doi:10.1016/j.jmb.2007.10.016 J. Mol. Biol. (2007) 374, 1186–1199

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Predicting Knot or Catenane Type of Site-Specific Recombination Products

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

[1]*Department of Mathematics and Center for Bioinformatics, Imperial College London, London, England SW7 2AZ, UK*

[2]*Department of Mathematics, Pomona College, Claremont, CA 91711, USA*

Received 31 July 2007; received in revised form 3 October 2007; accepted 5 October 2007
Available online 13 October 2007

Edited by J. Karn

Site-specific recombination on supercoiled circular DNA yields a variety of knotted or catenated products. Here, we present a topological model of this process and characterize all possible products of the most common substrates: unknots, unlinks, and torus knots and catenanes. This model tightly prescribes the knot or catenane type of previously uncharacterized data. We also discuss how the model helps to distinguish products of distributive recombination and, in some cases, determine the order of processive recombination products.

Keywords: site-specific recombination; DNA knots; serine recombinases; tyrosine recombinases; DNA topology

Introduction

Since their discovery in the late 1960s, DNA knots and catenanes (also known as links) have been implicated in a number of cellular processes (see Refs. 1,2 and references therein). They can occur as a result of replication and recombination, and as products of enzyme actions, notably with topoisomerases, recombinases, and transposases.[2,3] Most prevalently, DNA knots and catenanes arise as products of topological enzymology experiments on artificially constructed small (5–10 kb) circular DNA plasmids.[4–13] These product knots and catenanes can help determine the binding and mechanism of the enzyme being studied. The variety of DNA knots and catenanes observed has made biologically separating and distinguishing these molecules a critical issue.

Experimentally, DNA knots and catenanes can be resolved via either electron microscopy or electrophoretic migration.[14–16] Electron microscopy, after coating the DNA molecules with RecA to visualize crossings, can definitively determine the precise knot or catenane type. However, this process can be laborious and difficult, particularly in producing a relatively large amount of product and in deciphering the sign of crossings. Alternately, gel electrophoresis will stratify nicked DNA knots and catenanes of a given molecular mass and charge. This is straightforward and requires relatively small amounts of DNA. Typically, the distance a given knot or catenane migrates through the gel is proportional to the minimal crossing number (MCN; the fewest number of crossings with which it can be drawn; see Products Are Even More Restricted When Recombination Adds One Crossing for more details). Knots of greater MCN migrate more rapidly than those with lesser MCN†.[17–19]

However, there are 1,701,936 knots with MCN ≤ 16; hence, a better stratification is needed to positively identify a particular knot.[20] Recent work has shown that two-dimensional gel electrophoresis can separate some prime knots with the same MCN.[15] Unfortunately, there is no clear relationship between relative migration of knots with the same MCN in the second dimension. In some cases, for DNA of a given length (one-dimensional), gel electrophoresis can separate some knots with the same MCN. For example, the five- and seven-crossing torus knots

*Corresponding author. E-mail address: d.buck@imperial.ac.uk.

Abbreviation used: MCN, minimal crossing number.

† However, there are gel conditions where, for example, the unknot will migrate ahead of the trefoil.

migrate more slowly than the corresponding five- and seven-crossing twist knots.[17,21] This has not been generalized, although recent experiments indicate that knots and catenanes may migrate linearly with respect to the average crossing number of a particular conformation—the *ideal configuration*‡ of the knot or catenane.[21,23,24]

For gel electrophoresis, one must also construct an appropriate knot ladder as a control to determine the exact DNA knot or catenane since adjacent bands determine only relative MCN or average crossing number, not precise values. While this can be done in some cases (e.g., T4 topoisomerase will produce a ladder of twist knots[25]), generating such a ladder of known knots and catenanes from DNA of the same length as the unknown knots is highly nontrivial.

Thus, topological techniques, such as those presented here, can aid experimentalists in characterizing DNA knots and catenanes, in particular by restricting the topology of the products observed in gel bands.

In this work, we focus on knots and catenanes that arise from site-specific recombination. *Site-specific recombinases* mediate a rearrangement of the genome (see, e.g., Refs. 26,27 for a more thorough introduction). The recombinases bind and synapse two small DNA segments and then cleave (by nucleophilic displacement of a DNA hydroxyl by a protein side chain), exchange, and reseal the backbones before releasing the DNA. 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.

The result of site-specific recombination can be excision, integration, or inversion of DNA. This corresponds to a wide variety of physiological processes, including integration of viral DNA into the host genome, bacterial gene replication, and plasmid copy number control.[27] If the substrate DNA contains supercoils or if synapsis introduces DNA crossings, these crossings can become knot or catenane nodes in the product.

Topological techniques have already played a significant role in characterizing knotted and catenated products of site-specific recombination. For example, several approaches have been developed to determine a particular DNA knot or catenane type, including utilizing the node number for knots,[2] the Jones polynomial for catenanes,[28] Schubert's classification of 4-plats,[29] and the HOMFLY polynomial.[30] Perhaps most famously, Ernst and Sumners have developed the *tangle model* of recombination to describe the action of particular site-specific recombinases in terms of tangle sums.[31] The tangle model has since been used to determine various features of protein–DNA interactions for a number of specific proteins.[10,32–40]

With the exception of Ref. 38 discussed below, the previous topological treatments began with the precise, biologically determined knot or catenane types of (at least some of) the products. This input was then harnessed in topological arguments that probed various features of the pathway and/or mechanism (e.g., in Ref. 36, this helped determine that the recombinase Flp aligned the crossover sites in antiparallel orientation within the recombinase complex).

Here, we consider the alternate paradigm: Given a few assumptions about the mechanism, we predict which knots and catenanes are putative products.

More specifically, rather than focusing on a specific recombinase as many earlier studies have done, we present a topological model that predicts which knots and catenanes can occur as products of site-specific recombination in general. We do this by describing the topology of how DNA knots and catenanes are formed as a result of a single recombination event (or multiple rounds of processive recombination events), given a plectonemically supercoiled unknot, unlink, or $T(2, m)$ torus knot or catenane substrate (see Fig. 2 for illustrations of each substrate type§). Note that 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 an initial substrate, with several intermediate exiting points for the reaction.

Our model is independent of the size of the substrate, the sign of the supercoils, and the site orientation. We also do not model recombination in terms of tangles; thus, the results here are independent of the restrictions of the tangle model (see Ref. 41 for more details). Our model relies on three assumptions, and we provide biological evidence for each. Given these assumptions, we predict that products arising from site-specific recombination must be members of a single family of products (illustrated in Fig. 5). In Ref. 41, we provide the technical proofs for the model developed here, whose nascent form we sketched in Ref. 42.

This article complements earlier work of Sumners *et al.*, which used the tangle model and several biologically reasonable assumptions to solve tangle equations.[38] They then determined which 4-plat knots and catenanes arise as a result of (possibly processive) site-specific recombination on the unknot for the serine subfamily of recombinases (see Background and Terminology for a discussion of the two subfamilies). (For the particular case of the recombinase Gin, they considered four different knotted substrates in addition to an unknotted substrate.) The current work goes further in several ways. In addition to an unknotted substrate for a generic recombinase, we allow substrates that are unlinks with one site on each component, as well as $T(2, m)$ torus knots and catenanes. Also, our assumptions are based exclusively on the biology

‡ *Ideal geometric configurations* of knots are the conformations that allow maximal radial expansion of a virtual tube of uniform diameter centered on the knot.[22,23]

§ Note: All figures represent the axis of duplex DNA.

of the recombination process. In particular, we do not assume that the tangle model holds or that the products must be 4-plats. This is particularly important as (distributive) recombination has produced composite knots and catenanes, and such knots and catenanes are not 4-plats (see Table 1).

This article is organized as follows. We begin by explaining the terminology and giving background in Background and Terminology. In The Assumptions of Our Model, we state the three assumptions of our model and give supporting experimental evidence for each. In Results, we explain how, given an unknot, unlink, or $T(2, m)$ torus knot or catenane substrate, all possible knotted or catenated products fall into a single characterized family. We also consider the (common) case of substrates, which are $T(2, m)$ torus knots and catenanes whose products have MCN = $m+1$, and explain how the product knot or catenane type is even more tightly prescribed. (The technical proofs of the results in Results can be found in Ref. 41.) Finally, in Discussion and Applications, we discuss how the model can, in some cases, distinguish products of distributive recombination, determine the order of products of processive recombination, and narrow the possible knot or catenane type for previously uncharacterized experimental data.

Background and terminology

We consider DNA that is covalently closed, duplex, and (positively or negatively) plectonemically supercoiled possibly with branch points. Roughly speaking, a circular DNA molecule is plectonemically supercoiled if there is a second-order helix formed by the DNA axis itself (see Ref. 1 for a more complete description). Negatively supercoiled DNA is the typical form of DNA *in vivo*.[28] Negatively supercoiled DNA is often branched—branched DNA structures within supercoiled plasmids *in vitro* have been visualized by electron microscopy.[28,43] Additionally, atomic force microscopy *in situ* illuminates branched plectonemic superhelices at physiological conditions.[44] *In vivo*, there is evidence from several more indirect experiments that branched supercoiled DNA is ubiquitous.[6]

Minimally, site-specific recombination requires both the site-specific recombinase and two short (30–50 bp), identical DNA segments, the *crossover sites*, cloned into one or two small circular DNA plasmids.[27] These crossover sites have nonpalindromic subsequences; thus, each site can be assigned an orientation. Hence, if there are two crossover sites on a single molecule of circular DNA, they can be in either *direct* orientation (head to tail; e.g., …ATGC…ATGC) or *inverted* orientation (head to head; e.g., …ATGC…CGTA). Larger site-specific recombination systems can also require additional proteins (e.g., accessory proteins) and sites (e.g., enhancer sequences).

