Modeling protein-DNA complexes with tangles

Isabel K. Darcy
Mathematics Department, University of Iowa
Iowa City, IA 52242, USA
idarcy@math.uiowa.edu

Abstract

The mathematics of tangles has been applied to model protein-DNA binding. A tangle consists of strings properly embedded in a 3-dimensional ball. The protein complex can be thought of as a 3D ball while the DNA segments bound by the protein complex can be thought of as strings embedded within the ball. This simple model can be used to determine the topology of protein-bound DNA. We review some tangle models and related software for modeling recombinase and topoisomerase action.

keywords: tangle; topoisomerase; recombinase; DNA knot; unknotting number

1. Introduction

A tangle consists of strings properly embedded within a 3-dimensional (3D) ball. Tangles were first used to model protein-DNA complexes by Ernst and Sumners [1]. Since then many others have also used tangles to model protein-DNA complexes (for example, [2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. We illustrate with an example [5, 6]. In this example our protein complex will be Mu transposase (Mu), but any protein which binds DNA can be used. Two different models of Mu bound to DNA are shown in Fig. 1. The grey ball represents Mu. The Mu protein complex binds to three DNA sequences. These segments of DNA are referred to as E, L, and R in Fig. 1. The grey ball combined with these three segments is a tangle. The older model shown in Fig. 1A was a common way to represent one stage in the Mu transposition reaction. Before [5], there was little information on how to draw the DNA within this protein-DNA complex, so normally very simple figures were drawn. Tangle analysis was used in [5] to show that a more complicated DNA configuration is more accurate. In reality the three DNA segments within the protein complex should be drawn containing about five crossings as shown in Fig. 1B.

The model in Fig. 1B was determined by solving tangle equations. Fig. 2A shows an example of one of these tangle equations. The light blue ball in Fig. 2A represents the unknown shape of DNA bound by Mu. We will call this tangle unknown $\mathbf{M}$. The black loops coming out of this light blue ball represent the double stranded DNA not bound within the Mu-DNA complex. If we remove the protein complex, then we can see that the DNA is a four crossing link. Hence we have the tangle equation: DNA configuration not bound by protein plus the unknown DNA configuration bound by protein equals a four crossing link. Solving the tangle equation in Fig. 2A for the tangle unknown $\mathbf{M}$ is equivalent to determining possible topological conformations of the DNA bound by Mu. Observe that the five crossing tangle in Fig. 2B is a solution to the tangle equation in Fig. 2A as shown in Fig. 2C. The three crossing tangle

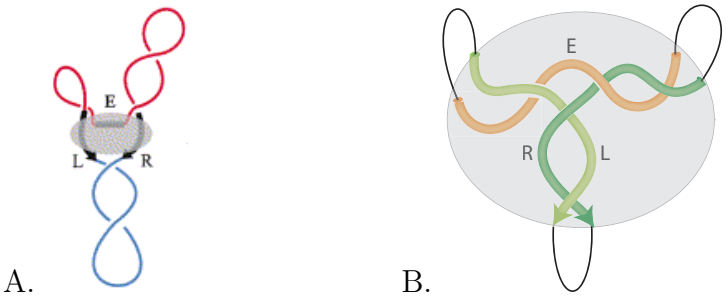

Figure 1: Two models of Mu bound to DNA. A) A model before tangle analysis was applied (from Fig. 1 in [5]). B.) A model resulting from tangle analysis [5]. Figure courtesy of Rasika Harshey and Makkuni Jayaram (grey circle added).

in Fig. 2D is not a solution as shown in Fig. 2E. In fact, the equation in Fig. 2A implies that any solution must have at least four crossings. We leave it to the reader to find other solutions to this tangle equation with one tangle unknown.

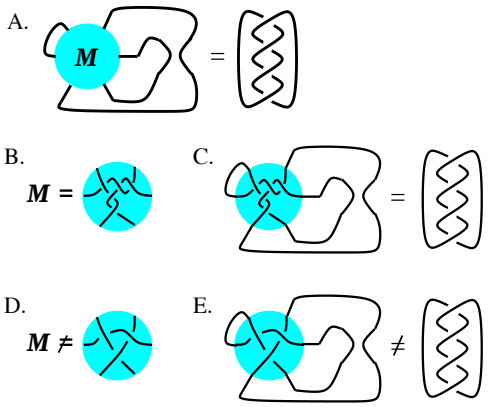

Figure 2: A.) An example of a tangle equation. B.) A five crossing tangle solution to this tangle equation. C.) We plug this five crossing tangle into the tangle equation. Since the equality holds, this five crossing tangle is a solution. D.) A three crossing tangle which is not a solution to this equation. E.) When we plug this three crossing tangle into the tangle equation, we see that the equality does not hold. Hence this three crossing tangle is not a solution.

The equations in this paper are all topological. Two objects are topologically equivalent if one can be smoothly deformed into the other (breaking and/or puncturing the objects is not allowed). For example, in Fig. 2C, the conformation on the left-side of the equality has five crossings. However, if we remove the protein complex, one of these crossings can be removed. The other four crossings cannot be eliminated without breaking the DNA. Hence the five crossing conformation on the left-side of the equality in Fig. 2C is topologically equivalent to the four crossing link on the right-side. Similarly our solutions are topological. We do not distinguish between two conformations that are geometrically different as long as they are topologically equivalent. Hence we will consider the two tangles in Fig. 3A and B to be the same.

When determining tangle equivalence, the endpoints of the strings must remain in their fixed locations on the boundary of the ball. Hence the tangle in Fig. 3C is topologically different than the tangle in Fig. 3A, B.

The main focus of this paper will be a review on using 2-string tangles to model proteins which can change circular DNA topology, but we will give a brief introduction near the end on how tangles are now being used to model more general protein-DNA complexes such as in Figs 1, 2. For some protein-DNA complexes, there are an infinite number of mathematically possible tangle models, the vast majority of which are not biologically plausible. We review software for finding, visualizing, and manipulating these models. We will start with an introduction to 2-string tangles (section 2). Topoisomerases and recombinases are two classes of proteins that will knot and catenate circular DNA. In section 3 we will model topoisomerase action, while in section 4 we will discuss recombination. In sections 3 and 4, we will also review applicable software.

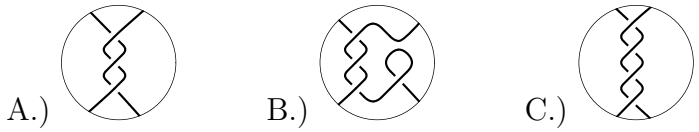

Figure 3: Equivalent and non-equivalent tangles. The tangles in (A) and (B) are geometrically different, but since one can be deformed into the other without breaking the strings or moving the endpoints of the strings, they are topologically equivalent. We will not distinguish between topologically equivalent tangles. The tangle in (C) is topologically different from the tangles in (A) and (B) since we are not allowed to cut and rejoin the strings or move the string endpoints.

2. Introduction to 2-string tangles

In this section, we define mathematical notation used in many 2-string tangle equations. However in the following sections, we will normally provide figures corresponding to any tangle notation used. Hence the reader is welcome to skim (or skip) this section and refer back as needed. For a more detailed introduction to tangles see [12, 13, 14, 15, 16]

Some examples of 2-string tangles are shown in Fig. 4. Since a string has two endpoints, a 2-string tangle has four endpoints. We will refer to these four endpoints as NW, NE, SW, and SE as indicated in Fig. 4. The zero tangle is the tangle containing no crossings where NW is connected to NE. The infinity tangle is the tangle containing no crossings where NW is connected to SW.

