Iowa Research Online
The University of Iowa's Institutional Repository

University of Iowa
Iowa Research Online

Theses and Dissertations

2010

A 4-string tangle analysis of DNA-protein complexes based on difference topology

Soojeong Kim
University of Iowa

Copyright 2010 Soojeong Kim

This dissertation is available at Iowa Research Online: http://ir.uiowa.edu/etd/528

Recommended Citation
Kim, Soojeong. "A 4-string tangle analysis of DNA-protein complexes based on difference topology." PhD (Doctor of Philosophy) thesis, University of Iowa, 2010.
http://ir.uiowa.edu/etd/528.

Follow this and additional works at: http://ir.uiowa.edu/etd

Part of the Applied Mathematics Commons

A 4-STRING TANGLE ANALYSIS OF DNA-PROTEIN COMPLEXES BASED ON DIFFERENCE TOPOLOGY

by

Soojeong Kim

An Abstract

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Applied Mathematical and Computational Sciences
in the Graduate College of
The University of Iowa

May 2010

Thesis Supervisor: Associate Professor Isabel K. Darcy

ABSTRACT

An n-string tangle is a three dimensional ball with n-strings properly embedded in it. In late the 1980′s, C. Ernst and D. W. Sumner's introduced a tangle model for protein-DNA complexes. This model assumes that the protein is a 3-dimensional ball and the protein-bound DNA are strings embedded inside the ball.

Originally the tangle model was applied to proteins such as Cre recombinase which binds two DNA segments. This protein breaks and rejoins DNA segments and can create knotted DNA. We can use a 2-string tangle model for this complex. More recently, Pathania, Jayaram and Harshey predicted that the topological structure within a Mu protein complex consists of three DNA segments containing five crossings. Since Mu binds DNA sequences at 3 sites, this Mu protein-DNA complex can be modeled by a 3-string tangle. Darcy, Leucke and Vazquez analyzed Pathania et al′s experimental results by using 3-string tangle analysis.

Based on the 3-string tangle analysis of a Mu protein-DNA complex, we addressed the possibility that a protein binds DNA sequences at four sites. Such a protein complex bound to a circular DNA molecule is modeled by a 4-string tangle with four loops outside of the tangle. In this thesis, we determined a biologically relevant 4-string tangle model. We also developed mathematics for solving tangle equations to predict the topology of DNA within the protein complex.

Abstract Approved: _____

Thesis Supervisor

_____

Title and Department

_____

Date

A 4-STRING TANGLE ANALYSIS OF DNA-PROTEIN COMPLEXES BASED ON DIFFERENCE TOPOLOGY

by

Soojeong Kim

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Applied Mathematical and Computational Sciences
in the Graduate College of
The University of Iowa

May 2010

Thesis Supervisor: Associate Professor Isabel K. Darcy

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Soojeong Kim

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Applied Mathematical and Computational Sciences at the May 2010 graduation.

Thesis Committee: _____
Isabel K. Darcy, Thesis Supervisor

_____
Richard Randell

_____
Maggy Tomova

_____
Rodica Curtu

_____
Keiko Kawamuro

ACKNOWLEDGEMENTS

I have received a lot of help and support from many people and institutions. As I finish my thesis, I would like to give my thanks to them.

I want to express my heartfelt gratitude to my advisor, Dr. Isabel K. Darcy, for warm, positive and energetic advise throughout my graduate studies. Her unstinted encouragement led me to finish this thesis. I believe I am a really lucky person to be a student of her.

A special thanks to Dr. John Luecke at the University of Texas and Dr. Mariel Vazquez at the San Franscisco State University for inviting me and giving me a chance to present my research. I sincerely appreciate their interest in my work and very helpful and deep discussions.

I also thanks to my master thesis advisor, Dr. Chan Huh, and other professors, Dr. Jeong-Rae Cho and Dr. Yang Lee, at the Pusan National University in Korea. They always inspired me to further effort in my study and completion of the degree.

I cannot forget to say thank you to my lab colleagues and all my friends in the US and Korea.

Thanks to IMA, mathematics and AMCS department of the University of Iowa for providing assistantships throughout my entire graduate study.

