/α yields the 4-plat $b(\alpha, \beta)$; and the numerator construction applied to the tangle $\beta/\alpha \geqslant 1$ yields the 4-plat $b(\beta, -\alpha)$.

Our first lemma is a calculation (in terms of classifying rational numbers) for the 4-plat which results when the numerator construction is applied to the sum of two rational tangles. A 4-plat is the closure of a braid on 4 strings, in which only the first 3 strings form crossovers. The calculation below is based on the calculation in [7], p. 186, in which a 4-plat in S^3 is written in terms of generators $\{\sigma_1, \sigma_2\}$ of the braid group B_3. We write S^3 as two 3-balls $\{B_1, B_2\}$ connected by $S^2 \times I$, and with the braid contained in $S^2 \times I$. Lifting this picture to the 2-fold branched cyclic cover, the 3-balls lift to two solid tori $\{T_1, T_2\}$, connected by $(S^1 \times S^1) \times I$. The braid generators lift to Dehn twists on $S^1 \times S^1$, and in terms of oriented (meridian, longitude) generators for
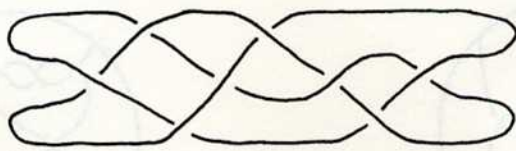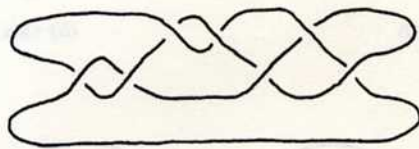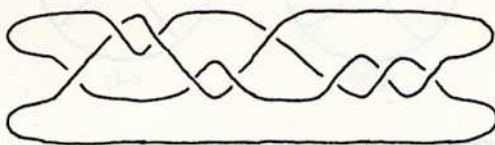
(a) b(19,8)

(b) $\langle 2, 2, 1, 1, 1 \rangle = b(19, 8)$

(c) $\langle 1, 2, -2, -1, 3 \rangle = b(19, 8)$

Fig. 4. 4-Plats.

$H_1(S^1 \times S^1, \mathbb{Z})$, the Dehn twists are expressed by elementary 2×2 integer matrices. Every tangle β/α admits a (perhaps non-canonical) vector representation of even length $(a_1, \ldots, a_{2k})$. If the tangle is non-exceptional, one such vector representation can be obtained from the Conway symbol by relaxing the requirement that $|a_1| > 1$. We take the following vector representations for the four exceptional tangles: $(0) = (1, -1)$, $\infty = (0, 0)$, $(1) = (1, 0)$, and $(-1) = (-1, 0)$. An even-length vector representative for the tangle β/α determines a 2×2 matrix representative via the following equation:

$$\begin{pmatrix} u & v' \\ v & u' \end{pmatrix} = \begin{pmatrix} 1 & a_{2k} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ a_{2k-1} & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 \\ a_1 & 1 \end{pmatrix}. \tag{1}$$

Equation (1) expresses a homeomorphism (determined by the vector representative for the tangle β/α) from the boundary of the 2-fold branched cyclic cover of β/α to the boundary of a reference solid torus. In the above equation, the rational number $u/v = \beta/\alpha$ classifies the tangle, but it may not necessarily be in normal form. Note that $u'/v' = a_n + 1/(1_{n-1} + 1/(\ldots + 1/a_2))$.

LEMMA 2·1. *Given two rational tangles $A_1 = \beta_1/\alpha_1$ and $A_2 = \beta_2/\alpha_2$, then $N(A_1 + A_2)$ is a 4-plat which is equal to $b(\alpha, \beta)$, where $\alpha = |\alpha_1 \beta_2 + \alpha_2 \beta_1|$ 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 uniquely determined by the following: $0 < \beta < \alpha$ and $\beta \equiv \sigma(\alpha_1 \alpha_2' + \beta_1 \beta_2') \pmod{\alpha}$, where $\sigma = \text{sign}(\alpha_1 \beta_2 + \alpha_2 \beta_1)$ and α_2' and β_2' are the entries in the second column of any matrix representative for the tangle β_2/α_2.*

Proof. First we observe that conditions (i)–(iii) above imply that β is uniquely determined. In case (iii), substituting $\beta'_2 = (1 + \alpha_2 \alpha'_2)/\beta_2$ into the equation for β, we obtain $\beta \equiv (\sigma\beta_1 + \alpha'_2 \alpha)/\beta_2 (\mathrm{mod}\,\alpha)$. Since $(\alpha_2, \beta_2) = 1$, all possible values for α'_2 are of the form $c + k\beta_2$, with c and k integers, and c a constant. This means that there is exactly one choice for β such that $0 < \beta < \alpha$.

Represent A_1 and A_2 by vectors of even length: $A_1 = (a_1, \ldots, a_{2k})$ and $A_2 = (b_1, \ldots, b_{2r})$. Then $N(A_1 + A_2)$ is a 4-plat $b(\alpha, \beta)$, for some values of α and β. This 4-plat has a (possibly non-canonical) representation by one of the vectors $\langle b_1, \ldots, b_{2r-1}, (b_{2r} + a_{2k}), a_{2k-1}, \ldots, a_1 \rangle$, or $\langle a_1, \ldots, a_{2k-1}, (a_{2k} + b_{2r}), b_{2r-1}, \ldots, b_1 \rangle$, as shown in Figure 5. Figure 5(a) shows one of the main steps in this geometric calculation, in which the horizontal twists at the right-hand end of A_2 are converted to horizontal twists at the left-hand end of A_2 in order to facilitate addition. Note that the signs of a_i and b_i may differ, and that $(b_{2r} + a_{2k})$ may be zero. The 2-fold branched cyclic cover of $b(\alpha, \beta)$ is the lens space $L(\alpha, \beta)$, which is obtained by glueing together two solid tori $\{T_1, T_2\}$ by a homeomorphism $f : \partial T_1 \to \partial T_2$, where f is the product of Dehn twists given by a vector representation for the 4-plat. With respect to the (meridian, longitude) basis (μ_i, λ_i) for $H_1(\partial T_i, Z)$, for $i = 1, 2$, values (not necessarily normalized) for α and β are determined by the following matrix representation for f_*:

$$\begin{pmatrix} \beta & \alpha' \\ \alpha & \beta' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ b_1 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & b_{2r} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & a_{2k} \\ 0 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 \\ a_1 & 1 \end{pmatrix}. \tag{2}$$

We have $\beta\beta' - \alpha\alpha' = 1$.

Applying Equation (1) above to the chosen vector representations for β_1/α_1 and β_2/α_2, we obtain

$$\begin{pmatrix} \beta & \alpha' \\ \alpha & \beta' \end{pmatrix} = \begin{pmatrix} \beta'_2 & \alpha'_2 \\ \alpha_2 & \beta_2 \end{pmatrix} \begin{pmatrix} \beta_1 & \alpha'_1 \\ \alpha_1 & \beta'_1 \end{pmatrix}. \tag{3}$$

From the matrix product in (3), we have

$$\alpha = \alpha_1 \beta_2 + \alpha_2 \beta_1, \quad \beta = \alpha_1 \alpha'_2 + \beta_1 \beta'_2.$$

We can now determine normalized values for α and β as follows. If $\alpha_1 \beta_2 + \alpha_2 \beta_1 > 1$, since $L(\alpha, \beta) = L(\alpha, \beta + k\alpha)$ where $k \in \mathbb{Z}$, we set

$$\alpha = \alpha_1 \beta_2 + \alpha_2 \beta_1 \quad \text{and} \quad \beta \equiv (\alpha_1 \alpha'_2 + \beta_1 \beta'_2)(\mathrm{mod}\,\alpha).$$

The conditions $0 < \beta < \alpha$, $\beta_2 \beta'_2 - \alpha_2 \alpha'_2 = 1$ and $(\alpha_2, \beta_2) = 1$ yield a unique choice for β.

If $-(\alpha_1 \beta_2 + \alpha_2 \beta_1) > 1$, since $L(\alpha, \beta) = L(-\alpha, -\beta + k\alpha)$ where $k \in \mathbb{Z}$, we set

$$\alpha = -(\alpha_1 \beta_2 + \alpha_2 \beta_1) \quad \text{and} \quad \beta \equiv -(\alpha_1 \alpha'_2 + \beta_1 \beta'_2)(\mathrm{mod}\,\alpha).$$

If $|\alpha_1 \beta_2 + \alpha_2 \beta_1| = 1$, since $L(\pm 1, \beta) = S^3$, we set $\alpha = 1$ and $\beta = 1$. If $\alpha_1 \beta_2 + \alpha_2 \beta_1 = 0$, then $N(A_1 + A_2)$ is the unlink of two unknotted components, with 2-fold branched cyclic cover $L(0, 1) = S^1 \times S^2$. In this case we set $\alpha = 0$ and $\beta = 1$. ∎

