# TOPOLOGICAL ANALYSIS OF DNA-PROTEIN COMPLEXES

SOOJEONG KIM AND ISABEL K. DARCY*

**Abstract.** A tangle consists of strings properly embedded in a 3-dimensional ball. Tangles have been used to model protein-bound DNA. The protein is represented by the 3D ball and the protein-bound DNA is represented by the strings embedded in the 3D ball. We review tangle analysis of protein-DNA complexes involving three or four segments of DNA.

**Key words.** Site-specific recombination, Difference topology, Tangle.

**AMS(MOS) subject classifications.** Primary 57M25, 92C40, 92E10.

**1. Introduction.** An $n$-string tangle is a three dimensional ball with $n$-strings properly embedded in it. Tangles were studied by Conway in the 1960's [3]. In the 1980's, Ernst and Sumners introduced a mathematical tangle model for protein-bound DNA complexes [8]. In this model, the protein is modeled by a three dimensional ball and the protein-bound DNA is modeled by strings. They used a 2-string tangle model to analyze experimental results for Tn3 resolvase and phage $\lambda$ integrase.

This work was motivated by Nick Cozzarelli [7, 16, 19, 21]. Some proteins can break and rejoin DNA segments and will knot circular DNA molecules. The knot types of the products can be used to determine information regarding how these proteins act. Nick Cozzarelli also used such proteins to study other protein-DNA complexes [13]. Type II topoisomerases will knot circular DNA by cutting the DNA, allowing a segment of DNA to pass through the break before resealing the DNA. In order to study the protein 13S condensin, DNA was first incubated with 13S condensin allowing the condensin to bind the DNA. Topoisomerase was then added. A spectrum of knots resulted which was different than that when topoisomerase acts on DNA in the absence of condensin. The difference in the knot spectrum in the presence versus absence of condensin was used to determine the manner in which 13S condensin is bound to DNA.

Pathania, Jayaram and Harshey extended these methods to derive the number of DNA crossings trapped in an unknown protein-DNA complex involving multiple DNA segments [15]. This methodology, called difference

*Department of Mathematics, University of Iowa, Iowa City, Iowa 52242.

topology, was used to determine the topological structure within the Mu protein complex, which consists of three DNA segments containing five crossings. Since Mu binds DNA sequences at 3 sites, the Mu protein DNA complex can be modeled by a 3-string tangle. 3-string tangle analysis is much more complicated than 2-string tangle analysis. The experimental results in [15] were mathematically [6] and computationally [5] analyzed by using a 3-string tangle model. We address a 4-string tangle model for a protein-DNA complex which binds four DNA segments.

In section 2, we introduce basic concepts of DNA recombination. We focus on site-specific recombination since this is a very important concept for understanding difference topology. In section 3, we introduce tangle analysis of protein-DNA complexes. In section 4, we explain the methodology of difference topology and its application to a Mu protein-DNA complex. In section 5, we summarize the 3-string tangle analysis of the Mu protein-DNA complex in [6]. Finally, in section 6, we introduce a 4-string tangle model for a protein which binds four DNA segments. We conclude that a 4-string tangle (with small number of crossings) which satisfies certain experimental conditions must be $R$-standard.

**2. DNA Recombination.** DNA recombination refers to a process in which DNA is rearranged within a genome. This is one of the biological processes which can change topological properties of DNA. We are interested in DNA recombination where two specific short DNA sequences are exchanged. This process is called *site-specific recombination* and the specific sequences are called *target sites*. This reaction requires specialized proteins, called *recombinases*, to recognize these sites and to catalyze the recombination reaction at these sites.

Site-specific recombination can result in either the inversion or deletion of a DNA segment. As one can see from Figure 1(a), if the orientation of target sites are opposite to one another (inverted repeat), then recombination leads to the inversion of the DNA segment between the two target sites. On the other hand, if the orientation of target sites are the same with respect to one another (directed repeat), then recombination leads to the deletion of the DNA segment between the two target sites, see Figure 1(b).

Note that the number of components is the same after inversion. But it is different after deletion, since the DNA sequence between the two target sites are deleted from the original DNA sequences. In particular, when the initial DNA is circular, inversion results in a knot and deletion results in a link as one can see from the following example.

*Cre* is a site-specific recombinases. The target sites of *Cre* are called *loxP*. *Cre* can catalyzes both DNA inversion and deletion. The recombination products depend on the relative orientation of the *loxP* sites, the target sites of *Cre*. When the DNA is circular, the products of DNA inversion and deletion by *Cre* are knots and catenanes, respectively (see Figure

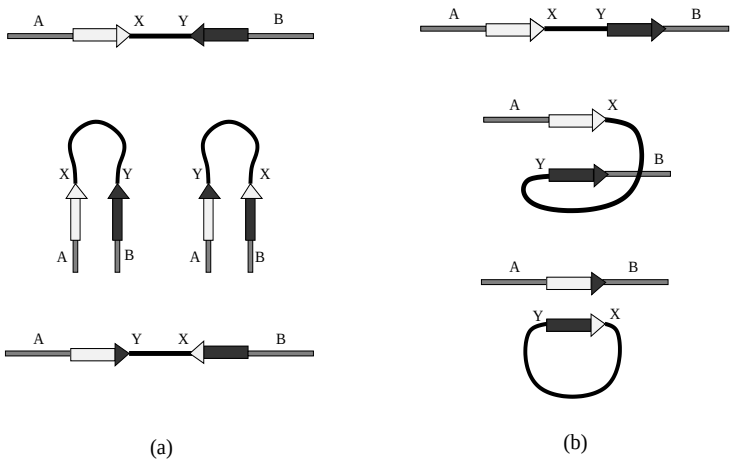


FIG. 1. *(a.) Inversion. (b.) Deletion.*
This figure is redrawn from http://www.mun.ca/biochem/courses/3107/Lectures/Topics/Site-specific-Recomb.html.

2).

