MATHEMATICAL METHODS IN DNA TOPOLOGY: APPLICATIONS TO CHROMOSOME ORGANIZATION AND SITE-SPECIFIC RECOMBINATION

JAVIER ARSUAGA*, YUANAN DIAO†, AND MARIEL VAZQUEZ‡

Abstract. In recent years, knot theory and low-dimensional topology have been effectively used to study the topology and geometry of DNA under different spatial constraints, and to solve the topological mechanisms of enzymes such as site-specific recombinases and topoisomerases. Through continuous collaboration and close interaction with experimental biologists, many problems approached and the solutions proposed remain relevant to the biological community, while being mathematically and computationally interesting. In this paper, we illustrate the use of mathematical and computational methods in a variety of DNA topology problems. This is by no means an exhaustive description of techniques and applications, but is rather intended to introduce the reader to the exciting applications of topology to the study of DNA. Many more examples will be found throughout this book.

Key words. DNA knots, bacteriophage P4, DNA packing, random knots, site-specific recombination, Xer, tangles.

AMS(MOS) subject classifications. Primary 57M25, secondary 92B99.

Motivation. DNA presents high levels of condensation in all organisms. Volume reduction, defined as the ratio between the volume occupied by a given genome and the volume occupied by a random walk of the same length as the genome, ranges from 10^2 in *Escherichia coli* to 10^4 in humans[50].

These large condensation values lead to questions such as how the DNA is packed inside the eukaryotic cell nucleus, the prokaryotic cell, as well as inside other organisms such as DNA viruses. The complexity of the packing problem is magnified when one considers that the DNA molecule needs to be readily available to multiple biological processes essential to the proper functioning of the organism, such as DNA replication, transcription, recombination and repair. The cell has evolved tools to remove unwanted DNA entanglement and solve other topological problems, such as DNA un-(or over-)winding, knotting or linking, and formation of multimers, that may interfere with its functions. DNA topology, the study of geometrical

*Department of Mathematics, San Francisco State University, San Francisco, CA 94132, USA (jarsuaga@sfsu.edu). This research was supported in part by the Institute for Mathematics and its Applications (IMA). J. Arsuaga is partially supported by NIH grant NIH-2S06GM52588-12.

†Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA (ydiao@uncc.edu). Y. Diao is partially supported by NSF grant DMS-0712958.

‡Department of Mathematics, San Francisco State University, San Francisco, CA 94132, USA (mariel@math.sfsu.edu). M. Vazquez is partially supported by the Institute for Mathematics and its Applications (IMA), and by NIH grants 2S06GM052588 and 20123550 U56 Mentoring core.

C.J. Benham et al. (eds.), *Mathematics of DNA Structure, Function and Interactions*,
The IMA Volumes in Mathematics and its Applications 150,
DOI 10.1007/978-1-4419-0670-0_2,

(supercoiling) and topological (knotting) properties of circular DNA, provides the necessary experimental and computational techniques to describe and quantify these problems and their solutions.

The paper is divided into two parts. In each part we present an important problem in DNA topology, and the mathematical and computational tools used to address it.

In Part I we discuss the formation of knots in bacteriophages and its implications for phage packing geometry. Bacteriophages are viruses that propagate in bacteria. Most dsDNA bacteriophages pack their genome in a similar way inside the capsid, a proteinic enclosure with icosahedral symmetry. In the 1980s Liu and colleagues found that DNA extracted from bacteriophages P4 and P2 capsids was mostly knotted [63, 64]. The origin of these knots and whether they contained any information about the organization of the DNA inside the capsid remained unexplored. Here we will describe our current knowledge on how these DNA knots are formed, in particular we will focus on different mathematical models that have been proposed to explain their formation. We will also emphasize how this problem has been amenable to interdisciplinary studies and has generated new mathematics [2, 7, 66].

Part II deals with the resolution of topological obstructions arising during replication of the E. coli chromosome. The bacterial chromosome, a $4.6Mbp$ double-stranded DNA circle, is condensed 10^3 times inside the nucleoid. The two DNA strands are wrapped around each other an average of $420,000$ times in the supercoiled bacterial chromosome and therefore the DNA double-helix must be unwound in order to be copied. Interwinding of newly replicated sister chromosomes in a partially replicated chromosome forms precatenanes, which become catenanes (links) upon completion of replication. Without careful management by cellular machinery, replication of the bacterial chromosome would lead to two sister molecules highly linked together. The cell must solve the topological problem of separating the two linked sister chromosomes to ensure proper segregation at cell division. Unlinking of replication catenanes is mainly achieved by the type II topoisomerase Topo IV (reviewed in [41, 81]).

Furthermore, stalled or broken replication forks are repaired by homologous recombination. Occasionally crossing-over by homologous recombination generates DNA dimers, which may be knotted [84]. The dimers are resolved by Xer recombination. The Xer system consists of enzymes XerC and XerD, which act cooperatively and co-localize at the septum with the protein FtsK. FtsK plays an essential role in dimer resolution, coordinates chromosome segregation and cell division (reviewed in [9]). Recent experimental evidence shows that XerCD-FtsK recombination can unlink catenanes formed by site-specific recombination *in vitro* [52], as well as catenanes formed by replication *in vivo* [47]. Here we will review the tangle method for site-specific recombination. We will illustrate the method

with applications to Xer recombination. The analysis will lead to several possible topological pathways followed by the enzymes. The question is posed as to whether the different pathways are simple planar projections of the same 3-dimensional topological mechanism.

Part I. DNA Knotting in Bacteriophages.

In this part of the paper we present the problem of DNA knotting in bacteriophage P4 as well as the various tools from the theory of random knotting used to approach this problem. Bacteriophage P4 knots are formed by random cyclization. In section 1 we introduce the problem of random cyclization of DNA in free solution. We discuss several computational methods currently used to simulate this process, as well as the corresponding analytical results to estimate the knotting probability of a random polygonal curve in $\Re^3$. This work is used as a framework to study the problem of DNA knotting in bacteriophages. Section 2 deals with cyclization of DNA in confined volumes. First, we review some of the experimental results on DNA knots found in the bacteriophage P4 system. This is followed by the description of three computational models and how these models have been used to address the biological problem. In section 3 we discuss the limitations of these approaches and future directions.

1. Cyclization of DNA molecules in free solution.

1.1. Experimental studies on random cyclization of DNA molecules. Random cyclization of long linear DNA molecules with sticky (i.e. complementary) ends produces knots with non-trivial probability. This knotting probability was independently estimated by Rybenkov et al. [78] and by Shaw and Wang [83]. Both groups showed that the formation of these knots depends on the length of the DNA molecule and on the ionic conditions of the solution (i.e. the effective diameter of the DNA molecule). In [78] it was found that the knotting probability for P4 DNA molecules circularized in solution was 3% and that the trefoil was the prevalent knot population followed by smaller amounts of the four crossing knot and even smaller amounts of the five crossing knots. Monte-Carlo simulations of idealized polymer chains (e.g [58, 66]) and analytical results [34] support these experimental results as explained below.

1.2. Simulations of Gaussian and equilateral random polygons without confinement. The wormlike chain is the most accurate polymer model for simulating DNA in solution. However other models such as the equilateral random polygon (ERP) or the Gaussian random polygon are good for estimating properties of long DNA molecules in the bulk and at the same time are more amenable to the development of rigorous analytical results. Several algorithms have been proposed for generating samples of equilateral random polygons. These include the crankshaft algorithm [58, 68], the hedgehog algorithm [58] and the pairwise rotation

TABLE 1

Summary results concerning various random polygons. Results in parenthesis are numerical results and – is the case when the numerical results vary because of the choices of parameters. GP_n stands for a Gaussian random polygon of n edges, EP_n stands for an equilateral random polygon of n edges, CEP_n stands for an equilateral random polygon of n edges within a confined space (usually a sphere with a predetermined radius), R_n stands for a uniform random polygon of n edges and SP_n for spooling random model.

	mean ACN	leading coeff.	knotting prob
GP_n	$O(n \ln n)$	$1/2\pi$	$\geq 1 - e^{-n^\epsilon} \to 1$
EP_n	$O(n \ln n)$	$3/16$	$\geq 1 - e^{-n^\epsilon} \to 1$
CEP_n	$O(n^2)$	$--$	$(\geq 1 - e^{-n^\epsilon} \to 1)$
R_n	$O(n^2)$	$(.115)$	$(\geq 1 - e^{-n^\epsilon} \to 1)$
SP_n	$O(n^2)$	$--$	$(\geq 1 - e^{-n^\epsilon} \to 1)$

algorithm [73]. The crankshaft algorithm is fairly popular among some researchers. In this algorithm two vertices of the polygon are selected at random, dividing the polygon into two subchains. One of the two subchains is selected at random (with equal probabilities for each subchain), and the selected subchain is rotated through a random angle around the axis connecting the two end vertices of the subchain. In the hedgehog algorithm an ERP is first generated and at each step two vectors are selected at random, rotated with respect to their sum and placed back in the polygon. The crankshaft algorithm generates an ergodic Markov chain in the space of all ERPs of fixed length [68]. However the main drawback with this algorithm is that the correlation between samples is very high and therefore many configurations need to be generated in between any two samples in the Markov chain. The hedgehog algorithm on the other hand generates independent samples of polygons however it is unknown whether it is ergodic or not. These algorithms have helped to estimate some of the biologically relevant properties, such as the knotting probability and the mean of the Average Crossing Number distribution (ACN), for equilateral polygons as shown in Table 1. The leading coefficient 3/16 in the equilateral random polygon case means the mean ACN of equilateral random polygons of length n is of the form $(3/16)n \ln n + O(n)$. This number is $1/2\pi$ for the case of Gaussian random polygons.