Most of all, I am thankful from all my heart to my parents, my husband Dongsu, my daughter Elina(Yoonju) and all my family in Korea . They aided me both materially and spiritually and were always with me during this long hard journey.

ABSTRACT

An n-string tangle is a three dimensional ball with n-strings properly embedded in it. In late the 1980′s, C. Ernst and D. W. Sumner's introduced a tangle model for protein-DNA complexes. This model assumes that the protein is a 3-dimensional ball and the protein-bound DNA are strings embedded inside the ball.

Originally the tangle model was applied to proteins such as Cre recombinase which binds two DNA segments. This protein breaks and rejoins DNA segments and can create knotted DNA. We can use a 2-string tangle model for this complex. More recently, Pathania, Jayaram and Harshey predicted that the topological structure within a Mu protein complex consists of three DNA segments containing five crossings. Since Mu binds DNA sequences at 3 sites, this Mu protein-DNA complex can be modeled by a 3-string tangle. Darcy, Leucke and Vazquez analyzed Pathania et al′s experimental results by using 3-string tangle analysis.

Based on the 3-string tangle analysis of a Mu protein-DNA complex, we addressed the possibility that a protein binds DNA sequences at four sites. Such a protein complex bound to a circular DNA molecule is modeled by a 4-string tangle with four loops outside of the tangle. In this thesis, we determined a biologically relevant 4-string tangle model. We also developed mathematics for solving tangle equations to predict the topology of DNA within the protein complex.

TABLE OF CONTENTS

LIST OF FIGURES v

CHAPTER

1 INTRODUCTION 1

1.1 Introduction 1
1.2 Organization of thesis 2

2 MATHEMATICAL PRELIMINARIES 3

2.1 Knots and links 3
2.2 Tangles 5
2.3 Low dimensional topology 8

3 BACKGROUND 10

3.1 DNA recombination 10
3.2 DNA topology and the tangle model 11
3.3 Difference topology and its application to Mu 14
3.4 3-String tangle analysis 18

4 4-STRING TANGLE ANALYSIS ON DNA-PROTEIN COMPLEXES 21

4.1 Introduction 21
4.2 4-string tangle analysis 28
4.2.1 Wagon wheel graph G 29
4.2.2 Pentahedral graph $\hat{\mathrm{G}}$ 34
4.2.3 4-valent graph $\Gamma(T)$ 35
4.2.4 Handle addition lemma and montesinos tangles 42
4.2.5 Proof of main theorem 50
4.3 Branched supercoiled DNA solutions 55
4.3.1 Simple branched solution tangles 55
4.3.2 Discussion On complicated branched solution tangles . . . 59
4.3.3 Generalized R-standard solution tangles 61

5 FUTURE DIRECTION 66

REFERENCES 68

LIST OF FIGURES

2.1 Crossings on a knot diagram 4

2.2 Reidemeister moves 5

2.3 Examples of 3-string tangles 6

2.4 Examples of 2-string tangles 7

2.5 Sum of two tangles, $T_1 + T_2$ 8

2.6 Numerator closure of 2-string tangle 9

3.1 Site-specific recombinations 11

3.2 Cre recombination 12

3.3 Tangle models of DNA-protein complexes 13

3.4 Examples of 3-string tangles 13

3.5 Mu transposition. Courtesy: Pathania et al.[27] 15

3.6 Cre recombination on the DNA-Mu protein complex 16

3.7 Tangle model of Mu transpososome 18

3.8 Tangle model of Mu transpososome 19

3.9 3-string tangle solutions 20

4.1 4-string tangle model of DNA-protein complexes 22

4.2 Examples of Cre recombination at *loxP* sites on two outside loops 23

4.3 4-string solution tangle 25

4.4 A standard tnagle and a standard graph 26

4.5 Examples of standard tangle and R-standard tangle 27

4.6 An example of a weighted graph G_R for an R-standard tangle T. 28

4.7 Abstract wagon wheel graph and a wagon wheel graph 30

4.8 $N(G)$ is a regular neighborhood of G in the 3-ball. 31

4.9 An example of a tangle carried by a wagon wheel graph. 31

4.10 A solution tangle carried bya wagon wheel graph and a Cre recombination on a solution tangle 33

4.11 $\hat{\mathrm{G}}$ is a graph abstractly homeomorphic to the 1-skeleton of a pentahedron 35