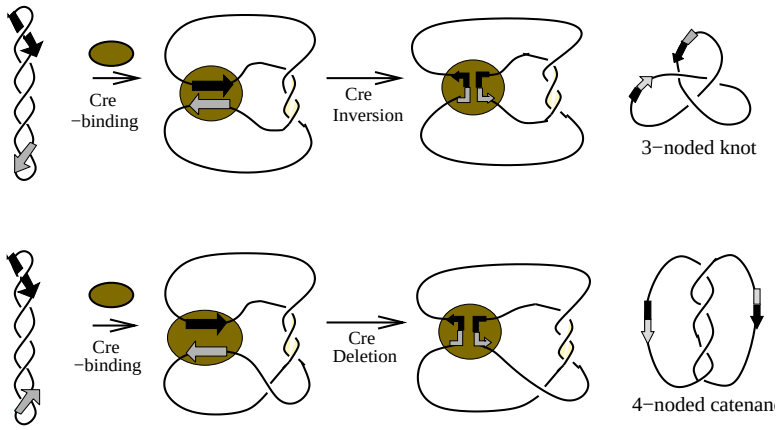


FIG. 2. *Cre recombination.*

**3. DNA Topology and the Tangle Model.** An *n-string tangle* is a three dimensional ball with n-strings properly embedded in it. The tangle model of a protein-DNA complex was developed by C. Ernst and D. W. Sumners [8]. This model assumes the protein is a three dimensional

ball and the protein-bound DNA are strings embedded inside the ball. See Figure 3.

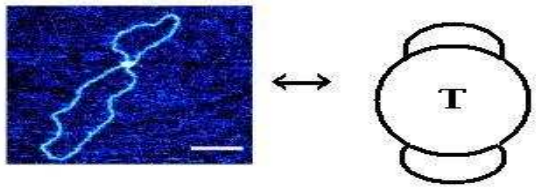


FIG. 3. *AFM image of a Cre synaptic complex formed with circular DNA [18] and a corresponding 2-string tangle model.*

Examples of 3-string tangles are given in Figure 4. A *rational tangles* is ambient isotopic to a tangle which has no crossings if we allow the boundary of the three ball to move. A tangle is rational if and only if its strings can be pushed to lie on the boundary of the 3D ball so that no string crosses over another string on the boundary of this ball. If the DNA wraps around the protein "ball" so that the DNA does not cross itself on the boundary of this protein ball, then the tangle modeling it is rational. Also, in nature, circular DNA is supercoiled. Protein-bound DNA is also often supercoiled. Hence rational tangles are generally believed to be the most biologically reasonable models for protein-bound DNA.

EXAMPLE 1. *Figure 4 (a)∼(e) give examples of 3-string tangles. Among those, (a), (c) and (d) are examples of rational 3-string tangles.*

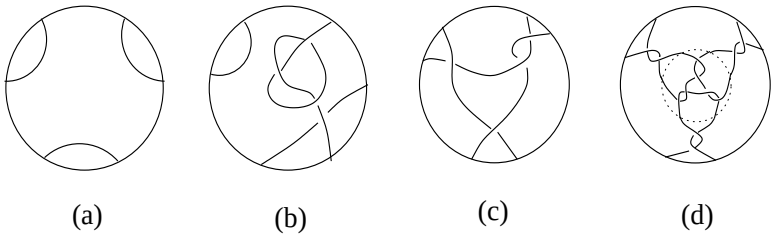


FIG. 4. *Examples of 3-string tangles*

The original tangle model was applied to proteins which bind two segments of DNA and which will break and rejoin segments of DNA creating knotted DNA. For a review of 2-string tangle analysis, see for example [8, 9, 17, 4]. Software has been developed to solve $n$-string tangle equations [5]. This software was used to search through all tangles up through 8 crossings which satisfy the experimental results of [5]. But computational software which can only solve one system of equations at a time

lacks the ability mathematical theories can provide for analyzing real and hypothetical experiments.

In the next section we will discuss the biological model for a Mu protein-DNA complex given in [15], while in section 5 we will summarize the mathematical tangle analysis given in [6].

**4. Difference Topology and Its Application to Mu.** DNA transposition results in the movement of a DNA segment from one location to another in a genome (http://research.utu.fi/celgenmol/molepid/savilahti.html). Bacteriophage Mu is a virus which uses transposition efficiently to replicate its DNA. During the transposition process, Mu proteins bind to 3 target sites including an enhancer sequence and two Mu ends (attL and attR) (see Figure 5). The enhancer sequence will be denoted by E, the attL site by L and the attR site by R. The protein-DNA complex consisting of Mu proteins along with these three DNA sequences is called the *transpososome*. The structure of the transpososome is very important for understanding the transposition pathway.

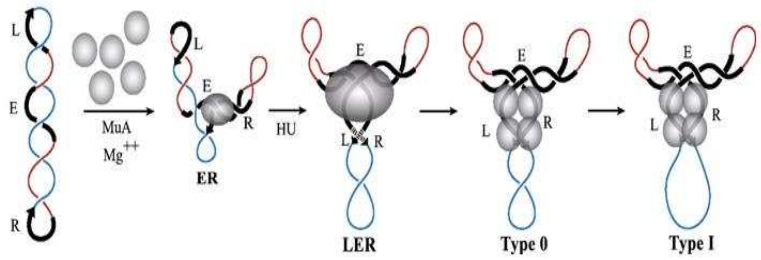


FIG. 5. *Mu transposition. Courtesy: Pathania el al.[23]*

The reaction pathway shown in Figure 6 was the model before the structure of the transpososome was determined in [15]. The new reaction pathway is shown in Figure 5. Note that in the older model, since there was no information available regarding the DNA shape bound by Mu, a very simple structure was assumed. The protein-bound DNA conformation in Figure 5, determined via difference topology, can be used to determine what DNA sequences are likely to be close to each other and therefore may interact [15]. Difference topology was also used to detect a new intermediate (denoted by ER in Figure 5) in the reaction pathway [14, 10, 23]. Difference topology was also used to investigate the role of supercoiling [22]. For additional applications of difference topology see [11].

