

RESEARCH STATEMENT

SOOJEONG KIM

1. INTRODUCTION

My research interests are in application of topology to biology. Currently I work on DNA topology. DNA topology is a study of the topological conformation and path of the DNA sequences in three dimensional space. We predict or confirm this DNA topology by mathematical theory. Especially, I am interested in the topology of the DNA within a protein-DNA complex. In my thesis, I used a tangle analysis, knot theory, graph theory and low dimensional topology to predict the topological shape of DNA in the complex.

2. MOTIVATION

DNA recombination refers to a process in which DNA is rearranged within a genome. There is a 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*. Recombinases recognize the target sites and bind to them. The DNA within the protein-bound DNA complex is cleaved and strands are exchanged before the DNA is resealed. During this recombination process, the topology of DNA is often changed. Thus to know topology of DNA within the protein-DNA complex is very important to understand the recombination and also for successful recombination.

An *n-string tangle* is a three dimensional ball with n -strings properly embedded in it. In late 80's, C. Ernst and D. Summers introduced a tangle model of protein-DNA complexes [4]. This model assumes that the protein is a 3-dimensional ball and protein-bound DNA are strings embedded inside the ball. See Figure 1.

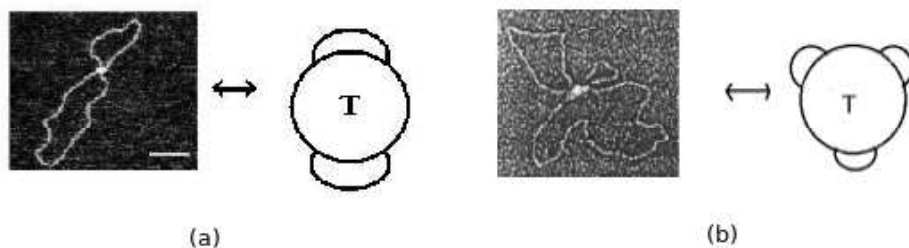


FIGURE 1. (a) AFM image of a Cre synapatic complex with circular DNA [7] and a corresponding 2-string tangle model; (b) Electron microscope image of Hin invertasome formed with circular DNA [5] and a corresponding 3-string tangle model.

Originally tangle model applied to a protein such as Cre which binds two DNA segments. The protein breaks and rejoins the DNA segments and then creates knotted DNA [4]. When this kind of protein complex bounds a circular DNA, after protein binds two DNA segments, there will be two DNA loops outside of DNA-protein complex. Hence we can use 2-string tangle model for this complex (Figure 1 (a)). More recently, Pathania, Jayaram and Harshey predicted that the topological structure within the Mu protein complex consists of three DNA segments containing five crossings [6]. Since Mu binds DNA sequences at 3 sites, the Mu protein-DNA complex can be modeled by a 3-string tangle. In general, 3-string tangle analysis is much complicated than 2-string tangle analysis. Darcy, Leucke and Vazquez analyzed the experimental results of [6] by using 3-tangle analysis [3]. And this is also computationally analyzed in [2].

3. DIFFERENCE TOPOLOGY AND ITS APPLICATION ON MU PROTEIN COMPLEX

Cre is a site-specific recombinases which binds to two target sites. Cre cuts these target sites and changes the topology of the DNA before resealing it again. The local reaction of Cre can be modeled by 2-string tangle as shown in Figure 2.

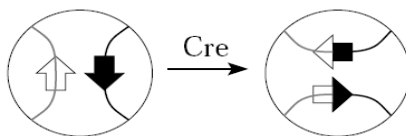


FIGURE 2. A 2-string tangle model for the local action of Cre at target sites.
Figure from [3]

If Cre acts on unknotted DNA not bound by any proteins except for Cre, then the main products of Cre recombinations are unknots and 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 [6] to determine the topological conformation of the DNA bound by Mu. This methodology is called difference topology. Note that every product topology of Mu protein complex after Cre recombination experiments was $(2,p)$ -torus knot or link.

4. MY WORK

From the 3-string tangle analysis of Mu protein-DNA complex, we addressed a possibility that a protein binds DNA sequences at four sites. 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. Such a protein complex bound to a circular DNA is modeled by a 4-string tangle with four loops outside of the tangle (Figure 3 (a)).

In nature, circular DNA is negatively supercoiled and it is plectonemic if the DNA does not wrap a protein. Plectonemic supercoiled DNA is usually branched [1]. Figure 3 (b) is an example of biologically reasonable model of DNA-protein complex involving four DNA

segments. We can generalize this to a mathematical tangle model in Figure 3 (c), where n_i 's are the number of left handed half twists. We call a tangle in Figure 3 (c) a *standard tangle*. Furthermore, we addressed a possibility that a pair of supercoiled DNA branches can be twisted. An example of tangle model of this case is shown in Figure 3 (d). A tangle which is obtained from standard tangle by twisting pairs of branches called *R-standard tangle*.