If A and B are tangles, and $N(A + B) = K$ (a knot or 2-component link), then we say that A and B are summands of K. If both A and B are rational tangles, and $N(A + B) = K$, then K is a 4-plat. Suppose, however, that A is a rational tangle, K is a 4-plat, and that we wish to solve the equation $N(X + A) = K$ for the unknown tangle X. Unfortunately, this data does not force X to be a rational tangle. For
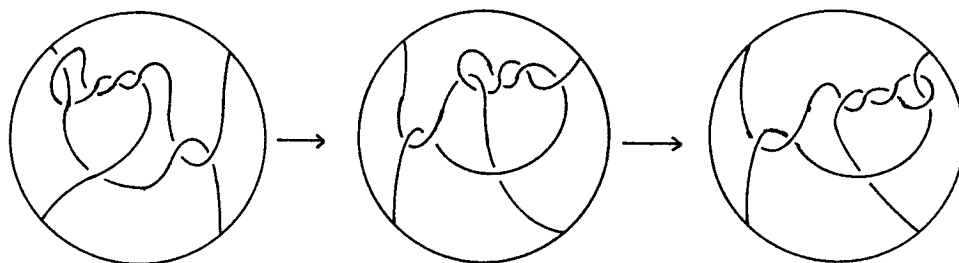
(a) Isotropy of the tangle $(-2, -3, -1, -2) = -25/9$
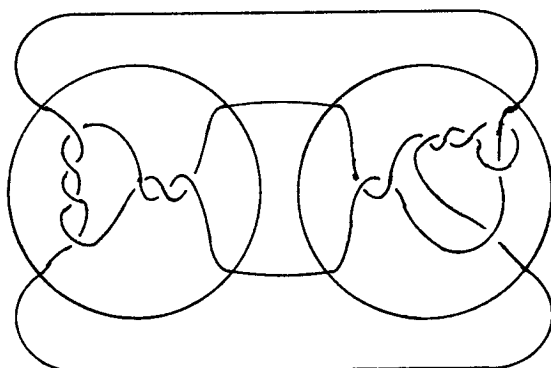
(b) $N(13/4 + -25/9) = N((4, 3) + (-2, -3, -1, -2))$
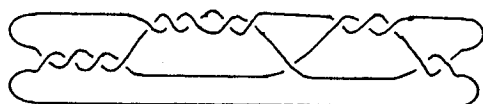
(c) $\langle 4, 1, -1, -3, -2 \rangle$

Fig. 5. Tangle addition.

example if A is any integral tangle (a horizontal row of half-twists), and $K = $ unknot, the equation $N(X + A) = $ unknot has infinitely many distinct prime tangle solutions of the form $(B + (-A))$, where B denotes any prime tangle with the property that $N(B + (0)) = $ unknot. One such prime tangle is shown in Figure 1(b). Since B is prime, then $(B + (-A))$ is prime (see [16]), and

$$N((B + (-A)) + A) = N(B + (0)) = \text{unknot}.$$

Although prime tangle solutions to these equations are in general difficult to enumerate, all rational tangle solutions to equations of this type are given by the following theorem:

THEOREM 2·2. *Let* $A = \beta/\alpha = (a_1, \ldots, a_{2n})$ *be a rational tangle and* $K = \langle c_1, \ldots, c_{2k+1} \rangle$ $\mp \langle 0 \rangle$ *be a 4-plat. The rational tangle solutions to the equation* $N(X + A) = K$ *are the following*:

$$X = (c_1, \ldots, c_{2k+1}, r, -a_1, \ldots, -a_{2n}), \quad or \quad X = (c_{2k+1}, \ldots, c_1, r, -a_1, \ldots, -a_{2n}),$$

with r *any integer. If* $K = \langle 0 \rangle$, *then* $X = (-a_1, \ldots, -a_{2n})$ *is the unique solution.*

Proof. It is easily seen (from Lemma 2·1, or by drawing a picture), that for arbitrary r, each of the tangles X described above is a solution to the equation. The negative string of a_i's cancels out the tangle A, leaving

$$D(c_1, \ldots, c_{2k+1}, r) = \langle c_1, \ldots, c_{2k+1} \rangle = \langle c_{2k+1}, \ldots, c_1 \rangle = D(c_{2k+1}, \ldots, c_1, r).$$

We must show that the set of solutions described above is complete. The 4-plat $K = b(p, q)$ is represented by the rational number q/p, where either $0 < q < p$ are coprime integers, or $p = 1, q = 1$, or $p = 0, q = 1$. Suppose that $0 < q < p$. Let q' be the unique integer such that $0 < q' < p$ and $qq' \equiv 1 \pmod{p}$. Given a Conway symbol for K (either $\langle c_1, \ldots, c_{2k+1} \rangle$ or $\langle c_{2k+1}, \ldots, c_1 \rangle$), the integers $\{p, q, q'\}$ (and matrix representatives for K) are determined by one of the matrix equations below:

$$\begin{pmatrix} q & p' \\ p & q' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ c_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & c_2 \\ 0 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 \\ c_{2k+1} & 1 \end{pmatrix}; \tag{4}$$

$$\begin{pmatrix} q' & p' \\ p & q \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ c_{2k+1} & 1 \end{pmatrix} \begin{pmatrix} 1 & c_{2k-1} \\ 0 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 \\ c_1 & 1 \end{pmatrix}. \tag{5}$$

From the above matrix equations, we have $qq' - pp' = 1$. If $K = \langle 0 \rangle$ or $\langle 1 \rangle$, we use Equation (4) to produce a matrix representative. We wish to write down a general expression for the family of all 2×2 matrices which describe glueings which give rise to a lens space orientation-preserving homeomorphic to $L(p, q) = L(p, q')$. The lens space $L(p, q)$ is the result of glueing the solid torus T_1 to the solid torus T_2 by the glueing homeomorphism $f \colon \partial T_1 \to \partial T_2$. Suppose that the left-hand sides of Equations (4), (5) represent f with respect to a (meridian, longitude) basis for $\{T_1, T_2\}$. Then, all other glueings can be viewed as perturbations on these matrices, the perturbations coming from orientation-preserving change-of-coordinates on each of ∂T_1 and ∂T_2. That is, if $\{r, s\}$ are arbitrary integers, then from the left-hand-side of Equation (4) we obtain a 2-parameter family of glueing matrices:

$$\{G_q(r, s)\} = \epsilon \begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix} \begin{pmatrix} q & p' \\ p & q' \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \quad (\epsilon = \pm 1). \tag{6}$$

Similarly, using the left-hand side of (5), we obtain a 2-parameter family of glueing matrices:

$$\{G_{q'}(r, s)\} = \epsilon \begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix} \begin{pmatrix} q' & p' \\ p & q \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \quad (\epsilon = \pm 1). \tag{7}$$

Suppose now that X is a rational tangle such that $N(X + A) = K$. Suppose that tangles A and X admit the following matrix representatives:

$$A : \begin{pmatrix} \beta & \alpha' \\ \alpha & \beta' \end{pmatrix}, \quad \text{and} \quad X : \begin{pmatrix} u & v' \\ v & u' \end{pmatrix}.$$

Since $N(X + A) = K$, a glueing map for $L(p, q)$ is described by the matrix product (as in Equation (3))

$$\begin{pmatrix} \beta' & \alpha' \\ \alpha & \beta \end{pmatrix} \begin{pmatrix} u & v' \\ v & u' \end{pmatrix}.$$

This means that either

$$\begin{pmatrix} u & v' \\ v & u' \end{pmatrix} = \epsilon \begin{pmatrix} \beta & -\alpha' \\ -\alpha & \beta' \end{pmatrix} \begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix} \begin{pmatrix} q & p' \\ p & q' \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \quad (\epsilon = \pm 1) \tag{8}$$

or

$$\begin{pmatrix} u & v' \\ v & u' \end{pmatrix} = \epsilon \begin{pmatrix} \beta & -\alpha' \\ -\alpha & \beta' \end{pmatrix} \begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix} \begin{pmatrix} q' & p' \\ p & q \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \quad (\epsilon = \pm 1). \tag{9}$$

By multiplying out the right-hand sides of (8) and (9), we observe that the parameter s is irrelevant in the determination of $\{u, v\}$, so without loss of generality we take $s = 0$. The right-hand side of (8) (with $s = 0$) can be obtained from the vector representation $X = (c_1, \ldots, c_{2k+1}, r, -a_1, \ldots, -a_{2n})$. The right-hand side of (9) (with $s = 0$) can be obtained from the vector representation $X = (c_{2k+1}, \ldots, c_1, r, -a_1, \ldots, -a_{2n})$.

The 4-plat $K = \langle 0 \rangle$ is represented by the identity 2×2 matrix. In this case, Equations (8), (9) specify that $u/v = -\beta/\alpha$, so that $X = (-a_1, \ldots, -a_{2n})$ is the unique solution. ∎

In terms of classifying rational numbers, the rational solutions for X in the above theorem are given by the fractions

$$u/v = (\beta q - \alpha' p + r\beta p)/(\beta' p - \alpha q - r\alpha p)$$

and

$$u/v = (\beta q' - \alpha' p + r\beta p)/(\beta' p - \alpha q' - r\alpha p).$$

Although Theorem 2·2 says that one equation in one unknown has infinitely many rational solutions ($K \neq \langle 0 \rangle$), the next result says that two equations in one unknown have at most two rational solutions.

COROLLARY 2·3. *Let A_1, A_2 be distinct rational tangles, and K_1 and K_2 be 4-plats. There are at most two distinct rational tangle solutions to the equations* (i) $N(X + A_1) = K_1$, (ii) $N(X + A_2) = K_2$.

Proof. Let $X = u/v$, $A_1 = \beta_1/\alpha_1$, $A_2 = \beta_2/\alpha_2$, $K_1 = b(\alpha, \beta)$, and $K_2 = b(\alpha', \beta')$. Then, by Lemma 2·1, we have

$$\alpha = |u\alpha_1 + v\beta_1| \quad \text{and} \quad \alpha' = |u\alpha_2 + v\beta_2|.$$

In the (u, v)-plane, these equations describe two pairs of parallel straight lines. These lines intersect in at most 4 points. Since $u/v = -u/-v$, these 4 points of intersection describe at most two distinct rational tangle solutions for the equations in the hypothesis. ∎

Corollary 2·3 is sharp, as can be seen by taking $A_1 = 1/3$, $A_2 = 5/17$, $K_1 = b(5, 3)$, and $K_2 = b(29, 17)$. The two solutions for X are $X = -70/239$ and $X = -75/254$. It may happen that two equations of the above form have no rational solutions. For example, the pair of equations

$$\{N(X + (0)) = \langle 1 \rangle, \quad N(X + (1)) = \langle 1, 1, 1, 1, 1 \rangle\}$$

has no solutions of any kind (prime, rational, or locally knotted), as will be argued in Section 3.