During site-specific recombination, a recombinase dimer first binds to each crossover site. The two crossover sites are then brought together within a

Table 1. All experimentally characterized products of site-specific recombination on supercoiled unknotted, unlinked, or $T(2, n)$ (torus knot or catenane) substrates

Recombinase	Substrate	Product(s)	Subfamily	Reference
Cre	Unknot (inverted)	Unknot, $T(2, 3)$, $T(2, 5)$, $T(2, 7)$, $T(2, 9)$, $T(2, 11)$	1	9
	Unknot (direct)	Unlink, $T(2, 2)$, $T(2, 4)$, $T(2, 6)$, $T(2, 8)$, $T(2, 10)$, $T(2, 12)$	1	9
	$T(2, 2)$ (direct)	Unknot, $T(2, 3)$	1	10
Flp	Unknot (inverted)	Unknot, $T(2, 3)$, $T(2, 5)$, $T(2, 7)$, $T(2, 9)$, $T(2, 11)$	1	10,11
	Unknot (direct)	Unlink, $T(2, 2)$, $T(2, 4)$, $T(2, 6)$, $T(2, 8)$, $T(2, 10)$, $T(2, 12)$	1	10,11
λ Int	Unknot (PB inverted)	Unknot, $T(2, 3)$, $T(2, 5)$, $T(2, 7)$, $T(2, 9)$, $T(2, 11)$, $T(2, 13)$, $T(2, 15)$, $T(2, 17)$, $T(2, 19)$	1	12
	Unknot (PB direct)	$T(2, 4)$, $T(2, 6)$	1	12
	$T(2, 2)$ (PB direct)	$C(3, -2)$, $C(5, -2)$, $C(7, -2)$, $C(9, -2)$	2	10
	Unknot (LR inverted)	$T(2, 3)$, $T(2, 5)$, $T(2, 7)$	1	10
	Unknot (LR direct)	$T(2, 4)$	1	10
	$T(2, 2)$ (LR direct)	Unknot, $T(2, 3)$, $C(3, -2)$, $C(5, -2)$, $C(7, -2)$, $C(9, -2)$, $C(11, -2)$	1, 2	10
Xer	Unknot (direct)	$T(2, 4)$	1	13
Gin	Unknot (inverted)	Unknot → $T(2, 3)$ → $C(2, -2)$	1, 2	7
	$T(2, 3)$ (inverted)	$T(2, 3)\#T(2, 3)$	4	7
	Unknot (direct)	$T(2, 3)$ → $C(3, -2)$ → $C(5, -2)$	1, 2	7
	$T(2, 3)$ (direct)	$T(2, 3)\#T(2, 3)$	4	7
Gin mutant	Unknot (inverted)	$T(2, 3)$, $C(2, -2)$, $T(2, 5)$, $C(3, -2)$, $T(2, 7)$, $C(5, -2)$, $T(2, 9)$, $C(7, -2)$	1, 2	8
Hin	Unknot (inverted)	Unknot → $T(2, 3)$ → $C(2, -2)$ → $C(3, -2)$	1, 2	6
Hin mutant	Unknot (inverted)	$T(2, 3)$, $C(3, -2)$	1, 2	5
	$T(2, 3)$	$T(2, 3)\#T(2, 3)$	4	5
Tn3, γδ	Unknot (direct)	$T(2, 2)$ → $C(2, -2)$ → $F(2, -2, -2, 0)$ → $T(2, 6)$	1, 2, $F(p, q, r, s)$	4

Consistent with our model, these products all fall within the single family of Fig. 5. Two crossover sites on a single molecule of circular DNA can be in either *direct* orientation (head to tail; e.g., …ATGC…ATGC) or *inverted* orientation (head to head; e.g., …ATGC…CGTA). A substrate with direct sites usually yields different products from the same substrate with inverted sites, as shown above. PB and LR are two different pairs of crossover sites recognized by λ Int. In the Product(s) column, commas between knot and catenane types indicate a spectrum of simultaneous products, while "→" indicates processive recombination. For example, Gin acting on an unknot with sites in inverted repeats yields an unknot in the first round of processive recombination, $T(2, 3)$ in the second round, and $C(2, -2)$ in the third round.

recombinase complex, *B*: the smallest convex region containing the four bound recombinase molecules and the two crossover sites.[45] Hence, *B* is a *topological ball* (i.e., it can be deformed to a round ball). The crossover sites can be located either on the outside, separated by the catalytic domains (e.g., with γδ and Tn3 resolvase), or inside the four recombinase subunits.[6,46–52] The *recombinase–DNA complex* is *B* together with the substrate. If the recombinase complex *B* meets the substrate in precisely the two crossover sites, it is a *productive synapse*.

The existence of a productive synapse for recombinases is in contrast with some transposases such as Tn5 and Tn10, whose enhancer sequences are intertwined with the active transposition sites,[53] preventing the existence of a productive synapse. Figure 1 demonstrates two examples where the recombinase complex *B* is a productive synapse and one where *B* is not.

Site-specific recombinases fall into two families—the serine (also known as the resolvase) and tyrosine (also known as the integrase) recombinases—based on sequence homology and catalytic residues.[27] The serine and tyrosine recombinases also differ in their mechanism of cutting and rejoining DNA at the crossover sites. Both families are large: A phylogenetic analysis has been performed on 72 serine recombinases[54] and a recent iterative Position-Specific Iterated BLAST search documents approximately 1000 related sequences of putative tyrosine recombinases.[55]

The diverse family of serine recombinases is composed of resolvases (such as Tn3 and γδ), invertases (such as Gin, Hin, Pin, and Min), large serine recombinases (also called large resolvases), and insertion sequence elements.[54] These recombinases may trap a fixed number of supercoils before initiating recombination.[27] For example, Tn3 resolvase requires three negative supercoils to be trapped by the binding of (nonactive) resolvase molecules. These trapped supercoils (outside of the recombinase complex), together with the recombinase complex itself, are known as the *synaptic complex*.[4,45] Likewise, the invertases also require a fixed number of supercoils trapped outside the recombinase complex. Rather than using additional recombinase molecules, they rely on accessory proteins and enhancer sequences, which facilitate the organization of a unique stereospecific synapse that promotes DNA cleavage. (In the Hin and Gin systems, these bound supercoils, together with the recombinase complex, are referred to as the *invertasome*.)[5,6] With serine recombinases, recombination proceeds through a concerted four-strand cleaving and rejoining reaction.[27] Serine recombinases can perform processive recombination (discussed in more detail after Assumption 3).

In contrast, tyrosine recombinases first cleave, exchange, and reseal two sugar-phosphate backbones. The DNA–protein complex thus proceeds through an intermediary structure (a Holliday junction) before repeating the process with the other two DNA backbones.[27,56] Most tyrosine recombinases, including Flp, λ Int, and Cre, tolerate varying numbers of supercoils outside of the recombinase complex. However, there are exceptions, most notably XerCD, which trap a fixed number of supercoils using accessory proteins before initiating cleavage.[57] Like serine recombinases, tyrosine recombinases can also employ accessory proteins to help assemble the synaptic complex and to drive the overall reactions (e.g., XerCD and λ Int).[57,58]

Finally, we define several knot theoretic terms. Figure 2 presents examples and diagrams of these terms. In particular, Fig. 2 illustrates the convention of crossings of the following forms: +1, −1, 0, +2, and −2 vertical crossings and +2 and −2 horizontal crossings. It also shows typical conformations of the substrates we are considering: an unknot, an unlink, and a torus knot or catenane, $T(2, m)$, described below. The *components* of a catenane are the separate rings of the catenane. (Hence, a knot is considered to be a catenane with only one component.) A $T(2, m)$ torus knot or catenane is one that can be drawn so that all of its crossings occur as a row of m (positive or negative plectonemic) twists, as illustrated in Fig. 2. We will denote a knot or catenane of this form by $T(2, m)$. Note that if m is odd, then $T(2, m)$ is a knot, and if m is even, then $T(2, m)$ is a catenane. Finally, given two knots or catenanes K and J, their *composite knot or catenane*, written $K\#J$, is obtained by removing an unknotted arc from each and gluing the resulting two endpoints of K to the two endpoints of J without introducing any additional knotting.

