IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL

J. Phys. A: Math. Theor. **40** (2007) 12377–12395 doi:10.1088/1751-8113/40/41/008

A topological characterization of knots and links arising from site-specific recombination

Dorothy Buck[1] **and Erica Flapan**[2]

[1] Department of Mathematics and Centre for Bioinformatics, Imperial College London, UK
[2] Department of Mathematics, Pomona College, Claremont, CA 91711, USA

E-mail: d.buck@imperial.ac.uk and eflapan@pomona.edu

Received 1 August 2007, in final form 3 September 2007
Published 25 September 2007
Online at stacks.iop.org/JPhysA/40/12377

Abstract

We develop a topological model of knots and links arising from a single (or multiple processive) round(s) of recombination starting with an unknot, unlink, or $(2, m)$-torus knot or link substrate. We show that all knotted or linked products fall into a single family, and prove that the size of this family grows linearly with the cube of the minimum number of crossings. Additionally, we prove that the only possible nontrivial products of an unknot substrate are $(2, m)$-torus knots and links and those knots and links which consist of two non-adjacent rows of crossings. (In the special case where one row contains only two crossings, these are the well-known *twist* knots and links.) In the (common) case of $(2, m)$-torus knot or link substrates whose products have minimal crossing number $m + 1$, we prove that the types of products are tightly prescribed, and use this to examine previously uncharacterized experimental data. Finally, we illustrate how the model can help determine the sequence of products in multiple rounds of processive recombination.

PACS numbers: 87.14.gk, 02.40.Sf, 87.15.kj
Mathematics Subject Classification: 92C40, 92E10, 57M25

1. Introduction

Molecular biologists are interested in DNA knots and links, because they have been implicated in a number of cellular processes. The axis of DNA molecules can become knotted or linked as a result of many reactions, including replication and recombination. The wide variety of DNA knots and links observed has made separating and characterizing these molecules a critical issue. Experimentally, this is most conclusively accomplished via electron microscopy [17]. However, this is a laborious and difficult process. Thus topological techniques, such as those presented here, can aid experimentalists in characterizing DNA knots and links by restricting the types of knots or links that can arise in a particular context.

This work focuses on one such DNA knotting and linking process, *site-specific recombination*, mediated by a protein, known as a *site-specific recombinase*. Site-specific recombination is important because of its key role in a wide variety of biological processes. (See e.g. [4] or [9] for more information.) In addition, pharmaceutical and agricultural industries have become increasingly involved in genetically modifying organisms or testing whether a mutation in a particular gene leads to a disease. As a result, these industries are now interested in site-specific recombinases as tools for precisely manipulating DNA (e.g. [13]). Experimentally, site-specific recombination is performed on small circular DNA molecules (5 000–10 000 basepairs, short enough to allow agarose gel electrophoresis), which are artificially created by cloning in recombination site(s).

Site-specific recombination roughly has three stages. Two recombinase molecules first bind to each of two specific sites on one or two molecules of covalently closed circular DNA (known as the *substrate*) and then bring the sites close together. We shall refer to these DNA sites as the *crossover sites*. Next, the sites are cleaved, exchanged and resealed. The precise nature of this intermediary step is determined by which of the two recombinase subfamilies the particular protein belongs to (see assumption 3 below for more details). And finally, the rearranged DNA, called the *product*, is released.

Multiple rounds of strand exchange can occur before releasing the DNA—this process is known as *processive* recombination. This is in contrast to *distributive* recombination, where the entire process of recombination (including releasing and rebinding) occurs multiple times. In the case of processive recombination, we use the term *substrate* exclusively for the DNA molecule prior to the first cleavage (before the first round of recombination). We treat processive recombination as one extended process, given a single substrate, with several intermediate exiting points for the reaction.

Understanding precisely which knots and links arise during site-specific recombination can help one understand the details of this process (e.g. [14]). In this paper, we begin by developing a model that predicts all possible knots and links which can arise as products of a single round of recombination, or multiple rounds of (processive) recombination, starting with substrate(s) consisting of an unknot, an unlink, or a $(2, m)$-torus knot or link (henceforth denoted by $T(2, m)$). Our model is independent of site orientation, and we make no assumption on the size (number of basepairs) of the substrate molecule(s). This model rigorously develops and extends ideas that we originally sketched in [5]. Of all possible knotted or linked substrates, we have chosen to focus on $T(2, m)$, because $T(2, m)$ are the most commonly occurring knots and links in DNA. Our model rests on three assumptions that we justify biologically in [4]. Building on these assumptions, we use knot theoretic techniques to prove that all products fall within a single family, illustrated in figure 10.

We will use the following terminology and notation. Let J denote a substrate which is either an unknot, an unlink or $T(2, m)$ (illustrated in figure 5). We use the term *recombinase complex, B,* to refer to the convex hull of the four bound recombinase molecules together with the two crossover sites, and use the term *recombinase–DNA complex* to refer to B together with the substrate J. If the recombinase complex meets the substrate in precisely the two crossover sites then we say the recombinase complex is a *productive synapse*. In particular, for recombinases that utilize an enhancer sequence or accessory proteins, the accessory sites or enhancer sequence must be sequestered from the crossover sites in order to form a productive synapse—see [4] for further details. In figure 1, we illustrate two examples where the recombinase complex B is a productive synapse, and one where B is not. Finally, we let $C = \mathrm{cl}(\mathbb{R}^3 - B)$.

This paper complements earlier work by Sumners, Ernst, Cozzarelli and Spengler [20], which used the tangle model [11] and several biologically reasonable assumptions to solve

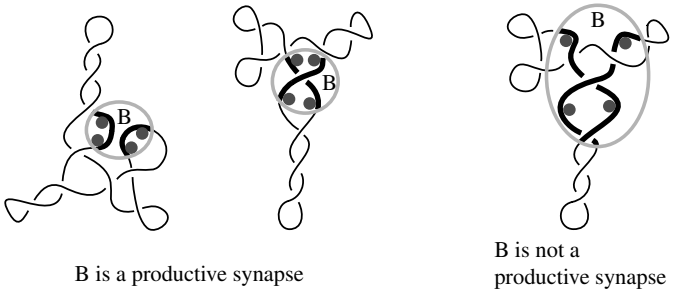

Figure 1. The two examples on the left have a productive synapse and the one on the right does not. The crossover sites are highlighted.

tangle equations. They then determined which 4-plat knots and links arise as a result of (possibly processive) site-specific recombination on the unknot for the serine subfamily of recombinases (see just before assumption 3 for a discussion of the two subfamilies). For the particular case of the recombinase Gin, they considered the knots $3_1, 4_1, 5_2$ or 6_1 as well as unknotted substrates. Our paper goes further in several ways. In addition to allowing an unknotted substrate for an arbitrary recombinase, we allow substrates that are unlinks with one site on each component, as well as any $T(2, m)$. Also, our assumptions are based exclusively on the biology of the recombination process. For example, we do not assume that all products must be 4-plats. Allowing products which are not 4-plats is important because recombination has been seen to produce knots and links which are connected sums (see [4]). Furthermore, we do not assume that the tangle model holds. The tangles used in the tangle model are all required to be 2-string tangles (B, t), where t consists precisely of two arcs [11]. In contrast, in our model, while $(B, B \cap J)$ and $(C, C \cap J)$ can be 2-string tangles, this is not always the case. We prove in lemma 1 that $C \cap J$ can have one of four forms, which are illustrated in figure 8. Observe that forms C3 and C4 could each contain a simple closed curve as well as two arcs, and thus are not 2-string tangles.

The structure of the paper is as follows. In section 2, we state our three assumptions about the recombinase–DNA complex, and use our assumptions to determine the pre-recombinant and post-recombinant forms of $B \cap J$. In section 3, we characterize the forms of $C \cap J$ for each of our substrates. In section 4, we glue each of the post-recombinant forms of $B \cap J$ to each form of $C \cap J$ to determine all possible knotted or linked products predicted by our model. In section 5, we prove that the number of product knots and links predicted by our model grows linearly with the cube of the minimal crossing number. We further prove that the product knot or link type is tightly prescribed when the substrate is $T(2, m)$ and the product has minimal crossing number $m + 1$. Finally, in section 6, we discuss several applications of our model. In particular we explain how the model can, in some cases, determine the order of products of multiple rounds of processive recombination. We also use the results of section 5 to predict the knot or link type of previously uncharacterized experimental data.