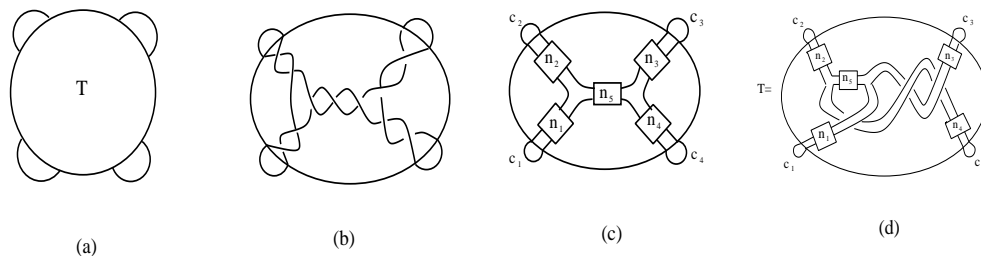


FIGURE 3. (a) A 4-string tangle model of a DNA-protein complex. (b) Examples of 4-string tangle model which are biologically relevant. (c) Standard tangle. (d) An example of R-standard tangle.

For Cre recombination, we need to put target 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. Based on difference topology experiments, we focus on a 4-string tangle model such that all six product topology is $(2,p)$ -torus knot or link.

We define a 4-string tangle T as a *solution tangle* if all six product topology of protein-bound DNA after doing Cre recombination on a pair of outside loops is $(2,p)$ -torus knot or link. See Figure 4.

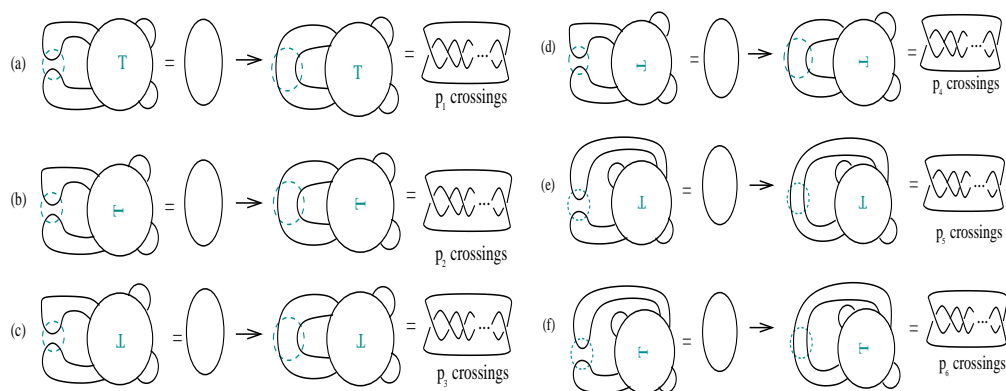


FIGURE 4. A 4-string solution tangle. In (b)~(f), T is rotated. The dotted circle represents a Cre recombinases.

Fianlly, let me introduce our main result and some definitions to understand that:

Definition 1. Two tangles are freely isotopic to each other if they are ambient isotopic allowing the boundary to move.

Definition 2. A *rational tangles* is ambient isotopic to a tangle which has no crossings if we allow the boundary of the three ball to move.

Theorem 1. *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 4 and has less than 8 crossings up to free isotopy, T can be represented by an R-standard tangle. Since all rational tangles are freely isotopic to a tangle which has no crossings, we can find all rational solutions.

Remark. 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.

5. FUTURE PLANS

First, I plan to complete the work started with my dissertation. That is, I would like to extend the number of crossings in Theorem 1. Even though rational tangle is the most biologically relevant tangle model of DNA-protein complexes, mathematically, it is still interesting problem.

Besides that, I’m working on a double branched cover of 3-string tangle. For 2-string tangle, a double branched cover of 3-ball along the tangle is a torus. Thus, the fundamental group (π_1) of a torus is $\mathbb{Z} \otimes \mathbb{Z}$, abelian. However, for 3-string tangle, since a double branched cover is a torus with genus two, π_1 is not abelian and very complicated. I am interested in finding π_1 of double branched cover of general rational 3-tangle, and analyze the DNA conformation within Mu-protein complex by using this theory.

REFERENCES

- [1] A.D. Bates and A. Maxwell. *DNA Topology*. IRL Press, Oxford, 1993.
- [2] 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.
- [3] I.K. Darcy, J. Luecke, and M. Vazquez. Tangle analysis of difference topology experiments: Applications to a Mu protein-DNA complex. Preprint.
- [4] C. Ernst and D. W. Sumners. A calculus for rational tangles: applications to DNA recombination. *Math. Proc. Camb. Phil. Soc.*, 108:489–515, 1990.
- [5] R. Harshey and M. Jayaram. The Mu transpososome through a topological lens. *Crit Rev Biochem Mol Biol.*, 41(6):387–405, 2006.
- [6] 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.
- [7] 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.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF IOWA, IOWA CITY, IOWA 52242.

E-mail address: soojkim@math.uiowa.edu