LEMMA 2·4. *If R is a rational tangle, then $R + R$ is locally unknotted. Moreover, if $R + R$ is rational, then R is an integral tangle. Conversely, if R is not integral and does not have the parity of ∞, then $R + R$ is prime.*

Proof. The tangle $R+R$ ends in a number (possibly zero) of horizontal half-twists. By adding (if necessary) an extra half-twist at the right-hand end of $R+R$, we obtain $(R+R)_1$, which has the property that $N((R+R)_1)$ is a 4-plat link with two unknotted components, so $R+R$ is locally unknotted. If $R+R$ is rational, then $D(R+R)$ is a 4-plat, hence prime (or unknotted, or the unlink of two unknotted components). If R is not an integral tangle (a horizontal row of half-twists), then $D(R+R)$ is composite. ∎

3. *Detecting rationality*

Suppose that A and B are tangles, and that $N(A+B) = K$, where K is a 4-plat. If K is either a 2-component link or the unknot, then both A and B are locally unknotted. If K is a non-trivial knot, in general we cannot conclude that both A and B are locally unknotted. Since 4-plats are prime, at most one of the two tangles can be locally knotted, and if so, the other tangle can be either prime or rational.

LEMMA 3·1 ([2, 16]). *Suppose that A and B are locally unknotted tangles, and that $N(A+B) = K$, where K is a 4-plat. Then at least one of $\{A, B\}$ is a rational tangle.*

Proof. A tangle X is prime if and only if its 2-fold branched cyclic cover X' is irreducible and boundary-irreducible, and rational if and only if its 2-fold branched cyclic cover is a solid torus (see [16]). If X is prime, then $\pi_1(\partial X')$ injects into $\pi_1(X')$ under inclusion. If both A and B are prime tangles, then the 2-fold branched cyclic cover of K is $K' = A' \cup B'$, identified along their common incompressible torus boundary. This means that $\pi_1(K') \supseteq \mathbb{Z} \oplus \mathbb{Z}$, which is never the case for the 2-fold branched cyclic cover of a 4-plat. ∎

It turns out that the cyclic surgery theorem [9] is very useful in proving that certain tangles which arise in models for DNA enzyme action are rational tangles. The strategy is to use the cyclic surgery theorem to prove that the 2-fold branched cover M of the tangle is a Seifert fibre space (or SFS) [22], and then to use facts about the lens space results of Dehn surgeries on M to prove that M must be a solid torus. Let M be a compact, connected, irreducible, orientable 3-manifold with ∂M a torus. The unoriented isotopy class of a non-trivial simple closed curve in ∂M will be called its slope (r). For any slope r, a closed 3-manifold $M(r)$ may be constructed by attaching a solid torus J to M, so that a curve of slope r bounds a disk in J. Given any two slopes $\{r, s\}$, let $\Delta(r, s)$ denote the minimal geometric intersection number for embedded representatives of r and s.

CYCLIC SURGERY THEOREM ([9]). *If M is not a Seifert fibre space (SFS), and if $\{r, s\}$ are slopes such that $\pi_1(M(r))$ and $\pi_1(M(s))$ are cyclic groups, then $\Delta(r, s) \leqslant 1$. Hence there are at most three slopes r such that $\pi_1(M(r))$ is cyclic.*

One corollary of the cyclic surgery theorem concerns the case when M is a knot complement. In this case, there is a unique coordinate system on ∂M, namely (meridian, longitude) on the boundary of the tube neighbourhood of the knot. If a slope is represented by (p, q) in this coordinate system, then the slopes on ∂M are parametrized by $p/q \in \mathbb{Q} \cup \{1/0\}$. Moreover M is a SFS if and only if M is the complement of a torus knot.

THEOREM ([9]). (i) *If M is the complement of a non-torus knot, then $\pi_1(M(r))$ can be non-trivial and cyclic only if r is an integer. Moreover there are at most two such integers r, and if there are two, then they must be successive.* (ii) *If M is the complement of a non-trivial knot and $r \neq \pm 1$, then M(r) is not simply connected. Moreover at most one of M(1) and M(−1) is simply-connected.*

These cyclic surgery results are useful to us in the situation of the hypothesis of Lemma 3·1. One of the tangles (say B) must be rational. Let K', A', and B' denote the 2-fold branched cyclic covers of K, A, and B, respectively. The lens space K' is equal to $A' \cup B'$, with B' a solid torus. In other words, one obtains a lens space by attaching a solid torus to A'. The cyclic surgery theory says that if A' is not a SFS, then there are very few ways to do this.

LEMMA 3·2. *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 3 equations hold:*

$N(X + A_1) = b(1, 1)$ *(the unknot)*;

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

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

If $|\alpha - \alpha'| > 1$, then the 2-fold branched cyclic cover X' is a torus knot complement.

Proof. Since $N(X + A_1)$ is the unknot, both X and A_1 are locally unknotted, and one of them must be rational. If X is rational, then X' is a solid torus and the proof is accomplished. Otherwise X is prime, and A_i is rational for $1 \leqslant i \leqslant 3$. The 2-fold branched cyclic cover X' is the bounded complement of a strongly invertible knot in S^3: see [3]. If X' is not a SFS, then the Dehn surgeries represented by A_2 and A_3 must be integral. The results of two integral Dehn surgeries along the knot which defines X' yields the Lens spaces $L(\alpha, \beta)$ and $L(\alpha', \beta')$, so the slopes of the surgeries are α and α'. Since the slopes are not successive integers by hypothesis, we conclude that X' is a SFS, and a torus knot complement. ∎

THEOREM 3·3. *Let M be a SFS with $\partial M = S^1 \times S^1$, and bounded orbit surface S. Suppose that there exists a boundary slope r such that M(r) is a lens space. Then S is a disk with at most two exceptional fibres.*

Proof. Let k be the number of exceptional fibres of M, and S be the bounded orbit surface of M. If a curve C of boundary slope r is homologous to a fibre of ∂M, then $M(r)$ is a non-trivial connected sum if either $k > 1$ or S is not a disk. This is a slight generalization of proposition 2 of [14] (it includes the case of S non-orientable). Therefore if C is homologous to a fibre, then $k \leqslant 1$ and S is a disk. Otherwise C is not homologous to a fibre, and the Seifert fibration extends over the solid torus J. Let $S(r)$ be the orbit surface of $M(r)$. Then $S(r) = S \cup$ disk. Since $\pi_1(M(r))$ is finite, this means that $S(r)$ is either a projective plane with at most one exceptional fibre, or a sphere with at most 3 exceptional fibres: see [22]. Since $\pi_1(M(r))$ is non-abelian if $S(r)$ is a projective plane or a sphere with 3 exceptional fibres, we conclude that $S(r)$ is a sphere with at most 2 exceptional fibres, so S is a disk with at most two exceptional fibres. ∎

The next two lemmas deal with proving that a SFS is actually a solid torus, given information about various lens spaces obtained from it by Dehn surgery. These results will be used in the next sections to analyse the results of some DNA recombination experiments.

LEMMA 3·4. *Let M be a torus knot complement. Suppose there is a boundary slope r for M such that M(r) is one of $L(2, x)$, $L(3, x)$, $L(4, x)$ for some integer x. Then M is a solid torus.*

Proof. The results of Dehn surgery on torus knot complements are well known: see [18]. If one performs p/q surgery along an (r, s) torus knot, then one obtains a lens space if and only if $|rsq+p| = 1$. In this case, $L(|p|, qs^2)$ is the lens space obtained by the Dehn surgery. Now r, s are coprime positive integers, with $0 < s < r$, and a necessary condition for the torus knot to be non-trivial is $s \geq 2$, and so $rs \geq 6$. The condition $|rsq+p| = 1$ means that $|p| = rs|q| \pm 1$, so $|p| \in \{2, 3, 4\}$ is impossible if the torus knot is non-trivial. The hypothesis forces the torus knot to be trivial, and so its complement M is a solid torus. ∎

LEMMA 3·5. *Let M be a SFS with $\partial M = S^1 \times S^1$, and with bounded orbit surface S. If there exist boundary slopes r_j for $j = 1, 2, 3$ such that $M(r_1) = L(p, q_1)$, $M(r_2) = L(p+2, q_2)$, and $M(r_3) = L(p+4, q_3)$ for integers p, q_j, then M is a solid torus.*

