# Topological tangle modeling of difference topology experiments: tangle analysis of DNA-protein complexes

**Soojeong Kim**
Assistant Professor, Yonsei University, International Campus,
85 Songdogwahak-ro, Yeonsu-gu, Incheon 406-840, Republic of Korea
E-mail: soojkim@yonsei.ac.kr

**Abstract.** An n-string tangle is a three dimensional ball with n strings which are properly embedded in it. Topological tangle analysis of DNA-protein complexes based on difference topology experiments are studied for last 30 years. In this paper, A 2-string tangle analysis of DNA-Tn3 complex [5, 6] and a 3-string tangle analysis of DNA-Mu complex [3, 13] and a biologically relevant 4-string tangle analysis [8, 10, 11] is reviewed. And a biologically relevant n-string tangle analysis is introduced.

## 1. Introduction

Mathematical $n$-string tangle is a three dimensional ball $B$ with $n$-strings which are properly embedded in $B$. I.e, each string's two end points are on the boundary of the ball $B$. In the early 1990's, C. Ernst and D. Sumners used 2-string tangle to analyze the conformation of DNA segments within the Tn3 and Phase$\lambda$ proteins based on N. Cozzarelli's experiments [4, 15, 18, 19]

In late 2000's, a 3-string tangle analysis of DNA topology within Mu-protein is introduced by I. Darcy et. al.[3]. Their work is motivated by S. Pathania et. al.'s difference experiment of Mu-transpososome [13]. I. Darcy et. al. figured out the 3-branched and 5 crossings conformation of DNA within Mu-proteins.

Recently, the author and I. Darcy developed 4-string tangle analysis of DNA-protein complexes [10] and this theory is generalized to n-string tangle analysis which is introduced in this paper.

Some basic concept of difference topology experience of DNA-protein complexes is explained in section 2. In section 3, each of 2, 3 and 4-string tangle modeling of DNA topology and analysis of difference topology experiment is explained. A generalized $n$-string tangle analysis is introduced at the end of this section 3.

## 2. Difference topology experiments of DNA-protein complexes

As widely known, every organism has its genetic information in DNA. In 1953, J. Watson and F. Crick discovered that the structure of DNA is double helical [20]. Here, 'double' means that one DNA string consists of two strands and they are connected to each other. 'Helical' means that the double stranded DNA is coiled around the helical axis at regular interval. Simply speaking, a three dimensional topology of DNA is twisted ladder as in Figure 1(a). Sometimes, to investigate

a three dimensional shape(topology) of DNA, a double stranded DNA is represented by a string (see Figure 1(b)).

In nature, DNA is involved in many biological process including replication, transposition, recombination etc. Due to the double helical structure of it, a long DNA molecule contort into another shape. This is a supercoiling. A simple example of supercoiled DNA is shown in Figure 1(c).

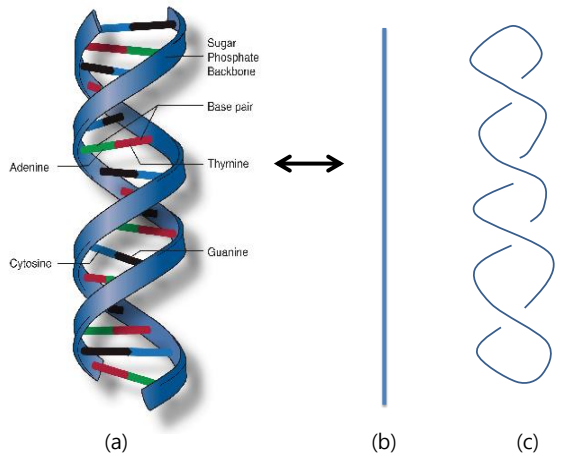


**Figure 1.** (a) The double helical structure of DNA. Figure Courtesy: National Human Genome Research Institute. (b) A double stranded DNA is simply represented by a string. (c) Supercolied DNA.

For performing a biological process, DNA used to bind with some proteins or enzymes. In this process, it is not easy to observe the topological shape of DNA which is bound by proteins. Scientists uses another protein to figure out the conformation of DNA within the proteins. For example, one can use a protein Cre. Cre is a protein which binds to DNA at two specific sites and cut both two sites and rejoin them in different way as in Figure 2. This type of action is called DNA recombination. In Figure 2, initially there is a supercoiled circular DNA with two Cre binding sites on it. There are two different type of DNA recombination reactions are occurred. First, the supercoiled DNA is unwounded and cre binds to its two binding sites. After Cre recombination, DNA conformation is turned out to unlink(two separated rings). Second, a unknown protein binds to the supercoiled DNA , and then Cre binds to its two binding sites. After Cre recombination, the conformation of DNA is changed to Hopf link (two rings with two crossings which cannot be removed by moving each component). Since two different types of Cre recombination experiments results in different DNA conformations, one can notice that the unknown protein used in second experiment wrapped some crossings of DNA. Furthermore, one can guess the number of crossings wrapped is two. This result is from the difference of two DNA topology between two experiments, so it is called 'difference topology experiment'.