Pathania, Jayaram and Harshey used *Cre* inversion and deletion to determine the topological structure of DNA within the Mu transpososome [15]. If Cre acts on unknotted DNA not bound by any proteins except for Cre, then the main products of Cre inversion and deletion are unknots and

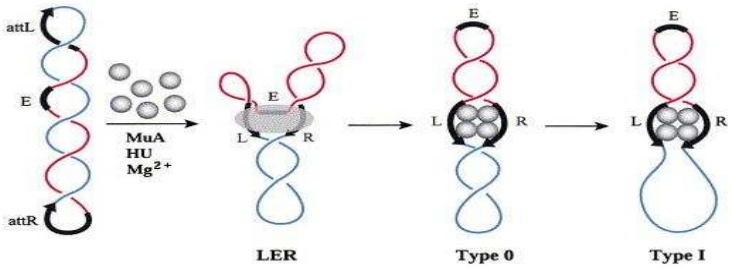


FIG. 6. *An older model of Mu transposition. Courtesy: Pathania el al.[15]*

unlinks, respectively. If, however, a protein complex such as Mu binds the DNA before Cre acts, the products can be more complicated. This difference in products was used in [15] to determine the topological conformation of the DNA bound by Mu. This methodology is called *difference topology*.

Pathania *et al.* first performed *Cre* inversion with two *loxP* sites lying on either side of E, isolating this site from L and R. In Figure 7 (a), the *loxP* sites are inversely repeated. *Cre* cuts these target sites and changes the topology of the DNA before resealing it again. The product topology in this case was a three noded knot. Those three crossings resulted from E crossing R and L three times. Note that the crossings between R and L can be untwisted and thus have no affect on the topology of the product.

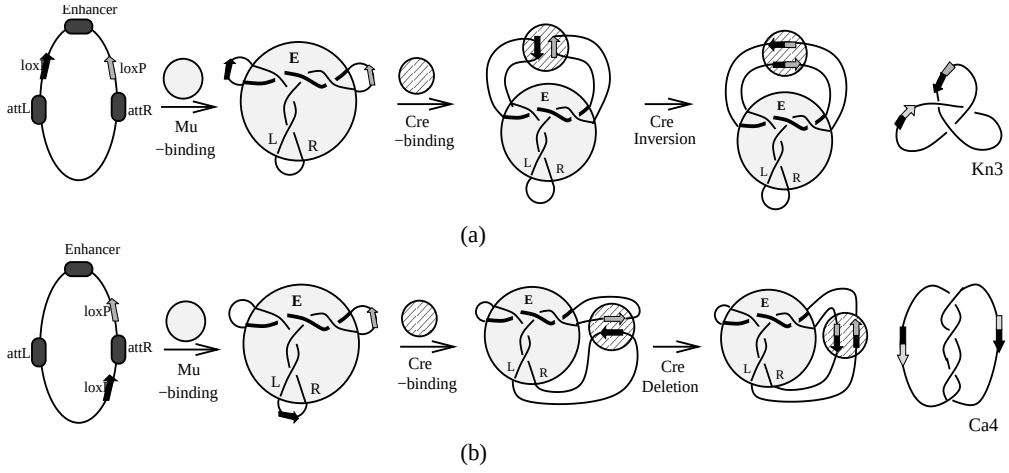


FIG. 7. *Cre recombination on the DNA-Mu protein complex.*

If the *loxP* sites are placed on the loops indicated in Figure 7, but

directly repeated instead of inversely repeated, than an extra crossing not bound by either Mu or Cre is necessary to properly orient the *loxP* sites within the Cre-DNA complex. In this case the product of Cre recombination is a 4-crossing link. Note that this product has one more crossing than the product when the *loxP* sites were placed on the same pair of loops, but in inverse orientation. There are three pairs of loops on which to place the *loxP* sites. In each case the number of crossings in the product differed by one when comparing inversely repeated versus directly repeated *loxP* sites on the same pair of loops. It was assumed that the smaller crossing product corresponded to the tangle equation where no extra crossing is needed to properly orient the *loxP* sites within the Cre-DNA complex. The equations corresponding to the smaller crossing product when comparing *loxP* sites on the same pair of loops is shown in Figures 8a, 9a. In Figure 8a, the solution found in [15] is shown while Figure 9a shows the equations where the tangle corresponding to the Mu transpososome is unknown. One can prove that the solution set for $T$ to the system of three equations in Figure 9a is the same as the solution set for $T$ if all six experiments are considered [5, 6].

To determine the number of DNA crossings within the Mu transpososome, we are interested in how many crossings are between E and R, R and L, L and E. Note that the protein-bound DNA conformation shown in Figure 8 consists of supercoiled DNA with three branches: one branch contains one crossing while the other two branches each contain two crossing. The solution found in [15] was obtained by assuming the protein-bound DNA conformation is a 3-branched supercoiled structure. Let $x$ be the number of crossings between E and R, $y$ the number of crossing between R and L, and $z$ the number of crossings between L and E. If the DNA conformation bound by Mu is supercoiled with three branches, then $x, y, z$ represent the number of crossings in each of the three branches. In this case, the equations in Figure 9a correspond to the equations $x + z = 3$, $x + y = 3$ and $y + z = 4$. Since we have three unknown variables and three linear equations, we can easily solve this linear system. The solution is that $x = 1$, $y = 2$, and $z = 2$. This implies that there is one crossing between E and R, two crossings between R and L, and two crossings between L and E. Thus if the DNA conformation bound by Mu is supercoiled with three branches, then the Mu transpososome has the five crossing configuration shown in Figure 9b [15].

**5. 3-String Tangle Analysis.** Mathematically, the Mu proteins can be modeled by a three dimensional ball and the three DNA segments can be modeled by 3 strings in the ball. Pathania *et al.* found a solution to the system of equations in Figure 9a in which the DNA bound by Mu consists of supercoiled DNA with 3 branches and 5 crossings ([15], see section 4). Pathania *et al.*'s experimental data [15] was mathematically analyzed by using 3-string tangle analysis [6] without the assumption that the tangle $T$