1.3. Analytical results for knotting of polygons without confinement: Gaussian and equilateral random polygon models. A Gaussian random vector $X = (x, y, z)$ is a random point whose coordinates x, y and z are independent standard normal random variables (with mean $= 0$ and variance $= 1$). The pdf (probability density function) of X is the joint pdf of x, y and z, which is

$$f(X) = \left(\frac{1}{\sqrt{2\pi}}\right)^3 e^{-\frac{x^2+y^2+z^2}{2}} = \left(\frac{1}{\sqrt{2\pi}}\right)^3 e^{-\frac{|X|^2}{2}}.$$

A Gaussian random walk of n steps (denoted by GW_n) consists of $n+1$ consecutive points $X_0 = (0,0,0) = O,\ X_1,\ X_2,\ \ldots\ ,\ X_n$ such that $Y_{k+1} = X_{k+1} - X_k$ $(k = 0, 1, ..., n-1)$ are independent Gaussian random vectors. It follows that the joint pdf for all the vertices is

$$f(X_1, X_2, ..., X_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^{3n} e^{-\frac{1}{2}(|Y_1|^2+|Y_2|^2+\cdots+|Y_n|^2)}$$
$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{3n} e^{-\frac{1}{2}(|X_1|^2+|X_2-X_1|^2+\cdots+|X_n-X_{n-1}|^2)}.$$

A Gaussian random polygon GP_n is a conditioned GW_n of n steps such that the last vertex X_n coincides with the starting point $X_0 = O$. Thus, if we let $g_n(X_n)$ be the pdf of X_n for a GW_n,

then the joint pdf of $X_1,\ X_2,\ ...,\ X_{n-1}$ of a GP_n is

$$h(X_1, X_2, ..., X_n) = f(X_1, X_2, ..., X_n)/g_n(O).$$

The one advantage of the Gaussian random polygons (over other random polygon models) is that the joint probability density function of its vertices is of an explicitly nice form. This enabled the derivation of the following result concerning the knotting probability of a GP_n [34].

THEOREM 1.1. *[34] Let $\mathcal{K}$ be any knot type, then there exists a positive constant ϵ such that GP_n contains $\mathcal{K}$ as a connected sum component with a probability at least $1 - \exp(-n^{\epsilon})$ provided that n is large enough.*

One can obtain a similar result for equilateral random polygons.

Suppose $Y_1,\ Y_2,\ \ldots\ ,\ Y_n$ are n independent random vectors uniformly distributed on S^2. An equilateral random walk of n steps, denoted by EW_n, is defined as the sequence of points in the three dimensional space $\mathbf{R}^3$: $X_0 = O,\ X_k = Y_1 + Y_2 + \cdots + Y_k,\ k = 1, 2, ..., n$. Each X_k is called a vertex of the EW_n and the line segment joining X_k and X_{k+1} is called an edge of EW_n (which is of unit length). Notice that the coordinates of each point are not independent from each other due to the fact that the distance between consecutive points in the polymer needs to be one. If the last vertex X_n of EW_n is fixed, then we have a conditioned random walk $EW_n|X_n$. In particular, EW_n becomes a polygon if $X_n = O$. In this case, it is called an equilateral random polygon and is denoted by EP_n. The joint probability density function $f(X_1, X_2, ..., X_n)$ of the vertices of an EW_n is $f(X_1, X_2, ..., X_n) = \varphi(U_1)\varphi(U_2)\cdots\varphi(U_n) = \varphi(X_1)\varphi(X_2 - X_1)\cdots\varphi(X_n - X_{n-1})$. Where $\varphi(U_i)$ is the density function of selecting a random point over the surface of the sphere.

Let X_k be the k-th vertex of an EW_n $(n \geq k > 1)$, its density function is defined by

$$f_k(X_k) = \int\int\cdots\int \varphi(X_1)\varphi(X_2 - X_1)\cdots \varphi(X_k - X_{k-1})dX_1dX_2\cdots dX_{k-1} \tag{1.1}$$

and it has the closed form $f_k(X_k) = \frac{1}{2\pi^2 r}\int_0^\infty x \sin rx \left(\frac{\sin x}{x}\right)^k dx$ [74]. In the case of EP_n, the density function of the vertex X_k can be approximated by a Gaussian distribution, as given in the following lemma [30, 33, 34].

LEMMA 1.1. *Let X_k be the k-th vertex of an EP_n and let h_k be its density function, then*

$$h_k(X_k) = \left(\sqrt{\frac{3}{2\pi\sigma_{nk}^2}}\right)^3 \exp\left(-\frac{3|X_k|^2}{2\sigma_{nk}^2}\right) + O\left(\frac{1}{k^{5/2}} + \frac{1}{(n-k)^{5/2}}\right), \quad (1.2)$$

where $\sigma_{nk}^2 = \frac{k(n-k)}{n}$.

In other words the density of the k step of an EP_n can be approximated by a Gaussian distribution with mean 0 and a standard deviation that depends on The vertex number k and on the distance from the vertex to the origin (or first vertex in the polygon).

This lemma provided the key link to apply the methods used in [34] for the Gaussian random polygons to the equilateral random polygons, which leads to the following theorem.

THEOREM 1.2. *[30] Let $\mathcal{K}$ be any knot type, then there exists a positive constant ϵ such that EP_n contains $\mathcal{K}$ as a connected sum component with a probability at least $1 - \exp(-n^\epsilon)$, provided that n is large enough.*

Numerical studies on EP_n suggest a scaling law of $1-\exp(-n/a)$ with $a = 244 \pm 5$ (see [66] and references therein).

The above two theorems imply that a long GP_n or EP_n contains many connected sum components (with a high probability), which makes it highly unlikely for the polygon to be achiral. This is stated in the following corollary. However, this only provides reason for the long GP_n and EP_n to favor chiral knots than achiral ones. For relatively short polygons, this is not clear.

COROLLARY 1.1. *[30, 34] There exists some constant $\theta > 0$ such that the probability that a GP_n or an EP_n is a chiral knot is at least $1 - \frac{1}{n^\theta}$.*

The determination of the knot type of a circular molecule can tell us its topological (minimum) crossing number, i.e., the minimum number of crossings one will see no matter how this molecule is artificially stretched, twisted, or bent. However, the average crossing number (ACN), defined as the average of crossing numbers over all orthogonal projections of the molecule, is a more natural geometric measure of the molecule entanglement as it refers to the actual number of crossings that can be perceived while observing a non-perturbed trajectory of a given molecule [55]. Furthermore, it is believed that DNA knots migrate in gel eletrophoresis accordingly with their ACN [99].

The following theorems are presented in [31, 32] and establish the $O(n \ln n)$ behavior of the mean ACN for the Gaussian and equilateral random polygons as illustrated in Table 1.

THEOREM 1.3. *Let χ_n be the ACN of an equilateral random walk of n steps; then*

$$E(\chi_n) = \frac{3}{16} n \ln n + O(n).$$

On the other hand, if χ'_n is the ACN of a Gaussian random polygon of n steps, then

$$E(\chi'_n) = \frac{1}{2\pi} n \ln n + O(n).$$

2. Cyclization of DNA molecules in confined volumes: DNA knotting in bacteriophage P4 capsids.

2.1. Experimental studies on DNA knots in bacteriophages. In dsDNA bacteriophages the volume of the bacteriophage genome is reduced 100 times inside the capsid [53]. This volume reduction imposes severe physical constraints on the DNA molecule. For instance the DNA molecule is under (at least) 50 atmospheres of pressure [42, 93] and at a concentration of 800mg/ml [56]. Despite these conditions the dsDNA molecule is believed to preserve its double helical structure [8] and not to have sequence-specific associations with the protein capsid. A number of models have been proposed to describe the organization of the viral chromosome under such extreme conditions of condensation. These include coaxial and concentric spooling models [4, 20, 35, 76, 82], coaxial models [10], toroidal models [51, 72], and liquid-crystaline models [61].

Bacteriophage P4 is an icosahedral phage of radius $r = 180$Å and a linear dsDNA genome of 11.5 kb ($l = 120 \times 10^3$Å). The genome is flanked by two 16*bp* long single stranded complementary sequences of DNA called *cos* sites. During phage morphogenesis a protein enclosure called capsid is assembled first. This is followed by the packing of a single linear DNA molecule into the capsid through the portal vertex. Infective viruses keep at least one of their *cos* sites attached near the portal [21]. This attachment prevents the two *cos* ends from meeting within the capsid and circularizing the chromosome. However in the experiments performed by Liu and colleagues it was found that most of the DNA molecules phenol extracted from bacteriophage P4 are circular and non-trivially knotted [63, 64].

Recent work [5, 6, 91, 92] reproduced and extended the results of Liu et al. Figure 1 shows a two dimensional gel of DNA knots from bacteriophage P4 in which different conditions are used in each dimension [91]. In this figure the top spot corresponds to the unknotted molecule followed, along a bell-shaped curve, by the trefoil knot, the figure eight (four-crossing) knot, and so on. The spot ahead of the bell is the linear chain. The most remarkable fact about this distribution is that about 95% of the DNA molecules are knotted and only about 2% are knots between 3 and 10 crossing knots.

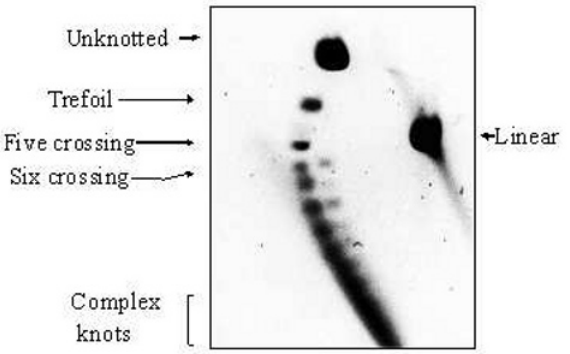

FIG. 1. *Two dimensional gel of knots extracted from bacteriophage P4.*

Furthermore the large majority of the population consist of knots with 30 crossings or more [5]. These results are in high contrast with those obtained by random cyclization of P4 DNA molecules in free solution (see Section 1.1 [78]) and suggested that knots extracted from bacteriophage P4 are formed inside the phage capsid and therefore may be used as reporters for chromosome organization in P4. Despite the small percentage of knots with less than 8 crossings (i.e. those that can be separated by gel electrophoresis) two important properties of the knot distribution were revealed. First the four crossing knot is mostly absent and second the torus knots 5_1 and the 7_1 are more probable than the twist knots 5_2 and 7_2 (contrary to what is expected in free solution). The theoretical work described next aims at explaining these experimental results. In this review we will focus only on the problems of knotting probability and complexity.

2.2. Random knotting within a confined space. A simple approach to study the knotting probability and complexity of P4 knots is by generating ensembles of random polygons inside different convex volumes. Next we describe three models: the Confined Equilateral Random Polygon (CEP_n), the Uniform Random Polygon (URP_n) and the Random Spooling (SP_n).

The Confined Equilateral Random Polygon.

In this model we consider ERPs confined to spheres of certain radius r and use CEP_n to denote such a polygon of length n. Figure 2 shows an example of such polygon.