Proof. Since at least one Dehn surgery on M produces a lens space, by Theorem 3·3 we know that S is a disk and that M has at most two exceptional fibres. If M has zero or one exceptional fibre, then M is a solid torus. Suppose then that M has two exceptional fibres of orders (α_i, β_i) with $0 < \beta_i < \alpha_i$ for $i = 1, 2$, and with $\{\alpha_i, \beta_i\}$ coprime integers. Since no lens space admits a Seifert fibration with three exceptional fibres, the Seifert fibration must extend over the solid torus J which is glued to M to form the spaces $M(r_j), j = 1, 2, 3$. If we take a basis for ∂M consisting of F (a fibre in the SFS on ∂M) and Q (a curve on ∂M with intersection number $+1$ with F), then the glueing for each of the $M(r_j)$ is such that a meridian of J is homologous to $Q+b_j F$, for some integers b_j. In terms of the Seifert symbols (see [22]), we have $M(r_j) = (O, o, 0 \,|\, b_j, (\alpha_1, \beta_1), (\alpha_2, \beta_2))$ for $j = 1, 2, 3$. We have $|H_1(M(r_j))| = |b_j \alpha_1 \alpha_2 + \alpha_1 \beta_2 + \alpha_2 \beta_1|$. At least two of the b_j (say b' and b'') must have the property that the numbers $(b' \alpha_1 \alpha_2 + \alpha_1 \beta_2 + \alpha_2 \beta_1)$ and $(b'' \alpha_1 \alpha_2 + \alpha_1 \beta_2 + \alpha_2 \beta_1)$ have the same sign (or that one of them is zero). Let r' and r'' denote the boundary slopes corresponding to b' and b''. Now $||H_1(M(r'))| - |H_1(M(r))|| = |b'-b''| \alpha_1 \alpha_2 \in \{2, 4\}$. Since $\alpha_i \geq 2$, the only possibilities (after renumbering if necessary) are $\alpha_1 = 2$, $\alpha_2 = 2$, $\beta_1 = 1$, and $\beta_2 = 1$. However there are no integers $\{x, y\}$ such that $||4x+4| - |4y+4|| = 2$. ∎

The following example shows that Lemma 3·5 is sharp. The values $\alpha_1 = 5$, $\beta_1 = 3$, $\alpha_2 = 3$, $\beta_2 = 1$, $b_1 = -3$, $b_2 = 1$ produce a pair of lens spaces with $|H_1| \in \{29, 31\}$.

In the following we derive necessary algebraic conditions for two equations to have solutions. The next lemma is a generalization of a result of Lickorish [17]. Related results also appear in [5].

LEMMA 3·6. *Let X be any tangle, T and β/α be rational tangles, and $b(p, q)$ be a 4-plat, such that $N(X+T) = \langle 1 \rangle$ and $N(X+\beta/\alpha) = b(p, q)$.*

(1) *If $T = \infty$ then $L(p, q)$ can be obtained by $(\beta+s\alpha)/\alpha$ surgery along a knot in S^3, where s an integer and $p = \pm(\beta+s\alpha)$.*

(2) *If $T = (0)$ then $L(p, q)$ can be obtained by $(\alpha+s\beta)/\beta$ surgery along a knot in S^3, where s an integer and $p = \pm(\alpha+s\beta)$.*

Proof. The 2-fold branched cyclic cover of $\langle 1 \rangle$ is S^3, the 2-fold branched cyclic cover T' of T is a solid torus, so the 2-fold branched cyclic cover X' of X is a knot complement.

For the moment let us assume $T = \infty$. Then the arcs NW to SW and SW to SE on

∂T lift respectively to a meridian μ' and a longitude λ' on $\partial T'$. The first equation implies $X \cup_g (\infty) = S^3$, where $g : \partial(\infty) \to \partial X$ is a glueing map. Lifting to 2-fold branched covers we have $X' \cup_{g'} T' = S^3$, where $g' : \partial T' \to \partial X'$. Choose a meridian $\mu = g'(\mu')$ and a longitude λ on $\partial X'$. Then λ is isotopic to the curve $g'(\mu') + sg'(\lambda')$ for some integer s. There exist (see [19]) orientation-preserving homeomorphisms ψ and F, where

(i) F maps the ∞ tangle to the β/α tangle,

(ii) $\psi : \partial T' \to \partial T'$ sends the meridian μ' to a curve isotopic to $\beta\mu' + \alpha\lambda'$,

(iii) the maps $F|_\partial$ and ψ commute with the covering map $p|_\partial : \partial T' \to \partial T$, that is, $(p|_\partial)(F|_\partial) = \psi(p|_\partial)$.

Using the second equation $N(X + \beta/\alpha) = b(p,q)$, the 2-fold branched cover $L(p,q)$ of $b(p,q)$ can be constructed as $X' \cup_{g'\psi} T'$. The glueing map $g'\psi : \partial T' \to \partial X'$ maps μ' to a curve isotopic to $(\beta + s\alpha)\mu + \alpha\lambda$. Hence $L(p,q)$ is obtained by $(\beta + s\alpha)/\alpha$ surgery on the knot complement X'. The group $H_1(L(p,q)) = \mathbb{Z}_p$ is generated by the meridian μ, and so $p = \pm(\beta + s\alpha)$.

The result for the case $T = (0)$ is proved in the same way. The only difference is that the map ψ sends the meridian μ' (the lift of the SW SE arc) to a curve isotopic to $\alpha\mu' + \beta\lambda'$. ∎

THEOREM 3·7. *Let X be any tangle, let T and β/α be rational tangles, and let $b(p,q)$ be a 4-plat, where $N(X + T) = \langle 1 \rangle$ and $N(X + \beta/\alpha) = b(p,q)$.*

If $T = \infty$ then $q \equiv \pm\alpha t^2 \pmod{p}$ for some integer t.

If $T = (0)$ then $q \equiv \pm\beta t^2 \pmod{p}$ for some integer t.

Proof. Let us recall the following facts.

(i) $L(p,q)$ is obtained by p/q surgery on the unknot in S^3 (see [20]).

(ii) Suppose that M is a 3-manifold and $H_1(M) = \mathbb{Z}_p$. If M is obtained by p/q surgery on a knot k in S^3, then the linking form $L : H_1(M) \times H_1(M) \to \mathbb{Q}/\mathbb{Z}$ is such that $L(g,g) = q/p$ where g is a generator of $H_1(M)$ representing a meridian of the knot k (see [17]).

By (i) and (ii) there exists a generator ζ of $H_1(L(p,q))$ such that $L(\zeta, \zeta) = q/p$. By Lemma 3·6 and (ii) there exists a generator ξ of $H_1(L(p,q))$ such that $L(\xi, \xi) = \alpha/(\beta + s\alpha) = \pm\alpha/p$ if $T = \infty$ and $L(\xi, \xi) = \beta/(\alpha + s\beta) = \pm\beta/p$ if $T = (0)$. Now $H_1(L(p,q))$ is cyclic, so that $\zeta = t\xi$ for some integer t, and $q/p = t^2 L(\xi, \xi)$ in $\mathbb{Q}/\mathbb{Z}$. ∎

The following corollary (and proof) is due to M. Boileau; it is of use in the analysis of site-specific recombination in Section 6.

COROLLARY 3·8 (Boileau). *If T is either ∞ or (0), then there is no tangle X which satisfies the equations* (i) $N(X + T) = \langle 1 \rangle$ *and* (ii) $N(X + (\pm 1)) = \langle 1, 1, 1, 1, 1 \rangle$.

Proof. Since $\langle 1, 1, 1, 1, 1 \rangle = b(8,5)$, this follows from Theorem 3·7 on taking $\alpha = \beta = \pm 1$, $p = 8$ and $q = 5$, since $5 \equiv \pm t^2 \pmod{8}$ has no solution for t. ∎

4. *Site-specific recombination*

We will now consider the situation of site-specific recombination enzymes operating on covalently closed circular duplex DNA. Duplex DNA consists of two linear backbones of sugar and phosphorus. Attached to each sugar is one of the four bases:

A = Adenine, T = Thymine, C = Cytosine, G = Guanine.

A ribbon surface is formed by hydrogen bonding between base pairs, where A binds with T, and C binds with G. In the classical Crick–Watson model for DNA, the ribbon is twisted in a right-handed helical fashion. If two DNA ribbons are twisted about each other in a helical fashion (like the twisting of the DNA backbones about each other), each full twist forms a DNA supercoil. If the axis of a DNA ribbon is planar (relaxed DNA), the pitch of the twisted ribbon is approximately 10·5 base pairs per full helical twist. If the axis of the DNA is non-planar, or the DNA is under stress, or bound to a protein, the helical pitch can change. Duplex DNA can exist in closed circular form, where the ribbon surface forms a twisted orientable band (instead of a twisted Möbius band). In the case of circular duplex DNA, the ribbon surface forms a mathematical framing for the normal bundle of the central axis of the molecule. Given a circular duplex DNA molecule, one can 'nick' (break) one of the two DNA backbones with an enzyme called dnase, and the nicked molecule then loses its desire to supercoil in 3-space and 'relaxes'. Nicking one of the backbone strands does not change the knot type of the axis of the molecule. Relaxed unknotted circular duplex DNA, for example, is believed to assume a configuration in which its axis is nearly planar. After relaxation by nicking, one can then repair the nick in the DNA backbone with another enzyme called ligase, obtaining a relaxed DNA molecule with no breaks in either backbone. If one orients the backbone strands in a parallel fashion, the relaxed-state linking number (Lk_0) of the backbone strands of the ligated relaxed molecule can be used as a reference for measuring the framing of the supercoiled molecule of the same knot type, using the linking number (Lk) of the backbone strands of the supercoiled molecule. For unknotted circular duplex DNA, the framing ($Lk_0 - Lk$) is an observable via gel electrophoresis, because a change of ± 1 in the framing is converted by the local stiffness of the ribbon (its desire to maintain a locally constant helical pitch) into ± 1 supercoils (see [24, 33]). The differential geometry of ribbons in 3-space [33] plays an important role in understanding supercoil formation, helical twist of duplex DNA, and the mechanism of topoisomerase enzymes which pass DNA through itself in order to reduce molecular entanglement (see [29]).