## 3. Tangle analysis of difference topology experiments

As I mentioned in section 2, an n-string tangle is a three dimensional ball with $n$-strings properly embedded in it. Especially, if $n$-strings are connected as in Figure 3(a), it is called a 0-tangle(or zero-tangle). Examples of 2-string tangle and 3-string tangle are shown in Figure 3(b) and (c), respectively. Those two tangles in Figure 3(b) and (c) can be isotoped to a 2-string or 3-string zero-tangle if the boundary of the strings are allowed to move along the boundary of the ball. In other words, those tangles are *freely isotopic* to zero-tangle. In general, a *rational tangle* is a tangle which is freely isotopic to the zero-tangle.

C. Ernst and D. Sumners first used 2-string tangle model to analyze DNA-protein complexes based on difference topology experiments [5]. In that model, proteins are represented by a three dimensional ball and DNA within the proteins is represented by strings. They predicted

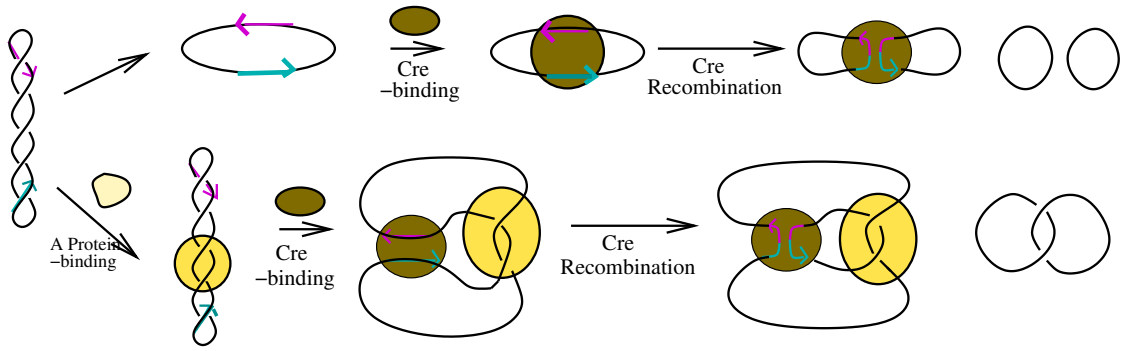


**Figure 2.** Example of difference topology experiments. Figure from [9]

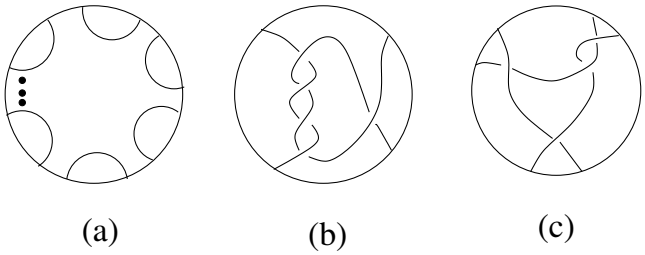


**Figure 3.** (a) $n$-string zero tangle (b) Example of 2-string tangle (c) Example of 3-string tangle

a conformation of DNA within a protein Phase$\lambda$ as in Theorem3.1. In this theorem, R is a two string tangle which represent DNA-Phase$\lambda$ complex. See [5, 6] for more detail.

**Theorem 3.1** *[5, 6] Suppose that 2-string tangles $O_b, R$ satisfy the three tangle equations in Figure 4, then the solution tangles for $O_b, R$ are as in Figure 5*

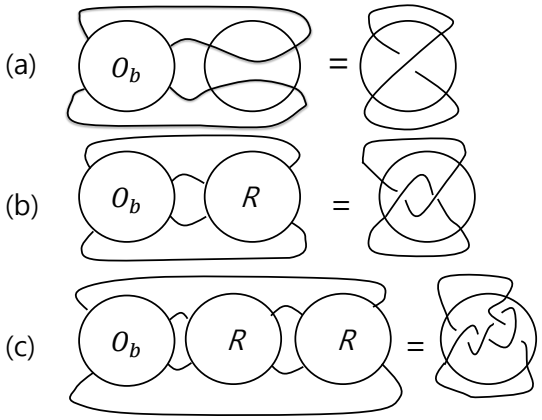


**Figure 4.**

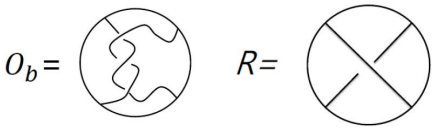


**Figure 5.**

Note that the tangle equations in Figure 4 imply processive Tn3 recombination experiments. For more detail, see [5, 6].