2. Our assumptions and $B \cap J$

2.1. The three assumptions

We make the following three assumptions about the recombinase–DNA complex, which we state in both biological and mathematical terms. In [4], we provide experimental evidence showing that each of these assumptions is biologically reasonable.

Figure 2. We fix a projection of J so that $B \cap J$ has one of these pre-recombinant forms in the initial substrate.

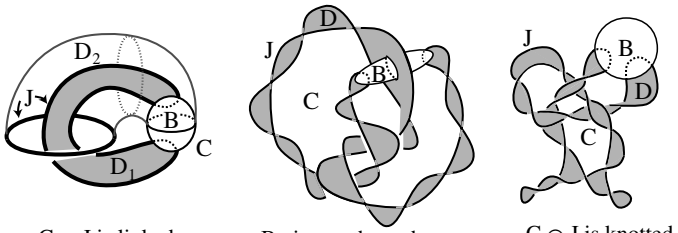

Figure 3. Left: a link is trapped in the branches on the outside of B. *Middle*: the productive synapse pierces through a supercoil in a nontrivial way. *Right*: a knot is trapped in the branches on the outside of B.

Assumption 1 (Biological). *The recombinase complex is a productive synapse, and there is a projection of the crossover sites which has at most one crossing between the sites and no crossings within a single site.*

Note that we allow the possibility of one crossing, because it is a probable biological configuration. See [4] for additional explanation and examples.

This is equivalent to:

Assumption 1 (Mathematical). *$B \cap J$ consists of two arcs and there is a projection of $B \cap J$ which has at most one crossing between the two arcs, and no crossings within a single arc.*

As a result of this assumption, we now fix a projection of the initial substrate J such that $B \cap J$ has one of the forms illustrated in figure 2. Observe that form B1 can be rotated by 90° to obtain form B2. However, we list form B1 and B2 as two different forms to make subsequent figures easier to follow (see for example figure 6). Recall that for processive recombination, the term substrate refers only to the initial configuration of DNA, before the first cleavage. So we are only requiring that $B \cap J$ have one of these three forms prior to the first cleavage.

Assumption 2 (Biological). *The productive synapse does not pierce through a supercoil or a branch point in a nontrivial way and the supercoiled segments are closely juxtaposed. Also, no persistent knots or links are trapped in the branches of the DNA on the outside of the productive synapse.*

Assumption 2 implies that the recombinase–DNA complex cannot resemble any of the illustrations in figure 3.

In order to restate assumption 2 mathematically, we first introduce some preliminary terminology. Consider a surface lying in a plane together with a finite number of arcs in the surface whose endpoints are on the boundary of the surface (see the illustration on the left in

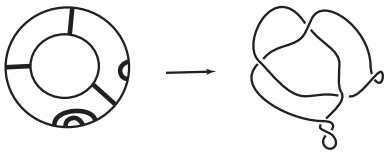

Figure 4. We obtain a *planar surface with twists* by replacing a neighborhood of each arc by a half-twisted band.

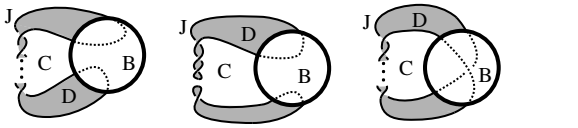

Figure 5. Examples of different substrates J and a spanning surface D bounded by J.

figure 4). We can use this planar surface with arcs to obtain a non-planar surface by replacing a neighborhood of each arc in the original surface by a half-twisted band and removing the top and bottom ends of the band. Figure 4 illustrates how such a surface can be obtained from an annulus together with a collection of arcs defining the twists. Any surface which can be obtained from a planar surface in this way is said to be a *planar surface with twists*.

We shall say that a surface D with boundary J is a *spanning surface* for J, if D is topologically equivalent to a disc, two disjoint discs, or a closed twisted band when J is an unknot, unlink, or $T(2, m)$, respectively. We can think of a spanning surface for J as a soap film surface with boundary J. Figure 5 gives some examples of the relationship between a spanning surface D and the productive synapse B. Observe that in each of the illustrations of figure 5, $D \cap \partial B$ consists of two arcs. By assumption 1, B contains precisely two arcs of $J = \partial D$. Hence ∂B meets J in precisely four points. It follows that the intersection of any spanning surface for J with ∂B contains exactly two arcs. According to biological assumption 2, B does not pierce through a supercoil or a branch point in a nontrivial way. Mathematically, this means that B does not pierce the interior of every spanning surface for J. In general, a spanning surface D is pierced by B if and only if $D \cap \partial B$ contains at least one circle in addition to the required two arcs. For example, in the two illustrations on the left in figure 3, no matter how the spanning surface D is chosen, $D \cap \partial B$ contains at least one circle as well as two arcs.

Suppose that D is a spanning surface for J. We know by biological assumption 2 that the supercoiled segments of J are closely juxtaposed. This implies that we can think of the spanning surface D as a narrow soap film surface. In particular, this means that the two arcs in $D \cap \partial B$ are each very short. So roughly speaking we can assume that the two arcs in $D \cap \partial B$ are co-planar (note that this does not mean that the crossover sites themselves are co-planar). This allows us to define the surface $D \cap C$ to be *unknotted rel* ∂B, if there is an ambient isotopy of C pointwise fixing ∂B which takes $D \cap C$ to a planar surface with twists, where the endpoints of the arcs defining the twists are disjoint from ∂B. For example, $D \cap C$ is unknotted rel ∂B for each of the spanning surfaces in figure 5. However, this is not the case for the surfaces $D \cap C$ in figure 3.

We now restate assumption 2 mathematically as follows.

Assumption 2 (Mathematical). *J has a spanning surface D such that $D \cap \partial B$ consists of two arcs which are co-planar and $D \cap C$ is unknotted rel ∂B.*

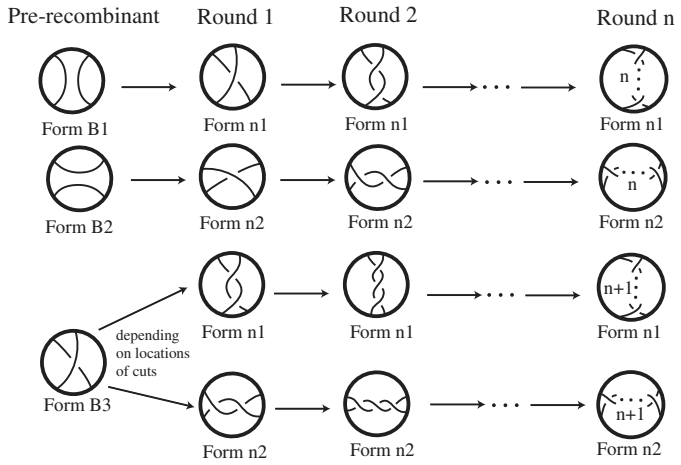

Figure 6. (Processive) recombination by a serine recombinase. On the left are the pre-recombinant forms of $B \cap J$ in the initial substrate. On the right, are the possible post-recombinant forms of $B \cap J$ after the first, second, ..., nth round of processive recombination.

Site-specific recombinases fall into two families—the serine and tyrosine recombinases. Assumption 3 addresses the mechanism of recombination according to which subfamily the recombinase is in. While the overall reactions of the two families of recombinases are the same, they differ in their precise mechanism of cutting and rejoining DNA at the crossover sites. We explain more of the biological details in [4].

Assumption 3 (Biological). ***Serine*** *recombinases perform recombination via the 'subunit exchange mechanism.' This mechanism involves making two simultaneous (double-stranded) breaks in the sites, rotating opposite sites together by* 180° *within the productive synapse and resealing opposite partners. In each subsequent round of processive recombination, the same set of subunits is exchanged and the sense of rotation remains constant. After recombination mediated by a* ***tyrosine*** *recombinase, there is a projection of the crossover sites which has at most one crossing.*

The mathematical restatement of assumption 3 is almost identical to the biological statement.