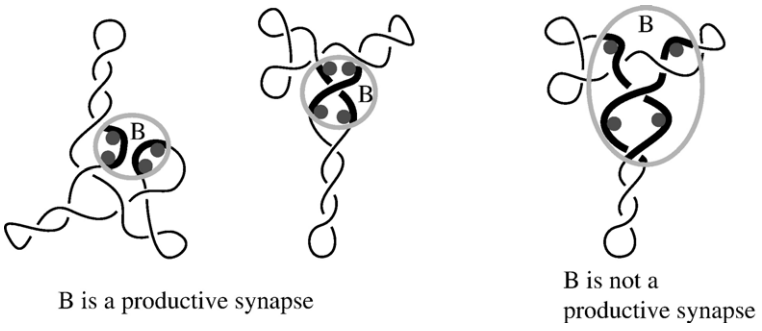

Fig. 1. Productive synapse. The smallest ball containing the four bound recombinase molecules and the two crossover sites is *B*, the recombinase complex. The crossover sites are highlighted in black; each recombinase monomer corresponds to a gray disc, and *B* is represented by a light gray circle. We assume that *B* is actually a *productive synapse*: that *B* contains only the two crossover sites and no other DNA. Left and middle panels: *B* is a productive synapse. Right panel: *B* is not a productive synapse since there is no way to draw *B* so that only the two crossover sites are inside *B* without also including the third (horizontal) strand.

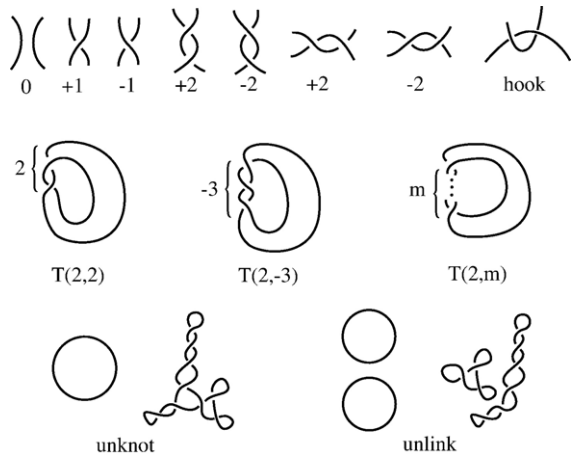

Fig. 2. Substrates and Assumption 1. The model considers substrates that are unknots, unlinks, or $T(2, m)$ torus knots or catenanes. Top row: Crossing sign conventions: $+1, -1, 0, +2$, and -2 vertical; $+2$ and -2 horizontal; and a hooked junction. Assumption 1 states that there is a projection of the productive synapse with at most one crossing. Hence, the productive synapse must have a projection that locally looks like one of the first three diagrams. Note that Assumption 1 does allow productive synapses like the hook junction, where there is a projection with one crossing but no projections with zero crossings. Middle row: $T(2, m)$ torus knots and catenanes, formed by closing a row of m plectonemic twists. Left panel: The catenane $T(2, 2)$. Center panel: The knot $T(2, -3)$, a trefoil knot. Right panel: The general form of $T(2, m)$. Note that $T(2, m)$ is a catenane if m is even and is a knot if m is odd. Bottom row: Possible conformations of the unknot and unlink.

Figure 5, subfamilies 4 and 5, gives examples of composite knots and catenanes (see Ref. 59 for a more thorough treatment, and beautiful diagrams, of knots and catenanes).

The Assumptions of Our Model

We begin with a fixed recombinase and an unknot, unlink, or torus knot or catenane $T(2, m)$ substrate. (If the substrate is an unlink, then we assume that one site is on each component, as otherwise this case reverts back to a single unknotted substrate.) Here, we state three assumptions about the recombinase–DNA complex and provide experimental evidence for each.

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

Figure 1 presents examples of recombinase complexes that are or are not productive synapses.

Note that we allow the possibility of one crossing between sites. For example, the hooked junction in Fig. 1 has projections where there is only one crossing but no projections where there are no crossings. Two protein-induced local DNA bends, such as those induced by Flp and Cre,[60] could create such a hooked junction, if the bends were toward the center of the productive synapse, as illustrated in Fig. 1. There are also many site-specific recombinases whose productive synapse is uncharacterized, and these conformations could be hooked. Finally, the products of recombination, according to our model, are not more complicated for a productive synapse with one crossing than for a productive synapse with zero crossings (see Ref. 41 for details). Thus, we do not make the stronger assumption of zero crossings, as it has no effect on the predictions of the products.

Evidence for Assumption 1. Below, we present a variety of experimental data that suggest that the recombinase complex is sufficiently dense both to form a proper productive synapse and to preclude extraneous crossings.

Most convincingly, recent crystal structures of several recombinase complexes support both of these assumptions.[49–52,61] Additionally, structures of a single site synapsed with either a dimer or a monomer indicate that there are no crossings within an individual site.[62–64] Furthermore, structures of two intermediate complexes—a synaptic γδ resolvase tetramer covalently linked to two *cleaved* DNA segments and the Flp recombinase–Holliday junction complex—have at most one crossing between sites and none within a site.[48,51] Thus, the large-scale conformational changes necessary to unwrap crossings during the reaction imply that it is unlikely that the crossover sites contain additional crossings at synapsis or that a productive synapse does not exist.

Also, there are significant DNA–protein and protein–protein interactions that appear to prevent additional crossings and extraneous strands from piercing the recombinase complex. With tyrosine recombinases, each domain flanking the crossover site DNA inserts a helix into a major groove, and the highly conserved C-terminal domain interacts with consecutive minor and major grooves on the opposite face of the DNA.[27] With serine recombinases, DNA binding involves the conserved H-T-H domain and a DNA-binding domain on the C or N terminus of the protein.[48,54,64] For both families, there are also significant protein–protein interfaces with the other proteins in an assembled tetrameric complex. Additionally, DNA itself has a geometric diameter of 2 nm and, depending on the ionic conditions, a much greater electrostatic diameter (e.g., $\simeq$5 nm at physiological conditions).[65]

Additional biochemical experiments support the existence of both a productive synapse and a bound on the number of crossings between or within the sites. Atomic force microscopy of both the Cre and Flp productive synapses concurs with the conclusions drawn from the crystal structures of the resolvases and integrases.[35] Also, the architecture of the γδ resolvase recombinase complex has been

determined to be a productive synapse with a single crossing, in experiments using constrained DNA.[46] Furthermore, solution structures from neutron and X-ray scattering data of hyperactive Tn3 resolvase mutants show that a productive synapse exists and that there is a projection of the sites with at most one crossing between sites and no crossings within a single site.[47] Additionally, recent cyclization experiments indicate that dimers of Flp and Cre each bend the DNA sites upon binding but not enough to introduce a crossing within a single site.[60] The steric and electrostatic constraints mentioned above imposed by the short length of the sites also putatively limit crossings between and/or within the sites.

Finally, for recombinases that utilize accessory proteins, we recall that an accessory site or enhancer sequence is neither a crossover site nor a part of a crossover site. Thus, in order for our assumption to hold, if a recombinase requires an enhancer sequence, then it must be sequestered from the crossover sites. In particular, we claim that when the enhancer or an accessory site loops around to form a specific recombinase complex, all crossings are trapped outside of the complex, although the recombinases might interact directly with the enhancer sequence. Supporting evidence is twofold. Firstly, the recent λ Int–DNA complex crystal structure includes the accessory sites, and it is clear that a productive synapse exists and has the required limited number of crossings.[52] Additionally, support for the invertase family comes from detailed biochemical experiments of the Hin system. The standard molecular model of the Hin invertasome, based on the cross-linked structure of the Hin–DNA co-complex, has two Fis dimers bound to the enhancer sequence and two Hin dimers bound at the recombination sites.[6] According to this model, the enhancer sequence is sequestered from the crossover sites, and the crossover sites are not interwound.

All of the above evidence indicates that it is reasonable to assume that a given recombinase–DNA complex satisfies Assumption 1.

Assumption 2. 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 catenanes are trapped in the DNA branches outside of the productive synapse.

Figure 3 demonstrates several scenarios that are allowed or forbidden in Assumption 2.

Evidence for Assumption 2. A variety of microscopy studies support Assumption 2. Atomic force microscopy revealed that at physiological conditions, supercoiled DNA adopts a compact plectonemic configuration with close juxtaposition of DNA segments in the loops, which makes it unlikely that a supercoiled domain could be penetrated.[44,66] Under conditions that minimize intersegmental repulsion, both electron and scanning force microscopy studies demonstrate that opposing segments of interwound supercoiled DNA are frequently close together.[28,43,67] Probabilistically then, it is thus unlikely that either a supercoiled domain or *B* could pierce through a supercoil.