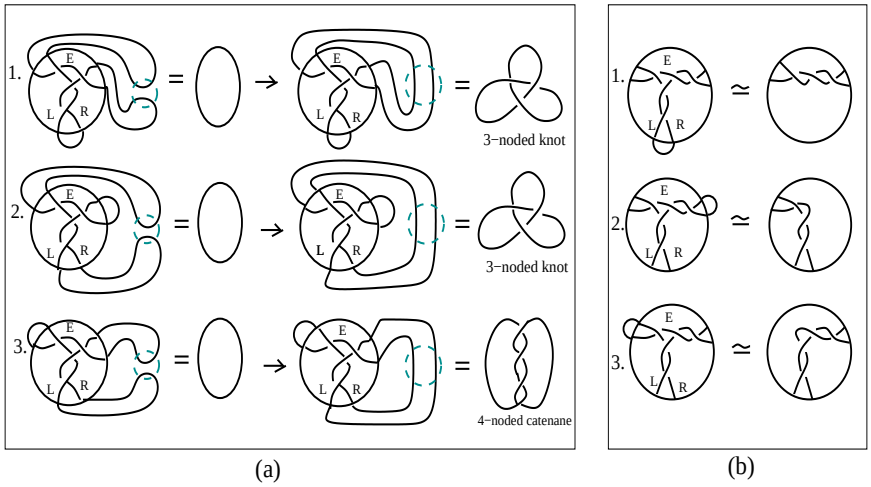


FIG. 8. *Tangle model of Mu transpososome.*

represents supercoiled DNA with three branches. If a tangle $T$ satisfies all the experimental data in [15], it can be a possible tangle model for the Mu transpososome. By using tangle theory, the following result was obtained:

PROPOSITION 5.1. *Let $T$ be a 3-string tangle which satisfies the system of tangle equations in Figure 9 (a). If $T$ can be freely isotopic to a projection with less than 8 crossings, then $T$ is the tangle in the Figure 9 (b).*

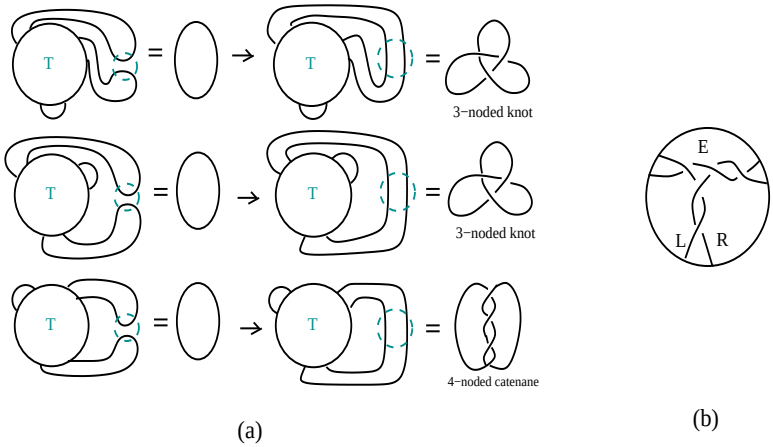


FIG. 9. *Tangle model of Mu transpososome.*

Two tangles are freely isotopic to each other if they are ambient iso-

topic allowing the boundary to move. For example, a rational tangle is freely isotopic to a tangle with no crossings. Thus Proposition 5.1 implies that the only rational tangle solution to the Figure 9 (a) equations is that given in Figure 9 (b).

An additional experiment not described here was used in [6] to rule out eight crossing solutions. The upper bound for the number of crossings which could be bound by Mu is unknown. However, since the solution found in [15] has five crossings, it is unlikely that a solution with more than eight crossings could be a model for the Mu transpososome. Thus the solution found in [15] is the only biologically reasonable solution.

**6. 4-String Tangle Analysis.** We do not currently have experimental data for a protein-DNA complex which binds four segments of DNA. In fact, we are not aware of such a complex. However there are a number of protein-DNA complexes, such as those involved in replicating and transcribing DNA, in which multiple proteins interact with each other and with multiple segments of DNA. Thus it is highly likely that protein-DNA complexes exist involving four or more DNA segments. We address a model for a protein complex which binds four DNA segments. Such a protein complex bound to circular DNA is modeled by a 4-string tangle with four loops outside of the tangle (Figure 10 (a)).

In nature, DNA is negatively supercoiled if it is circular or if the ends are constrained (http://www.cbs.dtu.dk/staff/dave/roanoke/genetics980213a.html). There are two kinds of DNA supercoiling, plectonemic and solenoi -dal. Plectonemic supercoils are frequently branched [1]. Figure 10 (b) shows an example of a branched supercoiled DNA-protein complex which would be a biologically reasonable model for a protein-DNA complex involving four segments of DNA. More generally, Figure 10 (c) shows a biologically natural tangle model of a 4-branched supercoiled DNA-protein complex, where the $n_i$'s are the number of right-handed half twists.

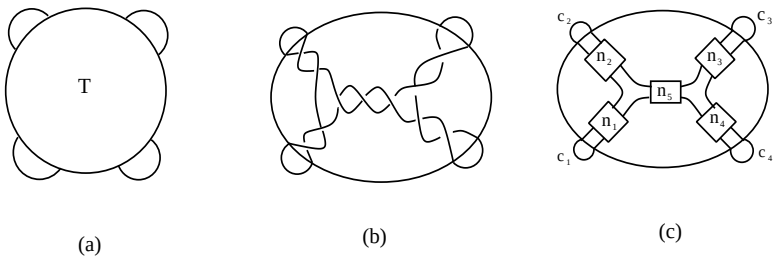


FIG. 10. *(a) A 4-string tangle model of a DNA-protein complex. (b), (c) Examples of 4-string tangle model which are biologically relevant.*

In this section, we would like to extend 3-string tangle analysis (section 5) to 4-string tangle analysis based on difference topology. For Cre

recombination, we need to put *loxP* sites on two of the outside loops. In the 3-string tangle model, there are three choices for a pair of loops on which to place Cre binding sites. On the other hand, in the 4-string tangle model, there are six different possible pairs of loops. In each case, there are two possible orientations for the Cre binding sites, directly or inversely repeated. Thus there are twelve possible Cre reactions for the 4-string tangle model (six different pairs of loops and two different orientation of *loxP* sites for each pair). By the prediction of difference topology (section 4), the crossing number of the knotted inversion product and the catenated deletion product will differ by one when the Cre binding sites are placed on the same pair of loops but in different orientations.