Site-specific recombination is one of the ways nature alters the genetic code of an organism, either by moving a block of DNA to another position on the molecule (a move performed by transposase), or by integrating a block of alien DNA into a host genome (a move preformed by integrase). An enzyme which mediates site-specific recombination on DNA is called a recombinase. A recombination site for a given recombinase is a short (10–15 base pairs) linear segment of DNA whose genetic coding is recognized by the recombinase. Site-specific recombination can occur when a pair of sites (on the same or on different DNA molecules) becomes juxtaposed in the presence of the recombinase. The pair of sites is aligned (brought close together), probably through enzyme manipulation or random thermal motion (or both), and both sites (and perhaps some contiguous DNA) are then bound by the enzyme. This stage of the reaction is called synapsis, and we will call the protein-DNA complex (in the biochemical sense) formed by the part of the substrate which is bound to the enzyme together with the enzyme itself the local synaptic complex. We will call the entire DNA molecule(s) involved in synapsis (which includes the parts of the DNA molecule(s) not bound to the enzyme) together with the enzyme itself the synaptic complex. After forming the local synaptic complex, the enzyme then performs two
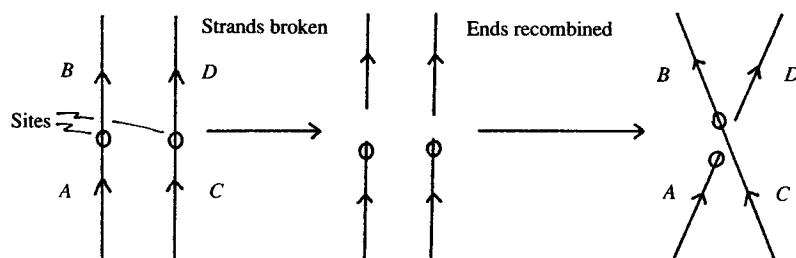
Fig. 6. A single recombination event.

double-stranded breaks at the sites, and recombines the ends by exchanging them in an enzyme-specific manner. The local synaptic complex then dissociates, and the DNA is released by the enzyme. We call the pre-recombination unbound DNA molecule(s) the substrate, and the post-recombination unbound DNA molecule(s) the product. The schematics of a single recombination event on a synaptic complex are shown in Figure 6. In these figures, double-stranded DNA is represented by a single line, and supercoiling is omitted.

The process of recombination clearly involves some interesting topological changes in the substrate. In order to trap these topological changes, one chooses to perform experiments on circular DNA substrate. One must perform an experiment on a large number of circular molecules in order to obtain an observable amount of product. Using cloning techniques, one can synthesize circular duplex DNA molecules which contain two copies of a recombination site. The linear base pair sequence induces a local orientation on each of the recombination sites. The orientation of each of the recombination sites induces an orientation on the ambient circular DNA molecule. If these induced orientations agree, this site configuration is called direct repeats, and if the induced orientations disagree, this site configuration is called inverted repeats. If the substrate is a single DNA circle with direct repeats, the recombination product is a pair of DNA circles, and can form a DNA link (or catenane) (Figure 7). If the substrate is a pair of DNA circles with one site each, the product is a single DNA circle (Figure 7 read in reverse), and can form a DNA knot. If the substrate is a single DNA circle with inverted repeats, the product is a single DNA circle, and can form a DNA knot (Figure 8).

The experimental strategy is to observe the enzyme-caused changes in the euclidean geometry (supercoiling) and topology (knotting and linking) of the DNA, and to deduce enzyme mechanism from these changes. This has been called the topological approach to enzymology [30]. The geometry and topology of the circular substrate are experimental control variables. The geometry and topology of the recombination reaction products are observables. It is a new observation technique (*rec A* enhanced electron microscopy) [15] which makes possible the detailed knot-theoretic analysis of reaction products. *Rec A* is an *E. coli* protein which binds to DNA. Naked duplex DNA is approximately 20 angstroms in diameter, and *rec A* coated DNA is approximately 100 angstroms in diameter. The process of *rec A* coating fattens, stiffens and stretches (untwists) the DNA. The enlarged diameter of *rec A* coated DNA means that crossovers in electron micrographs of *rec A* coated DNA circles can (usually) be unambiguously resolved. *In vitro* (in the laboratory) experiments usually proceed as follows. Circular substrate is prepared, with all of the

substrate molecules representing the same knot type (usually the unknot). The amount of supercoiling of the substrate molecules (the supercoiling density) is also a control variable. The substrate molecules are reacted with a high concentration of purified enzyme, and the reaction products are fractionated by gel electrophoresis. DNA molecules are naturally negatively charged, with the amount of negative charge proportional to the molecular weight. A gel is a protein medium through which the DNA molecules can be forced under the influence of an electric field. A DNA sample is placed at the top of a gel column, and similar molecules migrate in the electric field with similar velocities, forming discrete DNA bands in the gel when the electric field is turned off. Normally, gel electrophoresis discriminates among DNA molecules on the basis of molecular weight; given that all molecules are the same molecular weight (as is the case in these experiments), electrophoresis discriminates on the basis of subtle differences in the geometry (supercoiling) and topology (knot and link type) of the DNA molecules. For example, in unknotted DNA, gel electrophoresis discriminates on the basis of number of supercoils (framing), and can detect a difference of one in the number of supercoils. In gel electrophoresis of knotted and linked DNA, one must nick the reaction products prior to electrophoresis in order to relax the molecular knots and links, because supercoiling confounds the gel behaviour. For nicked DNA knots and links, under the proper conditions gel velocity is (surprisingly) determined by the crossover number of the knot or link; knots and links of the same crossover number migrate with the same gel velocities (see [10]). After running the gel, the gel bands are excised, and the DNA molecules are removed from the gel, and coated with *rec A* protein. The DNA molecules are then shadowed with platinum for viewing under the electron microscope. Electron micrographs of the reaction products (see Figures 9, 10) are made, and frequency distributions of knot types of the products are prepared.

5. *The tangle model*

In site-specific recombination on circular substrate, two kinds of geometric manipulation of the DNA occur. The first is a global move, in which the sites are juxtaposed, either through enzyme action or random collision (or some combination of these or other processes), and the enzyme binds to the DNA, forming the synaptic complex. In electron micrographs of synaptic complexes in various experiments, the enzyme can be seen as a small black mass attached to a large DNA circle. After synapsis is achieved, the next move is local, and entirely due to enzyme action. Within the region controlled (bound) by the enzyme, the enzyme breaks the molecular circle(s) at each site, and recombines the ends by exchanging them. We will model this local move, and focus on the pre-recombination and post-recombination structure of the local synaptic complex. We model the enzyme itself (or, if necessary, its sphere of influence) as an enzyme ball (B^3). The local synaptic complex consisting of the pair (enzyme, bound DNA) forms a 2-string tangle in the enzyme ball. We make the usual biological assumption that the enzyme mechanism is constant, independent of the variable euclidean geometry (supercoiling) and topology (knotting and linking) of the substrate population. That is, given any two pre-recombination copies of the local synaptic complex pair (enzyme, bound DNA), one can by rotation and translation superimpose one pair on the other. We likewise assume that all of the
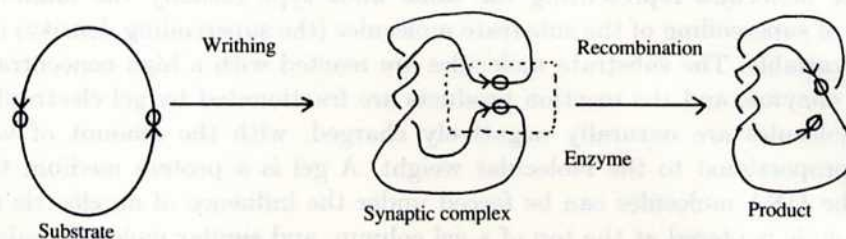
Fig. 7. Hypothetical recombination link synthesis.

copies of the post-recombination local synaptic complex are congruent. The enzyme ball does not come equipped with a preferred boundary parametrization; in order to do calculations, we must coordinatize the local synaptic complex by choosing one. We assume that this choice of boundary parametrization is enzyme-specific, in the sense that one makes the 'same' choice over a population of local synaptic complexes which may form during an experiment. At synapsis, we assume that the recombination action takes place entirely within the interior of the enzyme ball, and that the substrate configuration outside the ball remains fixed while the strands are being broken and recombined inside the enzyme ball.

For symmetry of mathematical exposition, we take the point of view that the reaction is taking place in the 3-sphere S^3. A single synaptic complex will involve one copy of the enzyme, and one or two substrate DNA circles. The boundary of the recombination ball (the enzyme S^2) divides the pair (S^3, DNA circle(s)) into two complementary tangles, glued together along their common boundary. In Figures 7, 8, the dotted circle represents an equatorial circle on the enzyme S^2. The enzyme S^2 divides the substrate into two complementary tangles, the outside tangle O, and the site tangle T. The tangle T is the pre-recombination local synaptic complex. The local effect of recombination is to perform tangle surgery, that is, to delete the site tangle T from the synaptic complex, and replace it with the recombinant tangle R. The tangle R is the post-recombination local synaptic complex. As in Figure 8, the knot types of the substrate and the product yield equations in the recombination variables O, T and R. Specifically, a single recombination event products two equations in three unknowns:

Substrate equation: $N(O+T)$ = substrate;

Product equation: $N(O+R)$ = product.