Unfortunately, the extra condition that confines the polygon to a sphere of radius r completely invalidates the approximation formula given in Lemma 1.2 for the vertex X_k of CEP_n. Intuitively, a CEP_n would be more likely to be knotted than an EP_n. Indeed, this is confirmed by numerical studies [5, 65, 66] which were pioneered by Michels and Wiegel [65].

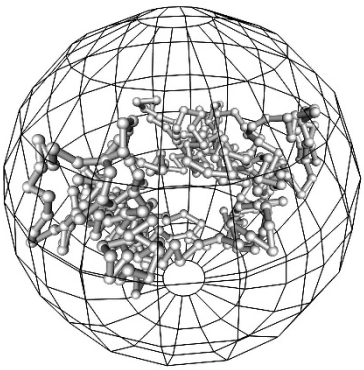

FIG. 2. *An equilateral random polygon inside a sphere.*

In their studies molecular dynamics algorithms were used to sample closed polygons and the knotting probability was computed. Michels and Wiegel found that the knotting probability of a polygon inside a sphere increases with respect to that in $\Re^3$ following a $\exp(N^\alpha/r^3)$ law with $\alpha = 2.28$. In more recent work [5, 66] large ensembles of CEP_n were generated by the crankshaft algorithm. In [66] the scaling law proposed by Michels and Wiegel was confirmed and the coefficient $\alpha = 2.15 \pm 0.04$ improved.

In [5], the combination of experimental and theoretical results led to propose that the effect of the confinement during the random cyclization process of the DNA molecule is one of the key drivers in the formation of knots in the P4 system. This argument has been extended by D. Smith′ group to explain the knotting of chains in confined volumes [75]

The Uniform Random Polygon.

Developing analytical results for CEP_n is a very difficult problem. An alternative model was proposed in [68] as a way to study the random behavior of circular DNA molecules packed in phage capsids that may provide clues about showing some of these analytical results. For $i = 1, 2, ..., n$, let $U_i = (u_{i1}, u_{i2}, u_{i3})$ be a three-dimensional random point that is uniformly distributed in the unit cube C^3 (or in a unit ball) such that U_1, U_2, ..., U_n are independent. Let e_i (called the i-th edge) be the line segment joining U_i and U_{i+1}, then the edges e_1, e_2, ..., e_n define a *uniform random polygon* R_n in the confined space (either the cube or the sphere), where e_n is the line segment joining U_n and U_1. A polygon of length n is denoted by URP_n.

While the knotting probability of an R_n has not been analytically determined (even in the case of $n \to \infty$), a numerical study carried out in [3] provided convincing data that the knotting probability of an R_n quickly approaches 1 as n approaches infinity. Figure 4 is the plot of the percentage of URPs with non-trivial determinant (i.e. those whose Alexander poly-

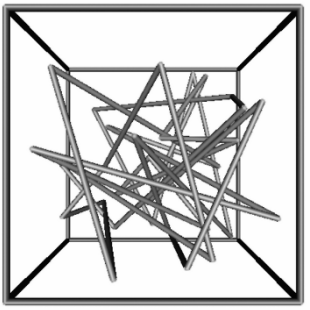

FIG. 3. *A uniform random polygon confined in the unit cube.*

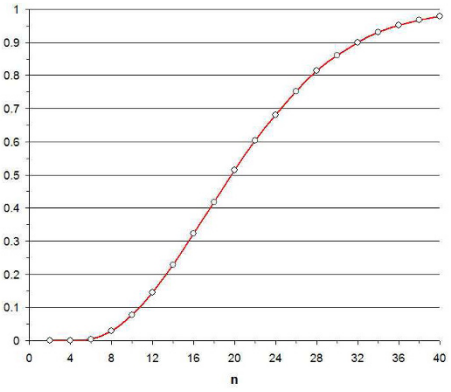

FIG. 4. *The lower bound of knotting probability for URPs up to 40 segments.*

nomial evaluated at $t = -1$ is non-trivial). Since the trivial knot has a trivial determinant, the results give a lower bound of the knotting probability. Figure 4 below is the plot of the data. The fitting curve used here is $1 - \exp(-0.000082n^3)$, although this is not to be expected as a general rule since the trivial knot probability of an R_n is at least of order $\exp(-n \ln n)$ as shown in [3].

It turns out that the mean ACN behavior for an R_n is much easier to determine, both analytically and numerically than for CEPs. Consider a uniform random polygon R_n with n edges e_1, e_2, ..., e_n in that consecutive order. Let $a(e_i, e_j)$ be the average crossing number between e_i and e_j, then the ACN of R_n is

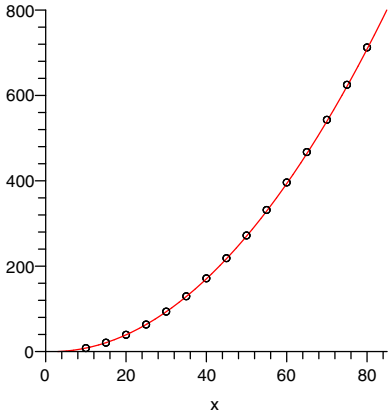

FIG. 5. *The mean ACN of uniform random polygons up to 80 vertices.*

$$\chi_n = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i-1,i,i+1} a(e_i, e_j).$$

It follows that the expected value of the average crossing number of R_n is

$$E(\chi_n) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i-1,i,i+1} E(a(e_i, e_j)) = p(n-3)n.$$

This establishes the $O(n^2)$ behavior of the mean ACN of an R_n as shown in Table 1. Numerical studies in [2] produced the following near perfect fit using $E(\chi_n) \approx 0.115(n-3)n$.

There have not been enough numerical studies on the knot types of the CEP_n's [5, 66] and R_n's [67] to indicate their bias against achiral knots, even though this is generally expected for long random polygons since achiral knots are much rarer than the achiral ones within large knots.

The Random Spooling model.

The last model we discuss is the *random spooling model*. This model incorporates features from the random knotting models (described above) into the spooling and toroidal models [4, 10, 20, 35, 51, 72, 76, 82]. In the standard spooling model DNA fibers spool around an axis forming coaxial spherical layers. In [59] the knot type of molecular dynamics generated spooling conformations was studied and it was found that most of these conformations were unknotted. These results together with the wide distributions of knots that are observed in P4 suggested that current theoretical models of DNA packing disregard the effect of random fluctuations which

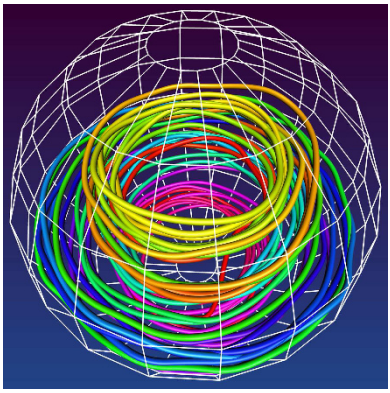

FIG. 6. *The random spooling model.*

in fact may play an important role in the packing of the viral chromosome. We recently proposed [7] that fibers follow spooling trajectories and at the same time they intermingle, as illustrated in the figure 6. This intermingling between fibers of different coaxial layers increases the knotting probability.

Some initial simulation and analytical results have been published [7]. For instance we have estimated the complexity of the average crossing number in the direction of the spooling axis as stated in the next theorem.

THEOREM 2.1. *[7] Let P_n^s be a spooling random polygon, then the average number of crossings in its projection to the xy-plane perpendicular to its center axis is of the order of $O(n^2)$.*

Although the knotting probability has not yet been shown to increase to 1 as suggested by the numerical results shown in Figure 7 a relationship between the writhe of the projection along the spooling axis and the knot type has been proven. The following theorem is a consequence of a theorem due to Morton[69]

THEOREM 2.2. *[7] Let $w(D_n)$ be the writhe of the projection in the direction of spooling axis and $\sigma(P_n^s)$ the number of times the spool goes around its axis. If $|w(D_n)| \geq \sigma(P_n^s)$, then P_n^s is a non-trivial knot. Furthermore, in this case P_n^s cannot be an achiral knot.*

This theorem shows that spooling conformations with high writhe are knotted. This agrees well with some of our results that relate DNA knotting and writhing in bacteriophage P4 [6, 11] and suggests that high writhe may also play an important role in the formation of knots in P4.

3. Conclusions. Here we have discussed the problem of knotting by random cyclization in free solution and in confined volumes. In both cases we have presented experimental, analytical and computational results. By comparing our experimental results with those obtained in free solution we

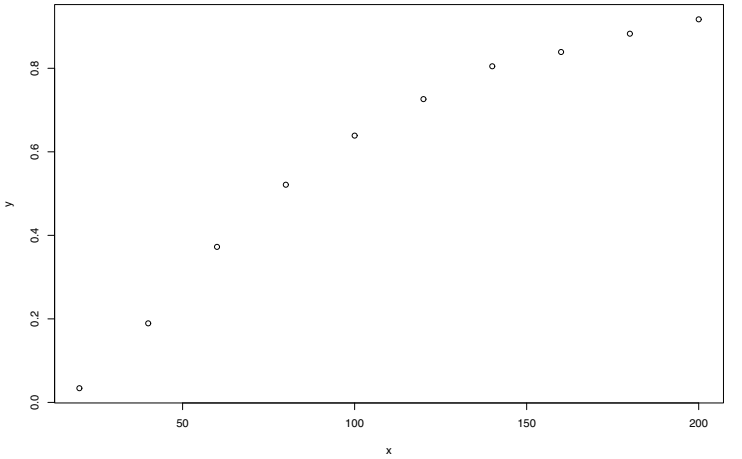

FIG. 7. *Knotting probability as a function of the length of the chain for the random spooling model.*

concluded that knotting in bacteriophage P4 occurs before, or very soon after, the disruption of the capsid and therefore P4 knots can be used as reporters of DNA packing. The large amount of knotting is still a feature that is not fully explained by current mathematical models. In this review we have presented three random knotting models: the confined equilateral polygon, the uniform random polygon and the spooling random polygon. All these models present consistent results however they do not reach the high levels of complexity found in bacteriophages. This is specially true if more accurate representations of the DNA molecule are taken into account. Nevertheless some information about the biological system has been extracted from these theoretical models. For instance our current simulations results suggest that DNA knotting in P4 is mainly driven by the confinement imposed by the capsid during the cyclization reaction, and perhaps also by possible biases introduced by the arrangement of the viral chromosome [6, 11]. Importantly none of the current idealized models proposed in the literature account for the formation of knots inside the capsid and previous simulations results failed to do so [59] thus suggesting that they may not reflect some important properties of the DNA packing. The random spooling model is our first attempt to address this issue. It remains to be seen if such models can reproduce the knot distributions observed experimentally. New experimental results have recently obtained for P4 deletion mutants whose genomes range between 5 and 8 *kb* [92]. These experiments hold a great promise for unveiling the properties that drive knotting in phage capsids as well as some of the essential features of the viral packing.