Many of the tangles involved in tangle equations modeling biological reactions belong to the class of tangles called rational. A tangle is rational if it can be obtained from the zero tangle by twisting the strings while allowing the endpoints of the string to move on the boundary of the 3-ball as shown in Fig 5. We can encode the shape of such a tangle by recording these twists. For example to form the rational tangle (2, 3, 4), we first add two horizontal half-twists to form the (2) tangle. We then add three vertical half twists to form the (2, 3, 0) tangle. This notation requires that

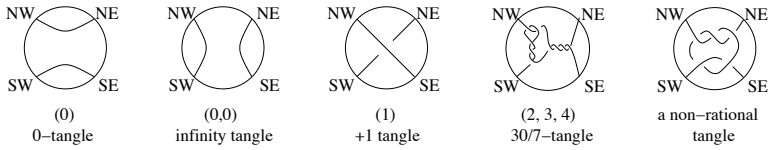

Figure 4: Some 2-string tangles.

we always end with number of horizontal twists. Since this tangle ends with zero horizontal twists (we just added three vertical twists), we have formed the (2, 3, 0) tangle. We can then add four horizontal half twists to form the (2, 3, 4) tangle. For a general tangle, $(c_1, ..., c_n)$, we will take n to be odd. The ith term represents c_i half twists which if i is odd, are horizontal and righthanded if $c_i > 0$ or lefthanded if $c_i < 0$; while if i is even, the c_i half twists are vertical and lefthanded if $c_i > 0$ or righthanded if $c_i < 0$.

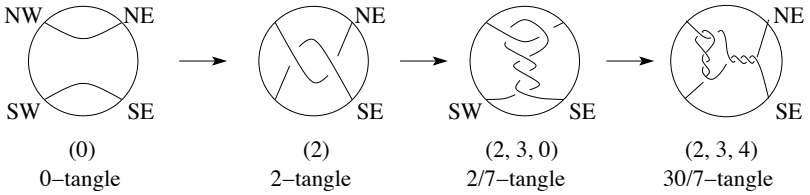

Figure 5: Any rational tangle can be obtained from the zero tangle by applying an alternating sequence of horizontal and vertical twists.

A rational tangle, $(c_1, ..., c_n)$, is uniquely identified by its continued fraction, $c_n + \frac{1}{c_{n-1}+...\frac{1}{c_2+\frac{1}{c_1}}}$ [12]. For example the two tangles drawn in Fig. 6 are equivalent since

$$(2,3,4) \;\rightarrow\; 4+\frac{1}{3+\frac{1}{2}} = \frac{30}{7} = 3+\frac{1}{1+\frac{1}{-4+\frac{1}{-1+\frac{1}{-1}}}} \;\leftarrow\; (-1,-1,-4,1,3)$$

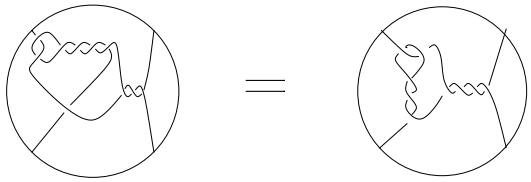

Figure 6: These two tangles are equivalent since their corresponding continued fractions are the same.

Knots and links can be formed from tangles by connecting the endpoints of the tangles. For example, given a tangle $\mathbf{A}$, its numerator closure, $N(\mathbf{A})$, is formed by

connecting the NW endpoint to the NE endpoints and connecting the SW endpoint to the SE endpoints as shown in Figs. 7, 8. Given a rational tangle, $\frac{\mathbf{a}}{\mathbf{b}}$, its numerator closure, $N(\frac{\mathbf{a}}{\mathbf{b}})$, is called a rational knot or link (also called 4-plat or 2-bridge).

Figure 7: Numerator closure of a tangle, $N(\mathbf{A})$, where $\mathbf{A}$ is an arbitrary tangle.

We can easily determine when two rational knots/links $N(\frac{\mathbf{a}}{\mathbf{b}})$ and $N(\frac{\mathbf{c}}{\mathbf{d}})$ are the same. Take $a, c \geq 0$. Then $N(\frac{\mathbf{a}}{\mathbf{b}}) = N(\frac{\mathbf{c}}{\mathbf{d}})$ if and only if $a = c$ and $bd^{\pm 1} = 1 \bmod a$ [17, 18]. For example, $N((2,3,0)) = N(0 + \frac{1}{3+\frac{1}{2}}) = N(\frac{\mathbf{2}}{\mathbf{7}})$ and $N((2)) = N(\frac{\mathbf{2}}{\mathbf{1}})$. Thus, $N((2,3,0)) = N(\frac{\mathbf{2}}{\mathbf{7}}) = N(\frac{\mathbf{2}}{\mathbf{1}}) = N((2))$ since $7 = 1 + 2(3) = 1 \bmod 2$ (see also Fig. 8).

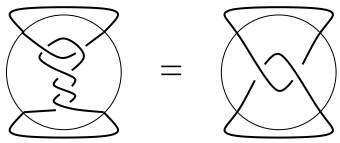

Figure 8: $N((2,3,0)) = N((2))$

We can add two tangles, $\mathbf{A}$ and $\mathbf{B}$ by connecting the NE and SE endpoints of $\mathbf{A}$ to the NW and SW endpoints of $\mathbf{B}$, respectively to form the tangle $\mathbf{A} + \mathbf{B}$ as shown in Fig. 9.

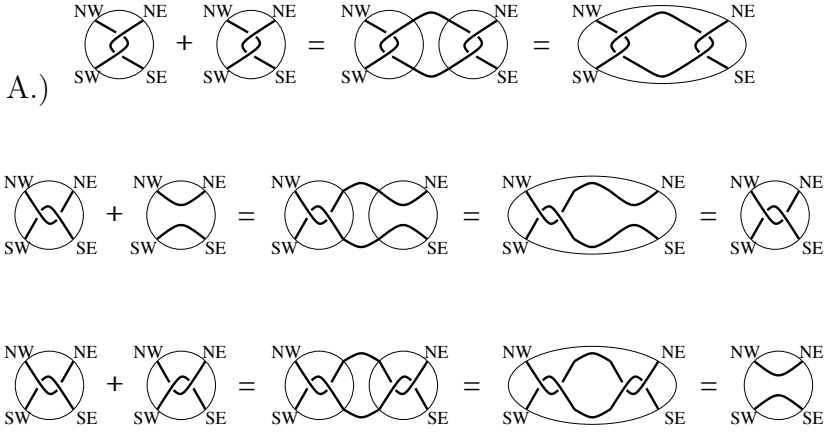

Figure 9: Adding tangles A.) Adding two tangles can result in the creation of a circular component B.) Observe that the zero tangle acts like an additive identity, $\mathbf{2} + \mathbf{0} = \mathbf{2}$. C.) $\mathbf{2} + \mathbf{-2} = \mathbf{0}$. If we add the $\mathbf{2}$ tangle and the $\mathbf{-2}$ tangle, we obtain the zero tangle (most tangles, however, do not have additive inverses).

3. Modeling topoisomerase action

Type II topoisomerases are proteins which cut one double-stranded DNA segment, allowing a second DNA segment to pass through before resealing the break [19]. This results in a crossing change as shown in Fig. 10. Topoisomerase is represented by the dotted green circle. On the left, topoisomerase is shown binding a three crossing knot, the trefoil knot. Topoisomerase performs a crossing change on this knot resulting in the unknot shown on the right (the unknot is the knot which can be drawn without any crossings – although here it is drawn with three removable crossings). Type II topoisomerases can knot and catenate double-stranded DNA [20]. Type I topoisomerases only break one strand of DNA. Type I topoisomerases change how the two DNA strands in double-stranded DNA are linked. They can also knot and catenate single-stranded DNA or nicked double-stranded DNA [21, 22].

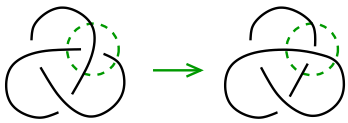

Figure 10: A model of a topoisomerase changing a DNA crossing. The green dotted circle represents topoisomerase.