As we mentioned at the beginning of this section, Figure 10 (c) is a biologically relevant 4-string tangle model. Assume two *loxP* sites are located on loops $c_1$ and $c_2$ of Figure 10 (c). After Cre recombination, the $n_1$ and $n_2$ crossings on two branches of the supercoiled DNA would be trapped, but the $n_3$, $n_4$ and $n_5$ crossings on the other three branches can be removed. The result is a $(2,\ n_1+n_2)$-torus knot if $n_1+n_2$ is odd or $(2,\ n_1+n_2)$-torus link if $n_1+n_2$ is even. For example, Cre recombination assuming the tangle model Figure 10 (b) results in $(2,4)$-torus link. See Figure 11 (a). Similarly, if two *loxP* sites are located on loops $c_2$ and $c_3$, Cre recombination results in the $(2,7)$-torus knot as shown in Figure 11 (b). For convenience, Cre is placed on the left side and the 4-string tangle is rotated 90° counterclockwise in Figure 11(b).

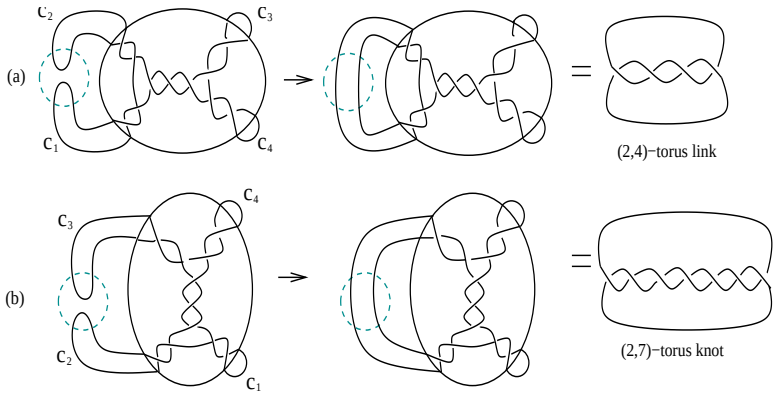


FIG. 11. *(a) If Cre acts on the loops $c_1$ and $c_2$, then Cre recombination for this protein-bound DNA conformation results in a $(2,4)$-torus link. (b) If the Cre binding sites are placed on loops $c_2$ and $c_3$, then Cre recombination for this protein-bound DNA conformation results in a $(2,7)$-torus knot.*

Hence if $T$ is a tangle of the form shown in Figure 10 (c), Cre recombination results in a $(2,p)$-torus knot if $p$ is odd or a $(2,p)$-torus link if $p$ is even. Note that the products of Cre recombination in the Mu/Cre

experiments were $(2, p)$-torus knots and links [15]. Thus for the 4-string tangle model, we focus on equations where we assume the products are $(2, p)$-torus knots and links. This process can be modeled by Figure 12. In this figure , the tangle $T$ represents a protein which binds to four DNA segments. The dotted circle represents Cre. For convenience, Cre is placed on the left side of $T$. $T$ is rotated by 90° in (b) and (f), 180° in (c) and (e), 270° in (d) counterclockwise. We can summarize all these assumptions with Figure 12 and define a tangle satisfying these conditions a *solution tangle.*

DEFINITION 6.1. *A solution tangle is a tangle $T$ which is a solution to the system of 12 difference topology equations where the products are $(2, p_i)$ torus knots/links. Six of these equations are shown in Figure 12.*

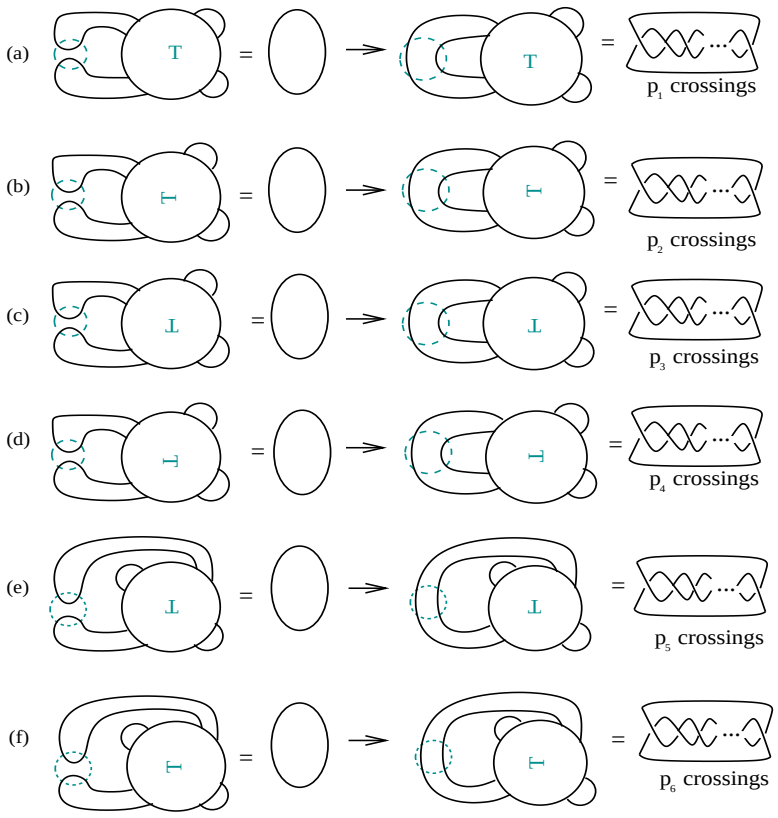


FIG. 12. *A 4-string solution tangle. In (b)∼(d) and (f), $T$ is rotated. The dotted circle represents a Cre recombinases.*

We will first discuss branched supercoiled DNA solutions. A 4-string tangle model of a branched supercoiled DNA-complex can be represented

by a weighted graph. For example, the 4-string tangle in Figure 10 (b) has 2 or 3 right handed half twists on each branch resulting in the graph shown in Figure 13 (b).