Assumption 3 (Mathematical). *A* ***serine*** *recombinase cuts each of the crossover sites, adds a crossing within B between the cut arcs on different sites, then reseals. Each subsequent round of processive recombination cuts, adds an identical crossing between the same two arcs, and reseals in exactly the same manner as the first round. After recombination mediated by a* ***tyrosine*** *recombinase, there is a projection of the crossover sites which has at most one crossing.*

2.2. The forms of $B \cap J$

As a result of assumption 1, we fixed a projection of J prior to cleavage such that $B \cap J$ has form B1, B2 or B3 (illustrated in figure 2). It follows from assumption 3 that after n recombination events with serine recombinases, we have added a row of n identical crossings. Thus after n recombination events our fixed projection of $B \cap J$ is isotopic fixing ∂B to one of

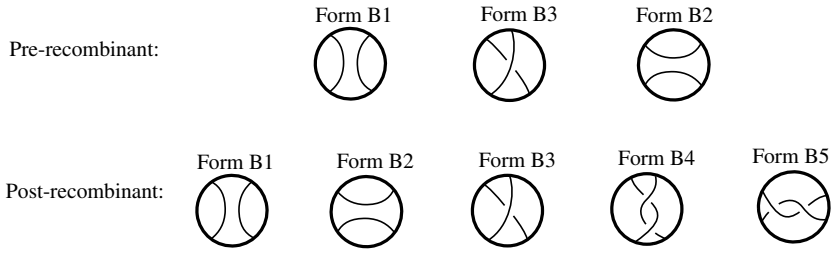

Figure 7. Recombination by a tyrosine recombinase. Above are the pre-recombinant forms of $B \cap J$ in the initial substrate. Below are the possible post-recombinant forms of $B \cap J$ in the product.

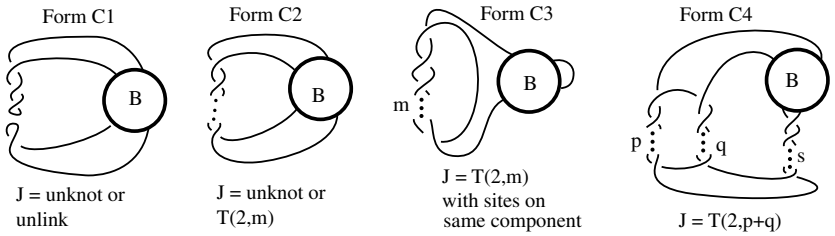

Figure 8. $C \cap J$ has a projection with one of these forms.

the forms illustrated in figure 6 (where the actual crossings can be positive, negative or zero). Note that we can obtain form n2 from form n1 by rotating by 90°. However, we list n1 and n2 as separate forms in order to make it easier to follow the use of figure 13 in the proof of theorem 2.

Also for tyrosine recombinases, we know from assumption 3 that after recombination there exists a projection of $B \cap J$ with at most one crossing. We are working with the pre-recombinant projection of J which we fixed as a result of assumption 1, and we cannot be sure that this particular projection of the post-recombinant $B \cap J$ will have at most one crossing. However, our post-recombinant projection must be ambient isotopic, fixing ∂B, to one of the forms illustrated in figure 7. So without loss of generality we will assume that the post-recombinant projection of $B \cap J$ has one of these forms. Note that forms B1 and B2 are equivalent by a 90° rotation, and forms B3 and B4 are equivalent by a 90° rotation. We list these separately to make it easier to follow the use of figure 12 in the proof of theorem 1.

3. The possible forms of $C \cap J$

Using assumption 2, we now prove the following lemma. The forms of $C \cap J$ referred to in the lemma are illustrated in figure 8. For simplicity, we shall say that $C \cap J$ *has a particular form* if $C \cap J$ is ambient isotopic, pointwise fixing ∂B to that form. Recall that we have already fixed a projection of $B \cap J$, which has one of the forms illustrated in figure 2.

Lemma 1. *Suppose that assumptions 1, 2 and 3 hold for a particular recombinase–DNA complex with substrate J. If J is an unknot, then either $C \cap J$ has form C1 and $B \cap J$ has form B1 or B3, or $C \cap J$ has form C2 and $B \cap J$ has form B2. If J is an unlink, then $C \cap J$*

has form C1 and $B \cap J$ has form B2. If J is $T(2, m)$, then $C \cap J$ can have form C2, C3, or C4. However, if $C \cap J$ has form C4 then $B \cap J$ has form B2.

Proof. We consider separate cases according to the knot or link type of J.

Case 1: J is the unknot

In this case, by assumption 2, we can choose a spanning surface D which is a disc such that $D \cap \partial B$ is two co-planar arcs and $D \cap C$ is unknotted rel ∂B. Since D is a disc, the two arcs of $D \cap \partial B$ separate D such that one of $B \cap D$ and $C \cap D$ is a strip and the other is a pair of disjoint discs. Furthermore, if $C \cap D$ is a strip it is not knotted. Thus, $C \cap D$ is either a pair of disjoint discs or an unknotted twisted strip. It follows that $C \cap J$ has form C1 or form C2. Furthermore, since J is the unknot, if $C \cap J$ has form C1 then $B \cap J$ must have form B1 or B3. Also, since we are assuming that J is supercoiled, if $C \cap J$ has form C2 it must have at least one crossing. In this case, since J is an unknot, $B \cap J$ must have form B2.

Case 2: J is the unlink

In this case, we assume that one site is on each component of J (or else the substrate was actually an unknot). Thus by assumption 2, we can choose a spanning surface D which is a pair of discs such that $D \cap \partial B$ is two co-planar arcs and $D \cap C$ is unknotted rel ∂B. Hence, $B \cap D$ and $C \cap D$ are each a pair of disjoint discs. It follows that $C \cap J$ has form C1. Furthermore, since J is an unlink, $B \cap J$ must have form B2.

Case 3: $J = T(2, m)$

In this case, by assumption 2, we can choose a spanning surface D to be a closed twisted band such that $D \cap \partial B$ is two co-planar arcs and $D \cap C$ is unknotted rel ∂B. Let A be the circle representing the core of D. Observe that regardless of whether D is a twisted annulus or a twisted Möbius band, an arc in D whose endpoints are in ∂D separates D into two components if and only if the arc intersects A in an even number of points. The following argument relies on this observation, and hence is independent of whether or not D is orientable.

Any circle in $\mathbb{R}^3$ must cross a sphere an even number of times (possibly zero). In particular, A must cross ∂B an even number of times. Each point where A crosses ∂B is contained in $D \cap \partial B$. Since the total number of points in $A \cap \partial B$ is even and $D \cap \partial B$ consists of two arcs, either A must intersect each of these two arcs an even number of times, or A must intersect each of the two arcs an odd number of times. If each arc of $D \cap \partial B$ intersects the core A an odd number of times, then each of these arcs cuts D into a strip. Hence the two arcs of $D \cap \partial B$ together cut D into a pair of strips. If each arc of $D \cap \partial B$ intersects the core A an even number of times, then each arc cuts off a disc from D. In this case, either the two arcs cut off disjoint discs in D, or one of the discs is contained inside of the other. In this latter case, the two arcs form the edges of a strip in D, on one side of which is a disc and on the other side of which is a closed twisted band. The three forms of $D \cap \partial B$ are illustrated on the top of figure 9. Note that the illustration on the right may have one, rather than two, rows of twists. Since $B \cap J$ contains at most one crossing, the component of D with almost all of the twists of $T(2, m)$ must be contained in C.

Since $C \cap D$ is unknotted rel ∂B, the abstract forms illustrated on the top of figure 9 yield the corresponding forms of $C \cap D$ which are illustrated in the bottom of figure 9 up to isotopy fixing ∂B. Observe that when $C \cap J$ has form C4, the projection of $B \cap J$ must have form B2 as illustrated. Also, in form C3, while there may be twists to the left of B, they are topologically insignificant, since they can be removed by rotating $D \cap C$ by some multiple of π. Similarly, in form C4, any twists which had occurred above B can be removed and added to the row of twists below B by rotating $D \cap C$ by some multiple of π. These rotations can

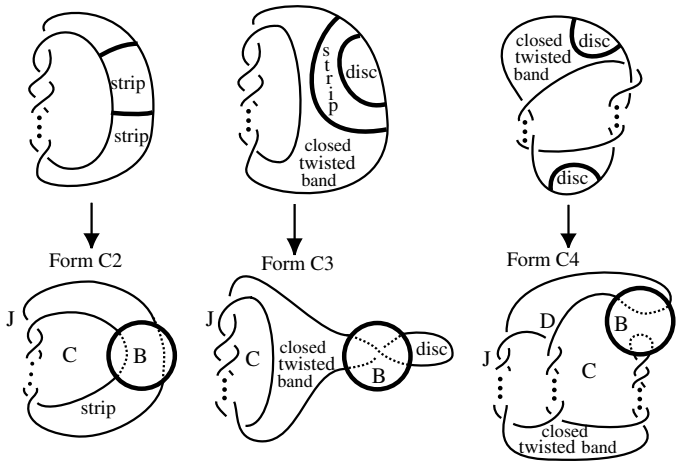

Figure 9. These are the forms of $C \cap J$ when $J = T(2, m)$. If $C \cap J$ has form C4, then B has form B2.