Part II. Enzymes that change the topology of DNA.

The cell has evolved a set of tools to remove unwanted DNA entanglement and solve other topological problems. Noteworthy are enzymes that change the topology of DNA such as site-specific recombinases and topoisomerases. For example, replication of the circular E. coli genome produces two catenated (linked) circles that cannot be separated without breaking the DNA chain; the type II topoisomerase topoIV plays an important role in unlinking the newly replicated genomes, thus ensuring proper segregation at cell division. Similarly, circular chromosome dimers arise occasionally due to crossing over by homologous recombination of newly replicated chromosomes. The site-specific recombination system XerCD coupled with the multifunctional protein FtsK resolves the chromosome dimers, thus allowing for stable inheritance. In Section 4 we introduce the biology of site-specific recombination. In Section 5 we briefly review the mathematics used in the tangle method, and the method itself, including the main assumptions. We finish the review by summarizing the tangle analysis of Xer recombination at *psi* and of XerCD-FtsK recombination at *dif*. We emphasize the 3-dimensional interpretation of the data.

4. Site-specific recombination. *Site-specific recombinases* are enzymes able to change the genetic code of their DNA substrate. They mediate important biological processes such as inversion of DNA segments, transposition of a DNA segment from one location to another along the genome, integration and excision of viral DNA into and out of its host genome, and resolution of multimeric DNA molecules [48, 79].

Site-specific recombination occurs in two steps. First, during *synapse* two short DNA segments of specific sequence are brought together by the site-specific recombinase(s) and any necessary accessory proteins. These segments are called *recombination sites*. Second, during *strand-exchange*, each recombination site is cleaved, the lose ends recombined and the recombined pieces are rejoined. The DNA sequence in a recombination site is usually non-palindromic, thus allowing to define a site orientation. When the DNA substrate consists of circular DNA molecules, the two recombination sites may occur in a single DNA circle or in two separate circles. In the first case, for *intramolecular recombination*, if the sites induce the same orientation on the circle they are said to be in *direct repeat*, otherwise they are in *inverted repeat* (Figure 8). Relative orientation of the sites is harder to characterize in the case of *intermolecular recombination* (i.e. two sites on separate DNA circles). In the simple case of $T(2, n)$ torus links with 4 or more crossings the sites are said to be in *parallel* or *anti-parallel* orientation with respect to each other (Figure 8).

Based on sequence homology and strand-passage mechanism, site-specific recombinases are divided into two families: serine recombinases and tyrosine recombinases [60, 71]. After synapse formation, the serine

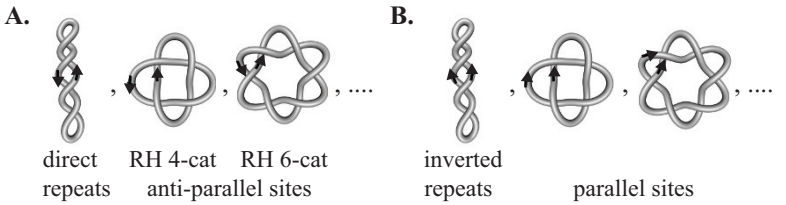

FIG. 8. *Site orientation. A. Negatively supercoiled circle with sites in direct repeat, right-hand 4-crossing torus link (RH 4-cat) with anti-parallel sites, right-hand 6-crossing torus link (RH 6-cat) with anti-parallel sites. B. Negatively supercoiled circle with sites in inverted repeats, RH 4-cat and RH 6-cat with parallel sites.*

recombinases introduce one double-stranded break on each recombination site and act by rotation of the synaptic complex around a dyad axis. The enzymes may iterate this process two or more times before releasing the sites, resulting in a *processive recombination* reaction. Enzymes in this family include the Tn3 resolvase and Gin of bacteriophage Mu [54, 87, 100]. Tyrosine recombinases go through a Holiday Junction intermediate performing single-stranded cleavage in two steps. Tyrosine recombinases are often represented by λ-Int from bacteriophage λ and include enzymes such as XerCD, Flp and Cre [71]. There is little evidence of the ability of tyrosine recombinases to act processively due to the fact that the HJ forces the synapse to reform at each step. However in Gourlay and Colloms [45] and in Grainge *et al.* [47], XerCD recombination is consistent with iterative cleavage and strand-exchange.

Site-specific recombinases can change the topology of circular DNA substrates [24, 87, 100]. Such changes can be quantified experimentally via gel electrophoresis and electron microscopy (*e.g.* [25]) . The experimental data can then be subjected to quantitative and mechanistic analysis to answer questions of DNA bending and strand-exchange. Close scrutiny of the data aided by geometrical models is used by biologists to understand the molecular mechanism of the enzyme(*e.g.* [87, 100]). Mathematicians can greatly improve the reliability of the results using rigorous mathematical tools. The most famous example of these is the tangle method proposed by Ernst and Sumners [39], which has been extensively used in the topological analysis of site-specific recombination reactions. Other more contemporary approaches include the topological characterization of site-specific recombination products ([13, 14, 15, 16]), the classification of 3-string tangles in order to generalize the 2-string tangle model [18, 19, 36], and novel methods to solve 2- and 3-string tangle equations for tangles that are not necessarily rational or Montesinos (*i.e.* sum of rational) [27, 28, 57].

5. The tangle method. The tangle method, based on knot theory and low-dimensional topology, is a mathematical method in which the ac-

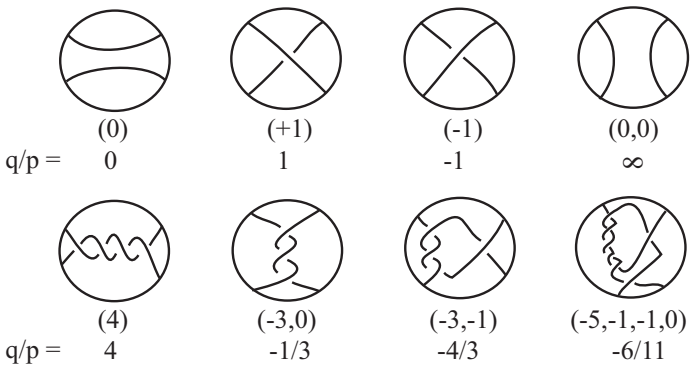

FIG. 9. *Rational tangles. The figure shows eight rational tangles, their Conway vectors and their corresponding classifying rational numbers (Conway Numbers). The top row illustrates the 4 trivial tangles.*

tion of site-specific recombinases on circular DNA substrates is modeled using 2-string tangles. Explaining the method requires a few basic definitions.

5.1. Tangles and knots. The following is a brief overview, more detailed definitions can be found in [17, 70, 77]. A 2-string tangle is an ordered pair (B, t), where B is a fixed oriented 3-ball in S^3 and t is a pair of non-oriented disjoint arcs properly embedded in B, whose endpoints lie on the bounding sphere. For each (B, t) there is an orientation-preserving homeomorphism

$$\Phi : \ (B, t) \to (D, t_\Phi)$$

that maps B onto the unit 3-ball D, and t onto two straight arcs t_Φ in D connecting the preferred equatorial points NE with SE, and NW with SW. The endpoints of t are mapped to the 4 special equatorial points {NW, NE, SE, SW}. Notice that in order to compare tangles defined in different 3-balls, we shall define a 2-string tangle more precisely as the triplet (B, t, Φ). In this way we may consider, without loss of generality, all tangles as defined on the unit 3-ball D with strings attached to the 4 special equatorial points. Two tangles (D, t_1) and (D, t_2) are equivalent if there is an ambient isotopy that takes t_1 to t_2 while fixing the endpoints.

A tangle diagram is a planar representation of the 3-dimensional tangle in the 3-ball, and may be obtained by projecting the arcs onto the equatorial disk (illustrated in Figure 9).

There are three different types of tangles: rational, locally knotted and prime. A tangle is rational if the strings can be continuously deformed to lie on the bounding 2-sphere [23]. Several rational tangles are illustrated in Figure 9. A tangle (D, t) is locally knotted if there is a 2-sphere S inside D

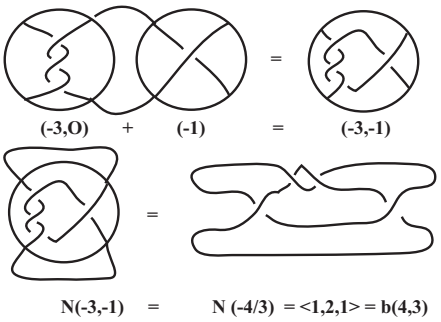

FIG. 10. *Tangle sum and numerator closure. The top diagram illustrates the sum of two rational tangles, yielding in this case another rational tangle with Conway vector* $(-3,-1)$. *The bottom diagram shows the numerator closure of* $(-3,-1)$ *which results in the 4-plat* $b(4,3)$ *described by the vector* $\langle 1,2,1\rangle$

that intersects one of the two arcs of t transversely in two points, and such that the 3-ball bounded by S contains t as a knotted arc with endpoints on S. If a tangle is neither rational nor locally knotted, then it is called a prime tangle [62].

There is a $1-1$ correspondence between equivalence classes of rational tangles and the extended rational numbers [23, 43]. Each equivalence class of rational tangles can be represented by its Conway symbol, a classifying vector $(a_1, a_2, , a_n)$ of integer entries such that $|a_1| > 1$, all entries are nonzero except possibly for a_n, and all entries have the same sign. This scheme applies to all but four exceptional rational tangles called trivial tangles. A unique extended rational number can be obtained from the classifying vector of a rational tangle via the following continuous fraction calculation :

$$\frac{q}{p} = a_n + \cfrac{1}{a_{n-1} + \cfrac{1}{a_{n-2} + \cfrac{1}{\cdots a_2 + \cfrac{1}{a_1}}}}$$

where $\frac{q}{p} \in \mathbf{Q} \cup \{\frac{1}{0}\}$, $q \in \mathbf{Z}$ and $\gcd(p,q) = 1$. Rational numbers associated to a few different rational tangles are shown in Figure 9.