DEFINITION 6.2. *A tangle of the form shown in Figure 13 (a) will be called standard, where $n_i$ is the number of right-handed half twists. Note that a 4-string standard tangle $T$ can be represented by a weighted graph $G$, where $G$ is as in Figure 13(b). Call this graph $G$ a standard graph.*

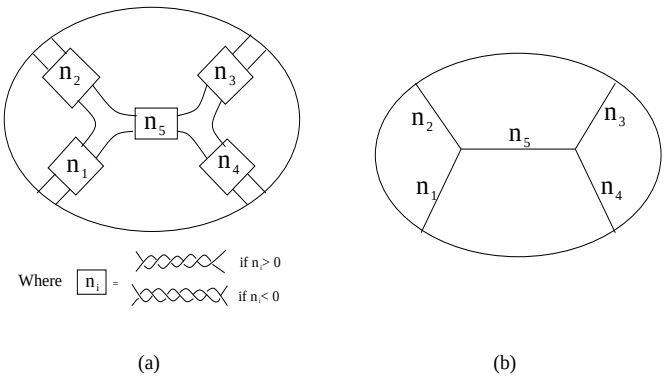


FIG. 13. *(a) Standard tangle. (b) A weighted graph $G$ representing a 4-string standard tangle.*

We will also address the possibility that a pair of supercoiled DNA branches can be twisted. In other words, what if a tangle model is isotopic to a standard tangle allowing boundary of the corresponding graph (see Definition 6.2) to move?

DEFINITION 6.3. *A weighted graph $G_R$ is an $R$-standard graph if it is isotopic to a standard graph $G$ allowing the boundary of $G$ to move. A tangle $T$ is $R$-standard if it corresponds to an $R$-standard graph $G_R$.*

EXAMPLE 2. *Examples of 4-string standard tangles are shown in Figure 14 (a), (b) and an example of a 4-string $R$-standard tangle is shown in (c).*

EXAMPLE 3. *Figure 15 (a) shows an example of a weighted graph $G_R$ which represents the $R$-standard tangle $T$ in Figure 15 (c).*

By extending 3-string tangle analysis of [6] to 4-string tangles, we determined that the biologically relevant solutions to the system of equations in Figure 12 must be $R$-standard:

THEOREM 6.1. *[12] Suppose $T$ is a 4-string tangle which has less than 8 crossings up to free isotopy. If $T$ is a solution tangle, then $T$ is $R$-standard.*

In other words, if a 4-string tangle $T$ satisfies all the equations of Figure 12 and has less than 8 crossings up to free isotopy, $T$ can be represented

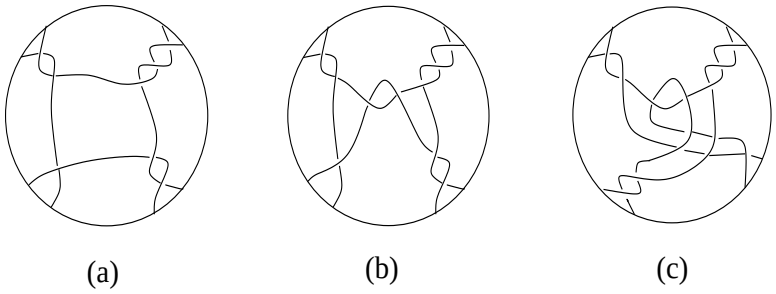


FIG. 14. *(a),(b) Examples of standard tangles; (c) Example of R-standard tangle.*

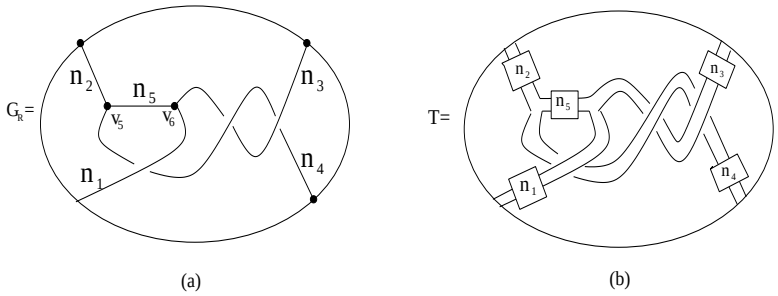


FIG. 15. *An example of a weighted graph $G_R$ for an R-standard tangle $T$.*

by an $R$-standard graph. Since all rational tangles are freely isotopic to a tangle which has no crossings, we can find all rational solutions.

We start with the following definition:

DEFINITION 6.4. *Let $G_R$ be a graph which corresponds to an R-standard tangle. There are two vertices in the interior of the ball and 4 vertices on the boundary of the ball. Let $v_1 = SW$, $v_2 = NW$, $v_3 = NE$, $v_4 = SE$ be the vertices on the boundary of the ball, and $v_5$ and $v_6$ be the vertices in the interior of the ball. $G_R$ is $(2, j)$-branched if $v_5$ connects $v_2 = NW$ and $v_j$ for some $1 \leqslant j \leqslant 4$, $j \neq 2$.*

The vertex $v_5$ can only be connected with $(v_1,\ v_2)$ or $(v_2,\ \ v_3)$ or $(v_2,\ v_4)$, hence there are 3 different $(i, j)$ branchings (Figure 16). For example, the graph $G_R$ in Figure 15 (a) is $(2, 4)$-branched. Note that $n_5 = 0$ if and only if $G_R$ is $(i, j)$-branched for all $(i, j)$.

Each edge of $G_R$ represents a branch of a branched supercoiled DNA molecule. This implies that an $(i, j)$-branched graph and a $(k, l)$-branched graph represent different geometries of a DNA molecule when $(i, j) \neq (k, l)$.

We will first focus on tangles of the form shown in Figure 17. Suppose a tangle whose corresponding graph has the form shown in Figure 17 (a)

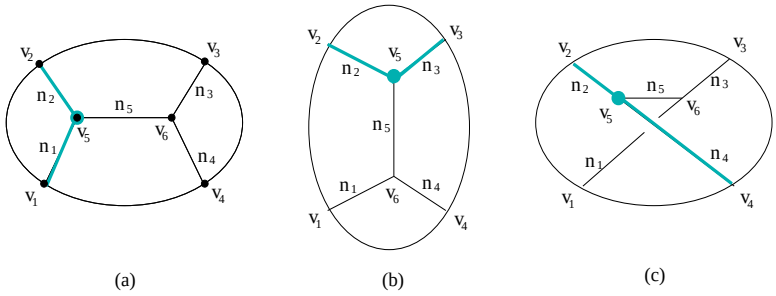


FIG. 16. *(a)(1,2)-branched; (b)(2,3)-branched; (c)(2,4)-branched weighted graph for R-standard tangle.*