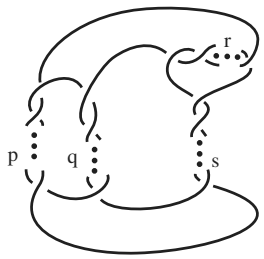

Figure 10. We show that all knotted and linked products are in this family.

occur while pointwise fixing B. Thus the four forms of $C \cap D$ illustrated in figure 8 are the only ones possible. □

4. Product knots and links predicted by our model

In this section, we suppose that the substrate is an unknot, an unlink, or $T(2, m)$ and that all three of our assumptions hold for a particular recombinase–DNA complex. Then we prove theorems 1 and 2, which characterize all possible knotted or linked products brought about by tyrosine recombinases and serine recombinases, respectively. If the substrate is an unknot or unlink we will also show that all nontrivial products are in the torus link family $T(2, n)$, or the family $C(r, s)$ consisting of one row of r crossings and a non-adjacent row of s crossings. We shall refer to knots and links in the latter family as *clasp* knots and links. Note that $C(r, \pm 2)$ is the well-known family of *twist* knots and links. If the substrate is $T(2, m)$, then all products are in the family of knots and links illustrated in figure 10.

Observe that in figure 10, p, q, r and s can be positive, negative or zero. Furthermore, by letting p, q, r and/or s equal 0 or 1 as appropriate, we obtain the five subfamilies illustrated in figure 11. In particular, this means that the products $T(2, m)$ and $C(p, q)$ are contained in the family illustrated in figure 10. Subfamily 3 is the family of *pretzel* knots or links $K(p, q, r)$

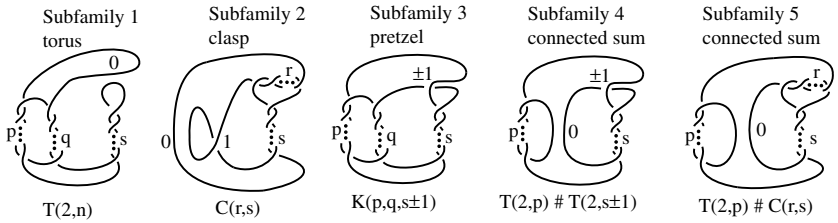

Figure 11. These subfamilies are contained in the family illustrated in figure 10.

with three non-andjacent rows containing p crossings, q crossings and r crossings respectively. Observe that subfamily 4 is a connected sum. However, if $q = 0, r = 1$ and $s = -1$, then it is a $T(2, p)$ together with an unlinked trivial component.

Note that the family illustrated in figure 10 is a subfamily of the family of Montesinos knots and links[3]. All knots which can be drawn with up to seven crossings are in the family of figure 10. However beginning with eight crossings, there are knots which are not in this family. For example, the knot 8_{18} which is not Montesinos, is not in this family. Similarly, beginning with seven crossings there are links which are not in the family of figure 10. In particular, the link 7^2_6, which is not Montesinos, is not in this family.

Theorem 1. *Suppose that assumptions 1, 2 and 3 hold for a particular tyrosine recombinase–DNA complex with substrate J. If J is an unknot then the only possible products are the unknot, the unlink, $T(2, n)$ and $C(p, 2)$. If J is an unlink, then the only possible products are the unknot, the unlink and $T(2, 2)$. If J is $T(2, m)$, then the only possible products are the unknot, the unlink and those knots and links illustrated in figure 10 where $|r| \leqslant 2$.*

Proof. We saw that as a result of assumption 3, after recombination with a tyrosine recombinase, $B \cap J$ has one of the five forms illustrated in figure 7. Also, by lemma 1, $C \cap J$ has one of the four forms illustrated in figure 8. For each of the four forms of $C \cap J$, the products of recombination with tyrosine recombinases are obtained by replacing B with each of the five post-recombinant forms of $B \cap J$ illustrated in figure 7. The resulting products are illustrated in figure 12.

More specifically, first suppose that J is an unknot. Then by lemma 1, $C \cap J$ must have form C1 or C2. Hence by figure 12, the only possible products are the unknot, the unlink, $T(2, m)$ and $C(p, 2)$. Next suppose that J is the unlink. Then by lemma 1, $C \cap J$ has form C1. Hence by figure 12, the only possible products are the unknot, the unlink and $T(2, 2)$. Finally, suppose that J is $T(2, m)$. Then by lemma 1, $C \cap J$ can have form C2, C3 or C4. Hence by figure 12, the possible products are the unknot, the unlink, $T(2, n)$ (possibly with an additional trivial component), $C(p, 2)$, $K(p, q, t)$, $T(2, m)\#T(2, 2)$ and those knots and links illustrated in figure 10 where $r = \pm 2$. Note that we obtain the unlink in the case where $C \cap J$ has form C2 with $p = 2$ and the post-recombinant $B \cap J$ has form B4 with two negative crossings. It follows from figure 11 that $T(2, n)$ (possibly with an additional trivial component), $C(p, 2)$, $K(p, q, t)$ and $T(2, m)\#T(2, 2)$ can all be obtained as knots or links in the family illustrated in figure 10 where $|r| \leqslant 2$. □

Since $T(2, n)$ and $C(p, 2)$ can be obtained as knots or links in the family illustrated in figure 10 with $|r| \leqslant 2$, it follows from theorem 1 that every product of recombination with

[3] Members of our family are obtained by closure of Montesinos tangles of the form $(\frac{1}{p}, \frac{1}{q}, \frac{r}{rs+1})$.

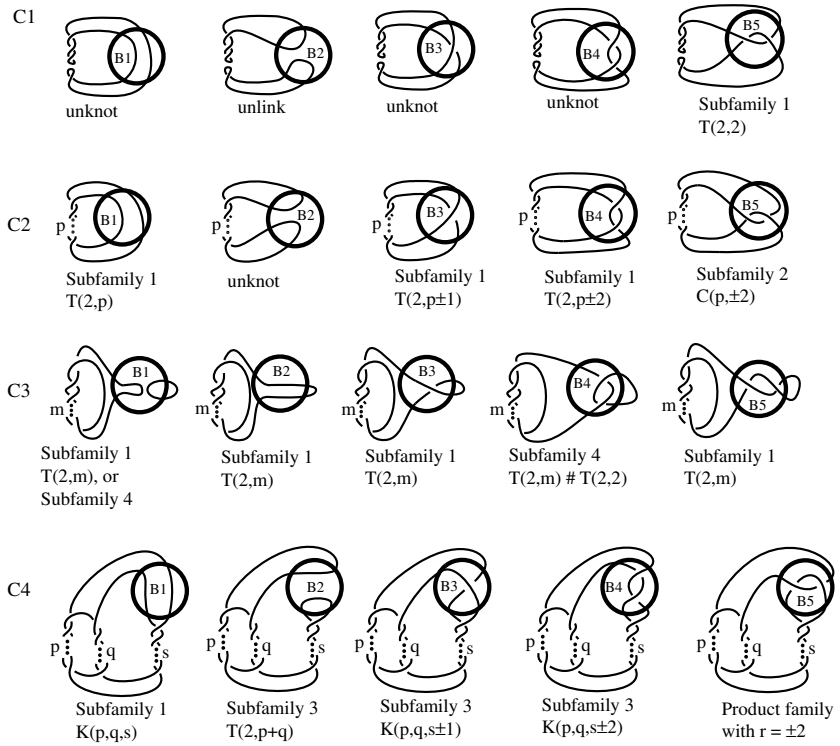

Figure 12. Products of recombination with tyrosine recombinases.

tyrosine recombinases is a member of the family in figure 10 with $|r| \leqslant 2$. Observe that there are knots and links in the family illustrated in figure 10 which do not have a projection with $|r| \leqslant 2$. For example, the knot 7_7 is in the family in figure 10 with $p = -1$, $q = -2$, $r = -3$ and $s = -2$. However, using the machinery developed in [2], it can be shown that there is no way to express the knot 7_7 as a member of this family with $|r| \leqslant 2$, and hence 7_7 is not a product of recombination with a tyrosine recombinase.