Two tangle operations, the sum and the numerator closure, are illustrated in Figure 10. The numerator closure converts a tangle into a knot or link. In particular, it relates rational tangles with the family of 4-plats.

A 4-plat is a knot or link that can be obtained by braiding 4 strings and capping off the ends as illustrated in Figure 10 (bottom right) [17]. To each 4-plat can be associated a classifying vector $\langle c_1, c_2, ..., c_{2k+1}\rangle$ (Conway vector), such that $c_i > 0$ for all i. Two 4-plats $\langle c_1, c_2, ..., c_{2k+1}\rangle$ and $\langle d_1, d_2, ..., d_{2k+1}\rangle$ are equivalent if and only if $c_i = d_i$ or $c_i = d_{2k+1-i}$ for all $i = 1, ..., 2k+1$.

From the Conway vector, we can obtain a classifying rational number for each 4-plat through a continued fraction:

$$\frac{\beta}{\alpha} = \cfrac{1}{c_1 + \cfrac{1}{c_2 + \cdots + \cfrac{1}{c_{2k+1}}}}$$

where $0 < \beta < \alpha$. Hence, we can also denote a 4-plat knot by its *Conway symbol* $b(\alpha, \beta)$. Furthermore, by the Classification Theorem of 4-plats [17], we have that $b(\alpha, \beta)$ and $b(\alpha', \beta')$ are equivalent non-oriented links if and only if $\alpha = \alpha'$ and $\beta^{\pm 1} = \beta' \bmod(\alpha)$. If A is a rational tangle, then the result of taking its numerator closure $N(A)$ is a 4-plat (Figure 10). If A and B are rational tangles, then $N(A + B)$ is a 4-plat [17, 39].

5.2. The tangle method. The tangle method, first proposed by Ernst and Sumners in 1990 [39], and reviewed in [70, 90, 101], uses tangles to model the changes in topology of the synaptic complex before and after recombination. The method relies on knowledge on the topology of substrate and products, and on a few justified assumptions.

The first assumption is that the enzymatic mechanism is constant and independent of the geometry and topology of the DNA substrate. Changes in the DNA substrate due to binding and stand-exchange occur inside a tangle E, which contains the recombination sites and any other DNA bound by the enzyme(s). Topological differences between substrates are detected in the tangle O_f, the exterior of E, and are to remain unchanged during recombination.

The second assumption is that E can be partitioned into two tangles O_b and P. O_b contains any DNA bound by the enzymes which is unchanged by recombination. P contains the core regions of the recombination sites, *i.e.* the locations of cleavage and strand-exchange. Site-specific recombination is modeled by tangle surgery where P is converted into R. Most serine recombinases, and some tyrosine recombinases such as XerCD, display topological selectivity (*i.e.* they distinguish between sites in different orientations and in one or two molecules) and specificity (*i.e.* the topology of the product is uniquely determined by the substrate topology) [22, 86].

When the enzyme has topological specificity then, given a substrate of fixed topology, the mechanisms of binding and strand-exchange are constant. In this case the assumptions of the tangle method imply that the tangles O_b, P and R are constant (*i.e.* enzyme-specific). Solving for the topological mechanism of the enzyme when substrate and product of recombination are known is equivalent to knowing who O_b, P and R are. A site-specific recombination event where a substrate of specific topology (K_1) is converted into a product of specific topology (K_2) is modeled as a system of tangle equations:

$$N(O_f + O_b + P) = K_1 = N(O + P)$$
$$N(O_f + O_b + R) = K_2 = N(O + R),$$

where $O = O_f + O_b$ is the outside tangle consisting of all DNA (bound and unbound) which is not changed by strand-exchange.

A processive recombination event where two or more rounds of recombination take place in a single reaction is modeled as a system of two or more tangle equations.

$$\begin{aligned} N(O + P) &= K_0 \\ N(O + R) &= K_1 \\ N(O + 2R) &= K_2 \\ \cdots &= \cdots \\ N(O + nR) &= K_n. \end{aligned}$$

Here K_0 is the substrate and K_i (i = 1, 2, ..., n) is the product of successive rounds of recombination. When substrate and products of recombination are 4-plats, systems of tangle equations can be solved for tangles that are rational or sums of rational tangles [39, 40]. Using tools from low-dimensional topology (e.g. Dehn surgery, cyclic surgery theorem) O, P and R can sometimes be shown to be rational or sums of two rational tangles [26, 39, 40, 97, 98]. In the absence of processive recombination the corresponding systems of two tangle equations often admit infinitely many computable solutions. When this is the case, reasonable assumptions can be made on P and R to limit the number of solutions to a small finite number.

Assumptions on P and R.

P is defined as the ball containing the core regions of the recombination sites (*i.e.* where the breakage and rejoining takes place). Usually these regions are very short DNA segments (e.g. 28bp for Tn4430 [95], 32 bp for XerCD [89]) and are thus unlikely to cause tangling inside P. Therefore, P can be any of the four trivial tangles with 0 or 1 crossings. We chose $P = (0)$. Any geometrical complexity induced on the DNA substrate by enzymatic binding is trapped in the O tangle (and more specifically in O_b).

Serine recombinases act by rotation of the sites around a dyad axis. If $P = (0)$ with parallel sites then $R = (k)$ for some integer k. Subsequent rounds of recombination in a processive event are modeled by $2R = (2k)$, $3R = (3k)$ and so on.

Members of the tyrosine family of site-specific recombinases (such as XerCD, Int, Cre, Flp, TnpI) catalyze recombination through a Holiday junction intermediate (HJ) [60, 71]. If $P = (0)$ with parallel sites and the enzyme recombines via a HJ, then $R = (+1)$ or (-1). If $P = (0)$ with antiparallel sites then $R = (0, 0)$. Should two or more rounds of recombination occur iteratively, then $R = (k)$ for $P = (0)$ parallel. In the case of anti-parallel sites the progression of R is harder to visualize (Shimokawa *et al.* preliminary report [85]).

3D considerations of the tangle model.

The orientation of the recombination sites is inherited into the tangle P. The two recombination sites in $P = (0)$ are in *parallel alignment* if both arrows point in the same direction in the tangle diagram, otherwise they are in *antiparallel alignment*.

However, the concept of parallel and antiparallel alignment represents a local geometric property of the recombination sites and is well-defined only in the tangle diagram, which corresponds to a planar projection of the 3-dimensional tangle. Unless the two sites are strictly coplanar, we can always obtain for the same 3-D tangle a planar projection with parallel alignment of the sites and another planar projection with antiparallel sites.

To take into account the most general situation, we will assume that the recombination sites are not coplanar in 3-dimensional space, and hence they can be in parallel or antiparallel alignment in the tangle diagram, based on the direction in which the projection is taken. Biological evidence suggests that enzymes in the tyrosine family present a pseudo-planar conformation at the synapse where the sites are presented in anti-parallel alignment [44, 94]. However co-planarity, in a strict mathematical sense, is unlikely to occur in nature (*cf.* argument and references in [97]).

The assumption of non-coplanarity followed by the assumptions on P implies that, without loss of generality, P can be assumed to be (0) with parallel sites.

5.3. Solving the tangle equations and TangleSolve. In sum, the tangle method models a site-specific recombination event as a system of two or more tangle equations on 3 unknowns O, P and R. If $P = (0)$, $R = (k)$ for some integer k, or $R = (0, 0)$, then the system can be solved for O rational or sum of rational tangles, and for the integer k [37, 38, 39, 40]. Systems of tangle equations corresponding to processive and non-processive recombination have been studied extensively, and in many cases all possible solutions have been characterized [13, 14, 15, 16, 26, 37, 38, 39, 40, 46, 90, 96, 97, 98, 101]. Computing the solutions is not mathematically challenging but can be very tedious. In Saka and Vazquez [80], we introduced TangleSolve, a user-friendly computer implementation of the tangle method. TangleSolve is a java stand-alone program and web-based applet which offers a user-friendly interface for analyzing and visualizing recombination mechanisms. The program and documentation can be accessed from http://bio.math.berkeley.edu/TangleSolve/. This program is also reviewed and illustrated in Zheng *et al.* [101]. Based on the assumptions outlined in the previous section, TangleSolve finds rational and sums of rational tangle solutions to systems of equations arising from processive and non-processive recombination. TangleSolve computes only solutions that are rational or sum of rational tangles. In some cases, using tools from low-dimensional

topology, it can be proven that all possible solutions to a system of tangle equations are rational or sums of two rational tangles. Such instances are highlighted in TangleSolve.

It is our goal to make this program available to the wider scientific community of mathematicians, molecular biologists and computational biologists. Therefore the calculations rest heavily on a graphical interface. Mathematical notation, or deep knowledge of the tangle method are not required to insert substrate and product topologies and compute enzymatic mechanisms.

TopoICE-R is another computer implementation of the tangle model which is available through KnotPlot [29]. TangleSolve and TopoICE-R have complementary features, noteworthy TopoICE-R provides a 3D rendition of the tangle equations and benefits from all the capabilities within Knotplot . TangleSolve has the ability to compute solutions for processive recombination reactions. Also the two applications use different sets of assumptions.