Tangle model strategy. The purpose of the tangle model [25, 26, 27] is to compute the structure of the pre-recombination and post-recombination synaptic complex, given information about the substrate and product. Ideally, we will treat each of O, T, R as recombination variables, and solve the above equations posed by experiment for these unknowns. Since a single recombinant event yields only two equations involving three unknowns, the best we can hope for, given only this information, is to solve for any two in terms of a third. The analysis is greatly simplified at this point by making the following biologically reasonable assumption:

Biological assumption. The enzyme mechanism (the change from T to R) is an enzyme-specific constant, independent of the geometry and topology of the substrate.
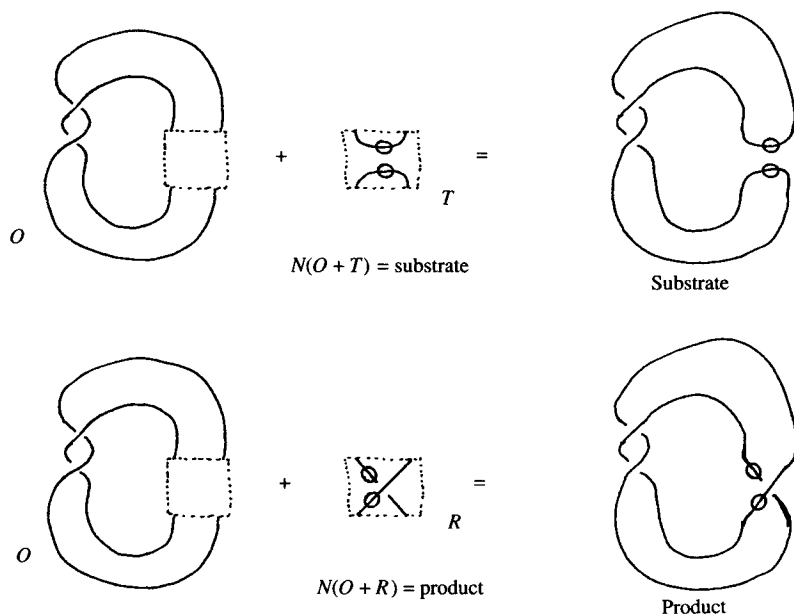
Fig. 8. The substrate and product equations.

In some experiments, the substrate may be a population of unknotted circular molecules, but these molecules come equipped with different amounts of supercoiling. Single recombination events can trap some of this euclidean geometry of unknots (as in Figures 7 and 8), producing a distribution of product knot (link) types from a single substrate knot type. This is what happens in the *phage* λ experiment, described in Section 7. In the *phage* λ experiment, then, the outside tangle O can vary over a number of configurations, but the tangles T and R appear in a number of pairs of equations, one for each product type.

In the *Tn3* resolvase experiment which is described in Section 6, multiple rounds of recombination occur, and some of the variables appear often enough to allow solution of the equations. We wish to model multiple rounds of recombination. There are two ways in which multiple recombination can happen. One way is called distributive recombination, in which the enzyme binds to the DNA, performs one or more recombination moves, releases the DNA, and then another copy of the enzyme attaches to the unbound DNA, performs recombination, etc. In distributive recombination, multiple rounds of recombination occur at multiple binding encounters between DNA and enzyme. A mathematically simpler and biologically significant situation is processive recombination, in which multiple rounds of recombination occur at a single binding encounter between DNA and enzyme, after which the DNA is released and undergoes no more recombination. In this situation, the synaptic complex stays intact through multiple recombination events. In order to model processive recombination, we make the following biologically reasonable assumption:

Mathematical assumption. Processive recombination acts by tangle addition; for each additional round of processive recombination, we add a copy of the recombinant tangle R.

Equivalently, for processive recombination, the transformation from processive recombinant round one to processive recombinant round two is to replace the tangle R by the tangle $(R+R)$, etc. This assumption of tangle addition allows us to solve the tangle equations, obtaining answers which agree with observations.

In the analysis of the phage λ and $Tn3$ resolvase experiments, we make use of recent results in Dehn surgery on 3-manifolds [9] to prove that many of the solutions to the tangle equations are rational tangles. If one makes the reasonable biological assumption that the enzyme itself forms a 3-ball, and that the two DNA strands are bound to the surface of the enzyme (both before and after recombination), then the pre- and post-recombination local synaptic complexes (the tangles T and R) will be rational tangles. We do not make this biological assumption in the analysis below, since the simplicity of the recombination products forces rationality of many of the tangle solutions.

Another mathematical treatment of site-specific recombination occurs in [34], in which the Jones polynomial is used. In [34] (in terms of the tangle notation of this paper) it is assumed that the site tangle $T = \langle 0 \rangle$, and that the recombinant tangle $R = \langle \pm 1 \rangle$. Knowledge of the topology of the substrate and product(s) is then used to compute the Jones polynomials of other products. This contrasts with our method, in which we solve tangle equations to obtain information about T, R and O.

6. $Tn3$ resolvase

$Tn3$ resolvase is a site-specific recombinase which reacts with certain circular duplex DNA substrate with directly repeated recombination sites [31, 32]. One begins with supersoiled unknotted DNA substrate, and treats it with resolvase. Most of the time, resolvase mediates one round of recombination, and releases the linked product. The principal product [31] of this reaction is known to be the Hopf link of Figure 7, the 4-plat $\langle 2 \rangle$. Moreover, when endowed with orientation inherited from the parent unknotted substrate, this recombination product has linking number -1: see [31]. Resolvase is known to act dispersively in this situation: to bind to the circular DNA, to mediate a single recombination event, and then to release the linked product. It is also known that resolvase and free (unbound) DNA links do not react, presumably because the enzyme cannot manage the juxtaposition of the sites when each site is on a separate molecule. However, in one in 20 encounters, resolvase acts processively: additional recombinant strand exchanges at a single binding encounter are promoted prior to the release of the product, with yield decreasing exponentially with increasing number of recombination rounds. Minor products are the figure 8 DNA knot (the 4-plat $\langle 2, 1, 1 \rangle$), believed to be the result of two rounds of processive recombination; the DNA Whitehead link (the 4-plat $\langle 1, 1, 1, 1, 1 \rangle$), believed to be the result of three rounds of processive recombination; and the 6-crossing DNA knot 6_2 (the 4-plat $\langle 1, 1, 1, 2, 1 \rangle$), believed to be the result of four rounds of processive recombination. The discovery of the DNA knot $\langle 1, 1, 1, 2, 1 \rangle$ substantiated a model for $Tn3$ resolvase mechanism: see [32]. Figure 9 shows the DNA knot $\langle 1, 1, 1, 2, 1 \rangle$ (from [32]).

We now prove that the experimental results of the first two rounds of processive $Tn3$ recombination yield tangle equations with four solutions for the tangles O and R. The third round of iterated recombination can then be used to discard three of
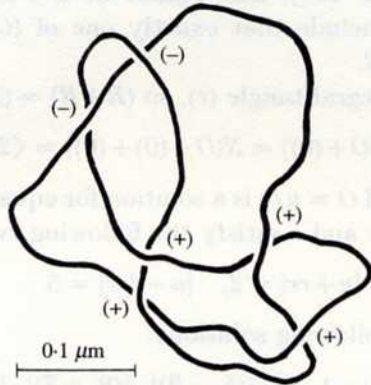
Fig. 9. The DNA knot $\langle 1, 1, 1, 2, 1 \rangle$ (from [32]).

these solutions, leaving us with the one believed to be biologically correct. This theorem can be viewed as a mathematical proof of resolvase synaptic complex structure: the model proposed in [**32**] is the only explanation for the first three observed products of $Tn3$ recombination, assuming that processive recombination acts by adding on copies of the recombinant tangle R. The first round of $Tn3$ recombination is that shown in Figure 7.

THEOREM 6.1. *Suppose that tangles O, T and R satisfy the following equations:* (i) $N(O+T) = \langle 1 \rangle$; (ii) $N(O+R) = \langle 2 \rangle$; (iii) $N(O+R+R) = \langle 2, 1, 1 \rangle$. *Then $\{O, R\}$ is one of* $\{(-3, 0), (1)\}$, $\{(3, 0), (-1)\}$, $\{(-2, -3, -1), (1)\}$ *or* $\{(2, 3, 1), (-1)\}$.

Proof. The first step in the proof is to argue that R must be a rational tangle. Now R, O and $(O+R)$ are locally unknotted, because $N(O+R)$ is the 4-plat $\langle 2 \rangle$. Likewise, T is locally unknotted, because $N(O+T)$ is the unknot. Moreover, if B is a prime tangle of parity (0) or (1), and A is any locally unknotted tangle, then $(A+B)$ is likewise a prime tangle (see [**16**]). Now if R has the parity of ∞, then $N(O+R+R)$ must be a link of either 2 or 3 components. Since $N(O+R+R)$ is a knot, then R has the parity of (0) or (1). Hence, if R is a prime tangle, then so is $(O+R)$. But $N((O+R)+R)$ is the 4-plat $\langle 2, 1, 1 \rangle$, so R cannot be a prime tangle (Lemma 3·1). This means that R is a rational tangle.

The next step is to show that O is a rational tangle. Suppose that O is a prime

tangle. Then T must be a rational tangle, because $N(O+T)$ is the unknot. The 2-fold branched cyclic cover O' is a knot complement. Now R is a rational tangle, so by Lemma 2·4, $R+R$ is locally unknotted. Since $N(O+(R+R)) = \langle 2,1,1 \rangle$, we conclude that $(R+R)$ is a rational tangle. Taking the 2-fold branched cyclic covers, we have that $N(O+R)' = L(2,1)$, and $N(O+R+R)' = L(5,3)$. Since both R and $R+R$ are rational tangles, this means that Dehn surgery on the knot complement O' produces $L(2,1)$ and $L(5,3)$. Application of Lemmas 3·2 and 3·4 shows that O' must be a solid torus, a contradiction to the assumption that O was a prime tangle. We conclude that O is a rational tangle.