Theorem 2. *Suppose that assumptions 1, 2 and 3 hold for a particular serine recombinase–DNA complex with substrate J. If J is an unknot then the only possible products are the unknot, the unlink, $T(2, n)$ and $C(p, q)$. If J is an unlink, then the only possible products are the unknot and $T(2, n)$. If J is $T(2, m)$ then the only possible products are the unknot, the unlink and any knot or link in the family illustrated in figure 10.*

Proof. We saw that as a result of assumption 3, after n recombination events with a serine recombinase, $B \cap J$ has form n1 or n2 as indicated in figure 6. Also, by lemma 1, $C \cap J$ has one of the four forms illustrated in figure 8. We obtain the products of recombination mediated by a serine recombinase from each of the forms of $C \cap J$ by replacing B with each of form n1 and form n2. The resulting products are illustrated in figure 13.

More specifically, first suppose that J is an unknot. Then by lemma 1, either $C \cap J$ has form C1 and $B \cap J$ has form B1 or B3, or $C \cap J$ has form C2 and $B \cap J$ has form B2. If $C \cap J$ has form C1 and $B \cap J$ has form B1 or B3, then by figure 6, the post-recombinant form of $B \cap J$ can be either form n1 or n2. Furthermore, if $B \cap J$ has form B3, then after one

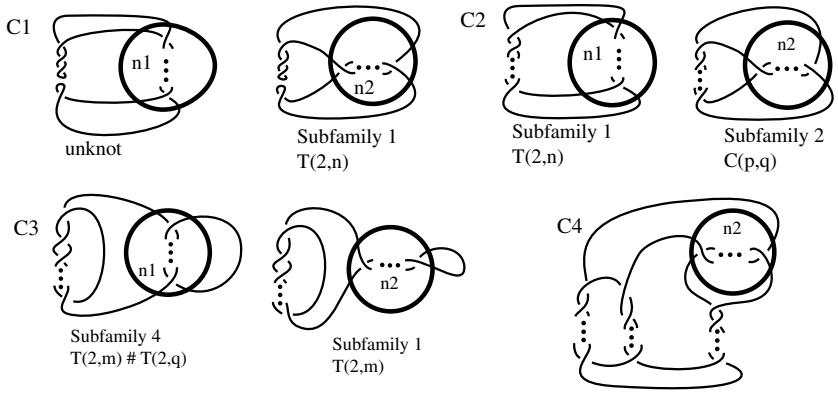

Figure 13. Products of recombination with a serine recombinase.

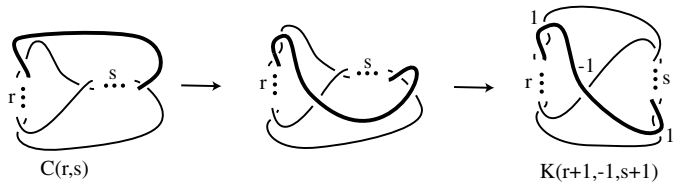

Figure 14. An isotopy from $C(r, s)$ to $K(r + 1, -1, s + 1)$.

round of recombination, form n1 or n2 may have zero crossings. In this case, it follows from figure 13 that the products can be either the unknot, the unlink, or $T(2, n)$. If $C \cap J$ has form C2 and $B \cap J$ has form B2, then the post-recombinant form of $B \cap J$ is form n2. In this case, it follows from figure 13 that the products are all $C(p, q)$.

Next, suppose that J is an unlink. Then by lemma 1, $C \cap J$ has form C1 and $B \cap J$ has form B2. It follows from figure 6 that the post-recombinant form of $B \cap J$ must be form n2. Also, we know by assumption 3 that each round of processive recombination cuts, adds an identical crossing between the same two arcs, and reseals in exactly the same manner as the first round. Since form B2 contains no crossings, after n rounds of recombination, $B \cap J$ contains n identical crossings, either all positive or all negative. Thus we see from form C2 in figure 13 that all products must be $T(2, n)$ with $|n| \geqslant 1$. This includes the unknot but not the unlink.

Finally, suppose that J is $T(2, m)$. Then according to lemma 1, $C \cap J$ can have form C2, C3 or C4. However when $C \cap J$ has form C4, then $B \cap J$ must have form B2. In this latter case, by figure 6, the post-recombinant form of $B \cap J$ must be of form n2. Since there are no other restriction on form n2, we see from form C4 in figure 13 that the products can be any knots or links in the family illustrated in figure 10. Note that we can obtain the unknot and unlink when $C \cap J$ has form C2, and the post-recombinant form of $B \cap J$ has form n1. □

Observe that in contrast with theorem 1, any knot or link in the family illustrated in figure 10 could occur as a consequence of theorem 2. Thus all of the knots and links in this family are possible products of recombination with a serine recombinase.

Table 1 summarizes the results of theorems 1 and 2.

Table 1. Nontrivial products predicted by our model.

Recombinase type	Substrate	Products
Tyrosine	Unknot	Unknot, unlink, $T(2,n)$, $C(2,n)$
	Unlink	Unknot, unlink, Hopf link $= T(2,2)$
	$T(2,m)$	Unknot, unlink, any in figure 10 with $\|r\| \leqslant 2$
Serine	Unknot	Unknot, unlink, $T(2,n)$, $C(p,q)$
	Unlink	Unknot, $T(2,n)$
	$T(2,m)$	Unknot, unlink, any in figure 10

5. The minimal crossing number and our model

5.1. Our family grows with n^3

The minimal crossing number of a DNA knot or link can be determined experimentally using gel electrophoresis. However, there are 1, 701, 936 knots with minimal crossing number less than or equal to 16 [15], and the number of knots and links with minimal crossing number n grows exponentially as a function of n [12]. By contrast, we will now prove that the total number of knots and links in our product family (Figure 10) grows linearly with n^3. Note that, while the knots and links in our family have at most four rows containing p, q, r and s signed crossings respectively, it does not follow that the minimal crossing number of such a knot or link is $|p| + |q| + |r| + |s|$. If the knot or link is not alternating, it is quite possible that the number of crossings can be significantly reduced. Thus there is no reason to believe *a priori* that the number of knots and links in our product family should grow linearly with n^3.

We begin with some results about minimal crossing number which will be used in our proof. We shall denote the *minimal crossing number* of a knot or link K by $\mathrm{MCN}(K)$.

Lemma 2. *Let $|r| > 1$ and $|s| > 1$. Then $C(r,s)$ is equivalent to both $K(r+1,-1,s+1)$ and $K(r-1,1,s-1)$. Furthermore, if r and s have the same sign then* $\mathrm{MCN}(C(r,s)) = |r|+|s|-1$, *and if r and s have opposite sign then* $\mathrm{MCN}(C(r,s)) = |r| + |s|$.

Proof. In figure 14, we show that $C(r,s)$ is ambient isotopic to $K(r+1,-1,s+1)$ by moving the highlighted strand in front of the diagram and then turning the horizontal row of s half-twists so that they become vertical. Analogously, by moving the highlighted strand behind rather than in front of the rest of the diagram, we see that $C(r,s)$ is also ambient isotopic to $K(r-1,1,s-1)$.

We evaluate $\mathrm{MCN}(C(r,s))$ as follows. Murasugi [19] and Thistlethwaite [22] proved that any reduced alternating diagram has a minimal number of crossings. Observe that if r and s have opposite signs, then the standard diagram of $C(r,s)$ is reduced and alternating. In this case, $\mathrm{MCN}(C(r,s)) = |r| + |s|$. If r and s have the same sign and $|r|, |s| > 1$, then either $r, s > 1$ or $r, s < -1$. If $r, s > 1$, then the diagram of $K(r-1,1,s-1)$ is reduced and alternating, since all three rows of crossings are positive. In this case, $\mathrm{MCN}(C(r,s)) = \mathrm{MCN}(K(r-1,1,s-1)) = r-1+1+s-1 = |r|+|s|-1$. If $r, s < -1$, then the diagram of $K(r+1,-1,s+1)$ is reduced and alternating, since all three rows of crossings are negative. In this case, $\mathrm{MCN}(C(r,s)) = \mathrm{MCN}(K(r+1,-1,s+1)) = -(r+1)+1-(s+1) = |r| + |s| - 1$. □

To prove our theorem, we will also make use of the following theorem of Lickorish and Thistlethwaite.

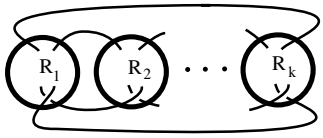

Figure 15. Each R_i is reduced, alternating and has at least two crossings.