6. Tangle analysis of Xer recombination.

6.1. Dimer resolution at *psi* sites. In Colloms *et al.* [22] it was shown that, when acting on unknotted DNA circles with two *psi* recombination sites in direct repeats, XerC/XerD yield products of unique topology b(4,3) (the right-hand 4-crossing torus catenane) and anti-parallel sites (Figure 8). The reaction requires two accessory proteins PepA and ArgR [12], and there is experimental evidence that the sites wrap around the accessory proteins approximately three times prior to recombination [1]. This reaction can be written as the following system of tangle equations:

$$N(O+P) = b(1,1),\ N(O+R) = b(4,3)$$

Using results on Dehn surgeries on strongly invertible Knots [49], one can show that O is rational [26, 97] and therefore all solutions to the XerCD-*psi* system of equations can be computed. If no assumptions are made on P and R, there are infinitely many rational solutions to the XerCD-*psi* system of equations, most of which are too complex to be biologically reasonable [26]. In [97] the assumptions on P and R stated in section 4.2 were used to extract biologically relevant solutions. Under these assumptions the solution set was dramatically reduced and relevant solutions were computed using tangle calculus [97]. After introducing experimental information on the relative orientation of the sites in the recombination product (RH 4-cat with anti-parallel sites), the following are the only two solutions for $P = (0)$ with parallel alignment, R integral, and O minimal:

$$S_1 :\ O = (-3, 0),\ R = (-1);$$
$$S_2 :\ O = (-5, 0),\ R = (+1).$$

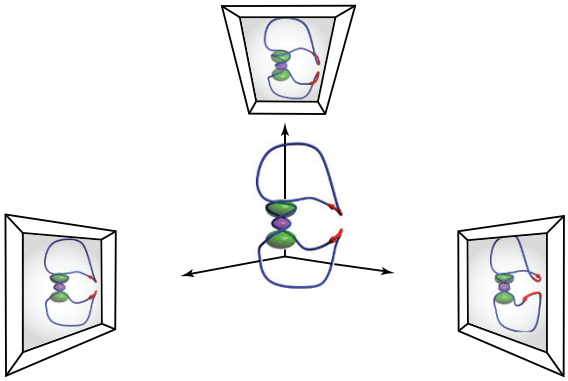

FIG. 11. *The center figure is a cartoon representation of the DNA substrate wrapped around the accessory proteins, and the two* psi *sites brought together in a pseudo-planar configuration by recombinases XerC/XerD. The surrounding images correspond to different planar projections of the center configuration. Each of these planar diagrams corresponds to one of the three tangle solutions for the Xer equations. From left to right these are:* $P = (0)$ *parallel and* $O = (-3, 0)$*;* $P = (0)$ *parallel and* $O = (-5, 0)$*; and* $P = (0)$ *anti-parallel and* $O = (-4, 0)$.

If P is assumed to be (0) with anti-parallel sites then, since XerC/XerD are tyrosine recombinases and recombine through a Holiday junction intermediate, the corresponding R is $(0, 0)$. In this case tangle calculus yields a unique solution:

$$S_3 : \ O = (-4, 0), \ R = (0, 0).$$

Solutions S_1, S_2 and S_3 are illustrated in Figure 11 left, top and right diagrams, respectively. The figure proposes a possible spatial relationship between the three solutions as outlined below.

It is worth noting that given a solution $S = \{O, P, R\}$ where O is rational, $P = (0)$ and $R = (+1)$ or $R = (-1)$, TangleSolve also displays a solution S' equivalent to S. S' is obtained by rotating the synapse so that $P = (0)$ becomes to $P = (0, 0)$, and by letting O have minimal number of crossings [80]. Interestingly, performing this simple transformation on S_1 and on S_2 yields the same solution $S' = \{P = (0, 0), O = (4), R = (0)\}$, where the sites in P are in anti-parallel alignment. Rotating S' in 3-space so that $P = (0)$ anti-parallel reveals S' to be the same as S_3.

In [97] we noted that there is a geometrical equivalence, obtained by rigid motion, between S_1 and S_3, and between S_2 and S_3. This equivalence suggests that a unique 3D representation of the synaptic complex, and topological mechanism of the enzymes, can be interpreted as three different tangle solutions when viewed from different spatial directions (see Figure 11). In [97] we presented a 3D cartoon model (as in Figure 11) and a 3D molecular model of XerCD/DNA which realized these spatial equivalences.

The molecular model was based on x-ray crystallographical data of the accessory proteins PepA and Arg R[88], and of the Cre/DNA complex [44, 94]. Cre is a tyrosine recombinase which shares high degree of homology with Xer.

This study indicates a limitation of the tangle model and suggests the need to consider equivalence classes of planar tangle diagrams related by 3D rigid motion (rotations and translations).

6.2. Unlinking by XerCD-FtsK. In [52] it was shown that XerCD-FtsK recombination at *dif* sites can unlink catenanes produced by λ-Int recombination with both parallel and anti-parallel sites (*cf* Figure 8). These results led to the hypothesis that *in vivo* XerCD-FtsK recombination at *dif* may work with topoIV to unlink catenanes produced by DNA replication. To test this hypothesis, supercoiled catenated plasmids with a *dif* site produced *in vivo* by replication in topoIV-deficient cells were incubated *in vitro* with XerCD-FtsK$_{50C}$. FtsK$_{50C}$ is a biochemically active form of FtsK [52]. In addition to catenanes with 2–14 crossings, a few dimeric knots were also extracted from the cells. The reaction, which was ATP dependent, efficiently produced unlinked circles. Experimental data suggested a stepwise reaction where crossings would be removed one at a time, thus converting catenanes into knots, into catenanes, iteratively until the two free circles would be obtained. A control experiment was done to demonstrate that XerCD-Ftsk$_{50C}$ recombination could convert knotted dimers of the type predicted above to free circles. These substrates were RH torus knots with two directly repeated *dif* sites produced by Cre-*loxP* recombination.

We used the tangle method to confirm that the recombination mechanism most consistent with the experimental data is one of stepwise unlinking where RH torus catenanes with parallel sites are converted to RH torus knots with directly repeated sites, and such knots are converted to RH torus catenanes until the reaction stops at two open circles. These results are summarized in Figure 12 and in [47]. The details of the mathematical analysis will be reported elsewhere [85].

Briefly, systems of tangle equations corresponding to the experimental data of Grainge et al. [47] were posed and solved using TangleSolve for tangles that are rational or sums of rational. For example, all possible systems of two equations converting a substrate of type RH 6-crossing torus catenane (6-cat) with parallel sites into a product with 5 or less crossings were considered. Under the tangle method assumptions outlined in section 4.2, only three solutions were found, all of which produced the RH 5-crossing torus knot with directly repeated sites. The solutions are shown in Figure 12. Iterating the mechanism shown in Figure 12.C. recombines the RH 6-cat into a RH 5-torus knot, into RH 4-cat, to + trefoil, to 2-cat, to unknot, to free circles, consistent with the experimental data (Figure 13).

7. Conclusion. In sections 3-5 we reviewed the tangle method for site-specific recombination, including certain computer implementations,

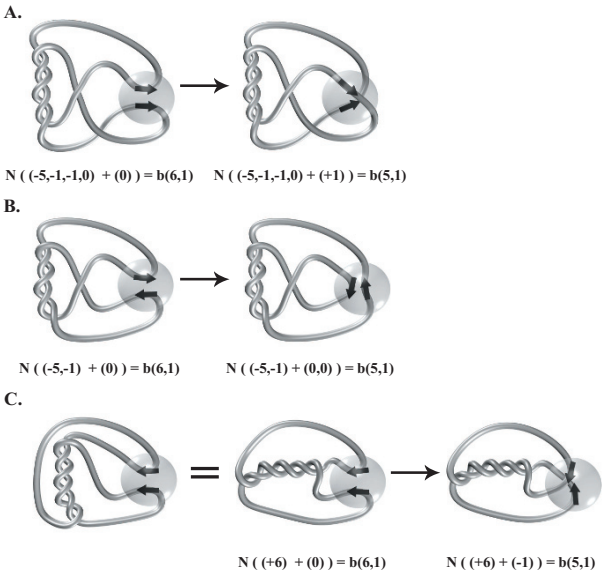

FIG. 12. *Solutions to the XerCD-FtsK tangle equations. Assuming a substrate of type RH 6-cat with parallel sites and a product knot or link with 5 or less crossings, we used TangleSolve to compute solutions that are rational or sum of rational tangles. We found that there are three biologically meaningful solutions, they are all rational and they produce the RH 5-crossing torus knot with sites in direct repeats. Panes A, B and C illustrate the three solutions. Each pane includes the synaptic complex before and after recombination, and shows how the P tangle is recombined into R. The three solutions are equivalent by rigid motion in 3 dimensions (results to be reported elsewhere [85]). Furthermore, iterating solution C results in the predicted gradual stepwise unlinking of the 6cat.*

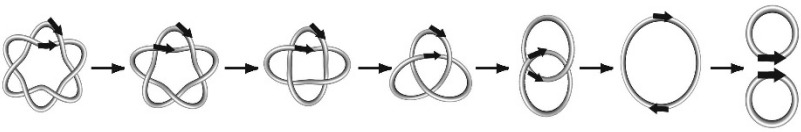

FIG. 13. *Stepwise unlinking by XerCD-*dif*-FtsK recombination.*

and we illustrated the method with the tangle analysis of Xer recombination. The tangle method is a powerful mathematical tool which uses contemporary pure mathematics to solve important biological questions. We highlighted the power of the method, as well as its limitations. In particular we pointed to the need to interpret the tangle results as 3-dimensional objects which may be related by rigid motions and thus equivalent from a biological (and geometrical) point of view.

Acknowledgements. The authors thank the following people: Jung Hun Koh for his help generating the spooling conformation, Rob Scharein for his assistance in generating figures using Knotplot, David Sherratt, Ian Grainge, Sean Colloms and Jonathan Bath for valuable discussions on the biology of Xer recombination, Kai Ishihara and Koya Shimokawa for their insights on the mathematical analysis of Xer. Special thanks go to De Witt L. Sumners for guiding the authors into this field. The three authors are very grateful to Nick Cozzarelli. All of them were fellows of the Program in Mathematics and Molecular Biology (PMMB). This program fostered the authors′ interdisciplinary interest. Y. Diao was inspired by the joint work of Nick Cozzarelli and De Witt L. Sumners on random knot theory. J. Arsuaga was a postdoctoral fellow with N. Cozzarelli and is grateful for the invaluable training and for his career advice. M. Vazquez is most grateful to Nick Cozzarelli for his vision of a world where topology could help biology (and vice versa), and for letting this vision materialize by promoting close interactions between mathematicians and molecular biologists.

REFERENCES

[1] C. Alén, D.J. Sherratt, and S.D. Colloms, *Direct interaction of aminopeptidase A with recombination site DNA in Xer site-specific recombination.*, EMBO J. **16** (1997), pp. 5188–5197.

[2] J. Arsuaga, T. Blackstone, Y. Diao, E. Karadayi, and M. Saito, *Linking of Uniform Random Polygons in Confined Spaces*, J. Physics A **40** (2007), pp. 1925–1936.

[3] J. Arsuaga, T. Blackstone, Y. Diao, E. Karadayi, and M. Saito, *Sampling Large Random Knots in a Confined Space*, J. Physics A **40** (2007), pp. 11697–11711.