4.12 4-valent graph. 36

4.13 If T is not split and has only $2n$ vertices, $\Gamma(T)$ has closed circle. 37

4.14 $\Gamma(T)$ with $2n+1$ vertices 38

4.15 There is only one type of $\Gamma(T)$ up to free isotopy after connecting v to v_i, v_j for some $1 \leq i, j \leq 2n$ 38

4.16 Example of two parallel strands 39

4.17 $\Gamma(T)$ with 6 or 7 vertices 40

4.18 Six cases of possible $\Gamma(T)$. Figure from [6] 41

4.19 If s_3 has non-alternating crossings, then T is freely isotopic to the zero tangle. 41

4.20 Possible 3-string tangles which corresponds $\Gamma(T)$ in Figure 4.18 41

4.21 F, F_1 and $\sigma(F_1; \varepsilon_2)$ 43

4.22 F_2 and $\sigma(F_2; J_3)$ 44

4.23 4 different cases of the position of D when $T - s_{ij}$ is split 48

4.24 A simple wagon wheel graph and a standard graph 51

4.25 $\hat{\mathrm{G}}$-b_{ij} can be 4 different graphs 53

4.26 $\hat{\mathrm{G}}$-$b_{ij} - b_{kl}$ can be 2 different graphs 53

4.27 (i, j)-branched weighted graphs for R-standard tangle 56

4.28 Examples of R-standard tangle model of a branched DNA-protein complex corresponding to a weighted graph 56

4.29 Example of R-standard tangle 60

4.30 Writhe and twist 61

4.31 Two ambient isotopic R-standard tangles 62

4.32 G_R 64

5.1 An working model of DNA-MukB protein complex with Topoisomerase II. Figure from [16] 66

CHAPTER 1
INTRODUCTION

1.1 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 [9]. 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 λ integrase. This work was motivated by Nick Cozzarelli [8, 20, 23, 24]. 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.

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 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.

From the 3-string tangle analysis of a Mu protein-DNA complex, we addressed the possibility that a protein binds DNA sequences at four sites. Such a protein complex bound to a circular DNA molecule is modeled by a 4-string tangle with four loops outside of the tangle. In this thesis, we determined a biologically relevant 4-string tangle model. We also developed mathematics for solving tangle equations to predict the topology of DNA within the protein complex.

1.2 Organization of thesis

In chapter 2, we state the necessary definitions and mathematical preliminaries. Background in biology and mathematics is given in chapter 3. In section 3.1, 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 3.2, we introduce tangle analysis of protein-DNA complexes. In section 3.3, we explain the methodology of difference topology and its application to a Mu protein-DNA complex. In section 3.4, we summarize the 3-string tangle analysis of the Mu protein-DNA complex in [6]. Finally, in chapter 4, 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 (See definition 4.3). Some future directions of this research are stated in Chapter 5.

CHAPTER 2
MATHEMATICAL PRELIMINARIES

In this chapter, I state definitions and mathematical preliminaries which I use throughout this thesis. They can be divided to three concepts: knot/link theory, tangle theory and low dimensional topology. We are working in the piecewise linear or smooth category.

2.1 Knots and links

Definition 2.1. [17] A subset K of a space X is a *knot* if K homeomorphic with a sphere S^p. More generally K is a link if K is homeomorphic with a disjoint union $S^{p_1} \cup \cdots \cup S^{p_r}$ of one or more spheres.

Definition 2.2. Let X and Y be topological spaces. A function $h : X \to Y$ is called a *homeomorphism* if h is 1-1, onto, continuous, and h^{-1} is continuous.

Definition 2.3. A function $i : X \to Y$ is an *embedding* if $i : X \to i(X)$ is a homeomorphism.