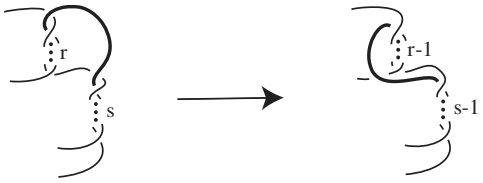

Figure 16. By moving a single strand we reduce from $|r| + |s|$ crossings originally to $(|r| - 1) + (|s| - 1) + 1$ crossings in the alternating diagram.

Theorem [18]**.** *Suppose that a knot or link L has a projection as in figure 15 with* $k \geqslant 3$*, and for each* i*,* $R_i \cap L$ *is a reduced alternating projection which contains a crossing between the two arcs at the bottom of* R_i *(as in figure 15) and at least one other crossing. Then the projection of L has a minimal number of crossings.*

We shall adopt the language of Lickorish and Thistlethwaite and refer to a projection of the form described by their theorem as a *reduced Montesinos diagram*. Thus by the theorem, any projection of a knot or link which is a reduced Montesinos diagram has a minimal number of crossings.

Theorem 3. *The number of distinct knots and links in the product family illustrated in figure 10 with MCN*$= n$ *grows linearly with* n^3*.*

Proof. We begin by fixing n, and supposing that K is a knot or link projection in the family of figure 10 which has MCN$= n$. This projection has $|p| + |q| + |r| + |s|$ crossings. If the given projection of K is reduced alternating or reduced Montesinos, then $|p| + |q| + |r| + |s| = n$. Otherwise, we show that K is ambient isotopic to one of 24 possible projections which have a minimal number of crossings. We will then show that there are at most $96n^3$ possible knots and links in our family with MCN$= n$.

The following example illustrates the type of strand move we shall use to reduce the number of crossings whenever the diagram is neither reduced alternating nor reduced Montesinos. Observe that the part of our knot or link consisting of the rows containing r and s crossings is alternating if and only if r and s have opposite signs. If r and s have the same sign, then by moving a single strand (as in figure 16), this part of the knot or link becomes alternating. This isotopy removes a crossing from both the r row and the s row and adds a single new crossing. Thus we reduce this part of the diagram from having $|r| + |s|$ crossings in a non-alternating form to having $(|r| - 1) + (|s| - 1) + 1$ crossings in an alternating form. All of the isotopies we use to get rid of unnecessary crossings involve moving at most three such strands.

Next we will discuss the one exceptional case where we cannot obtain a reduced alternating or reduced Montesinos diagram by moving some strands of K. This is the case when K is a knot or link in our family with $r > 1$, $p, q < -2$, and $s = 1$. In its original form, the projection has $-p - q + r + 1$ crossings. We can move a single strand of the diagram to obtain a projection

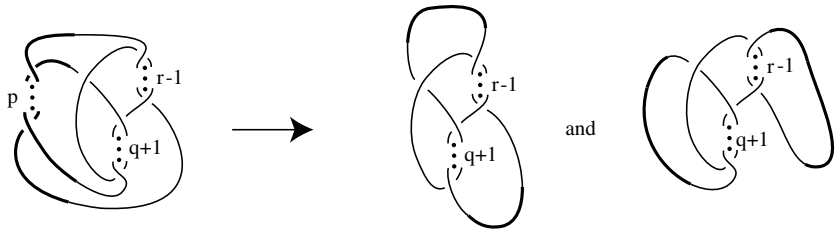

Figure 17. A projection of a knot or link is Hara–Yamamoto if when we cut off the row of p crossings on the left and reseal the strands in the two natural ways, then both resulting projections are reduced alternating.

with only $-p+(-q-1)+(r-1)+1$ crossings (illustrated on the left in figure 17). We define a *Hara–Yamamoto* projection as one in which there is a row of at least two crossings which has the property that if this row is cut off from the rest of the projection and the endpoints are resealed in the two natural ways, then both resulting projections are reduced alternating. The projection on the left of figure 17 is Hara–Yamamoto because the projections (on the right) obtained by resealing the endpoints are both reduced alternating. Hara and Yamamoto [16] have shown that any Hara–Yamamoto projection has a minimum number of crossings. Thus the projection on the left of figure 17 has a minimal number of crossings.

We will consider 27 cases according to the values of p, q, r and s, and show that in all but the above exceptional case K is isotopic to a diagram that is either reduced alternating or reduced Montesinos and hence has minimal crossing number. Since there are so many cases, we display the results in a table rather than discussing each case individually. We make the following notes about the table. In the second column we list the form of the knot or link which has a minimal number of crossings (e.g. reduced alternating). If the knot or link is isotopic to a clasp, pretzel, or torus knot or link we will list the specific form (e.g. $T(2,n)$). If the minimal crossing form is either a clasp $C(r,s)$ or a pretzel of the form $K(r,\pm 1,s)$ then (according to lemma 2) which one of these is the minimal crossing form depends on the signs of r and s. In this case, we just list one of these two forms though the one we list is not necessarily the form with the fewest number of crossings, as we do not know the signs or specific values of the variables. In this case, for the MCN we write an expression with $(-\mathbf{1}?)$ at the end to mean that depending on the relevant variables the MCN may be one smaller. If one of these knots or links contains a trivial component, we use the shorthand +O to indicate this in the table. We shall consider a knot or link and its mirror image to be of the same link type, and hence we will not count both. Thus without loss of generality, we shall assume that $r \geqslant 0$. Also, observe that the rows of crossings containing p and q crossings are intechangeable in figure 10, so we treat the variables p and q as interchangeable. We list the MCN as an unsimplified function of p, q, r and s to help the reader recreate the isotopy taking the original form to the minimal crossing form. Finally, apart from the cases where K reduces to $T(2,m)$ or $C(2,m)$, we obtain the upper bounds for the number of links in each case by expressing MCN$= n$ as a sum of nonnegative integers. This enables us to find an upper bound for the number of knots and links with MCN$= n$ in each case. Note that the upper bounds given are intended to be simple rather than as small as possible. In particular, a number of our cases overlap, and thus some knots and links are counted more than once. Also, for certain specific values of p, q, r and s, we may obtain a trivial knot or link. However, we do not specifically exclude these cases from our table.

There are 26 nontrivial cases in the table. However, all three instances of a $T(2,m)$ yield the same knot or link. Thus there are at most 24 distinct families of knots and links listed

in the table. The number of knots and link in each of these families is bounded above by $4n^3$ (in fact, for most of the cases there are significantly fewer than $4n^3$ knot and link types). It follows that for a given n, the number of distinct knots and links in our product family (Figure 10) which have MCN$= n$ is bounded above by $24 \times 4n^3 = 96n^3$. In particular, the number of distinct knots and links with the form of figure 10 which have MCN$= n$ grows linearly with n^3. □

It follows from theorem 3 that the proportion of all knots and links which are contained in our family decreases exponentially as n increases. Thus, for a knotted or linked product, knowing the MCN and that it is constrained to this family allows us to significantly narrow the possibilities for its precise knot or link type. The model described herein thus provides an important step in characterizing DNA knots and links which arise as products of site-specific recombination.

5.2. Products whose MCN is one more than the substrate

Finally, we prove a more directly applicable theorem as follows. Site-specific recombination often adds a single crossing to the MCN of a knotted or linked substrate. If the substrate is $T(2, m)$ and the product of a single recombination event has MCN $= m + 1$, then we can further restrict the resulting knot or link type.

Theorem 4. *Suppose that assumptions 1, 2 and 3 hold for a particular recombinase–DNA complex with substrate $T(2, m)$, with $m > 0$. Let L be the product of a single recombination event and suppose that* MCN$(L) = m + 1$. *Then L is either $T(2, m + 1)$, $C(-2, m - 1)$, or $K(s, t, 1)$ with $s, t > 0$ and $s + t = m$.*

Proof. For $m = 1$, $T(2, 1)$ is the trefoil knot 3_1, and hence L must be the figure eight knot 4_1 which can also be written as $K(2, 1, 1)$. Thus from now on we assume that $m \geqslant 2$. By assumption 1, there a projection of J such that $B \cap J$ has at most one crossing. Since $J = T(2, m)$, the proof of lemma 1 shows that $C \cap J$ has form C2, C3, or C4 (see figure 8). Furthermore, when $C \cap J$ has form C4, then $p + q = m$. By assumption 3 and figure 6, after a single recombination event with either a serine or tyrosine recombinase the post-recombinant form of $B \cap J$ is one of those illustrated in figure 7. Thus any knotted or linked product L has one of the forms illustrated in figure 12.