We will now use the equations (ii) and (iii) to compute simultaneously the rational tangles O and R. Since both $\langle 2 \rangle$ and $\langle 2,1,1 \rangle$ are achiral 4-plats (equal to their mirror images), given any solution $\{O,R\}$ for these two equations, then $\{-O, -R\}$ is likewise a solution. If both O and R are non-integral, then $N(O+R+R)$ is a non-trivial Montesinos knot (see [7], p. 196), and cannot be a 4-plat. Since $N(O+R+R) = \langle 2,1,1 \rangle$ is a 4-plat, we conclude that exactly one of $\{O,R\}$ must be an integral tangle, and that $O \neq \infty \neq R$.

Suppose that R is the integral tangle (r), so $(R+R) = (2r)$. If $r = 0$, then

$$\langle 2 \rangle = N(O+(0)) = N(O+(0)+(0)) = \langle 2,1,1 \rangle,$$

a contradiction. If $R = r$ and $O = u/v$ is a solution for equations (ii) and (iii), then by Lemma 2·1 the integers u, v and r satisfy the following two equalities:

$$|u+rv| = 2, \quad |u+2rv| = 5.$$

These equations have the following solutions:

$$\{(u, rv)\} = \{(-1,3)\}, \{(1,-3)\}, \{(9,-7)\}, \{(-9,7)\}.$$

From the first solution set, $u = -1$ and $rv = 3$, so as possible values for solutions we have

$$\{(u/v,r)\} = \{(-1/3,1)\}, \{(1/3,-1)\}, \{(-1,3)\}, \{(1,-3)\}.$$

The last two are not solutions to (iii), and are discarded. Working through all the values for $\{(u, rv)\}$ and discarding extraneous solutions, we find that we are left with the solutions stated in the theorem. This means that $\{(-3,0),(1)\}$, $\{(3,0),(-1)\}$, $\{(-2,-3,-1),(1)\}$ and $\{(2,3,1),(-1)\}$ form a complete set of solutions for $\{O,R\}$ in equations (ii) and (iii), if R is integral.

Suppose now that $O = (s)$ and $R = u/v$ (where $v > 1$) form a set of solutions to $\{(ii),(iii)\}$. By Lemma 2·1 we have $|u+sv| = 2$. Since O is integral, then

$$N((O+R)+R) = N(R+(R+O)) = \langle 2,1,1 \rangle.$$

The tangle $(R+O)$ is classified by the rational number $(s+u/v)$, so application of Lemma 2·1 gives

$$|2uv+sv^2| = v|2u+sv| = 5,$$

so $v = 5$. Simultaneous solution of the equations $\{|u+5s| = 2, |2u+5s| = 1\}$ yield solutions

$$(u,s) \in \{(-1,3/5),(1,-3/5),(-3,1),(3,-1)\}.$$

The first two solutions are discarded because s is not an integer; the last two solutions are discarded because they do not yield solutions to equation (iii). ∎

In order to determine which of the four solutions produced in the above theorem is the biologically correct one, we need more experimental evidence: in particular, we need a chiral (not equivalent to its mirror image) recombination product. Without a chiral product, there will always be an even number of solutions. Fortunately, such evidence is available. In the case of site-specific recombination with direct repeats, the result of three processive rounds of *Tn3* recombination is the Whitehead link, which is chiral (as an unoriented link). This chiral product determines the biologically correct solution for $\{O, R\}$, and predicts the result of four or more rounds of iterated recombination at a single binding encounter. The predicted fourth round of processive *Tn3* recombination (the DNA knot $\langle 1, 1, 1, 2, 1 \rangle$) has been observed [**32**].

COROLLARY 6·2. *Suppose that tangles O, T and R satisfy the following equations*:
(i) $N(O+T) = \langle 1 \rangle$; (ii) $N(O+R) = \langle 2 \rangle$; (iii) $N(O+R+R) = \langle 2, 1, 1 \rangle$; (iv) $N(O+R+R+R) = \langle 1, 1, 1, 1, 1 \rangle$. *Then* $\{O, R] = \{(-3, 0), (1)\}$, *and* $N(O+R+R+R+R) = \langle 1, 1, 1, 2, 1 \rangle$.

Proof. From Theorem 6·1, we know that $\{O, R\} = \{(-3, 0), (1)\}$, $\{(3, 0), (-1)\}$, $\{(-2, -3, -1), (1)\}$ and $\{(2, 3, 1), (-1)\}$ form a complete set of solutions to (i)–(iii). Of this group, it is easily verified that $\{(-3, 0), (1)\}$ is the only solution to (iv). ∎

There is another observed chiral product of processive *Tn3* recombination. For site-specific recombination on a circular substrate with direct repeats, the orientation of the circular substrate induces an orientation on the DNA link which is the result of one round of recombination. The observed linking number of the DNA Hopf link produced by one round of *Tn3* recombination is equal to -1 (see [**31**]). How can this information be used? In Corollary 6·2, for example, if one adds to the hypothesis the assumption that the parity of T is (0), and the linking number of $N(O+R)$ is -1, then the knowledge that the second round of processive recombination is the DNA figure 8 knot is enough to determine the biologically correct form of $\{O, R\}$, and to predict the third and all further rounds of processive recombination. On the other hand, knowledge of the correct structure of $\{O, R\}$ and the linking number of the first round of recombination can be used to gain information about the site tangle T.

COROLLARY 6·3. *Suppose that the tangle T satisfies the following equations*:
(i) $N((-3, 0)+T) = \langle 1 \rangle$ *and* (ii) $N((-3, 0)+(1)) = \langle 2 \rangle$ *(with linking number -1). Then T has the parity of (0). If T is rational, then $T = (2x, 3, 0)$ for some integer x.*

Proof. We attempt to solve the equation $N((-3, 0)+T) = \langle 1 \rangle$ for the unknown T. The tangle T cannot have the parity of (1), because then $N(O+T)$ would be a 2-component link. In the equation $N(O+T) = \langle 1 \rangle$, if the unknot is oriented, then the tangle O inherits an orientation. This induced orientation for O depends on the parity of T. The orientation for O in turn induces an orientation for the Hopf link in the second equation, and a linking number is then defined. If T has the parity of (0), this linking number is -1. If T has the parity of ∞, this linking number is $+1$. Therefore T has the parity of (0). If T is a rational tangle, then (by Theorem 2·2) $T = (2x, 3, 0)$ for some integer x. Note that $T = (0)$ if $x = 0$. ∎

The rational tangle calculus can be used to obtain lower bounds on the minimum number of recombinant events separating DNA configurations. For example, in [**28, 31**], it is asked if it is possible to convert the unknot to the Whitehead link

$\langle 1, 1, 1, 1, 1 \rangle$ in a single recombinant event. If one assumes that T is either (0) or ∞, and R is (± 1), then Corollary 3·8 answers this question in the negative.

7. *Phage λ integrase*

Bacteriophage lambda is a virus which attacks bacteria, inserting its own genetic material into that of the host. The genetic insertion mechanism is site-specific recombination by the enzyme *Int*. One prepares unknotted duplex circular substrate with two copies of the recombination site for the *Int* reaction. The first experiment we discuss [23] was performed with inverted repeats of the recombination sites, the second with direct repeats of the recombination sites. The enzyme *Int* does not perform iterated recombination, because the site sequences are severely disrupted by one round of recombination. One round of *Int* recombination on supercoiled unknotted substrate produces a remarkable family of reaction products. Specifically, for inverted repeats, the family of 4-plat DNA knots $\{ \langle -(2k+1) \rangle \mid 0 \leqslant k \leqslant 11 \}$ was observed. For direct repeats, the family of 4-plat DNA links $\{ \langle -(2k) \rangle \mid 0 \leqslant k \leqslant 11 \}$ was observed. The reason that a family of reaction products arises from a single knot type (the unknot) as the substrate is because the unknots come equipped with various degrees of supercoiling, and some of this supercoiling is trapped by the recombination reaction, as in Figure 8. The fact that more than one product was observed in each case means that the outside tangle O varies from substrate molecule to substrate molecule, depending on the degree of supercoiling. Figure 10 (from [23]) shows an electron micrograph of the *Int* knot $\langle -13 \rangle$.

Our first theorem deals with inverted repeats, and argues that the site and recombinant tangles, and one of the outside tangles for *Int* must be rational.

THEOREM 7·1. *Suppose that there exist tangles O_k $(k = 0, 1, 2)$, T and R, which satisfy the equations* (i) $N(O_k + T) = \langle 1 \rangle$ *and* (ii) $N(O_k + R) = \langle -(2k+1) \rangle$. *Then T, R and O_0 are rational tangles.*

Proof. Since $N(O_k + T) = \langle 1 \rangle$, both O_k and T are locally unknotted (for all k); and $N(O_0 + R) = \langle -1 \rangle = \langle 1 \rangle$ means that R is locally unknotted. If T is a prime tangle, then O_k is rational for all k. The 2-fold branched cyclic cover T' is a knot complement which admits 2 Dehn surgeries (for $k = 0, 1$) which produce S^3. These Dehn surgeries must have distinct slopes because of the second set of equations. The knot complements theorem [13] assert, that this is impossible unless the Dehn surgery is performed on the unknot, and hence the tangle T is rational. This result also follows from the fact that strongly invertible knots have property P: see [5]. A similar analysis of both equations for the case $k = 0$ shows that O_0 is rational. Moreover, if R is a prime tangle, then O_k is rational (for $k = 0, 1, 2$), and the 2-fold branched cyclic cover R' admits Dehn surgeries which produce S^3, $L(3, 2)$ and $L(5, 4)$. Application of Lemmas 3·2 and 3·4 proves that R is a rational tangle. █

Our next theorem is the analogue of Theorem 7·1, for the case of *Int* direct repeats.