In cells, topoisomerases are responsible for keeping DNA unknotted, unlinked, and properly supercoiled. They are required for DNA replication and transcription [23, 24]. Cancer cells replicate more rapidly than healthy cells. Thus topoisomerases are the targets of many anti-cancer drugs [25].

Topoisomerase action can be modeled using tangles. For mathematical convenience, we draw a dotted circle around the DNA not bound by topoisomerase (Fig. 11A). We redraw Fig. 11A as formal system of tangle equations in Fig. 11B. The top equation in Fig. 11B corresponds to the knotted substrate configuration before topoisomerase has acted. The bottom equation in Fig. 11B corresponds to the unknotted product configuration after topoisomerase has acted.

To model topoisomerase action, we solve the tangle equations $N(\mathbf{U_f} + \mathbf{1}) =$ substrate topology, $N(\mathbf{U_f} + (-\mathbf{1})) =$ product topology (Fig. 12). Substrate topology refers to the knot/link type of the starting configuration of the circular DNA which will be acted upon by topoisomerase. Product topology refers to the knot/link type of the circular DNA after it has been acted upon by topoisomerase. The tangle $\mathbf{U_f}$ refers to the DNA configuration outside of the topoisomerase action. In this case, we know what the protein does: topoisomerase changes DNA crossings. Hence our tangle unknown corresponds to the DNA not bound by protein. Determining the configuration of unbound DNA could be important for solving some topoisomerase mysteries. For example, some topoisomerases will unknot DNA well below thermodynamic equilibrium [26]. There is also some evidence that for some knotted substrates, topoisomerase can differentiate between which crossing changes will most efficiently change a knot into an unknot [27]. Although some models have been proposed [26, 28, 29, 30, 31, 32, 33], it is unclear how topoisomerase action can have such

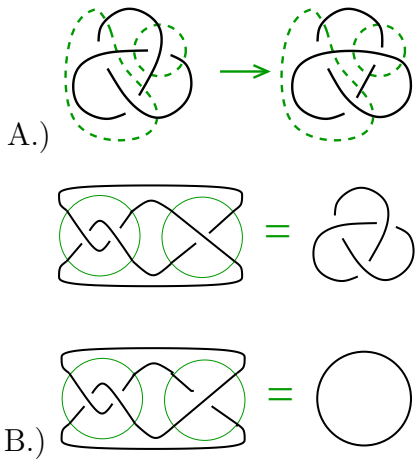

Figure 11: A.) A model of a topoisomerase changing a DNA crossing. In addition to the green circle modeling the topoisomerase protein, we have also drawn a dotted green circle around the DNA not bound by protein. B.) System of tangle equations corresponding to (A).

an effect on global topology when it only has local information as it performs DNA crossing changes.

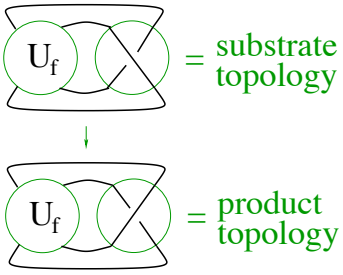

Figure 12: To model topoisomerase action, we solve the system of tangle equations, $N(\mathbf{U_f} + \mathbf{1}) =$ substrate topology, $N(\mathbf{U_f} + (-\mathbf{1})) =$ product topology for the tangle unknown $\mathbf{U_f}$, which represents the DNA conformation not bound by topoisomerase.

3.1. Mathematics and software for modeling topoisomerase action

We can completely solve the system of topoisomerase tangle equations in Fig. 12 if the knots/links involved in these equations belong to the rational family [34]. The mathematical results in [34] rely on [35, 36]. These results are implemented in TopoICE-X (Topological Interactive Construction Engine-X) [37]. TopoICE-X is a software tool within KnotPlot specifically designed for modeling topoisomerase action using tangles. KnotPlot is free software for visualizing and manipulating knots in 3D [38]. Thus one can use TopoICE-X to solve tangle equations and visualize and manipulate these solutions in 3D. KnotPlot can be downloaded at www.knotplot.com/download. Instructions for TopoICE-X can also be found at this website.

We can define a distance between two knots/links based upon topoisomerase action. Let the distance between K_1 and K_2, $d(K_1, K_2)$, be the minimum number of crossing changes needed to convert K_1 into K_2. This gives a mathematical lower bound for the minimum number of times topoisomerase needs to act to convert K_1 into K_2. Tables giving these distances are available in [39] for rational knots and in [40] for rational knots/links. In particular $d(K_1, K_2) > 1$ if and only if the system of topoisomerase equations, $N(\mathbf{U_f} + \mathbf{1}) = K_1$, $N(\mathbf{U_f} + (-\mathbf{1})) = K_2$ (Fig. 12), has no solution. TopoICE-X can be used to find all rational knots of distance one from a given rational knot. TopoICE-X will also output reaction pathways involving rational knots. The distance, $d(K_1, K_2)$, is a generalization of the unknotting number of a knot (= the minimum number of crossing changes needed to change a knot into the unknot). Hence any mathematics developed for calculating unknotting number can be useful for determining the minimum number of times topoisomerase acts. A quick search on MathSciNet indicates there are over 100 publications related to unknotting number.

4. Modeling recombination

Recombinases are proteins which cut two segments of DNA and interchange the ends resulting in the inversion or the deletion/insertion of a DNA segment. See Fig. 13. The two arrows represent binding sites for the recombinase. The recombinase protein complex binds to both these DNA segments, cuts both of them, and changes their connectivity by interchanging the cut ends before resealing the breaks. If the two DNA sequences are directly repeated on a circular DNA molecule as shown in Fig. 13A, then the result is a 2-component link. This is referred to as a deletion: part of the DNA sequence has been removed from the original circular DNA. If the two DNA recombinase binding sites start on two separate pieces of DNA, then recombination results in a joining of these two components. This is called an insertion: one piece of DNA has been inserted into another piece of DNA (read Fig. 13A from right to left). If the two DNA sequences are inversely repeated on a circular DNA molecule as shown in Fig. 13B, then the number of components doesn't change. Instead, the DNA sequence changes. Part of the DNA sequence is inverted with respect to the other part. Hence this is called an inversion. Recombinases are involved in gene rearrangement, viral integration, gene regulation, gene therapy, and in many more processes.

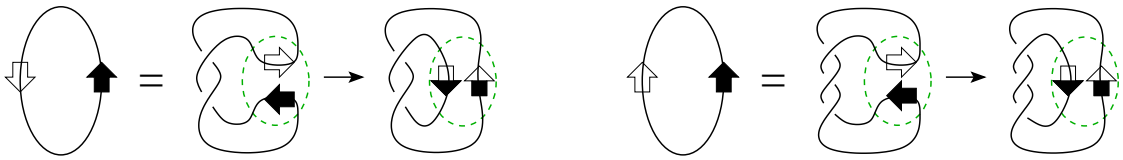

Figure 13: A.) directly repeated recombination binding sites resulting in a 2 component link B.) inversely repeated recombination binding sites resulting in a knot

There are a variety of recombinases, and they use a variety of mechanisms. Cre

recombination has been modeled as in Fig. 13 [41] while other recombinases are believed to bind several DNA crossings [1, 42, 2, 10]. the recombinase. We generally do not know the topology of the DNA bound within the protein complex either before or after recombination changes the DNA topology. Let the tangle $\mathbf{B}$ represent the DNA conformation bound within the recombinase protein complex before recombination (for example in Fig. 13, $\mathbf{B} = \mathbf{0}$). Note that in Fig. 13, recombination changes the topology of the protein-bound DNA. Let the tangle $\mathbf{E}$ represent the new protein-bound DNA conformation after recombination ($\mathbf{E}$ = infinity tangle in Fig. 13). Hence recombination is modeled by replacing the tangle $\mathbf{B}$ with the tangle $\mathbf{E}$. As before, we draw a circle around the DNA not bound by protein and represent the unbound DNA configuration with the tangle unknown $\mathbf{U_f}$. In general, we are solving the tangle equations $N(\mathbf{U_f} + \mathbf{B})$ = substrate topology, $N(\mathbf{U_f} + \mathbf{E})$ = product topology (Fig. 14). Hence we have a system of two equations with three unknowns, $\mathbf{U_f}$, $\mathbf{B}$, and $\mathbf{E}$. If we place no restrictions on these unknowns, this system of equations will always have an infinite number of solutions. Hence analyzing these solutions can be quite challenging. We illustrate with an example: Xer recombination.