First suppose that L has one of the forms illustrated when $C \cap J$ has form C2 or C3. We see that L cannot be $T(2, m)\#T(2, 2)$, since MCN$(T(2, m)\#T(2, 2)) = m + 2$. Certainly, L cannot be $T(2, m)$ with a trivial component. If $L = T(2, n)$ then $n = m + 1$, so we are done. If $L = C(-2, n)$ and $n > 1$, then $n = m - 1$, so again we are done. If $L = C(2, n)$ and $n > 1$, then $L = K(1, 1, n - 1)$. In this case $n = m$, and again we are done.

Now suppose that L has one of the forms illustrated when $C \cap J$ has form C4. If $L = K(p, q, a)$ for some value of a, then L has a projection in the product family illustrated in figure 10 with $r = 1$. Otherwise, L is a member of the product family with $r = \pm 2$. However, if $r = -2$, we can turn over the top loop to get $r = 2$ (this will also add a positive crossing to the s row). Thus we shall now assume that L has a projection in the product family (i.e., figure 10) with either $r = 1$ or $r = 2$. Table 2 lists all of the nontrivial knots and links in this family when $r \geqslant 0$. Thus all of the products that we are considering occur in table 2. We would like to know which of the cases in table 2 have $r = 1$ or 2, $p + q \geqslant 2$ and MCN$= p + q + 1$. Table 3 answers this question.

The only subtle case in table 3 is where $L = C(r \pm 1, p)$. In this case we must have $r = 2$, $|q| = 1$ and $p + q \geqslant 2$. It follows that $p \geqslant 1$. Since L is nontrivial, we must have

Table 2. The minimal crossing forms of knots and links in the family in figure 10.

Values of p, q, r, s for $r \geqslant 0$	Minimal crossing form	Strands moved	MCN written as a sum of nonnegative integers	Upper bound on # of links
$p = q = 0$	$C(r, s)+$ O	0	$r + \lvert s\rvert$ $(-\mathbf{1}?)$	$4n$
$r = 0$	$T(2, p + q)$	0	$\lvert p + q\rvert$	1
$r = 1, p \neq 0, q = 0$	$T(2, p)\#T(2, s + 1)$	0	$\lvert p\rvert + \lvert s + 1\rvert$	$2n$
$r = 1, p \neq 0, q \neq 0$	$K(p, q, s + 1)$	0	$\lvert p\rvert + \lvert q\rvert + \lvert s + 1\rvert$ $(-\mathbf{1}?)$	$8n^2$
$r > 1, p \neq 0, q = 0$	$T(2, p)\#C(r, s)$	0	$\lvert p\rvert + r - s$ $(-\mathbf{1}?)$	$8n^2$
$r > 1, pq = -1$	$T(2, r)$	0	r	1
$r > 1, pq = 1, s = 0$	$C(\pm 2, r)$	0	$2 + r$ $(-\mathbf{1}?)$	2
$r > 1, p \geqslant 1, q = 1, s > 0$	Reduced alternating	1	$p + (r - 1) + (s - 1) + 2$	n^2
$r > 1, p = q = 1, s < 0$	Reduced alternating	1	$r + (-s - 1) + 2$	n
$r > 1, p \leqslant -1, q = -1, s > 0$	Reduced alternating	2	$-p + (r - 1) + (s - 2) + 2$	n^2
$r > 1, p, q < 0, s \leqslant 0$	Reduced alternating	0	$-p - q + r - s$	n^3
$r > 1, p, q > 1, s = 0$	Reduced alternating	2	$(p - 1) + (q - 1) + (r - 2) + 2$	n^2
$r > 1, p < -1, q > 1, s = 0$	Reduced alternating	1	$-p + (q - 1) + (r - 1) + 1$	n^2
$r > 1, \lvert p\rvert > 1, \lvert q\rvert = 1, s = 0$	$C(r \pm 1, p)$	0	$-p + (r \pm 1)$ $(-\mathbf{1}?)$	$4n$
$r > 1, qs = -1$	$T(2, r + p \pm 1)$	1	$\lvert r + p \pm 1\rvert$	1
$r > 1, p > 0, q = 1, s < 0$	Reduced alternating	1	$p + r + (-s - 1) + 1$	n^2
$r > 1, p \leqslant -2, q = 1, s \leqslant -2$	Reduced alternating	1	$(-p - 2) + r + (-s - 2) + 1$	n^2
$r > 1, p, q > 0, s = 1$	Reduced alternating	1	$p + q + (r - 1) + 1$	n^2
$r > 1, p < -1, q > 0, s = 1$	Reduced alternating	1	$(-p - 1) + q + (r - 1)$	n^2
$r > 1, p < -1, q = 1, s > 1$	Reduced alternating	2	$(-p - 1) + (r - 1) + (s - 1) + 2$	n^2
$r > 1, p > 1, q = -1, s < 0$	Reduced alternating	1	$(p - 1) + r - s + 1$	n^2
$r > 1, p > 1, q = -1, s = 2$	Trivial	2	$0 \neq n$	0
$r > 1, p > 1, q = -1, s > 2$	Reduced alternating	3	$(p - 2) + (r - 1) + (s - 3) + 2$	n^2
$r > 1, \lvert p\rvert, \lvert q\rvert > 1, s < 0$	Reduced Montesinos	0	$\lvert p\rvert + \lvert q\rvert + r - s$	$4n^3$
$r > 1, \lvert p\rvert, \lvert q\rvert > 1, s > 1$	Reduced Montesinos	1	$\lvert p\rvert + \lvert q\rvert + (r - 1) + (s - 1) + 1$	$4n^3$
$r > 1, p < -1, q = -2, s = 1$	$K(p, 2, r - 1)$	1	$-p + 2 + (r - 1)$	n
$r > 1, p, q < -2, s = 1$	Hara–Yamamoto	1	$-p + (-q - 1) + (r - 1)$	n^2

$L = C(3, p) = K(2, 1, p - 1)$. Now $\mathrm{MCN}(L) = m + 1$ implies that $p - 1 + 2 = m = p + q$. Thus $L = K(s, t, 1)$ where $s + t = p + q$. Now from table 3, we can see that if $r = 1$ or 2, $p + q \geqslant 2$ and $\mathrm{MCN}(L) = p + q + 1$, then L is either $T(2, m + 1)$, $C(m - 1, -2)$, or $K(s, t, 1)$ with $s, q > 0$ and $s + t = m$. □

6. Applications of our model

We now discuss several applications of our model. Firstly, our model can be helpful in understanding processive recombination mediated by a serine recombinase. Table 1 (which summarizes the conclusions of theorems 1 and 2) can determine or narrow the possibilities for the sequence of products in multiple rounds of processive recombination. For example, suppose for an unlinked substrate, experimental conditions eliminate distributive recombination and products are an unlink and a (nontrivial) torus knot. Then table 1 determines that the order of recombination must be from an unlink substrate to the torus knot (product of the first round) to the unlink (product of the second round).

Secondly, we illustrate an application of theorem 4. Bath *et al* used the links 6_1^2 and 8_1^2 as the substrates for Xer recombination, yielding a knot with MCN = 7 and a knot with MCN = 9, respectively. These products have not been characterized beyond their minimal crossing number, and MCN is not sufficient to determine the knot type. In particular, there are seven knots with MCN = 7 and 49 knots with MCN = 9.

Table 3. Which products can have MCN$= p+q+1$?.