Definition 2.4. Let X and Y be two topological spaces. If f_1 and f_2 are continuous maps from X to Y, we say that f_1 is *homotopic* to f_2 if there exists a continuous map $F : X \times [0,1] \to Y$ such that $F(x,0) = f_1(x)$ and $F(x,1) = f_2(x)$ for each $x \in X$.

Definition 2.5. A homotopy $F : X \times [0,1] \to X$ is called an *ambient isotopy* if $F(x,0) =$ identity and each $F(x,t)$ is a homeomorphism for $t \in [0,1]$.

Definition 2.6. [17] Two knots or links K, K' in X are *isotopic* if there is an ambient isotopy $F : X \times [0,1] \to X$ such that $F(K,0) = K$ and $F(K,1) = K'$.

Definition 2.7. A *knot projection* is a projection of a knot into the 2-dimensional plane where under and over arcs are not specified. In this projection, no three points correspond to one point on the plane and arcs cross transversly.

Definition 2.8. A knot diagram is a projection where at each crossing the over and under arcs are specified as in Figure 2.1 (a).

If the diagram has an orientation, we can assign +1 or -1 at each crossing as in Figure 2.1

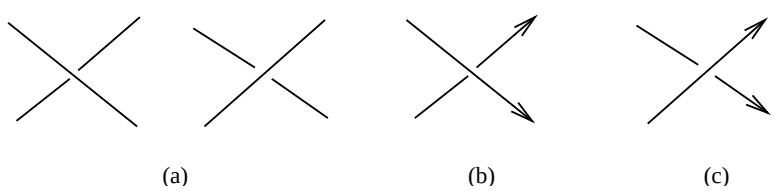

Figure 2.1: Crossings on a knot diagram

(a) Possible crossings on a knot diagram; (b) +1 crossing; (c) -1 crossing

Two knot diagrams are equivalent if they are related by Reidemeister moves (Figure 2.2).

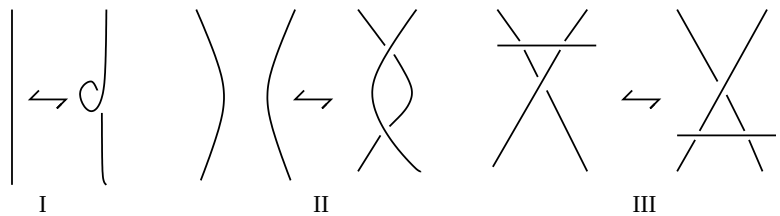

Figure 2.2: Reidemeister moves

2.2 Tangles

Definition 2.9. An *n-string tangle* is a pair (B, t), where B is a 3 dimensional ball and t is a set of arcs embedded in B. The two end points of each arc lie on the boundary of B.

Definition 2.10. Two tangles are *equivalent* if they are ambient isotopic keeping the boundary of B fixed.

Definition 2.11. Two tangles T_1, T_2 are *freely isotopic* if there is an isotopy of the 3-ball taking T_1 to T_2, which is not necessarily fixed on its boundary.

Definition 2.12. If a tangle $T = (B, t)$ is freely isotopic to a tangle with no crossing, then we say T is a *rational tangle*. T is *locally knotted* if there is a 2-sphere S in B that intersect one of the two strings transversely in two points and the string in S is knotted with end points on S. In the case that T is neither rational nor locally knotted, T is called a *prime tangle*.

Examples of 3-string tangles are shown in Figure 3.4.

Definition 2.13. [6] Let T be an n-string tangle. T is *split* if there is a properly embedded disk separating some of its strands from other strands (Figure 2.3(a)).

Strands s_1, s_2 of T are parallel if there is a disk D in B_3 such that $int(D)$ is disjoint from T and $\partial D = s_1 \cup \alpha \cup s_2 \cup \beta$ where $\alpha,\ \beta$ are arcs in ∂B^3 (Figure 2.3(b)).