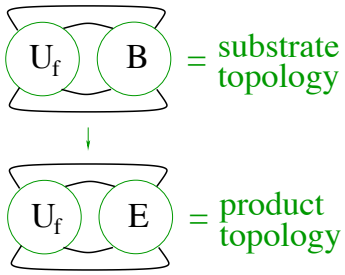

Figure 14: To model recombination, we solve the system of tangle equations, $N(\mathbf{U_f} + \mathbf{B})$ = substrate topology, $N(\mathbf{U_f} + \mathbf{E})$ = product topology. The tangle $\mathbf{U_f}$ represents the DNA conformation not bound by protein. The tangles $\mathbf{B}$ and $\mathbf{E}$ represent the DNA conformation bound by protein before and after recombination, respectively.

When acting on unknotted circular DNA, Xer recombination produces the four crossing link, $N(\frac{\mathbf{4}}{\mathbf{1}})$ [42]. Thus we have the system of equations in Fig. 15 where the tangles $\mathbf{B}$ and $\mathbf{E}$ represent the protein-bound DNA conformation before and after recombination, respectively, and the tangle $\mathbf{U_f}$ represents the unbound DNA topology. We would like to solve the system of equations in Fig. 15 for the three tangle unknowns, $\mathbf{B}$, $\mathbf{E}$, $\mathbf{U_f}$. There is much that we do not know about solutions for this system of tangle equations except that there are an infinite number of solutions as demonstrated below. In fact $\mathbf{B}$ can any rational tangle. For simplicity, let us assume that $\mathbf{B} = \frac{\mathbf{0}}{\mathbf{1}}$ and $\mathbf{E}$ is rational. Under these assumptions, $\mathbf{E} = \frac{\mathbf{3-4i}}{\mathbf{1+4i-k(3-4i)}}$, $\frac{\mathbf{1}}{\mathbf{k}}$, $\frac{\mathbf{3}}{\mathbf{1+3k}}$, or $\frac{\mathbf{5}}{\mathbf{1+5k}}$. where k and i can be any integers by [34]. In subsection 4.2, **Software for modeling recombination**, we will discuss software for finding and exploring these solutions.

Let us explore the case when $\mathbf{B} = \frac{\mathbf{0}}{\mathbf{1}}$ and $\mathbf{E} = \frac{\mathbf{9}}{\mathbf{5}}$ ($i = 3,\ k = -2$ in $\frac{\mathbf{3-4i}}{\mathbf{1+4i-k(3-4i)}}$). If we solve, $N(\mathbf{U_f} + \mathbf{0})$ = unknot, $N(\mathbf{U_f} + \frac{\mathbf{9}}{\mathbf{5}}) = N(\frac{\mathbf{4}}{\mathbf{1}})$, we find $\mathbf{U_f} = \mathbf{-1}$ (Fig. 16A). It is highly doubtful that Xer is capable of transforming a zero crossing tangle into the

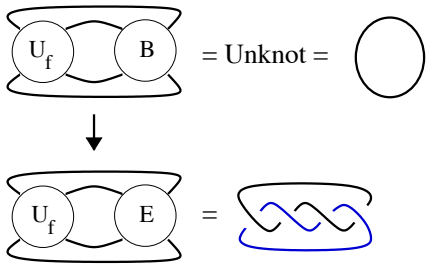

Figure 15: A system of tangle equations modeling Xer recombination: $N(\mathbf{U_f} + \mathbf{B}) =$ unknot, $N(\mathbf{U_f} + \mathbf{E}) = N(\frac{\mathbf{4}}{\mathbf{1}})$,

$\frac{\mathbf{9}}{\mathbf{5}}$ tangle. However this equation can be transformed into an equation that is a very likely model for Xer recombination. By pushing crossings into (or out of) $\mathbf{B}$, we can convert $\mathbf{B}$ into any rational tangle. The model in Fig. 16B is topologically equivalent to the model in Fig. 16A. In both Figs. 16A and 16B, we have the tangle equations, $N(-\mathbf{1} + \mathbf{0}) =$ unknot, $N(-\mathbf{1} + \frac{\mathbf{9}}{\mathbf{5}}) = N(\frac{\mathbf{4}}{\mathbf{1}})$. The only difference between Figs. 16A and 16B is that $\mathbf{B} = \mathbf{0}$ and $\mathbf{E} = \frac{\mathbf{9}}{\mathbf{5}}$ are both drawn with more crossings than necessary in Fig. 16B. For example, the tangle $\mathbf{B} =$ the zero tangle is drawn with six crossings in Fig. 16B. We can push some of these extraneous crossings from $\mathbf{B}$ and $\mathbf{E}$ into $\mathbf{U_f}$ (Fig. 16C). After simplifying the drawings in Fig. 16C, we obtain the tangle model in Fig. 16D. This model has been proposed for Xer recombination [42].

Given any recombinase tangle equation where $\mathbf{B}$ is some given rational tangle, we can move crossings between tangles and change $\mathbf{B}$ into any other rational tangle. Hence the system of tangle equations in Fig. 14 will always have an infinite number of solutions. Since $\mathbf{B}$ can be any rational tangle, we generally solve the system of equations in Fig. 14 for the tangles $\mathbf{U_f}$ and $\mathbf{E}$ in terms of $\mathbf{B}$ (and the substrate and product topologies). Since the complexity of the protein-bound DNA topology is limited, we will only consider solutions where $\mathbf{B}$ and $\mathbf{E}$ are rational as discussed below.

Recall that these models are topological. For example, the model in Fig 17 [43] is topologically equivalent to the model in Fig. 16D. In fact, both the model in Fig. 16D and Fig 17 could be different projections of the same 3D conformation [3]. Our tangle solutions are generally given as 2-dimensional projections of the 3-dimensional conformation. See http://bio.math.berkeley.edu/xeranim/ [3] for a possible 3-dimensional model of Xer recombination.

4.1. Assumptions in the tangle model

This tangle model makes two assumptions:

- The protein binds two segments of DNA (modeled by the 2-string tangle $\mathbf{B}$).
- The protein changes the topology within the tangle $\mathbf{B}$ leaving the topology of the unbound DNA (modeled by the tangle $\mathbf{U_f}$) unchanged.