Values of p, q, r, s for $r \geqslant 0$	Minimal crossing form	MCN written as a sum of nonnegative integers	Is $r = 1$ or 2 and $p+q \geqslant 2$?	Can MCN $= p+q+1$?
$p = q = 0$	$C(r, s)$+ O	$r + \lvert s\rvert$ $(-\mathbf{1}?)$	No	–
$r = 0$	$T(2, p+q)$	$\lvert p+q\rvert$	No	–
$r = 1, p \neq 0, q = 0$	$T(2, p)\#T(2, s+1)$	$\lvert p\rvert + \lvert s+1\rvert$	Yes	No
$r = 1, p \neq 0, q \neq 0$	$K(p, q, s+1)$	$\lvert p\rvert + \lvert q\rvert + \lvert s+1\rvert$ $(-\mathbf{1}?)$	Yes	If $s = 0$
$r > 1, p \neq 0, q = 0$	$T(2, p)\#C(r, s)$	$\lvert p\rvert + r + \lvert s\rvert$ $(-\mathbf{1}?)$	Yes	No
$r > 1, pq = -1$	$T(2, r)$	r	No	–
$r > 1, pq = 1, s = 0$	$C(\pm 2, r)$	$2 + r$ $(-\mathbf{1}?)$	Yes	If $r = p+q-1$
$r > 1, p \geqslant 1, q = 1, s > 0$	Reduced alternating	$p + (r-1) + (s-1) + 2$	Yes	No
$r > 1, p = q = 1, s < 0$	Reduced alternating	$r + (-s-1) + 2$	Yes	No
$r > 1, p \leqslant -1, q = -1, s > 0$	Reduced alternating	$-p + (r-1) + (s-2) + 2$	No	–
$r > 1, p, q < 0, s \leqslant 0$	Reduced alternating	$-p - q + r - s$	No	–
$r > 1, p, q > 1, s = 0$	Reduced alternating	$(p-1) + (q-1) + (r-2) + 2$	Yes	Only if $r = 3$
$r > 1, p < -1, q > 1, s = 0$	Reduced alternating	$-p + (q-1) + (r-1) + 1$	Yes	No
$r > 1, \lvert p\rvert > 1, \lvert q\rvert = 1, s = 0$	$C(r \pm 1, p)$	$-p + (r \pm 1)$ $(-\mathbf{1}?)$	Yes	Yes
$r > 1, qs = -1$	$T(2, r + p \pm 1)$	$\lvert r + p \pm 1\rvert$	Yes	Yes
$r > 1, p > 0, q = 1, s < 0$	Reduced alternating	$p + r + (-s-1) + 1$	Yes	No
$r > 1, p \leqslant -2, q = 1, s \leqslant -2$	Reduced alternating	$(-p-2) + r + (-s-2) + 1$	No	–
$r > 1, p, q > 0, s = 1$	Reduced alternating	$p + q + (r-1) + 1$	Yes	No
$r > 1, p < -1, q > 0, s = 1$	Reduced alternating	$(-p-1) + q + (r-1)$	Yes	No
$r > 1, p < -1, q = 1, s > 1$	Reduced alternating	$(-p-1) + (r-1) + (s-1) + 2$	No	–
$r > 1, p > 1, q = -1, s < 0$	Reduced alternating	$(p-1) + r - s + 1$	Yes	No
$r > 1, p > 1, q = -1, s = 2$	Trivial	$0 \neq n$	Yes	No
$r > 1, p > 1, q = -1, s > 2$	Reduced alternating	$(p-2) + (r-1) + (s-3) + 2$	Yes	No
$r > 1, \lvert p\rvert, \lvert q\rvert > 1, s < 0$	Reduced Montesinos	$\lvert p\rvert + \lvert q\rvert + r - s$	Yes	No
$r > 1, \lvert p\rvert, \lvert q\rvert > 1, s > 1$	Reduced Montesinos	$\lvert p\rvert + \lvert q\rvert + (r-1) + (s-1) + 1$	Yes	No
$r > 1, p < -1, q = -2, s = 1$	$K(p, 2, r-1)$	$-p + 2 + (r-1)$	No	–
$r > 1, p, q < -2, s = 1$	Hara–Yamamoto	$-p + (-q-1) + (r-1)$	No	–

Theorem 4 significantly reduces the number of possibilities for each of these products. In particular, it follows from theorem 4 that the seven-crossing products of Xer must be $7_1 = T(2, 7)$, $7_2 = C(5, -2)$ or $7_4 = K(3, 3, 1)$; and the nine-crossing products of Xer must be $9_1 = T(2, 9)$, $9_2 = C(7, -2)$, or $9_5 = K(5, 3, 1)$. All of these possible products are actually 4-plats. This example shows how our model complements the work of [10], which restricts attention to the tangle model and assumes that all products are 4-plats. In [3], building on earlier work of [6, 7, 10, 23], we use our model together with tangle calculus to completely classify all tangle solutions to these Int-Xer equations.

Finally, we note that a number of further applications of our model are discussed in [4]. In some cases, our model can distinguish between products of distributive recombination and products of processive recombination. For example, we determine which products arise from processive and which arise from distributive Tn3 resolvase-mediated recombination on a plasmid with four sites, as used in experiments of Benjamin *et al* [1].

Acknowledgments

The authors wish to thank Ken Millett, Andrzej Stasiak, De Witt Sumners, Alex Vologodskii and Stu Whittington for helpful conversations. We also thank the referees for their careful reading and useful comments—especially for asking the following two interesting questions: how could our model order products of processive recombination and what are the smallest knot and link not in our family (the latter also asked by Ken Millett). Dorothy Buck was partially supported by Grant # DMS-0102057 from the National Science Foundation's Division of

Mathematical Sciences. Erica Flapan was partially supported by an Association for Women in Mathematics Michler Collaborative Research Grant.

References

[1] Benjamin H W, Matzuk M M, Krasnow M A and Cozzarelli N R 1985 Recombination site selection by Tn3 Resolvase: topological tests of a tracking mechanism *Cell* **40** 147–58
[2] Bonahon F and Siebenmann L New geometric splittings of classical knots, and arborescent knots *Preprint*
[3] Buck D The hybrid Int-Xer system and its tangle solutions *Preprint*
[4] Buck D and Flapan E 2007 Predicting knot or Catenane type of site-specific recombination products *Preprint* arXiv:0707.3775v1
[5] Buck D and Flapan E 2007 A model of DNA knotting and linking *Knot Theory for Scientific Objects Proc. Int. Workshop on Knot Theory for Scientific Objects: OCAMI Studies vol 1* ed A Kawauchi pp 75–83
[6] Buck D and Verjovsky Marcotte C 2005 Tangle solutions for a family of DNA-rearranging proteins *Math. Proc. Camb. Phil. Soc.* **139** 59–80
[7] Buck D and Verjovsky Marcotte C 2007 Classification of tangle solutions for integrases, a protein family that changes DNA topology *J. Knot Theory Ramifications*, at press
[8] Colloms S D, Bath J and Sherratt D J 1997 Topological selectivity in Xer site-specific recombination *Cell* **88** 855–64
[9] Craig N, Craigie R, Gellert M and Lambowitz A (ed) 2002 *Mobile DNA II* (Washington, DC: ASM)
[10] Darcy I K 2001 Biological distances on DNA knots and links: applications to Xer recombination *Knots in Hellas '98, J. Knot Theory Ramifications* **10** 269–94
[11] Ernst C and Sumners D W 1990 A calculus for rational tangles: applications to DNA recombination *Math. Proc. Camb. Phil. Soc.* **108** 489–515
[12] Ernst C and Sumners D W 1987 The growth in the number of prime knots *Math. Proc. Camb. Phil. Soc.* **102** 303–15
[13] Feil R 2007 Conditional somatic mutagenesis in the mouse using site-specific recombinases *Handb. Exp. Pharmacol.* **178** 3–28
[14] Grainge I, Buck D and Jayaram M 2000 Geometry of site alignment during Int family recombination: antiparallel synapsis by the Flp recombinase *J. Mol. Biol.* **298** 749–64
[15] Hoste J, Thistlethwaite M and Weeks J 1998 The first 1, 701, 936 knots *Math. Intelligencer* **20** 33–48
[16] Hara M and Yamamoto M 1992 Some links with nonadequate minimal-crossing number *Math. Proc. Camb. Philos. Soc.* **111** 283–9
[17] Krasnow M A, Stasiak A, Spengler S J, Dean F, Koller T and Cozzarelli N R 1983 Determination of the absolute handedness of knots and catenanes of DNA *Nature* **304** 559–60
[18] Lickorish W B R and Thistlethwaite M B Some links with nontrivial polynomials and their crossing number *Comment. Math. Helvetici* **63** 527–39
[19] Murasugi K 1987 Jones polynomials and classical conjectures in knot theory *Topology* **26** 187–94
[20] Sumners D W, Ernst C, Spengler S J and Cozzarelli N R 1995 Analysis of the mechanism of DNA recombination using tangles *Q. Rev. Biophys.* **28** 253–313
[21] http://www.math.utk.edu/ morwen/png/link_stats.png
[22] Thistlethwaite M A 1987 A spanning tree expansion of the Jones polynomial *Topology* **26** 297–309
[23] Vazquez M, Colloms S D and Sumners D W 2005 Tangle analysis of Xer recombination reveals only three solutions, all consistent with a single 3-dimensional topological pathway *J. Mol. Biol.* **346** 493–504