2-string tangle theories are well studied by many mathematicians. However, 3-string tangle analysis is much complicate. For example, a most biologically relevant 2-string tangle model is rational tangle and all 2-string rational tangle is classified. On the other hand, 3-string rational tangles have not been classified yet.

In 2002, Pathania et. al. performed difference topology experiments on DNA-Mu complexes [13]. Mu is a protein which binds to DNA at three specific sites. When Mu binds to a circular DNA with three Mu binding sites, DNA-Mu complex can be modeled by 3-string tangle with three outside loops as in Figure 6(a). Darcy et. al. used 3-string tangle model to support Pathania et. al.'s difference topology experiments of Mu-transpososome as in Theorem 3.2:

**Theorem 3.2** *[3] Let $T$ be a 3-string tangle that satisfies the system of tangle equations in Figure 6(a). If $T$ is freely isotopic to a projection with less than 8 crossings, then $T$ is the tangle in the Figure 6(b).*

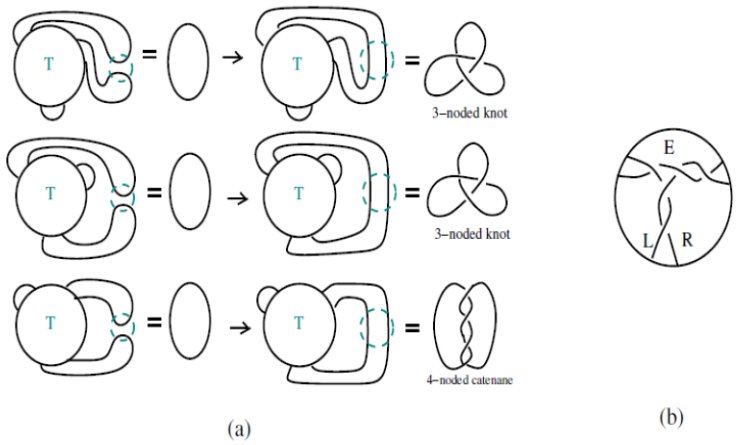


**Figure 6.**

Note that the tangle equations in Figure 6(a) imply difference experiments of Mu-transpososome with Cre recombination, and Figure 6(b) is a 3-string tangle model of DNA within Mu. For more detail, see [13, 3].

Mativated by 3-string tangle analysis, the author and I. Darcy developed a biologically relevant 4-string tangle analysis:

**Theorem 3.3** *[8, 10, 11] Suppose $T$ is 4-string tangle which has less than 8 crossings up to free isotopy. If $T$ is a tangle which satisfies all tangle equations in Figure 7, then $T$ is an R-standard tangle as in Figure 8 after allowing each pair of boundary points to move.*

Recently, the author extended a biologically relevant 3 or 4-string tangle analysis in Theorem 3.3 to $n$-string tangle analysis. The main theory is stated in Theorem 3.4:

**Theorem 3.4** *Suppose $T$ is an $n$-string rational tangle where $n \geq 5$. Let $G$ be a graph with edges $e_i, c_j$ and strings of $T$, as in Figure 9. If $T \cup_{1\leq k\leq n} c_k$ is unknot and $T \cup_{k\neq i,j} c_k$ is $(2,p)$-torus knot/link, then $G$ is planar. I.e, a loop formed by all $c_i$'s and strings of $T$ bounds a disk.*

In this Theorem 3.4, two assumptions about $T(T \cup_{1\leq k\leq n} c_k$ is unknot and $T \cup_{k\neq i,j} c_k$ is $(2,p)$-torus knot/link) imply difference topology experiments of a DNA-protein complex when the protein binds to DNA at $n$ sites. Even though there is no biological experimental data