In actuality, the changes brought about by the protein could have an effect on the unbound DNA. During formation of the protein-DNA complex, DNA twists or

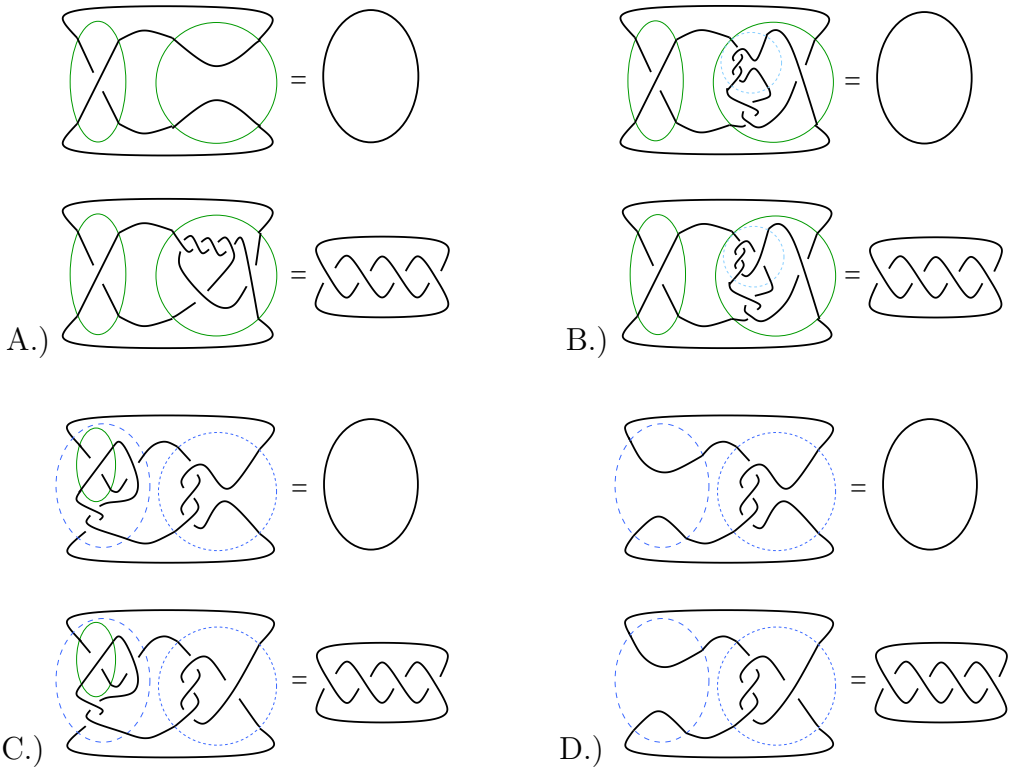

Figure 16: A.) $N(-\mathbf{1}+\mathbf{0})$ = unknot, $N(-\mathbf{1}+\frac{\mathbf{9}}{\mathbf{5}}) = N(\frac{\mathbf{4}}{\mathbf{1}})$. B.) $N(-\mathbf{1}+\mathbf{0})$ = unknot, $N(-\mathbf{1}+\frac{\mathbf{9}}{\mathbf{5}}) = N(\frac{\mathbf{4}}{\mathbf{1}})$. In this case both the zero tangle and the $\frac{\mathbf{9}}{\mathbf{5}}$ tangle are drawn with more crossings than necessary. C.) We push some of the extraneous crossings in the equations in (B) from $\mathbf{B}$ and $\mathbf{E}$ into $\mathbf{U_f}$. Hence we now have the tangle equations, $N(\mathbf{0}+\frac{\mathbf{-1}}{\mathbf{3}})$ = unknot $N(\mathbf{0}+\frac{\mathbf{-4}}{\mathbf{3}}) = N(\frac{\mathbf{4}}{\mathbf{1}})$. D.) The tangle equations, $N(\mathbf{0}+\frac{\mathbf{-1}}{\mathbf{3}}) =$ unknot $N(\mathbf{0}+\frac{\mathbf{-4}}{\mathbf{3}}) = N(\frac{\mathbf{4}}{\mathbf{1}})$, in simplified form.

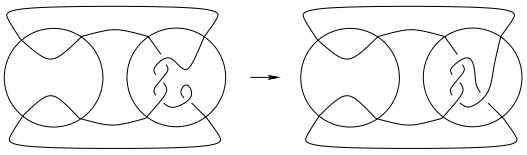

Figure 17: $N(\mathbf{0}+\frac{\mathbf{-1}}{\mathbf{3}})$ = unknot $N(\mathbf{0}+\frac{\mathbf{-4}}{\mathbf{3}}) = N(\frac{\mathbf{4}}{\mathbf{1}})$.

crossings could be introduced into the unbound DNA as the DNA sites come together; or changes to the DNA topology caused by the protein action may be transferred to outside the protein complex. The tangle model is a model. But by finding solutions for tangle equations modeling protein action, one can later decide how to manipulate the solutions to better model the real action.

If additional tangle equations are available, one generally tries to prove that the tangles **B** and **E** are rational. But with the system of only two equations with three unknowns in Fig. 14, this is not possible. In order to reduce the number of solutions and to make the problem mathematically tractable, we will assume

- **B** and **E** are rational tangles.

In most cases, the length of DNA bound within a protein-DNA complex is fairly short. Hence complicated tangles are unlikely to model protein-DNA complexes. The smallest non-rational tangle with two strings and no circular components has five crossings. Rational tangles also have the property that they can be formed via a sequence of twists from the zero tangle. Also, one can push the strings of a rational tangle to lie on the boundary of the 3-ball in which they are embedded so that the strings do not cross each other. Neither of these properties holds for non-rational tangles. Hence rational tangles are more likely models for protein-DNA complexes when considering the manner in which DNA sites may be brought together or how DNA may wrap around a protein complex.

Additional assumptions are needed to further reduce the number of solutions. In the original tangle model of Ernst and Sumners [1], the protein-bound DNA is partitioned into the sum of two tangles. It is assumed that the tangle $\mathbf{B} = \mathbf{U_b} + \mathbf{P}$ and the tangle $\mathbf{E} = \mathbf{U_b} + \mathbf{R}$ (Fig. 18). The tangle $\mathbf{U_b}$ represents the portion of the protein-bound DNA which is not changed by the protein action. The local action within the protein complex is modeled by replacing the tangle **P** by the tangle **R**.

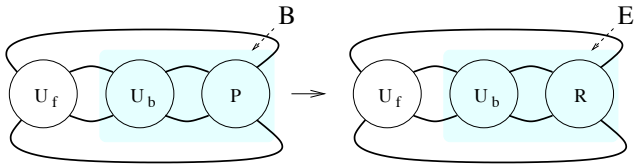

Figure 18: Ernst and Sumners tangle model from [1]:

Sometimes one looses potentially relevant solutions by adding additional assumptions. For example, to find a conformation like that in Fig. 19, one needs a more general tangle model like the one in Fig. 20 [8]. The model in Fig. 18 assumes that the tangles **P** and **R** modeling the local action can be separated from the remaining protein-bound DNA (represented by $\mathbf{U_b}$) via a tangle sum. This is not the case in Fig. 19. In order to reduce the number of solutions to a manageable size, assumptions often must be made. But the solutions found under these assumptions can then be manipulated to find other solutions which are not restricted by the additional assumptions.

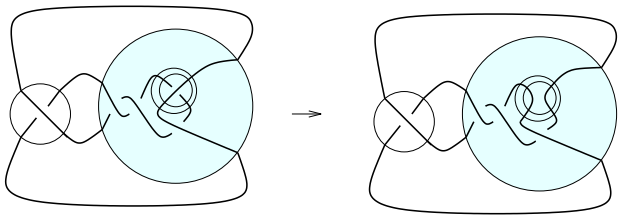

Figure 19: A solution using the model in Fig. 20. The local action is modeled by replacing the **-1** tangle with the infinity tangle.