Also, experimental work coupled with Metropolis Monte Carlo simulations indicate that, on average, supercoiled DNA helices are separated by

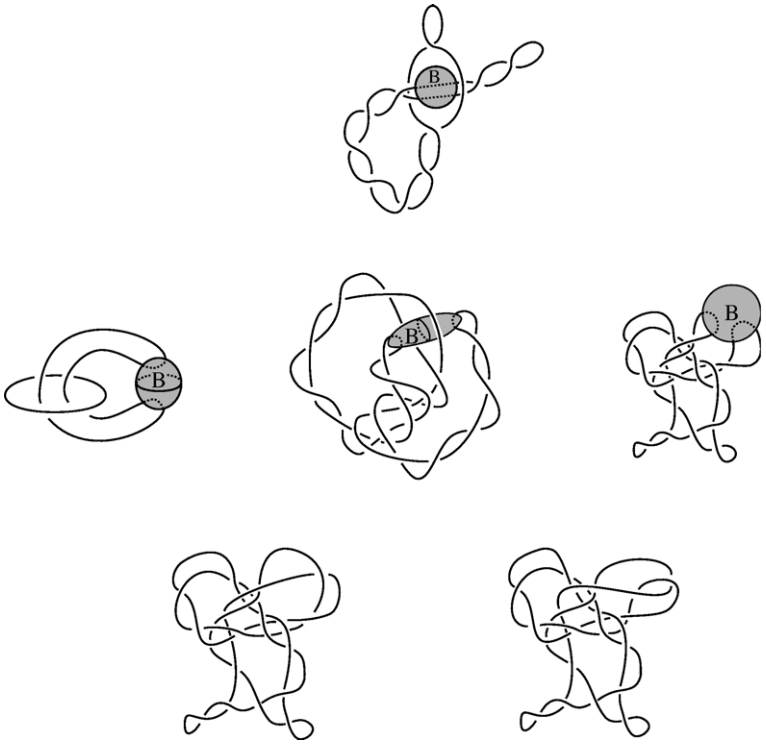

Fig. 3. Scenarios for Assumption 2. Top panel: The productive synapse *B* trivially pierces through a supercoil. This is allowed. Middle row, left panel: A catenane is trapped in the DNA branches outside of *B*. Middle row, center panel: *B* pierces through a supercoil in a nontrivial way. Middle row, right panel: A knot is trapped in the branches on the outside of *B*. These three scenarios are forbidden. Bottom row: If knots were trapped within the branches of the substrate(s) on the outside of the productive synapse, as in the middle row, then recombination would at least occasionally produce "doubly knotted" knots and catenanes such as these. There are no known products of this type.

10 nm—only five times the width of the DNA diameter itself.[28,65] Since, as discussed above, the *electrostatic* diameter of DNA in physiological conditions is closer to 5 nm, it seems quite unlikely that the productive synapse could pass through a supercoil (as in Fig. 3).[68–70] Brownian dynamics simulations of site juxtaposition support these findings.[71]

Additionally, the probability of one duplex (linear) invading a supercoiled domain has been shown to be quite low, by both experiments and Metropolis Monte Carlo simulations.[68] This frequency may be even lower if a supercoiled domain, rather than a linear duplex, invades another supercoiled domain. Therefore, trapping a persistent knot or catenane within the branches of the DNA during synapsis is also unlikely.

Also, the steric and electrostatic constraints arising from protein–DNA interactions discussed in the Evidence for Assumption 1 would appear to preclude piercing of the productive synapse by nonsite DNA.

Finally, if persistent knots or catenanes could be trapped within the branches of the substrate(s) on the outside of the productive synapse, then we would expect to see (at least occasionally) "doubly knotted" products like those illustrated in Fig. 3 (such knots and catenanes are known as *satellites*). However, no products like these have thus far been observed (see Table 1). This indicates that knotting or catenating of the branches is unlikely to occur.

All of the above evidence indicates that it is reasonable to assume that a given recombinase–DNA complex satisfies Assumption 2.

Assumption 3 for Serine Recombinases:

Serine recombinase performs 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.

Figure 4 illustrates this assumption.

Recall that with 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 an initial substrate, with several intermediate exiting points for the reaction (again, see Fig. 4).

Assumption 3 for Tyrosine Recombinases:

After recombination mediated by a tyrosine recombinase, there is a projection of the crossover sites that has at most one crossing.

Note that in the post-recombinant synapse, as with the productive synapse in Assumption 1, we are allowing hook junctions, as illustrated in Fig. 1. We allow these for the same reasons discussed above.

Evidence for Assumptions 3. *Serine recombinases*: A large number of *in vitro* topology studies performed on DNA invertases and resolvases have provided

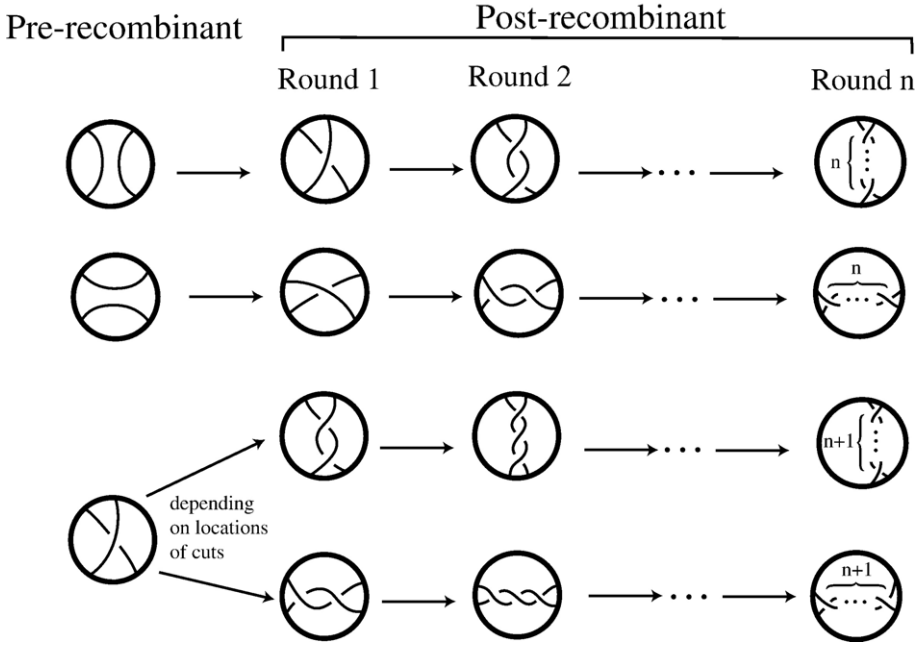

Fig. 4. Assumption 3 for Serine Recombinases. The productive synapse in processive recombination. On the left, a projection of the pre-recombinant productive synapse with zero or one crossing. On the right, the post-recombinant synapse at each round of processive recombination.

solid support for the "subunit exchange mechanism," where one set of recombinase subunits, each covalently associated with the 5′ ends of the cleaved recombination sites, switches places, resulting in a 180° rotation of DNA strands (see Ref. 27 and references therein). This is supported by a recent crystal structure of a synaptic tetramer of γδ resolvase covalently catenated to two cleaved DNA molecules, indicating a subunit rotation of 180°.[48]

Additional experiments involving Tn3, Hin, and Gin lend credence to the idea that during each round of processive recombination, the same set of subunits is exchanged and the sense of rotation remains constant.[2,5,7] For example, Heichman *et al.* demonstrate that there are multiple rounds of exclusively clockwise subunit rotation of one set of Hin subunits after DNA cleavage.[5]

Tyrosine recombinases: While there are no known post-recombinant crystal structures, there are synaptic intermediary crystal co-complexes for Flp,[51] Cre,[50] and λ Int[52] (this also includes accessory sites in addition to the typical λ Int crossover sites). These structures indicate that at the earlier stages of recombination—namely, after the first cleavage, exchange, or within a Holliday junction intermediate—there exists a projection with at most one crossing. They also highlight particular features of the productive synapse that may impede the large-scale conformational changes needed to introduce crossings.

As mentioned above, the protein–DNA interface is a large hydrogen-bonded network. Flp, Cre, and λ Int all form a C-shaped clamp around the DNA substrate, and the C-terminal domains interact with consecutive minor and major grooves on the opposite face of the DNA.[27] Additionally, there are significant protein–protein interactions; for example, the catalytic domains interact by swapping part of the C terminus with a neighboring protomer.

Also, the post-recombinant complex is formed from the Holliday junction intermediate by, first, an isomerization of the intermediary complex so that the inactive monomers become active and *vice versa* and, then, a repeated strand cleavage where the new 5′ ends migrate over and attack their partners' 3′ phosphotyrosine linkages. This second round of strand transfer completes the reaction. Particularly given the 2-fold symmetry of the reaction, it thus seems unlikely that, in the final stage of recombination, there is enough motion of the DNA arms to generate multiple additional crossings between sites or a crossing within a single site.

In vitro studies also suggest that tyrosine recombinases that mobilize the gene cassettes of integrons may preferentially bind DNA hairpins, which would constrain the number of crossings.[72] Finally, given the steric and electrostatic constraints of short DNA arms discussed for Assumption 1, it is probable that there exists a projection of the sites containing at most one crossing between sites and no crossings of a single site within the post-recombinant complex.

All of the above evidence indicates that it is reasonable to assume that a given recombinase–DNA complex satisfies Assumption 3.

Results

All products of unknots, unlinks, or *T*(2, *m*) torus knot and catenane substrates fall within a single family

Here, we state two theorems, given the three assumptions above, stating that all products of unknots, unlinks, and $T(2, m)$ torus knots and catenanes fall within the family of knots and catenanes $F(p, q, r, s)$ illustrated in Fig. 5. The technical proofs of these results can be found in Ref. 41.

Members of the family $F(p,q,r,s)$ can be characterized in terms of four variables: p, q, r, and s, which describe the number of crossings between two strands in the knot or catenane (see Fig. 5). Note that these crossings, p, q, r, and s, can each be positive, negative, or zero (see Fig. 2 for crossing conventions). Also, by letting p, q, r, and/or s be 0 or 1 as appropriate, we obtain five subfamilies, as illustrated in Fig. 5.