[4] J. Arsuaga, R. Tan, M. Vazquez, D.W. Sumners, and S.C. Harvey, *Investigation of viral DNA packaging using molecular mechanics models*, Biophys. Chem. **101** (2002), pp. 475–484.

[5] J. Arsuaga, M. Vazquez, S. Trigueros, D.W. Sumners, and J. Roca, *Knotting probability of DNA molecules confined in restricted volumes: DNA knotting in phage capsids.* Proc. Natl. Acad. Sci. USA **99** (2002), pp. 5373–5377.

[6] J. Arsuaga, M. Vazquez, P. McGuirk, S. Trigueros, D.W. Sumners, and J. Roca, *DNA knots reveal a chiral organization of DNA in phage capsids*, Proc. Natl. Acad. Sci. USA **102** (2005), pp. 9165–9169.

[7] J. Arsuaga and Y. Diao, *DNA Knotting in Spooling Like Conformations in Bacteriophages*, Computational and Mathematical Methods in Medicine **9**(3) (2008), pp. 303–316.

[8] K. Aubrey, S. Casjens, and G. Thomas, *Secondary structure and interactions of the packaged dsDNA genome of bacteriophage P22 investigated by Raman difference spectroscopy*, Biochemistry **31** (1992), pp. 11835–11842.

[9] F.X. Barre and D.J. Sherratt, *Chromosome dimer resolution.* In The Bacterial Chromosome (Higgins, N.P., ed.), Washington, DC: ASM Press (2005), pp. 513–524.

[10] L. Black,W. Newcomb, J. Boring, and J. Brown, *Ion etching bacteriophage T4: support for a spiral-fold model of packaged DNA*, Proc. Natl. Acad. Sci. USA **82** (1985), pp. 7960–7964.

[11] T. Blackstone,P. McGuirck,C. Laing, M. Vazquez, J. Roca, and J. Arsuaga, *The role of writhe in DNA condensation*, Proceedings of International Workshop on Knot Theory for Scientific Objects. OCAMI Studies Volume **1** (2007). Osaka Municipal Universities Press; pp. 239–250.

[12] M. Bregu, D.J. Sherratt, and S.D. Colloms, *Accessory factors determine the order of strand exchange in Xer recombination at psi.*, EMBO J. **21** (2002), pp. 3888–3897.

[13] D. Buck and E. Flapan, *Predicting Knot or Catenane Type of Site-Specific Recombination Products*, J Molecular Biology **374**(5) (2007), pp. 1186-1199.

[14] D. Buck and E. Flapan, *A topological characterization of knots and links arising from site-specific recombination.*, J. Phys. A: Math. Gen. **40** (2007), pp. 12377–12395.

[15] D. Buck and C. Verjovsky-Marcotte, *Tangle-solutions for a family of DNA rearranging proteins.*, Math Proc Camb Phil Soc **139** (2005), pp. 59–80.

[16] D. Buck and C. Verjovsky-Marcotte, *Classification of Tangle Solutions for Integrases, A Protein Family that Changes DNA Topology.*, J. Knot. Theory Ramifications **16** (2007), pp. 969–995.

[17] G. Burde and H. Zieschang, *Knots.*, vol. 5, In de Gruytier Studies in Mathematics (Gabriel, P., ed.) Walter de Gruyter, Berlin., 1985.

[18] H. Cabrera Ibarra, *On the classification of rational 3-tangles*, J. Knot Theory Ramifications **12**(7) (2003), pp. 921-946.

[19] H. Cabrera Ibarra, *Results on the classification of rational 3-tangles*, J. Knot Theory Ramifications **13**(2) (2004), pp. 175–192.

[20] K. Cerritelli, N. Cheng, A. Rosenberg, C. McPherson, F. Booy, and A. Steven, *Encapsidated conformation of bacteriophage T7 DNA*, Cell **91** (1997), pp. 271–280.

[21] D.K. Chattoraj and R.B. Inman, *Location of DNA ends in P2, 186, P4 and lambda bacteriophage heads*, J. Mol. Biol. **87** (1974), pp. 11–22.

[22] S.D. Colloms, J. Bath, and D.J. Sherratt, *Topological selectivity in Xer site-specific recombination*, Cell **88** (1997), pp. 855–864.

[23] J.H. Conway, *An enumeration of knots and links, and some of their algebraic properties.*, Computational Problems in Abstract Algebra, Pergamon, Oxford, UK (1967), pp. 329–358.

[24] N.R. Cozzarelli, M.A. Kraznow, S.P. Gerrard, and J.H. White, *A topological treatment of recombination and topoisomerases.*, Cold Spring Harbor Symp. Quant. Biol. **49** (1984), pp. 383–400.

[25] N.J. Crisona, R.L. Weinberg, B.J. Peter, D.W. Sumners, and N.R. Cozzarelli, *The topological mechanism of phage lambda integrase.*, J. Mol. Biol. **289** (1999), pp. 747–775.

[26] I. Darcy, *Biological distances on DNA knots and links: applications to Xer recombination.*, J. Knot Theory Ramification **10** (2001), pp. 269–294.

[27] I.K. Darcy, J. Chang, N. Druivenga, C. McKinney, R.K. Medikonduri, S. Mills, J. Navarra-Madsen, A. Ponnusamy, J. Sweet, and T. Thompson, sl Coloring the Mu transpososome., BMC Bioinformatics **7** (2006), pp. 435.

[28] I.K. Darcy, J. Luecke, and M. Vazquez *Tangle analysis of difference topology experiments: applications to a Mu-DNA protein complex*, IMA preprint series, (2008), https://www.ima.umn.edu/preprints/oct2007/2177.pdf.

[29] I.K. Darcy and R.G. Scharein, *TopoICE-R: 3D visualization modeling the topology of DNA recombination.*, Bioinformatics **22**(14) (2006), pp. 1790–1791.

[30] Y. Diao, *The Knotting of Equilateral Polygons in* $\boldsymbol{R}^3$, Journal of Knot Theory and its Ramifications, **4**(2) (1995), pp. 189–196.

[31] Y. Diao, A. Dobay, R.B. Kusner, K. Millet, and A. Stasiak, *The Average Crossing Number of Equilateral Random Polygons* J. Physics A **36**(46) (2003), pp. 11561–11574.

[32] Y. Diao and C. Ernst, *The Average Crossing Number of Gaussian Random Walks and Polygons*, Physical and numerical models in knot theory, J.A. Calvo, K.C. Millett, E.J. Rawdon, and A. Stasiak, editors, Series on Knots and Everything **36** (2005), World Scientific, pp. 275–292.

[33] Y. Diao, J. Nardo, and Y. Sun, *Global Knotting in Equilateral Random Polygons*; Journal of Knot Theory and its Ramifications, **10**(4) (2001), pp. 597–607.

[34] Y. Diao, N. Pippenger, and D.W. Sumners, *On Random Knots*, Journal of Knot Theory and its Ramifications, **3**(3) (1994), pp. 419–429.

[35] W.C. Earnshaw and S.R. Casjens, *DNA packaging by the double-stranded DNA bacteriophages*, Cell **21** (1980), pp. 319–331.

[36] J. Emert and C. Ernst, *N-string tangles*, J. Knot. Theory Ramifications **9(8)**(2000), pp. 987–1004.

[37] C. Ernst, *Tangle equations*, J. Knot. Theory Ramifications **5**(1996), pp. 145–159.

[38] C. Ernst, *Tangle equations II*, J. Knot. Theory Ramifications **6** (1997), pp. 1–11.

[39] C. Ernst and D.W. Sumners, *A calculus for rational tangles: applications to DNA recombination*, Math. Proc. Cambridge Phil. Soc. **108** (1990), pp. 489–515.

[40] C. Ernst and D.W. Sumners, *Solving tangle equations arising in a DNA recombination model*, Math. Proc. Cambridge Phil. Soc. **126** (1999), pp. 23–36.

[41] O. Espeli and K.J. Marians,, *Untangling intracellular DNA topology*, Mol Microbiol **52** (2004), pp. 925–931.

[42] A. Evilevitch, L. Lavelle, C.M. Knobler, E. Raspaud, and W.M. Gelbart, *Osmotic pressure inhibition of DNA ejection from phage*, Proc. Natl. Acad. Sci. USA **100** (2003), pp. 9292–9295.

[43] J.R. Goldman and L.H. Kauffman, *Rational tangles*, Advan. Appl. Math. **18** (1997), pp. 300–332.

[44] D.N. Gopaul, F. Guo, and G.D. Van Duyne, *Structure of the Holliday junction intermediate in Cre–loxP site-specific recombination*, EMBO J. **17** (1998), pp. 4175–4187.

[45] S.C. Gourlay and S.D. Colloms, *Control of Cre recombination by regulatory elements from Xer recombination systems*, Mol. Microbiol. **52** (2004), pp. 53–65.

[46] I. Grainge, D. Buck, and M. Jayaram, *Geometry of site alignment during Int family recombination: antiparallel synapsis by the FLP recombinase*, J. Mol. Biol. **298** (2000), pp. 749–764.

[47] I. Grainge, M. Bregu, M. Vazquez, V. Sivanathan, S.C. Ip, and D.J. Sherratt, *Unlinking chromosomes catenated in vivo by site-specific recombination*, EMBO J **26**(19) (2007), pp. 4228–4238.

[48] B. Hallet and D.J. Sherratt, *Transposition and site-specific recombination adapting DNA cut-and paste mechanism to a variety of genetic rearrangements*, FEMS Microbiol. Rev. (1997), p. 21.

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

[50] V.F. Holmes and N.R. Cozzarelli, *Closing the ring: links between SMC proteins and chromosome partitioning, condensation, and supercoiling*, Proc. Natl. Acad. Sci. USA **97** (2000), pp. 1322–1324.

[51] N. Hud, *Double-stranded DNA organization in bacteriophage heads: an alternative toroid-based model*, Biophys. J. **69** (1995), pp. 1355–1362.

[52] S.C. Ip, M. Bregu, F.X. Barre, and D.J. Sherratt, *Decatenation of DNA circles by FtsK-dependent Xer site-specific recombination.*, EMBO J **22** (2003), pp. 6399–6407.

[53] P.J. Jardine and D.L. Anderson, *DNA packaging in double-stranded DNA phages* The bacteriophages (2006), Ed. Richard Calendar, Oxford University Press, pp. 49–65.