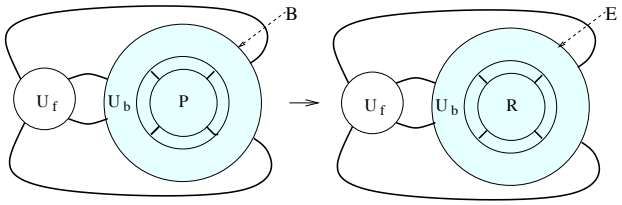

Figure 20: More general model

One can also reduce the number of solutions with additional experimental data. For example, AFM has been used to experimentally determine $\mathbf{U_f}$ [4]. Additional tangle equations obtained by using a different substrate [43] or as a result of processive or distributive recombination can eliminate some solutions. If a protein remains bound as it acts multiple times, then the protein acts processively. Processive recombination has been modeled by the system of equations [1]:

$$N(\mathbf{U_f} + \mathbf{U_b} + \mathbf{P}) = \ substrate$$

$$N(\mathbf{U_f} + \mathbf{U_b} + \mathbf{R}) = \ product\ 1$$

$$N(\mathbf{U_f} + \mathbf{U_b} + \mathbf{R} + \mathbf{R}) = \ product\ 2$$

$$N(\mathbf{U_f} + \mathbf{U_b} + \mathbf{R} + ... + \mathbf{R}) = \ product\ n$$

Observe that the unbound DNA represented by $\mathbf{U_f}$, does not change between protein action since the protein complex remains bound to the DNA. A system of processive tangle equations will sometimes result in a unique solution [1, 36, 44, 45, 10]. Another model for processive recombination is given in [46].

We call the action distributive if the protein releases the DNA between actions. Since the DNA is released between protein action, the unbound DNA conformation can change. Distributive recombination is modeled by tangle equations such as

$$N(\mathbf{U_f} + \mathbf{B}) = \ substrate$$

$$N(\mathbf{U_f} + \mathbf{E}) = \ product\ 1$$

$$N(\mathbf{U_f}(\mathbf{2}) + \mathbf{B}) = \ product\ 1$$

$$N(\mathbf{U_f}(\mathbf{2}) + \mathbf{E}) = \ product\ 2$$

If the action is distributive or if multiple substrates are used, we can still push crossings into and out of $\mathbf{B}$ and $\mathbf{E}$ as in Fig. 16, so there will still be an infinite number of solutions.

4.2. Software for modeling recombination.

There are two programs one can use to solve the tangle equations in Fig. 14: TangleSolve [47] and TopoICE-R [48]. Both programs assume the substrate and product are rational knots/links. They also assume $\mathbf{B}$ and $\mathbf{E}$ are rational tangles. Let $\mathbf{B} = \frac{\mathbf{f_1}}{\mathbf{g_1}}$, $\mathbf{E} = \frac{\mathbf{f_2}}{\mathbf{g_2}}$. Suppose also that the substrate and product are rational knots/links, $N(\frac{\mathbf{a}}{\mathbf{b}})$, $N(\frac{\mathbf{z}}{\mathbf{v}})$, respectively. Thus the equations in Fig. 14 become $N(\mathbf{U_f} + \frac{\mathbf{f_1}}{\mathbf{g_1}}) = N(\frac{\mathbf{a}}{\mathbf{b}})$, $N(\mathbf{U_f} + \frac{\mathbf{f_2}}{\mathbf{g_2}}) = N(\frac{\mathbf{z}}{\mathbf{v}})$. In the rest of this subsection, we will focus on this special case.

TopoICE-R is freely available by downloading KnotPlot (www.knotplot.com/download). TopoICE-R has some subroutines in common with TopoICE-X, but TopoICE-R is not restricted to tangle equations involving crossing changes. TopoICE-R, like TopoICE-X, is embedded within KnotPlot and hence has many tools for visualizing and manipulating solutions in 3D. The user can enter the rational knot/link type of the substrate/product by either entering a rational number corresponding to the knot/link or by clicking on a diagram of the knot/link. Similarly the user can enter the tangle $\mathbf{B} = \frac{\mathbf{f_1}}{\mathbf{g_1}}$ by entering its fraction or clicking on a diagram of the tangle. At this point the user can either ask TopoICE-R to find all solutions for $\mathbf{U_f}$ and $\mathbf{E}$ given this $\mathbf{B}$, or the user can restrict the value of $\mathbf{E}$ to a particular rational tangle. The later method can sometimes be used to rule out potential models.

Recall we are assuming $\mathbf{B} = \frac{\mathbf{f_1}}{\mathbf{g_1}}$, $\mathbf{E} = \frac{\mathbf{f_2}}{\mathbf{g_2}}$. If $|f_1 g_2 - g_1 f_2| > 1$, then TopoICE-R finds all solutions. However, if $f_1 g_2 - g_1 f_2 = \pm 1$, then TopoICE-R can miss some solutions. For example, TopoICE-R does not find any solution to the system of equations in Fig. 21A, but there is a solution as shown in Fig. 21B. In this figure, $\mathbf{B} = \mathbf{0} = \frac{\mathbf{0}}{\mathbf{1}}$ and $\mathbf{E} = \mathbf{1} = \frac{\mathbf{1}}{\mathbf{1}}$. Hence $|f_1 g_2 - g_1 f_2| = |0(1) - (1)(1)| = 1$, so there was no expectation that TopoICE-R would find all solutions to this system of equations.

Figure 21: A.) $N(\mathbf{U} + \frac{\mathbf{0}}{\mathbf{1}}) =$, $N(\mathbf{U} + \frac{\mathbf{1}}{\mathbf{1}}) = N(\frac{2}{1})$. Since $|f_1 g_2 - g_1 f_2| = 1$, TopoICE-R may not find all solutions to this system of equations. B.) A solution to the system of equations in (A). TopoICE-R does not find this solution (or any other solution).

TopoICE-R finds all solutions for which $\mathbf{B}$, $\mathbf{E}$ are rational and $\mathbf{U_f}$ is isotopic to a sum of rational tangles [34]. Frequently, solutions must be of this form [35, 36, 49, 50], but not always as illustrated in Fig. 21.

Although it can be difficult to analyze an infinite number of solutions to a system of tangle equations, these solutions can be used to rule out potential models. For example, the system of tangle equations in Fig. 22 has no solution. In this figure, $\mathbf{B} = \frac{-1}{3}$ and $\mathbf{E} = \frac{2}{3}$. Hence $|f_1g_2 - g_1f_2| = |(-1)(3) - (3)(2)| = -9 > 1$, so TopoICE-R is guaranteed to find all solutions to this system of equations, if there had been any. Since there are no solutions, this action cannot transform the unknot into $N(\frac{4}{1})$, and thus $\mathbf{B} = \frac{-1}{3}$, $\mathbf{E} = \frac{2}{3}$ is not a possible model for Xer recombination.

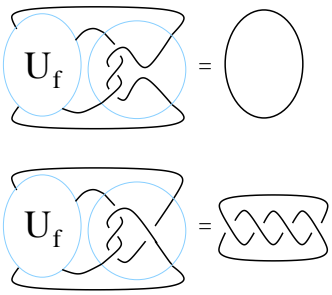

Figure 22: This system of equations has no solution. Hence Xer does not use this mechanism.

On the other hand, if one enters $N(\frac{a}{b}) = N(\frac{1}{0})$ = unknot, $N(\frac{z}{v}) = N(\frac{4}{1})$, $\mathbf{B} = \frac{0}{1}$ and $\mathbf{E} = \frac{9}{5}$ into TopoICE-R, one finds that $\mathbf{U_f} = -\mathbf{1}$ (Fig. 16A). One can similarly use TopoICE-R to find the solution in Fig. 16D. Hence $\mathbf{B} = \frac{0}{1}$, $\mathbf{E} = \frac{9}{5}$ is a mathematically possible model for Xer recombination, while $\mathbf{B} = \frac{-1}{3}$, $\mathbf{E} = \frac{-4}{3}$ is both a mathematically possible as well as a biologically plausible model for Xer recombination.