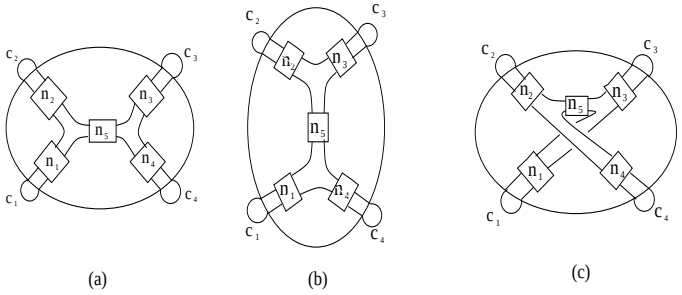


FIG. 17. *(a) Example of R-standard tangle model of a branched DNA-protein complex corresponding to a weighted graph which is (a) (1,2)-branched; (b) (2,3)-branched; (c) (2,4)-branched.*

is a solution to the system of equations in Figure 12. Then we have the equations in 6.1. The values $p_1, \cdots, p_6$ in Figure 12 must be determined experimentally. Our goal is to find $n_1, \cdots n_6$ in terms of the $p_i$'s.

$$\begin{aligned} n_1 + n_2 &= p_1 \\ n_2 + n_3 + n_5 &= p_2 \\ n_3 + n_4 &= p_3 \\ n_1 + n_4 + n_5 &= p_4 \\ n_1 + n_3 + n_5 &= p_5 \\ n_2 + n_4 + n_5 &= p_6. \end{aligned} \tag{6.1}$$

The solution to the system of equations 6.1 is the following:

$$n_1 = \frac{p_1 + p_4 - p_6}{2}, \quad n_2 = \frac{p_1 - p_4 + p_6}{2}$$

$$n_3 = \frac{p_2 + p_3 - p_6}{2}, \quad n_4 = \frac{-p_2 + p_3 - p_6}{2}$$
$$n_5 = \frac{-p_1 + p_2 - p_3 + p_4}{2}, \quad p_2 + p_4 = p_5 + p_6. \tag{6.2}$$

We can solve similar equations for tangles corresponding to the graphs in Figure 17 (b) and (c). The summary of the results are the following:

- The solution to the Figure 12 equations is the following if the solution is of the form Figure 17 (b):

$$n_1 = \frac{-p_3 + p_4 + p_5}{2}, \quad n_2 = \frac{p_1 + p_2 - p_5}{2}$$
$$n_3 = \frac{-p_1 + p_2 + p_5}{2}, \quad n_4 = \frac{p_2 + p_3 - p_5}{2} \tag{6.3}$$
$$n_5 = \frac{p_1 - p_2 + p_3 - p_4}{2}, \quad p_1 + p_3 = p_5 + p_6.$$

- The solution to the Figure 12 equations is the following if the solution is of the form Figure 17 (c):

$$n_1 = \frac{p_1 - p_2 + p_5}{2}, \quad n_2 = \frac{p_1 - p_4 + p_6}{2}$$
$$n_3 = \frac{-p_1 + p_2 + p_5}{2}, \quad n_4 = \frac{-p_1 + p_4 + p_6}{2} \tag{6.4}$$
$$n_5 = \frac{-p_2 + p_4 - p_5 - p_6}{2}, \quad p_1 + p_3 = p_2 + p_4.$$

Note that the $n_i$ must be integral. To have an integer solution set $\{n_1, \cdots, n_5\}$, all numerators of Equations 6.2, 6.3 and 6.4 should be even. In fact, there are eight possible cases to have an integer solution set for equation 6.2, shown in the following table.

| | $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ | $p_6$ |
|---|---|---|---|---|---|---|
| 1 | even | even | even | even | even | even |
| 2 | odd | odd | even | even | even | odd |
| 3 | even | odd | even | odd | odd | odd |
| 4 | odd | even | even | odd | odd | even |
| 5 | even | odd | odd | even | odd | even |
| 6 | odd | even | odd | even | odd | odd |
| 7 | even | even | odd | odd | even | odd |
| 8 | odd | odd | odd | odd | even | even |

Equations 6.3 and 6.4 have an integer solution set for the same eight cases. Thus the different ways of branching can only be distinguished by the last equation given in Equations 6.2, 6.3 and 6.4:

LEMMA 6.1. *The graph $G_R$ corresponding to an $R$-standard tangle can only be branched in three different ways, (1,2), (2,3) or (2,4)-branched. The $(i, j)$ branching of a solution can be determined as follows:*

- *If $p_2 + p_4 = p_5 + p_6$ holds, $G_R$ is (1,2)-branched.*
- *If $p_1 + p_3 = p_5 + p_6$ holds, $G_R$ is (2,3)-branched.*
- *If $p_1 + p_3 = p_2 + p_4$ holds, $G_R$ is (2,4)-branched.*

*In addition, $n_5 = 0$ if and only if $G_R$ is $(i, j)$-branched for all $(i, j)$.*

We have only proved Lemma 6.1 for tangles corresponding to the graphs shown in Figure 17. However, Lemma 6.1 can be extended to $R$-standard tangles as discussed in the next section.

**6.1. Discussion On Complicated Branched Solution Tangles.** We will now consider a more complicate branched solution tangle like that in Figure 18.

EXAMPLE 4. *Let $G$ be a graph which corresponds to the standard graph in Figure 18(a). After doing one counterclockwise half twist of $v_1$ and $v_4$ and two clockwise half twists of $v_3$ and $v_4$ moving the boundary of 3-ball, one obtains the weighted graph $G_R$ (Figure 18 (b)). Then $G_R$ is the weighted graph which corresponds to the $R$-standard tangle $T$ in Figure 18 (b). Since $v_5$ is connected to $v_2 = NW$ and $v_4$, $G_R$ is (2,4)-branched.*