The first subfamily is the familiar torus knots and catenanes, $T(2, p+q)$. The second subfamily is the *clasp* knots and catenanes, $C(r, s)$: those consisting of one row of r crossings and a nonadjacent row of s crossings. [By *nonadjacent* rows of r and s crossings, we mean that the two rows cannot be considered as a single row of $r+s$ crossings, as they can in the torus knot and catenane case, $T(2, p+q)$.] If r or s equals 2, then $C(r, s)$ is also commonly referred to as a *twist* knot or catenane (see the example for subfamily 2 drawn in Fig. 5). The third subfamily is the family of *pretzel* knots and catenanes, $K(p, q, r)$, with three nonadjacent rows containing p crossings, q crossings, and r crossings (illustrated in subfamily 3, where $r=s\pm1$). The fourth and fifth subfamilies are both *composite knots or catenanes*: *K#J*: formed by removing an unknotted arc from each of two knots, *K* and *J*, and then gluing the two knots together at their endpoints. The fourth subfamily comprises composite knots and catenanes, $T(2, p)\#T(2, s\pm1)$, formed from two torus knots. The fifth subfamily comprises composite knots and catenanes, $T(2,p)\#C(r, s)$, formed from a torus knot and clasp knot. The general form of each of these five subfamilies, as well as an example of each, is illustrated in Fig. 5.

Theorem 1. Tyrosine recombinases

Suppose that Assumptions 1, 2, *and* 3 *hold for a particular tyrosine recombinase–DNA complex. Then, recombination on an unknot yields an unknot, unlink, T(2,n), or C(p, 2). Recombination on an unlink yields an unknot, unlink, or T(2, 2). Recombination on T(2, m) yields an unknot, unlink, or F(p, q, r, s), with* $|r| \leq 2$. *For each of these three substrate types, any product whose knot or catenane type is not listed above must arise from distributive recombination.*

Theorem 2. Serine recombinases

Suppose that Assumptions 1, 2, *and* 3 *hold for a particular serine recombinase–DNA complex. Then,*

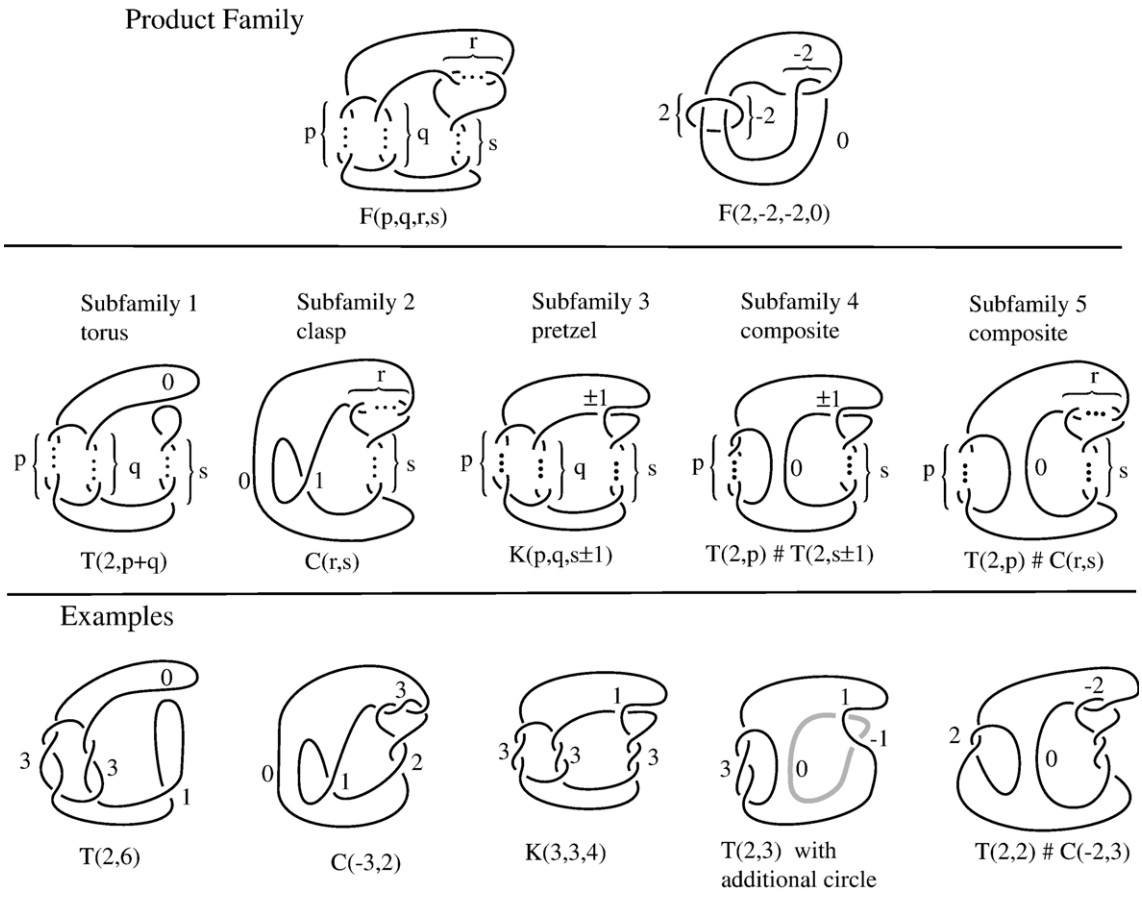

Fig. 5. Product family. Top row: The model predicts that all products of recombination on an unknot, unlink, or $T(2, n)$ torus knot or catenane substrate fall in this family, $F(p, q, r, s)$; and an example, $F(2, -2, -2, 0)$. Middle row: The five subfamilies, obtained by setting p, q, r, and/or s equal to 0 or ±1. From left to right, the torus knots and catenanes, $T(2,p+q)$; the clasp knots and catenanes, $C(r, s)$; the pretzel knots and catenanes, $K(p, q, s\pm1)$; the composite knot or catenane of two torus knots, $T(2, p)\#T(2, s\pm1)$; and the composite knot or catenane of a torus and clasp knot, $T(2, p)\#C(r, s)$. Bottom row: Examples of each of the subfamilies mentioned earlier. Note that the example for subfamily 2 is a member of the well-known family of twist knots, $C(r, 2)$, and that the example for subfamily 4 has an uncatenated additional circle.

recombination on an unknot yields an unknot, unlink, T(2, n), or C(p, q). Recombination on an unlink yields an unknot or T(2, n). Recombination on T(2, m) yields an unknot, unlink, or F(p, q, r, s). For each of these three substrate types, any product whose knot or catenane type is not listed above must arise from distributive recombination.

The statements of Theorems 1 and 2 are illustrated in Fig. 6.

Thus, recombination on an unknotted, unlinked, or torus knot or catenane $T(2, n)$ substrate can give rise to only very specific types of products, all of which are members of the family of Fig. 5, $F(p, q, r, s)$. Any other types of products must be from distributive recombination.

Products are even more restricted when recombination adds one crossing

Knots and catenanes are tabulated according to the fewest number of crossings with which they can be drawn.[20,59] This number of crossings is called the MCN of the knot or catenane, K, and is denoted by MCN(K). For example, MCN(unknot) = 0 and MCN(unlink) = 0. MCN($T(2, 2)$) = 2, and MCN($T(2, 3)$) = 3. Figure 7 presents two drawings of $T(2, 3)$: one with three crossings and one with nine crossings, the latter of which can be deformed to a drawing with only three crossings by unwinding the supercoils. It can be shown that there is no drawing of $T(2, 3)$ with fewer than three crossings; thus, MCN($T(2, 3)$) = 3. Similarly, MCN($T(2, 2)$) = 2, and, in fact, MCN($T(2, m)$) = m for any positive integer m. As discussed in the Introduction, DNA knots and catenanes of equal length migrate proportionally to their MCN. Thus, gel electrophoresis can be used to determine the MCN of an unknown knot or catenane product.[19]

Often, recombination increases the MCN of a knotted or catenated substrate by 1 (e.g., Ref. 73). If the substrate is $T(2, m)$ and the product has MCN = $m + 1$, then we can further refine the results of Theorems 1 and 2 as follows to determine more specific possibilities for the products.