For a fixed rational $\mathbf{B}$, TopoICE-R can be used to find almost all rational solutions for $\mathbf{E}$. When $\mathbf{B} = \frac{f_1}{g_1}$, TopoICE-R will sometimes miss solutions for $\mathbf{E} = \frac{f_2}{g_2}$ for which $|f_1g_2 - g_1f_2| = 1$ as the example in Fig. 21 shows. However, TopoICE-R can find all solutions for the Xer equations in Fig 15 (assuming $\mathbf{B}$ and $\mathbf{E}$ are rational). When $\mathbf{B} = \frac{0}{1}$ and $\mathbf{E}$ is rational, $\mathbf{E} = \frac{3-4i}{1+4i-k(3-4i)}$, $\frac{1}{k}$, $\frac{3}{1+3k}$, or $\frac{5}{1+5k}$. In general when solving the system of equations in Fig. 14, even when $\mathbf{B}$ is fixed, there will be an infinite number of solutions for the tangle $\mathbf{E}$ [34]. One can visualize hundreds of these solutions in minutes via TopoICE-R by toggling through the different families of solutions as well as changing the values for i and k (although analyzing these solutions will take longer). But finding biologically relevant solutions can be a challenge.

Yuki Saka and Mariel Vazquez's TangleSolve [47] solves the system of tangle equations $N(\mathbf{U_f} + \mathbf{U_b} + \mathbf{P}) = N(\frac{a}{b})$, $N(\mathbf{U_f} + \mathbf{U_b} + \mathbf{R}) = N(\frac{z}{v})$ (Fig. 18). It is freely available at http://bio.math.berkeley.edu/TangleSolve. It assumes $\mathbf{U_f}$ and $\mathbf{U_b}$ are rational tangles, and $\mathbf{R}$ is an integral tangle (a rational tangle whose fraction corresponds to an integer such as $\frac{n}{1}$ and hence consists of only horizontal twists). It also assumes $\mathbf{P}$ is the zero tangle (but there is an option for displaying some solutions with $\mathbf{P}$ = infinity tangle). Note these are the most biologically relevant solutions. Other relevant solutions may also exist, but when handling an infinite number of possibilities, this is one of the best places to start. Recall that the local action of a recombinase involves breaking two DNA segments. Generally, this involves only a

few DNA base pairs in each of these two segments. Thus one would expect there would exist a projection of the 3D conformation in which these few base pairs would project to the zero tangle. In most cases, the local action of recombination has been modeled by adding at most one crossing or in the case of processive recombination, an integral number of crossings. Thus it is likely $\mathbf{R}$ can be modeled by an integral tangle. If $\mathbf{P}$ is the zero tangle, then $\mathbf{B} = \mathbf{U_b} + \mathbf{P} = \mathbf{U_b} + \mathbf{0} = \mathbf{U_b}$. Hence if $\mathbf{B}$ is rational, then $\mathbf{U_b}$ is rational. So far, no relevant nonrational solutions for $\mathbf{U_f}$ to the Fig. 18 tangle equations have been found when modeling recombination.

TangleSolve has a graphical interface for entering data. One can also enter multiple substrates and products as TangleSolve can find both the intersection and union of tangle solutions from multiple systems of equations. This is particularly nice when only the crossing number of a product is known. There are also subroutines specific for modeling processive and distributive recombination.

Relevant solutions found using TangleSolve can be entered into TopoICE-R to determine if there are additional solutions (although TopoICE-R may not find all solutions, such as Fig. 21B) and/or to manipulate these solutions in 3D. Although when we discussed TopoICE-R, we assumed that $\frac{\mathbf{f_1}}{\mathbf{g_1}}$, $\frac{\mathbf{f_2}}{\mathbf{g_2}}$ represent $\mathbf{B}$, $\mathbf{E}$, respectively, we can instead let $\frac{\mathbf{f_1}}{\mathbf{g_1}}$, $\frac{\mathbf{f_2}}{\mathbf{g_2}}$ represent the local action, $\mathbf{P}$, $\mathbf{R}$, respectively. Thus to find models for Xer recombination, one can toggle through solutions when $\frac{\mathbf{f_1}}{\mathbf{g_1}} = \mathbf{0}$, when focusing on the local action, while one can toggle through solutions when $\frac{\mathbf{f_1}}{\mathbf{g_1}} = \frac{\mathbf{-1}}{\mathbf{3}}$, for example, when focusing on the global conformation of the protein-bound DNA. KnotPlot's tangle calculator or sketch tools can also be used to draw a particular conformation.

KnotPlot has a number of features for manipulating tangle models in 3D. A variety of energies can be placed on the knots and tangles in KnotPlot [38]. Conformations can be simplified by minimizing these energies. The conformations in KnotPlot can be displayed using either smooth tubes or beads and sticks. By using the Edit panel in KnotPlot, one can move these beads and sticks to explore different geometries. One can output postscript pictures as well as 3D coordinates of conformations for use in other software. Help pages discussing these and other features are available within KnotPlot.

5. Additional applications of tangles to protein-DNA complexes

Topoisomerases and recombinases can be used to investigate other proteins which bind DNA such as Mu transposase (Fig. 1). The tangle models discussed in sections 3 and 4 specifically applied to proteins that create knotted/linked DNA molecules. However, tangle analysis can also be applied to proteins that do not change DNA topology. In order to obtain tangle equations, an experimental technique called *difference topology* has been used. [51, 52, 5, 53, 54, 55]. For example, the tangle model for Mu transposase obtained by [5] (shown in Fig. 1) was obtained by first incubating circular DNA with the proteins required to form the Mu-DNA complex. Cre recombinase was then added to create DNA knots/links. The conformation of the DNA within the Mu-DNA complex affects the types of knots/links which results from Cre recombination. This technique opens up tangle analysis to a much larger class

of proteins including proteins that bind more than two segments of DNA. For more information on this technique, see [5, 55, 56, 57].

Acknowledgements

This work was supported by a grant from the Joint DMS/NIGMS Initiative to Support Research in the Area of Mathematical Biology to ID and S. D. Levene. (NIH GM 67242). The author would like to thank Steve Levene, Robert Scharein, Andrzej Stasiak, Mariel Vazquez for many interesting discussions.

References

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

[2] N. Crisona, R. Weinberg, B. Peter, D. Sumners, N. Cozzarelli, The topological mechanism of phage lambda integrase, J Mol Biol. 289(4) (1999) 747–75.

[3] M. Vazquez, S. D. Colloms, D. Sumners, Tangle analysis of Xer recombination reveals only three solutions, all consistent with a single three-dimensional topological pathway., J Mol Biol. 346(2) (2005) 493–504.

[4] A. A. Vetcher, A. Y. Lushnikov, J. Navarra-Madsen, R. G. Scharein, Y. L. Lyubchenko, I. K. Darcy, S. D. Levene, DNA topology and geometry in Flp and Cre recombination, J Mol Biol. 357 (2006) 1089–1104.

[5] S. Pathania, M. Jayaram, R. Harshey, Path of DNA within the Mu transpososome. transposase interactions bridging two Mu ends and the enhancer trap five DNA supercoils, Cell 109(4) (2002) 425–436.

[6] R. Harshey, M. Jayaram, The Mu transpososome through a topological lens, Crit. Rev. Biochem. Mol. Biol. (preprint).

[7] D. Sumners, C. Ernst, S. Spengler, N. Cozzarelli, Analysis of the mechanism of DNA recombination using tangles, Quarterly Reviews of Biophysics 28 (1995) 253–313.

[8] I. K. Darcy, Biological distances on DNA knots and links: applications to XER recombination, J. Knot Theory Ramifications 10 (2001) 269–294, knots in Hellas '98, Vol. 2 (Delphi).

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

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

[11] D. Buck, C. Verjovsky Marcotte, Tangle solutions for a family of DNA-rearranging proteins, Math. Proc. Cambridge Philos. Soc. 139 (2005) 59–80.

[12] J. H. Conway, An enumeration of knots and links and some of their algebraic properties, Computational Problems in Abstract Algebra, J. Leech (ed.) (1986) 329–358.

[13] J. R. Goldman, L. H. Kauffman, Rational tangles, Adv. in Appl. Math. 3 (1997) 300–332.