THEOREM 7·2. *Suppose that there exist tangles O_k $(k = 0, 1, 2, 3)$, T and R, which satisfy the equations* (i) $N(O_k + T) = \langle 1 \rangle$ *and* (ii) $N(O_k + R) = \langle -(2k) \rangle$. *Then T, R and O_0 are rational tangles.*

Proof. As above, T, R and O_k are locally unknotted (for all k). The argument that
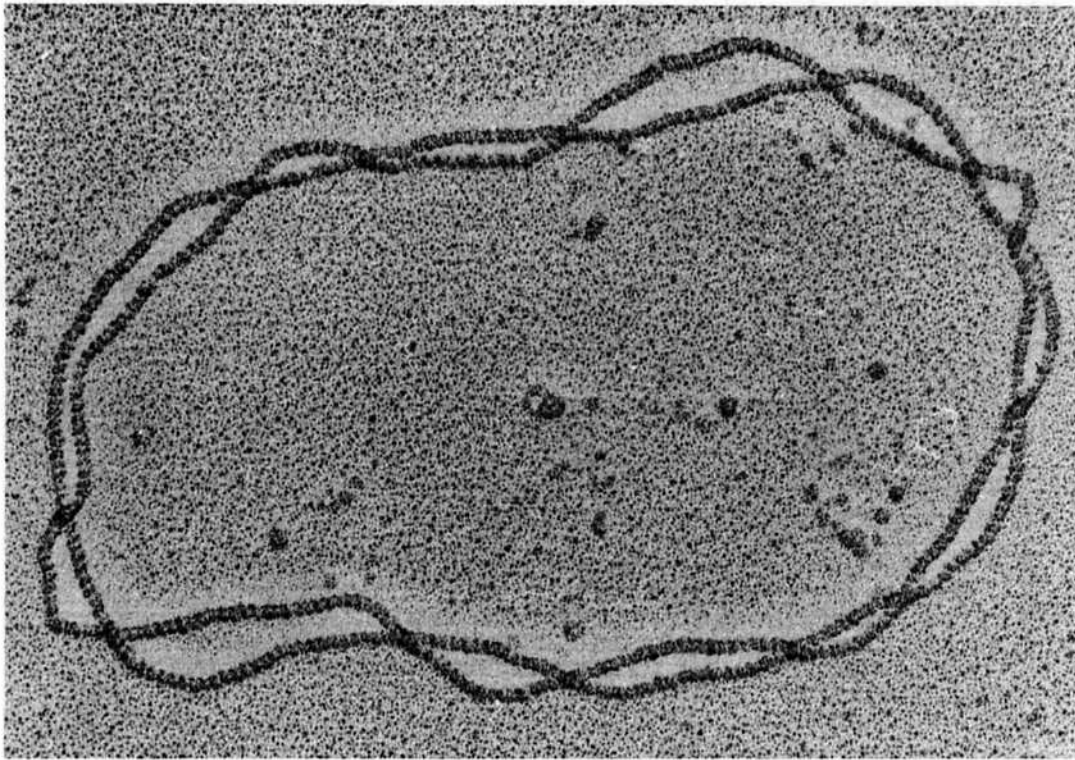
Fig. 10. The DNA knot $\langle -13 \rangle$ (from [23]).

T is a rational tangle is identical to that contained in Theorem 7·1. If R is a prime tangle, then there are 4 Dehn surgeries in the 2-fold branched cyclic cover R' which produce $L(0, 1) = S^1 \times S^2$, $L(2, 1)$, $L(4, 3)$, and $L(6, 5)$. Application of the cyclic surgery theorem yields that R' is a SFS. Application of Lemma 3·5 yields that R' is a solid torus, and so R is a rational tangle.

Since R is rational, the 2-fold cover O'_0 is a knot complement which admits a non-trivial Dehn surgery which produces $S^1 \times S^2$. By Property R for strongly invertible knots [4, 12], this means that O'_0 is a solid torus, and O_0 is a rational tangle. ∎

We would like to obtain some results on the structure of the tangles O_k, T and R. Indeed, one of the main conclusions of [23] was that the observed reaction products lead to the conclusion that the substrate supercoiling is plectonemic, or, in terms of tangles, that the substrate tangle is a row of vertical half-twists ($O_k = (n_k, 0)$ for some integer n_k). Without some assumptions about $\{T, O_k, R\}$, we cannot solve equations like those in the above theorems. One assumption generally made by biologists is that $T = (0)$ or $(0, 0)$ and $R = (\pm 1)$ or $(0, 0)$ (see for example [23]). Even these assumptions are not a great help, because one observes that the prime tangle B of Figure 1 (b) has the property that $N(B + (0)) =$ unknot, and $N(B + (-1)) = \langle -5 \rangle$. So, assuming simplicity of T and R for these equations does not in general rule out prime tangle solutions for O_k. If one additionally assumes that O_k is a rational tangle for all k, then indeed the equations can be solved, and for the answers biologically expected. Full discussion of these matters, as well as analysis of other enzyme experiments is to be found in [27]. Moreover a calculus for framed oriented rational tangles exists,

which utilizes the observed change in framing (when both substrate and a product are unknotted) to give more information on how the ribbons themselves are manipulated by the enzymes.

REFERENCES

[1] C. BANKWITZ and H. G. SCHUMANN. Über Viergeflechte. *Abh. Math. Sem. Univ. Hamburg* **10** (1934), 263–284.

[2] S. BLEILER. Knots prime on many strings. *Trans. Amer. Math. Soc.* **282** (1984), 385–401.

[3] S. BLEILER. Prime tangles and composite knots. In *Knot theory and manifolds. Proceedings, 1983*, Lecture Notes in Math. vol. 1144 (Springer-Verlag, 1985), pp. 1–13.

[4] S. BLEILER. Strongly invertible knots have property R. *Math. Z.* **189** (1985), 365–369.

[5] S. BLEILER and M. SCHARLEMANN. Tangles, property P and a problem of J. Martin. *Math. Ann.* **273** (1986), 215–225.

[6] F. BONAHON and L. C. SIEBENMANN. *New Geometric Splittings of Classical Knots*. London Math. Soc. Monographs. (To appear.)

[7] G. BURDE and H. ZIESCHANG. *Knots* (de Gruyter, 1985).

[8] J. CONWAY. On 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.

[9] M. C. CULLER, C. M. GORDON, J. LUECKE and P. B. SHALEN. Dehn surgery on knots. *Ann. of Math.* (2) **125** (1987), 237–300.

[10] F. B. DEAN, A. STASIAK, T. KOLLER and N. R. COZZARELLI. Duplex DNA knots produced by Escherichia Coli topoisomerase I. *J. Biol. Chem.* **260** (1985), 4795–4983.

[11] C. ERNST and D. W. SUMNERS. The growth of the number of prime knots. *Math. Proc. Cambridge Philos. Soc.* **102** (1987), 303–315.

[12] D. GABAI. Foliations and surgery on knots. *Bull. Amer. Math. Soc.* **15** (1986), 83–97.

[13] C. M. GORDON and J. LUECKE. Knots are determined by their complements. (Preprint, University of Texas, 1988.)

[14] W. HEIL. Elementary surgery of Seifert fiber spaces. *Yokohama Math. J.* **22** (1974), 135–139.

[15] M. A. KRASNOW, A. STASIAK, S. J. SPENGLER, F. DEAN, T. KOLLER and N. R. COZZARELLI. Determination of the absolute handedness of knots and catenanes of DNA. *Nature* **304** (1983), 559–560.

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

[17] W. B. R. LICKORISH. The unknotting number of a classical knot. *Contemp. Math.* **44** (1985), 117–121.

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

[19] J. M. MONTESINOS. *Revetements ramifies de nouds, Espaces fibres de Seifert et scindements de Heegard*. Publicaciones del Seminario Mathematico Garcia de Galdeano, Serie II, Seccion 3 (1984).

[20] D. ROLFSEN. *Knots and Links* (Publish or Perish, 1976).

[21] H. SCHUBERT. Knoten mit zwei Brücken. *Math. Z.* **65** (1956), 133–170.

[22] H. SEIFERT. Topologie dreidimensionaler gefaserter Räume. *Acta Math.* **60** (1933), 147–238.

[23] S. J. SPENGLER, A. STASIAK and N. R. COZZARELLI. The stereostructure of knots and catenanes produced by phase λ integrative recombination: implications for mechanism and DNA structure. *Cell* **42** (1985), 325–334.

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

[25] D. W. SUMNERS. Knots, macromolecules and chemical dynamics. In *Graph Theory and Topology in Chemistry* (Elsevier, 1987), pp. 3–22.

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

[27] D. W. SUMNERS, C. ERNST, N. R. COZZARELLI and S. J. SPENGLER. A topological model for site-specific recombination. (In preparation.)

[28] D. M. WALBA. Topological Stereochemistry. *Tetrahedron* **41** (1985), 3161–3212.

[29] J. W. WANG. DNA topoisomerases. *Scientific American* **247** (1982), 94–109.

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

[31] S. A. WASSERMAN and N. R. COZZARELLI. Determination of the stereostructure of the product of *Tn3* resolvase by a general method. *Proc. Nat. Acad. Sci. U.S.A.* **82** (1985), 1079–1083.

[32] S. A. WASSERMAN, J. M. DUNGAN and N. R. COZZARELLI. Discovery of a predicted DNA knot substantiates a model for site-specific recombination. *Science* **229** (1985), 171–174.

[33] J. H. WHITE. An introduction to the geometry and topology of DNA structure. In *Mathematical Methods for DNA Sequences* (CRC Press, 1989), pp. 225–253.

[34] J. H. WHITE, K. C. MILLETT and N. R. COZZARELLI. Description of the topological entanglement of DNA catenane and knots by a powerful method involving strand passage and recombination. *J. Mol. Biol.* **197** (1987), 585–603.