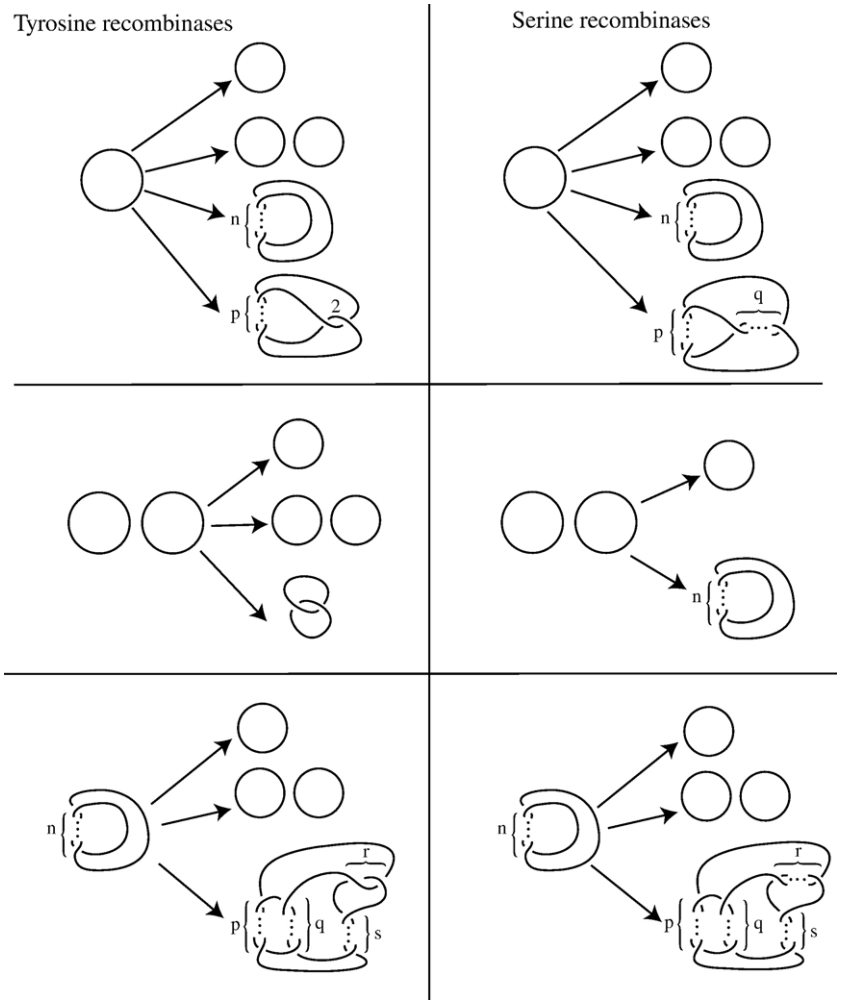

Fig. 6. Theorems 1 and 2. Left column: The products of recombination with a tyrosine recombinase (Theorem 1). Top panel: An unknotted substrate yields an unknot, unlink, $T(2, n)$ torus knot or catenane, or $C(p, 2)$ clasp knot or catenane (i.e., a twist knot or catenane). Middle panel: An unlinked substrate yields an unknot, unlink, or the catenane $T(2, 2)$. Bottom panel: A $T(2, n)$ substrate yields an unknot, unlink, or any knot or catenane from the family $F(p, q, r, s)$ of Fig. 5 that has $|r| \leq 2$. Right column: The products of recombination with a serine recombinase (Theorem 2). Top panel: An unknotted substrate yields an unknot, unlink, $T(2, n)$, or clasp knot $C(p, q)$. Middle panel: An unlinked substrate yields an unknot or $T(2, n)$. Bottom panel: A $T(2, n)$ substrate yields an unknot, unlink, or any knot or catenane from the family $F(p, q, r, s)$ of Fig. 5.

Theorem 3.

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

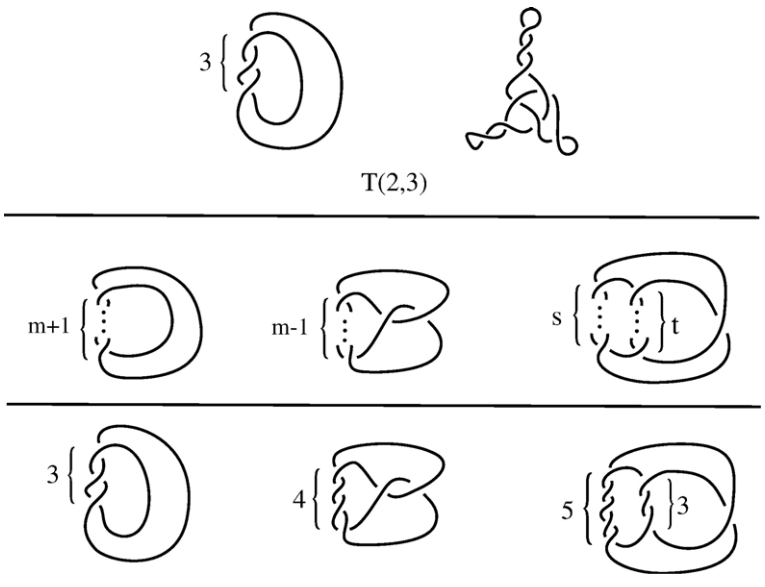

Fig. 7. Adding +1 to MCN greatly restricts products. MCN(K) is defined to be the minimal number of crossings, looking at all drawings of a knot or catenane, K. Top row: Two drawings of $T(2, 3)$, with three crossings (left panel) and with nine crossings (right panel). The right drawing can be deformed to the left, but the left cannot be simplified; thus, $MCN(T(2, 3))=3$. In general, $MCN(T(2, m))=m$. Middle row: Theorem 3 states that, given a $T(2,m)$ substrate, the only products possible with $MCN=m+1$ are the three illustrated here: $T(2, m+1)$, the clasp knot or catenane $C(m-1, -2)$, and the pretzel knot or catenane $K(s, t, 1)$ with $s+t=m$. Bottom row: Examples of each of the three types of possible products.

Theorem 3 is illustrated in Fig. 7.

Thus, there exist only three very special types of knots and catenanes that can be obtained when recombination adds 1 to the MCN, given a $T(2, m)$ substrate. Also, observe that $K(p, q, 1)$ is a knot if and only if at least one of p and q is odd. The proof of this theorem is given in Ref. 41 We discuss applications of this theorem in Applications to Uncharacterized Recombinant Products.

Discussion and Applications

All characterized recombinant products are in the predicted family

Table 1 summarizes the known products of recombinases starting with substrates, which are unknots, unlinks, or $T(2, m)$. As shown in Table 1, all products listed are members of the family $F(p,q,r,s)$ of Fig. 5. This provides further confirmation of the validity of our model.

Note that this table does not describe every product of site-specific recombination. For example, mutant Hin acting on a heterogenous population of $T(2, 3)$ and twist knots $C(3, -2)$ yields (double and, in multiple rounds of recombination, triple) composite knots and catenanes: $T(2, 3)\#T(2, 3)$, $T(2, 3)\#C(3, -2)$, $C(3, -2)\#C(3, -2)$, and $T(2, 3)\#T(2, 3)\#C(3, -2)$.[6] However, in our model, twist knots and two separate, uncatenated knots are not allowable as substrates. Hence, our model is not intended to predict products of these substrates.

Applications to uncharacterized recombinant products

The model presented here can aid experimentalists in several ways. First, as mentioned above, this model determines products that must arise from distributive recombination. For a given unknot, unlink, or $T(2, n)$ torus knot or catenane substrate, any product not explicitly listed in Theorems 1 and 2 (illustrated in Fig. 4) must be a result of distributive recombination. To illustrate, suppose recombination on an unlinked substrate with a tyrosine recombinase yields the unknot, a $T(2, 2)$ torus catenane, a $T(2, m)$ torus knot or catenane (for a fixed $m \neq 2$), and a $C(r, -2)$ clasp knot or catenane (for a fixed r) and $F(3, 4, 2, 1)$. Then, Theorem 1 states that the last three products listed must all arise from distributive recombination. Below, we will analyze experimental data for Tn3 resolvase and show how the model can help distinguish products of distributive recombination.

Second, our model can be helpful in understanding processive recombination mediated by a serine recombinase. Theorems 1 and 2 (illustrated in Fig. 6) 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 torus knot. Then, Fig. 6 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).

Finally, we now turn our attention to several experiments whose recombination products and/or type (processive or distributive) are unknown. For each, we will discuss how our model can help to restrict the knot types of these products.

Tn3

Benjamin *et al.* constructed a plasmid substrate for Tn3 resolvase with four directly repeated crossover sites.[74] Electron microscopy revealed $T(2, 2)$ as the primary product, and high-resolution gel electrophoresis of 7–8 days followed by electron microscopy revealed several additional products: $T(2,2)\#T(2,2)$ and two distinct four-component catenanes.

Theorem 2 predicts that, with this four-sited substrate, recombination must proceed from the unknot to $T(2, 2)$. It then utilizes this $T(2, 2)$ catenane as a substrate to yield the product of the composite catenanes, $T(2, 2)\#T(2, 2)$ (see Fig. 6). This composite catenane is then the substrate for the products of four-component catenanes. However, this composite is not one of the substrates that we consider. This would be akin to Tn3 performing multiple rounds of distributive recombination on a substrate with only two crossover sites. The current work thus supports Benjamin *et al.*'s hypothesis of neighboring-site recombination.

Xer

Creating a hybrid plasmid with both λ Int and Xer sites, Bath *et al.* generated the catenanes $T(2, 6)$ and $T(2, 8)$ as products of λ recombination.[73] These were then used 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 MCN. There are 7 knots with MCN = 7 and 49 knots with MCN = 9.

Theorem 3 significantly reduces the number of possibilities for each of these products. In particular, it follows from Theorem 3 that the seven-crossing products of Xer must be one of $T(2, 7)$, $C(5, -2)$, and $K(3, 3, 1)$ and that the nine-crossing products of Xer must be one of $T(2, 9)$, $C(7, -2)$, and $K(5, 3, 1)$. All of these knots are a special type, termed *4-plats*. This demonstrates how our model complements earlier work of Darcy, which assumes (rather than proves) that all products must be 4-plats and, hence, only considers seven-crossing products (since half of the nine-crossing knots are not 4-plats).[34] In Ref. 75, we use our model together with tangle calculus to completely classify all tangle solutions to these λ Int–Xer equations.