[54] R. Kanaar, A. Klippel, E. Shekhtman, J. M. Dungan, R. Kahmann, and N.R. Cozzarelli, *Processive recombination by the phage Mu Gin system: implications for the mechanisms of DNA strand-exchange, DNA site alignment, and enhancer action*, Cell **62** (1990), pp. 353–366

[55] V. Katritch, Bednar, D. Michoud, R. G. Scharein, J. Dubochet, and A. Stasiak, *Geometry and physics of knots*, Nature **384** (1996), pp. 142–145.

[56] E. Kellenberger, E. Carlemalm, J. Sechaud, A. Ryter, and G. Haller, *Considerations on the condensation and the degree of compactness in non-eukaryotic DNA-containing plasmas*, In Bacterial Chromatin: Proceedings of the Symposium "Selected Topics on Chromatin Structure and Function" (eds. C. Gualerzi and C. L. Pon), Springer, Berlin (1986), pp. 11–25.

[57] S. Kim and I.K. Darcy, *Topological analysis of DNA-protein complexes*, Included in this volume, Mathematics of DNA Structure, Function and Interactions (eds C.J. Benham, S. Harvey, W.K. Olson, D.W. Sumners and D. Swigon), Springer Science + Business Media, LLC, New York, (2009).

[58] K.V. Klenin, A.V. Vologodskii, V.V. Anshelevich, A.M. Dykhne, and M.D. Frank-Kamenetskii, *Effect of Excluded Volume on Topological Properties of Circular DNA,* J. Biomolec. Str. and Dyn. **5** (1988), pp. 1173–1185.

[59] J.C. LaMarque, T.L. Le, and S.C. Harvey, *Packaging double-helical DNA into viral capsids*, Biopolymers **73** (2004), pp. 348–355.

[60] A. Landy,*Coming or going its another pretty picture for the lambda-Int family album*, Proc. Natl Acad. Sci. USA **96** (1999), pp. 7122–7124.

[61] J. Lepault, J. Dubochet, W. Baschong, and E. Kellenberger, *Organization of double-stranded DNA in bacteriophages: a study by cryo-electron microscopy of vitrified samples* EMBO J. **6** (1987), pp. 1507–1512.

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

[63] L.F. Liu, J.L. Davis, and R. Calendar, *Novel topologically knotted DNA from bacteriophage P4 capsids: studies with DNA topoisomerases*, Nucleic Acids Res. **9** (1981), pp. 3979–3989.

[64] L.F. Liu, L. Perkocha, R. Calendar, and J.C. Wang, *Knotted DNA from bacteriophage capsids*, Proc. Natl. Acad. Sci. USA **78** (1981), pp. 5498–5502.

[65] J.P.J. Michels and F.W. Wiegel, *On the topology of a polymer ring*, Proc. R. Soc. London Ser A **403** (1986), pp. 269–284.

[66] C. Micheletti, D. Marenduzzo, E. Orlandini, and D.W. Sumners, *Knotting of random ring polymers in confined spaces*, J. Chem. Phys. **124** (2006), pp. 064903.1–10.

[67] K. Millett, *Knotting of regular polygons in 3-space*, Random knotting and linking (Vancouver, BC, 1993), World Sci. Publishing, Singapore (1994), pp. 31–46.

[68] K. Millett, *Monte Carlo Explorations of Polygonal Knot Spaces*, Knots in Hellas'98 (Delphi), Ser. Knots Everything **24** (2000), World Scientific, pp. 306–334.

[69] H.R. Morton, *Seifert circles and knot polynomials*, Math. Proc. Cambridge Phil. Soc. **99** (1986), pp. 107–109.

[70] K. Murasugi, *Knot Theory, Its Applications (Translated by B. Kurpita)*, Birkhauser, Boston, MA. 1996.

[71] S.E. Nunes-Duby, H.J. Kwon, R.S.T. Tirumalai, T. Ellenberger, and A. Landy, *Similarities and differences among 105 members of the Int family of site-specific recombinases*, Nucl. Acids Res. **26** (1998), pp. 391–406.

[72] A.S. Petrov, M.B. Boz, and S.C. Harvey, *The conformation of double-stranded DNA inside bacteriophages depends on capsid size and shape*, J Struct Biol. **160** (2007) pp. 241–248.

[73] P. Plunkett, M. Piatek, A. Dobay, J.C. Kern, K. Millet, A. Stasiak, and E. Rawdon, *Total curvature and total torsion of knotted polymers*, Macromolecules **40** (2007), pp. 3860–3867.

[74] L. Rayleigh, *On the problems of random vibrations, and of random flights in one, two, or three dimensions,* Phil. Mag. S. 6. **37**(220) (1919), pp. 321–347.
[75] D. Raymer and D. Smith, *Spontaneous knotting of an agitated string* Proc. Natl. Acad. Sci
[76] K. Richards, R. Williams, and R. Calendar, *Mode of DNA packing within bacteriophage heads*, J. Mol. Biol. **78** (1973), pp. 255–259.
[77] D. Rolfsen, *Knots* Mathematics Lecture Series **7**, Publish or Perish, Berkeley, CA., 1976.
[78] V.V. Rybenkov, N.R. Cozzarelli, and A.V. Vologodskii, *Probability of DNA knotting and the effective diameter of the DNA double helix,* Proc. Natl. Acad. Sci. USA **90** (1993), pp. 5307–5311.
[79] P.D. Sadowski, *Site-specific genetic recombination: hops, flips, and flops*, FASEB J. **7** (1993), pp. 760–767.
[80] Y. Saka and M. Vazquez, *TangleSolve: topological analysis of site-specific recombination*, Bioinformatics **18** (2002), pp. 1011–1012.
[81] J.B. Schvartzman and A. Stasiak, *A topological view of the replicon*, EMBO Rep. **5**(3) (2004), 256–261.
[82] P. Serwer, *Arrangement of double-stranded DNA packaged in bacteriophage capsids: An alternative model*, J. Mol. Biol. **190** (1986), pp. 509–512.
[83] S.Y. Shaw and J.C. Wang, *Knotting of a DNA chain during ring closure*, Science **260** (1993), pp. 533–536.
[84] Arciszewska, L.K. and D.J. Sherratt *Site-specific recombination and circular chromosome segregation*, Philos. Trans. R. Soc. Lond. B. Biol. Sci. **347** (1995), pp. 37–42.
[85] K. Shimokawa, K. Ishihara, I. Grainge, D.J. Sherratt, and M. Vazquez, *DNA unlinking by site-specific recombination: topological analysis of XerCD-FtsK action*, Preliminary report.
[86] W.M. Stark and M.R. Boocock, *Topological selectivity in site-specific recombination*, In Mobile Genetic Elements (Sherratt, D. J., ed.), IRL Press at Oxford University, Oxford (1995), pp. 101–129.
[87] W.M. Stark, D.J. Sherratt, and M.R. Boocock, *Site-specific recombination by Tn3 resolvase: topological changes in the forward and reverse reactions*, Cell **58** (1989), pp. 779–790.
[88] N. Strater, D.J. Sherratt, and S.D. Colloms, *X-ray structure of aminopeptidase A from Escherichia coli and a model for the nucleoprotein complex in Xer site-specific recombination.*, EMBO J. **18** (1999), pp. 4513–4522.
[89] D.K. Summers and D.J. Sherratt, *Multimerization of high copy number plasmids causes instability: ColE1 encodes a determinant essential for plasmid monomerization and stability*, Cell **36** (1984), pp. 1097–1103.
[90] D.W. Sumners, C. Ernst, N.R. Cozzarelli, and S.J. Spengler, *Mathematical analysis of the mechanisms of DNA recombination using tangles*, Quarterly Reviews of Biophysics **28** (1995), pp. 253–313.
[91] S. Trigueros, J. Arsuaga, M. Vazquez, D.W. Sumners, and J. Roca, *Novel display of knotted DNA molecules by two dimensional gel electrophoresis*, Nucleic Acids Research **29** (2001), e67.
[92] S. Trigueros and J. Roca, *Production of highly knotted DNA by means of cosmid circularization inside phage capsids*, BMC Biotechnol **7**(1) (2007), pp. 94.
[93] S. Tzill, J.K. Kindt, W.M. Gelbart, and A. Ben-Shaul, *Forces and Pressures in DNA Packaging and Release from Viral Capsids*, Biophys. J. **84** (2003), pp. 1616–1627.
[94] G.D. Van Duyne, *A structural view of Cre-loxP site-specific recombination*, Annu. Rev. Biophys. Biomol. Struct. **30** (2001), pp. 87–104.
[95] V. Vanhooff, C. Galloy, H. Agaisse, D. Lereclus, B. Revet, and B. Hallet, *Self-Control in DNA site-specific recombination mediated by the tyrosine recombinase TnpI*, Molecular Microbiology **60(3)** (2006), pp. 617–629.

[96] M. Vazquez, *Tangle analysis of site-specific recombination: Gin and Xer systems*, PhD dissertation in mathematics, Florida State University, Tallahassee, FL, 2000.

[97] M. Vazquez, S.D. Colloms, and D.W. Sumners, *Tangle analysis of Xer recombination reveals only three solutions, all consistent with a single 3-dimensional topological pathway*, J. Mol. Biol. **346** (2005), pp. 493–504.

[98] M. Vazquez and D.W. Sumners, *Tangle analysis of Gin site-specific recombination*, Math. Proc. Cambridge Phil. Soc. **136** (2004), pp. 565–582.

[99] A.V. Vologodskii, N.J. Crisona, B. Laurie, P. Pieranski, V. Katritch, J. Dubochet, and A. Stasiak, *Sedimentation and electrophoretic migration of DNA knots and catenanes*, J. Mol. Biol. **278** (1998), pp. 1–3.

[100] S.A. Wasserman, J.M. Dungan, and N.R. Cozzarelli, *Discovery of a predicted DNA knot substantiates a model for site-specific recombination*, Science **229** (1985), pp. 171–174.

[101] W. Zheng, C. Galloy, B. Hallet, and M. Vazquez, *The tangle model for site-specific recombination: a computer interface and the TnpI-IRS recombination system*, Knot Theory for Specific Objects, OCAMI studies **1**(2) (2007), pp. 251–271.

http://www.springer.com/978-1-4419-0669-4

Mathematics of DNA Structure, Function and Interactions

Benham, C.J.; Harvey, S.; Olson, W.K.; Sumners, D.W.; Swigon, D. (Eds.)

2009, XII, 356 p. 70 illus. in color., Hardcover
ISBN: 978-1-4419-0669-4