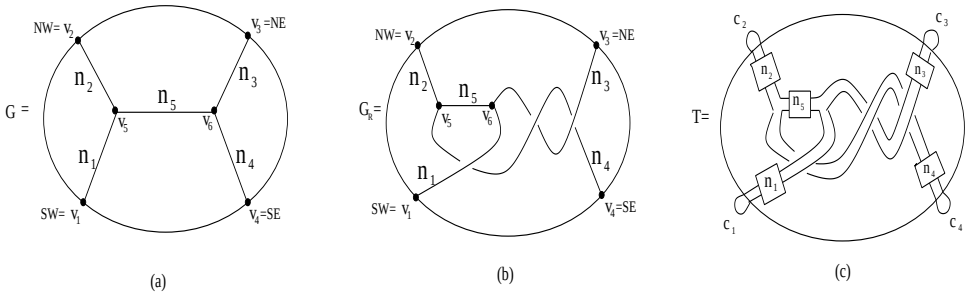


FIG. 18. *Example of R-standard tangle.*

Let's compare this example with the tangle in 17 (c), which we will call $T^{'}$. The link obtained from Cre recombination on $c_3$ and $c_4$ of $T$ has positive 2 writhe which can be converted to four half twists as shown in Figure 19 [2]. Hence $p_3 = n_3 + n_4 + n_5 + 4$ for $T$ while for $T^{'}$, $p_3 = n_3 + n_4 + n_5$. Similarly, $p_4 = n_1 + n_4 + n_5 - 2$ for $T$, while $p_4 = n_1 + n_4 + n_5$ for $T^{'}$. The remaining equations for $T$ are identical to the equations for $T^{'}$.

Note that a solution of the form $T$ will satisfy the first five equations in 6.4 if and only if a solution of the form $T^{'}$ satisfies these equations. This is because writhe of a link diagram can be converted to an even number of half twists. Thus, all numerators of Equations 6.2, 6.3 and 6.4 will still be even after adding or subtracting an even number. Hence the last equation in 6.4 determines if a tangle of the form $T$ or $T'$ can be a solution. For more information see [12].

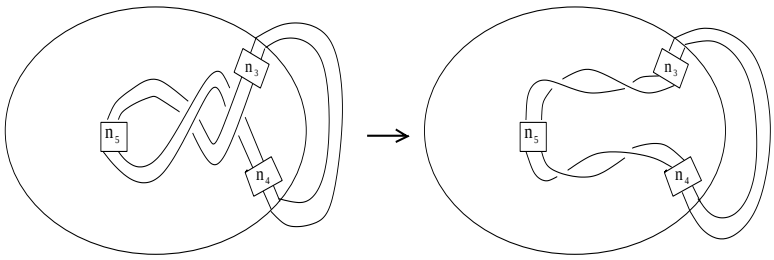


FIG. 19. *(a) A link obtained by doing Cre recombination in example 18 has writhe two. (b) The writhe two can be converted to 4 half twists.*

## REFERENCES


[1] A.D. Bates and A. Maxwell. *DNA Topology*. IRL Press, Oxford, 1993.

[2] W. R. Bauer, F. H. C. Crick, and J. H. White. Supercoiled DNA. *Scientific American*, 243:118–133, 1980.

[3] J.H. Conway. An enumeration of knots and links and some of their related properties. *Computational Problems in Abstract Algebra (John Leech, ed.) Pergamon Press, Oxford and New York*, 1969.

[4] I. Darcy, *Biological distances on DNA knots and links: Applications to Xer recombination*, Journal of Knot Theory and its Ramifications **10** (2001), 269–294.

[5] I. K. Darcy, A. Bhutra, J. Chang, N. Druivenga, C. McKinney, R. K. Medikonduri, S. Mills, J. Navarra Madsen, A. Ponnusamy, J. Sweet, and T. Thompson. Coloring the Mu transpososome. *BMC Bioinformatics*, 7:Art. No. 435, 2006.

[6] I.K. Darcy, J. Luecke, and M. Vazqez. Tangle analysis of difference topology experiments: Applications to a Mu protein-DNA complex. Preprint.

[7] F.B. Dean, A. Stasiak, T. Koller, and N.R. Cozzarelli. Duplex DNA knots produced by Escherichia Coli topoisomerase I. *I. Biol. Chem.*, 260:4795–4983, 1985.

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

[9] C. Ernst and D. W. Sumners. Solving tangle equations arising in a DNA recombination model. *Math. Proc. Camb. Phil. Soc.*, 126:23–36, 1999.

[10] R. Harshey and M. Jayaram. The Mu transpososome through a topological lens. *Crit Rev Biochem Mol Biol.*, 41(6):387–405, 2006.

[11] M. Jayaram and R. Harshey. Difference topology: analysis of high-order DNA-protein assemblies. Preprint.

[12] S. Kim and I. K. Darcy. An 4-string tangle analysis on DNA-protein complexes based on difference topology. Preprint.

[13] K. Kimura, V.V Rybenkov, N. J. Crisona, T. Hirano, and N.R. Cozzarelli. 13s condensin actively reconfigures DNA by introducing global positive writhe: implications for chromosome condensation. *Cell*, 98(2):239248, 1999.

[14] S. Pathania, M. Jayaram, and R. Harshey. A unique right end-enhancer complex precedes synapsis of Mu ends: the enhancer is sequestered within the transpososome throughout transposition. *EMBO J*, 22(14):3725–36, 2003.

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

[16] S.J. Spengler, A. Stasiak, and N.R. Cozzarelli. The stereostructure of knots and catenanes produced by phage $\lambda$ integrative recombination : implications for mechanism and DNA structure. *Cell*, 42:325–334, 1985.

[17] D. W. Sumners, C. Ernst, N.R. Cozzarelli, S.J. Spengler *Mathematical analysis of*

*the mechanisms of DNA recombination using tangles*, Quarterly Reviews of Biophysics **28** (1995).

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

[19] S.A. Wasserman and N.R. Cozzarelli. Determination of the stereostructure of the product of tn3 resolvase by a general method. *Proc. Nat. Acad. Sci. U.S.A.*, 82:1079–1083, 1985.

[20] S.A. Wasserman and N.R. Cozzarelli. Biochemical topology : applications to DNA recombination and replication. *Science*, 232:951–960, 1986.

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

[22] Z. Yin, M. Jayaram, S. Pathania, and R. Harshey. The mu transposase interwraps distant DNA sites within a functional transpososome in the absence of DNA supercoiling. *J Biol Chem.*, 280(7):6149–56, 2005.

[23] Z. Yina, A. Suzukia, Z. Loua, M. Jayarama, and R. M. Harshey. Interactions of phage Mu enhancer and termini that specify the assembly of a topologically unique interwrapped transpososome. *J. Mol. Biol.*, 372:382–396, 2007.