Cre111

Abremski *et al.* created the mutant Cre111, which yields products topologically distinct from those of

wild-type Cre.[76] Under their conditions, when Cre111 recombines an unknotted substrate, the knotted and catenated products are significantly more complex than those produced by wild-type Cre.

These knots and catenanes have thus far been uncharacterized. However, Theorem 1 predicts that these knots and catenanes must be of the form $T(2, n)$ or $C(p, -2)$. Thus, by running these products adjacent to a ladder of torus knots of the same length, experimentalists could determine the exact knot or catenane type.

Concluding Remarks

Here, we have developed a model of how DNA knots and catenanes are produced as a result of a recombinase acting on an unknot, unlink, or $T(2, n)$ torus knot or catenane substrate. Our model is based on three assumptions about site-specific recombination, and we have provided experimental evidence for each. Our model predicts that all knotted or catenated products of such enzyme actions will be in the family of Fig. 5, as described in Theorems 1 and 2 and illustrated in Fig. 6.

As mentioned above, the MCN of a DNA knot or catenane can be determined experimentally.[19] For small values of the MCN, there are not many knots or catenanes with a given value. However, the number of knots and catenanes with MCN = n grows exponentially as a function of n,[77] and there are 1,701,936 knots with MCN ≤ 16.[20] Thus, knowing the MCN is not sufficient to determine the knot or catenane.

However, we proved in Ref. 41 that the total number of knots and catenanes in the family $F(p, q, r, s)$ of Fig. 5 grows linearly with n^3. Hence, the proportion of all knots and catenanes that are contained in our family decreases exponentially as n increases. Thus, knowing the MCN of a product and knowing that the product is in one of our families allow us to significantly narrow the possibilities for its knot or catenane type. The model described herein thus provides an important step in characterizing DNA knots and catenanes, which arise as products of site-specific recombination.

Acknowledgements

We wish to thank J. Arsuaga, I. Grainge, R. Harshey, M. Jayaram, K. Millett, A. Stasiak, D.W. Sumners, A. Vologodskii, and S. Whittington for helpful conversations. We also thank the referees for their careful reading and useful comments. D.B. was partially supported by National Science Foundation grant no. DMS-0102057. E.F. was partially supported by an Association for Women in Mathematics Michler Collaborative Research Grant.

References

1. Maxwell, A. & Bates, A. D. (2005). *DNA Topology.* Oxford University Press, Oxford, England.
2. Cozzarelli, N. R., Krasnow, M. A., Gerrard, S. P. & White, J. H. (1984). A topological treatment of recombination and topoisomerases. *Cold Spring Harbor Symp. Quant. Biol.* **49**, 383–400.
3. Pathania, S., Jayaram, M. & Harshey, R. M. (2002). Path of DNA within the Mu transpososome. Transposase interactions bridging two Mu ends and the enhancer trap five DNA supercoils. *Cell*, **109**, 425–436.
4. Wasserman, S. A., Dungan, J. M. & Cozzarelli, N. R. (1985). Discovery of a predicted DNA knot substantiates a model for site-specific recombination. *Science*, **229**, 171–174.
5. Heichman, K. A., Moskowitz, I. P. & Johnson, R. C. (1991). Configuration of DNA strands and mechanism of strand exchange in the Hin invertasome as revealed by analysis of recombinant knots. *Genes Dev.* **5**, 1622–1634.
6. Merickel, S. K. & Johnson, R. C. (2004). Topological analysis of Hin-catalysed DNA recombination *in vivo* and *in vitro*. *Mol. Microbiol.* **51**, 1143–1154.
7. Kanaar, R., Klippel, A., Shekhtman, E., Dungan, J. M., Kahmann, R. & Cozzarelli, N. R. (1990). Processive recombination by the phage Mu Gin system: implications for the mechanisms of DNA strand exchange, DNA site alignment, and enhancer action. *Cell*, **62**, 353–366.
8. Crisona, N. J., Kanaar, R., Gonzalez, T. N., Zechiedrich, E. L., Klippel, A. & Cozzarelli, N. R. (1994). Processive recombination by wild-type Gin and an enhancer-independent mutant. Insight into the mechanisms of recombination selectivity and strand exchange. *J. Mol. Biol.* **243**, 437–457.
9. Hoess, R. H., Wierzbicki, A. & Abremski, K. (1986). The role of the loxP spacer region in P1 site-specific recombination. *Nucleic Acids Res.* **14**, 2287–2300.
10. Crisona, N. J., Weinberg, R. L., Peter, B. J., Sumners, D. W. & Cozzarelli, N. R. (1999). The topological mechanism of phage lambda integrase. *J. Mol. Biol.* **289**, 747–775.
11. Cox, M. M. (1989). DNA inversion in the 2 μm plasmid of *Saccharomyces cerevisiae*. In *Mobile DNA* (Berg, D. E. & Howe, M. M., eds), pp. 661–670, American Society for Microbiology, Washington, DC.
12. Spengler, S. J., Stasiak, A. & Cozzarelli, N. R. (1985). The stereostructure of knots and catenanes produced by phage λ integrative recombination: implications for mechanism and DNA structure. *Cell*, **42**, 325–334.
13. Colloms, S. D., Bath, J. & Sherratt, D. J. (1997). Topological selectivity in Xer site-specific recombination. *Cell*, **88**, 855–864.
14. Krasnow, M. A., Stasiak, A., Spengler, S. J., Dean, F., Koller, T. & Cozzarelli, N. R. (1983). Determination of the absolute handedness of knots and catenanes of DNA. *Nature*, **304**, 559–560.
15. Trigueros, S., Arsuaga, J., Vazquez, M. E., Sumners, D. W. & Roca, J. (2001). Novel display of knotted DNA molecules by two-dimensional gel electrophoresis. *Nucleic Acids Res.* **29**, E67.
16. Zechiedrich, E. L. & Crisona, N. J. (1989). Coating DNA with RecA protein to distinguish DNA path by electron microscopy. In *Methods in Molecular Biology: DNA Topoisomerase Protocols* (Bjornsti, M. &