[14] L. H. Kauffman, S. Lambropoulou, Classifying and applying rational knots and rational tangles, in Physical knots: knotting, linking, and folding geometric objects in R^3 (Las Vegas, NV, 2001), volume 304 of Contemp. Math., Amer. Math. Soc., Providence, RI, 2002, 223–259.

[15] L. H. Kauffman, S. Lambropoulou, On the classification of rational knots, Enseign. Math. (2) 49 (2003) 357–410.

[16] L. H. Kauffman, S. Lambropoulou, On the classification of rational tangles, Adv.

in Appl. Math. 33 (2004) 199–237.
[17] H. Schubert, Knoten mit zwei Brücken, Math. Z. 65 (1956) 133–170.
[18] G. Burde, H. Zieschang, Knots, Walter de Gruyter, De Gruyter Studies in Math. 5, 1985.
[19] L. Liu, C. Liu, B. Alberts, Type II DNA topoisomerases: enzymes that can unknot a topologically knotted DNA molecule via a reversible double-strand break, Cell 19(3) (1980) 697–707.
[20] T. Hsieh, Knotting of the circular duplex DNA by type II DNA topoisomerase from Drosophila melanogaster, J Biol Chem. 258(13) (1983) 8413–20.
[21] F. Dean, A. Stasiak, T. Koller, N. Cozzarelli, Duplex DNA knots produced by Escherichia coli topoisomerase I. structure and requirements for formation, J Biol Chem 260(8) (1985) 4975–83.
[22] F. Dean, N. Cozzarelli, Mechanism of strand passage by Escherichia coli topoisomerase I. the role of the required nick in catenation and knotting of duplex DNA., J Biol Chem 260(8) (1985) 4984–94.
[23] J. Wang, Cellular roles of DNA topoisomerases: a molecular perspective., Nat Rev Mol Cell Biol. 3(6) (2002) 430–440.
[24] A. D. Bates, A. Maxwell, DNA Topology, Oxford University Press, Oxford, 2005.
[25] A. Y. Chen, L. F. Liu, DNA topoisomerases: Essential enzymes and lethal targets, Annu. Rev. Pharmacol Toxicol. 34 (1995) 191–218.
[26] V. Rybenkov, C. Ullsperger, A. Vologodskii, N. Cozzarelli, Simplification of DNA topology below equilibrium values by type II topoisomerases, Science 277 (1997) 690–693.
[27] J. K. Mann, R. W. Deibler, D. W. Sumners, E. L. Zechiedrich, Unknotting by type II topoisomerases, Abstracts of papers presented to the American Mathematical Society 25:994-92-187.
[28] A. Bates, A. Maxwell, DNA topology: topoisomerases keep it simple, Curr Biol. 7(12) (1997) R778–81.
[29] J. Yan, M. Magnasco, J. Marko, A kinetic proofreading mechanism for disentanglement of DNA by topoisomerases, Nature 401(6756) (1999) 932–935.
[30] J. Roca, Varying levels of positive and negative supercoiling differently affect the efficiency with which topoisomerase II catenates and decatenates DNA, J Mol Biol. 305(3) (2001) 441–50.
[31] A. Vologodskii, W. Zhang, V. Rybenkov, A. Podtelezhnikov, D. Subramanian, J. Griffith, N. Cozzarelli, Mechanism of topology simplification by type II DNA topoisomerases, Proc Natl Acad Sci 98(6) (2001) 3045–9.
[32] G. Buck, E. Zechiedrich, DNA disentangling by type-2 topoisomerases, J Mol Biol. 340(5) (2004) 933–939.
[33] S. Trigueros, J. Salceda, I. Bermudez, X. Fernandez, J. Roca, Asymmetric removal of supercoils suggests how topoisomerase II simplifies DNA topology, J Mol Biol. 335(3) (2004) 723–31.
[34] I. K. Darcy, Solving unoriented tangle equations involving 4-plats, J. Knot Theory Ramifications 14 (2005) 993–1005.
[35] M. Culler, C. Gordan, J. Luecke, P. Shalen, Dehn surgery on knots, An. of Math 125 (1987) 237–300.
[36] C. Ernst, Tangle equations, J. Knot Theory Ramifications 5 (1996) 145–159.
[37] I. K. Darcy, R. G. Scharein, A. Stasiak, 3d visualization software to analyse topological outcomes of topoisomerase reactions. .
[38] R. G. Scharein, Interactive Topological Drawing, Ph.D. thesis, Department of Computer Science, The University of British Columbia (1998).
[39] I. K. Darcy, D. W. Sumners, Rational tangle distances on knots and links, Math.

Proc. Cambridge Philos. Soc. 128 (2000) 497–510.
[40] I. K. Darcy, Biological Metrics on DNA knots and catenanes, Ph.D. thesis, Department of Mathematics, Florida State University (1997).
[41] F. Guo, D. N. Gopaul, G. D. van Duyne, Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse, Nature 389 (1997) 40–46.
[42] S. Colloms, J. Bath, D. Sherratt, Topological selectivity in Xer site-specific recombination., Cell 88(6) (1997) 855–864.
[43] J. Bath, D. Sherratt, S. Colloms, Topology of Xer recombination on catenanes produced by lambda integrase., J Mol Biol. 289(4) (1999) 873–883.
[44] C. Ernst, Tangle equations. II, J. Knot Theory Ramifications 6 (1997) 1–11.
[45] C. Ernst, D. W. Sumners, Solving tangle equations arising in a DNA recombination model, Math. Proc. Camb. Phil. Soc. 126 (1999) 23–36.
[46] C. Cerf, A note on the tangle model for DNA recombination, Bull Math Biol. 60(1) (1998) 67–78.
[47] Y. Saka, M. Vazquez, TangleSolve: topological analysis of site-specific recombination., Bioinformatics 18 (2002) 1011–1012.
[48] I. K. Darcy, R. G. Scharein, TopoICE-R: 3D visualization modeling the topology of DNA recombination, Bioinformatics 22 (2006) 1790–1791.
[49] M. Hirasawa, K. Shimokawa, Dehn surgeries on strongly invertible knots which yield lens spaces, Proc. Amer. Math. Soc. 128 (2000) 3445–3451.
[50] P. Kronheimer, T. Mrowka, P. Ozsvath, Z. Szabo, Monopoles and lens space surgeries., `http://lanl.arxiv.org/abs/math.GT/0310164` .
[51] I. Grainge, D. Buck, M. Jayaram, Geometry of site-alignment during Int family recombination, J. Mol. Biol. 298 (2000) 749–764.
[52] E. Kilbride, M. Boocock, W. Stark, Topological selectivity of a hybrid site-specific recombination system with elements from Tn3 res/resolvase and bacteriophage PL lox*p*/Cre., J. Mol. Biol. 289 (1999) 1219–1230.
[53] S. Pathania, M. Jayaram, R. Harshey, A unique right end-enhancer complex precedes synapsis of Mu ends: the enhancer is sequestered within the transpososome throughout transposition, The EMBO journal 22(14) (2003) 3725–3736.
[54] Z. Yin, M. Jayaram, S. Pathania, R. Harshey, The Mu transposase interwraps distant DNA sites within a functional transpososome in the absence of DNA supercoiling, J Biol Chem. 280(7) (2005) 6149–6156.
[55] E. A. Kilbride, M. E. Burke, M. Boocock, W. Stark, Determinants of product topology in a hybrid Cre-Tn3 resolvase site-specific recombination system., J. Mol. Biol. 355(2) (2006) 185–95.
[56] I. K. Darcy, J. Luecke, M. Vazquez, A tangle analysis of the Mu transpososome protein complex which binds three DNA segments (preprint).
[57] I. K. Darcy, J. Chang, N. Druivenga, C. McKinney, R. K. Medikonduri, S. Mills, J. Navarra-Madsen, A. Ponnusamy, J. Sweet, T. Thompson, Coloring the Mu transpososome, BMC Bioinformatics (preprint).