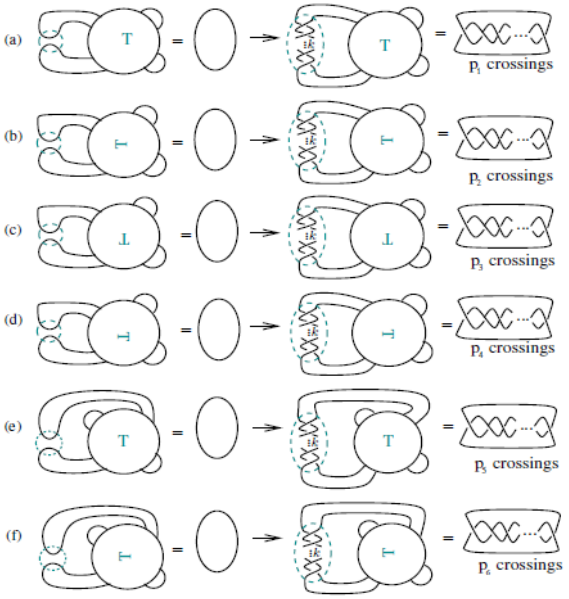


**Figure 7.**

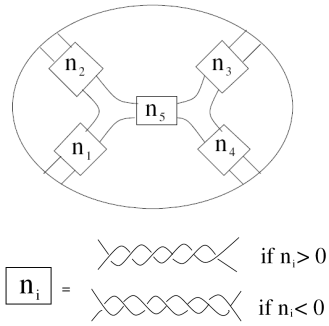


**Figure 8.**

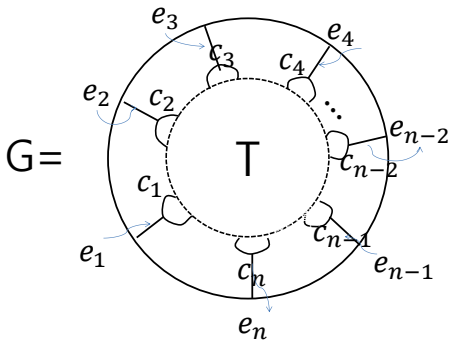


**Figure 9.**

for $n \geq 4$ yet, lots of biological processes (such as replication, transposition, translation, etc.) involve DNA-protein reaction at multiple sites. Hence developing $n$-string tangle analysis is greatly worth.

**Acknowledgments**
The author was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(NRF-2012R1A1A3015842).

## References

[1] Bates A and Maxwell A 1993 DNA topology *IRL press oxford*
[2] Bauer W, Crick F and White J 1980 *Scientific american* **243** 118–133.
[3] Darcy I, Luecke J and Vazquez M 2009 *Algebraic and geometric topology*, , **9** 2247–2309.
[4] Dean F, Stasiak A, Koller T and Cozzarelli N 1985 *J. Biol. Chem.* **260** 4795–4983.
[5] Ernst C and Sumners D 1990 *Math. Proc. Camb. Phil. Soc.* **108** 489–515.
[6] Ernst C and Sumners D 1990 *Math. Proc. Camb. Phil. Soc.* **126** 23–36.
[7] Hirasawa M and Shimokawa K 2000 *Proc. Amer. Math. Soc.* **128** 3445–3451.
[8] Kim S 2011 *J. Ksiam* **15(3)** 161–175.
[9] Kim S 2013 *J. Ksiam* **17(2)** 87–102.
[10] Kim S and Darcy I 2009 *Mathematics of DNA structure, function and interacctions, the IMA volumes in nathematics and its applications, springer science + business media, LLC, New York*
[11] Kim S 2009 *Ph.D thesis* ,
[12] Lewin B 1994 Genes *Oxford University Press, 5th ed., New York*
[13] Pathania S, Jayaram M and Harshey R 2002 *Cell* **109(4)** 425–436.
[14] Saka Y and Vazquez M 2002 *Bioinformatics* **18** 1011–1012.
[15] Spengler S, Stasiak A and Cozzarelli N 1985 *Cell* **42** 325–334.
[16] Sumners D, Ernst C, Cozzarelli N and Spengler S 1995 *Quarterly Reviews of Biophysics* **28**
[17] Vetcher A, Lushnikov A, Navarra-Madsen J, Scharein R, Lyubchenko Y, Darcy I and Levene S 2006 *J. Mol. Biol.*, **357**, 1089–1104.
[18] Wasserman S and Cozzarelli N 1985 *Proc. Nat. Acad. Sci. U.S.A* **82** 1079–1083.
[19] Wasserman S, Dungan J and Cozzarelli N 1985 *Science* **229** 171–174.
[20] Watson J and Crick F 1953 *Nature* **171** 737-738