Osheroff, N., eds), vol. 1, pp. 98–108, Humana Press, Totowa, NJ.
17. Katritch, V., Bednar, J., Michoud, D., Dubochet, J. & Stasiak, A. (1996). Electrophoretic mobility of DNA knots. *Nature*, **384**, 122.
18. Sundin, O. & Varshavsky, A. (1980). Terminal stages of SV40 DNA replication proceed via multiply intertwined catenated dimers. *Cell*, **21**, 103–114.
19. Levene, S. D. & Tsen, H. (1999). Analysis of DNA knots and catenanes by agarose-gel electrophoresis. *Methods Mol. Biol.* **94**, 75–85.
20. Hoste, J., Thistlethwaite, M. & Weeks, J. (1998). The first 1,701,936 knots. *Math. Intell.* **20**, 33–48.
21. Vologodskii, A. V., Crisona, N. J., Laurie, B., Pieranski, P., Katritch, V., Dubochet, J. & Stasiak, A. (1998). Sedimentation and electrophoretic migration of DNA knots and catenanes. *J. Mol. Biol.* **278**, 1–3.
22. Grosberg, A. Y., Feigel, A. & Rabin, Y. (1996). Flory-type theory of a knotted ring polymer. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **54**, 6618–6622.
23. Katritch, V., Bednar, J., Michoud, D., Scharein, R. G., Dubochet, J. & Stasiak, A. (1996). Geometry and physics of knots. *Nature*, **384**, 142–145.
24. Laurie, B., Katritch, V., Dubochet, J. & Stasiak, A. (1998). Geometry and physics of catenanes applied to the study of DNA Replication. *Biophys. J.* **74**(6), 2815–2822.
25. Wasserman, S. A. & Cozzarelli, N. R. (1991). Supercoiled DNA-directed knotting by T4 topoisomerase. *J. Biol. Chem.* **266**, 20567–20573.
26. Craig, N., Craigie, R., Gellert, M. & Lambowitz, A., (eds). *Mobile DNA*, ASM Press, Washington, DC, 2002.
27. Grindley, N. D., Whiteson, K. L. & Rice, P. A. (2006). Mechanisms of site-specific recombination. *Annu. Rev. Biochem.* **75**, 567–605.
28. Boles, T. C., White, J. H. & Cozzarelli, N. R. (1990). Structure of plectonemically supercoiled DNA. *J. Mol. Biol.* **213**, 931–951.
29. White, J. H. & Cozzarelli, N. R. (1984). A simple topological method for describing stereoisomers of DNA catenanes and knots. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 3322–3326.
30. White, J. H., Millett, K. C. & Cozzarelli, N. R. (1987). Description of the topological entanglement of DNA catenanes and knots by a powerful method involving strand passage and recombination. *J. Mol. Biol.* **197**, 585–603.
31. Ernst, C. & Sumners, D. W. (1990). A calculus for rational tangles: applications to DNA recombination. *Math. Proc. Cambridge Philos. Soc.* **108**, 489–515.
32. Buck, D. & Verjovsky Marcotte, C. (2005). Tangle solutions for a family of DNA-rearranging proteins. *Math. Proc. Cambridge Philos. Soc.* **139**, 59–80.
33. Buck, D. & Verjovsky Marcotte, C. (2007). Classification of tangle solutions for integrases, a protein family that changes DNA topology. *J. Knot Theory Ramif.* **16**(8), 969–995.
34. Darcy, I. K. (2001). Biological distances on DNA knots and links: applications to XER recombination. *J. Knot Theory Ramif.* **10**, 269–294; Knots in Hellas, 1998, vol. 2 (Delphi).
35. Vetcher, A. A., Lushnikov, A. Y., Navarra-Madsen, J., Scharein, R. G., Lyubchenko, Y. L., Darcy, I. K. & Levene, S. D. (2006). DNA topology and geometry in Flp and Cre recombination. *J. Mol. Biol.* **357**, 1089–1104.
36. Grainge, I., Buck, D. & Jayaram, M. (2000). Geometry of site alignment during Int family recombination: antiparallel synapsis by the Flp recombinase. *J. Mol. Biol.* **298**, 749–764.
37. Ernst, C. & Sumners, D. W. (1999). Solving tangle equations arising in a DNA recombination model. *Math. Proc. Cambridge Philos. Soc.* **126**, 23–36.
38. Sumners, D. W., Ernst, C., Spengler, S. J. & Cozzarelli, N. R. (1995). Analysis of the mechanism of DNA recombination using tangles. *Q. Rev. Biophys.* **28**, 253–313.
39. Vazquez, M. & Sumners, D. W. (2004). Tangle analysis of Gin site-specific recombination. *Math. Proc. Cambridge Philos. Soc.* **136**, 565–582.
40. Vazquez, M., Colloms, S. D. & Sumners, D. (2005). Tangle analysis of Xer recombination reveals only three solutions, all consistent with a single three-dimensional topological pathway. *J. Mol. Biol.* **346**, 493–504.
41. Buck, D. & Flapan, E. (2007). A topological characterization of knots and links arising from site-specific recombination. *J. Phys. A*, **40**, 12377–12395.
42. Buck, D. & Flapan, E. (2007). A model of DNA knotting and linking. *OCAMI Stud.* **1**, 75–83.
43. Adrian, M., ten Heggeler-Bordier, B., Wahli, W., Stasiak, A. Z., Stasiak, A. & Dubochet, J. (1990). Direct visualization of supercoiled DNA molecules in solution. *EMBO J.* **9**, 4551–4554.
44. Lyubchenko, Y. L. & Shlyakhtenko, L. S. (1997). Visualization of supercoiled DNA with atomic force microscopy *in situ*. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 496–501.
45. Stark, W. M., Boocock, M. R. & Sherratt, D. J. (1992). Catalysis by site-specific recombinases. *Trends Genet.* **8**, 432–439.
46. Leschziner, A. E. & Grindley, N. D. F. (2003). The architecture of the γδ resolvase crossover site synaptic complex revealed by using constrained DNA substrates. *Mol. Cell*, **12**, 775–781.
47. Nollmann, M., He, J., Byron, O. & Stark, W. M. (2004). Solution structure of the Tn3 resolvase-crossover site synaptic complex. *Mol. Cell*, **16**, 127–137.
48. Li, W., Kamtekar, S., Xiong, Y., Sarkis, G., Grindley, N. D. F. & Steitz, T. A. (2005). Structure of a synaptic resolvase tetramer covalently linked to two cleaved DNAs. *Science*, **309**, 1210–1215.
49. Ennifar, E., Meyer, J. E. W., Buchholz, F., Stewart, A. F. & Suck, D. (2003). Crystal structure of a wild-type Cre recombinase-loxP synapse reveals a novel spacer conformation suggesting an alternative mechanism for DNA cleavage activation. *Nucleic Acids Res.* **31**, 5449–5460.
50. Guo, F., Gopaul, D. N. & van Duyne, G. D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature*, **389**, 40–46.
51. Chen, J., Englund, P. T. & Cozzarelli, N. R. (1995). Changes in network topology during the replication of kinetoplast DNA. *EMBO J.* **14**, 6339–6347.
52. Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A. & Ellenberger, T. (2005). A structural basis for allosteric control of DNA recombination by λ integrase. *Nature*, **435**, 1059–1066.
53. Gueguen, E., Rousseau, P., Duval-Valentin, G. & Chandler, M. (2005). The transpososome: control of transposition at the level of catalysis. *Trends Microbiol.* **13**, 543–549.
54. Smith, M. C. & Thorpe, H. M. (2002). Diversity in the serine recombinases. *Mol. Microbiol.* **44**, 299–307.
55. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.

56. Grainge, I. & Jayaram, M. (1999). The integrase family of recombinase: organization and function of the active site. *Mol. Microbiol.* **33**, 449–456.
57. Alén, C., Sherratt, D. J. & Colloms, S. D. (1997). Direct interaction of aminopeptidase a with recombination site DNA in Xer site-specific recombination. *EMBO J.* **16**, 5188–5197.
58. Radman-Livaja, M., Biswas, T., Ellenberger, T., Landy, A. & Aihara, H. (2006). DNA arms do the legwork to ensure the directionality of lambda site-specific recombination. *Curr. Opin. Struct. Biol.* **16**, 42–50.
59. Rolfsen, D. (1971). *Knots and Links.* Publish or Perish Press, Berkeley, CA.
60. Du, Q., Livshits, A., Kwiatek, A., Jayaram, M. & Vologodskii, A. (2007). Protein-induced local DNA bends regulate global topology of recombination products. *J. Mol. Biol.* **368**, 170–182.
61. Rice, P. A. & Steitz, T. A. (1994). Model for a DNA-mediated synaptic complex suggested by crystal packing of γδ resolvase subunits. *EMBO J.* **13**, 1514–1524.
62. Blake, D. G., Boocock, M. R., Sherratt, D. J. & Stark, W. M. (1995). Cooperative binding of Tn3 resolvase monomers to a functionally asymmetric binding site. *Curr. Biol.* **5**, 1036–1046.
63. MacDonald, D., Demarre, G., Bouvier, M., Mazel, D. & Gopaul, D. N. (2006). Structural basis for broad DNA-specificity in integron recombination. *Nature*, **440**, 1157–1162.
64. Yang, W. & Steitz, T. A. (1995). Crystal structure of the site-specific recombinase γδ resolvase complexed with a 34 bp cleavage site. *Cell*, **82**, 193–207.
65. Vologodskii, A. V. & Cozzarelli, N. R. (1994). Conformational and thermodynamic properties of supercoiled DNA. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 609–643.
66. Shlyakhtenko, L. S., Hsieh, P., Grigoriev, M., Potaman, V. N., Sinden, R. R. & Lyubchenko, Y. L. (2000). A cruciform structural transition provides a molecular switch for chromosome structure and dynamics. *J. Mol. Biol.* **296**, 1169.
67. Cherny, D. I. & Jovin, T. M. (2001). Electron and scanning force microscopy studies of alterations in supercoiled DNA tertiary structure. *J. Mol. Biol.* **313**, 295–307.
68. Rybenkov, V. V., Vologodskii, A. V. & Cozzarelli, N. R. (1997). The effect of ionic conditions on DNA helical repeat, effective diameter and free energy of supercoiling. *Nucleic Acids Res.* **25**, 1412–1418.
69. Rybenkov, V. V., Cozzarelli, N. R. & Vologodskii, A. V. (1993). Probability of DNA knotting and the effective diameter of the DNA double helix. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 5307–5311.
70. Yarmola, E. G., Zarudnaya, M. I. & Lazurkin Yu, S. (1985). Osmotic pressure of DNA solutions and effective diameter of the double helix. *J. Biomol. Struct. Dyn.* **2**, 981–993.
71. Huang, J., Schlick, T. & Vologodskii, A. (1998). Internal motion of supercoiled DNA: Brownian dynamics simulations of site juxtaposition. *J. Mol. Biol.* **284**, 287–296.
72. Johansson, C., Kamali-Moghaddam, M. & Sundstrom, L. (2004). Integron integrase binds to bulged hairpin DNA. *Nucleic Acids Res.* **32**, 4033–4043.
73. Bath, J., Sherratt, D. J. & Colloms, S. D. (1999). Topology of Xer recombination on catenanes produced by λ integrase. *J. Mol. Biol.* **289**, 873–883.
74. Benjamin, H. W., Matzuk, M. M., Krasnow, M. A. & Cozzarelli, N. R. (1985). Recombination site selection by Tn3 resolvase: topological tests of a tracking mechanism. *Cell*, **40**, 147–158.
75. Buck, D. (2007). The hybrid Int–Xer system and its tangle solutions (preprint).
76. Abremski, K., Frommer, B., Wierzbicki, A. & Hoess, R. H. (1988). Properties of a mutant Cre protein that alters the topological linkage of recombination products. *J. Mol. Biol.* **202**, 59–66.
77. Ernst, C. & Sumners, D. W. (1987). The growth of the number of prime knots. *Math. Proc. Cambridge Philos. Soc.* **102**, 303–315.