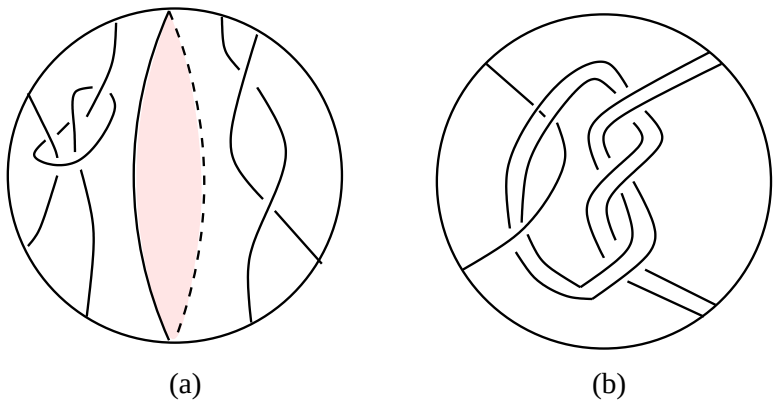

Figure 2.3: Examples of 3-string tangles

(a) Example of a split 3-tangle; (b) Example of a 3-tangle with two parallel strands..

In particular, when $n = 2$, much theory was developed which was initiated by John H. Conway [3]. For 2-string tangles, the zero tangle is the tangle in Figure 2.4 (a). A rational tangle can be obtained by adding alternating horizontal and vertical twists to the zero tangle. By using this fact, Conway developed a notation for the 2-string tangle, $(x_1, ..., x_n)$ where x_i's are integers. This notation means that we start from the zero tangle, add x_1 horizontal twists by moving NE and SE end points, add x_2 vertical twists by moving SW and SE end points, and add x_3 horizontal twists,

etc. Note that the last entry x_n of conway notation $(x_1, ..., x_n)$ must be the number of horizontal twists. A horizontal twist is positive if it is right-handed and a vertical twist is positive if it is left-handed. See an example in Figure 2.4 (b).

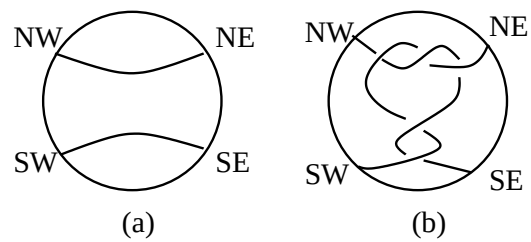

Figure 2.4: Examples of 2-string tangles

(a) The zero tangle; (b) The 2-string tangle, (-3,-2,0) In this figure, NW, NE, SW and SE refer to the four endpoints of the strings.

The unique extended rational number $\dfrac{p}{q} \in \mathbb{Q} \cup \infty$ corresponds to the conway notation of a 2-string rational tangle $(x_1, ..., x_n)$ by the continued fraction as follows:

$$\frac{p}{q} = x_n + \cfrac{1}{x_{n-1} + \cfrac{1}{x_{n-2} + \cfrac{1}{\ddots \; x_2 + \frac{1}{x_1}}}}$$

Conway proved that two rational 2-string tangles are equivalent if and only if the corresponding two extended rational numbers obtained by the continued fraction are the same. We denote a 2-string tangle which has a conway notation corresponding to the extended rational number $\frac{p}{q}$ by $T(\frac{p}{q})$.

We define an operation on 2-string tangles.

Definition 2.14. Let $T_1,\ T_2$ be two 2-string tangles. Then the *tangle sum* $T_1 + T_2$

is defined as in Figure 2.5. In this figure, NW, NE, SW and SE refer to the four endpoints of the strings.

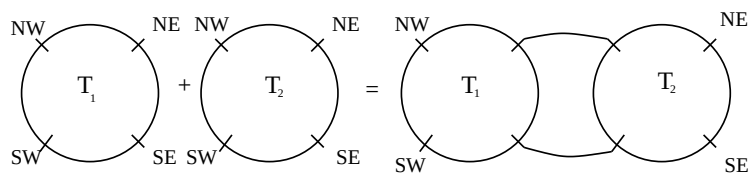

Figure 2.5: Sum of two tangles, $T_1 + T_2$

By using this operation, we can define a different kind of 2-string tangle:

Definition 2.15. [26] A *montesinos tangle* $T(r_1, ..., r_n)$ is a tangle obtained by taking sums of 2-string rational tangles, $T(r_1), ..., T(r_n)$. We always assume r_i is a non-integer.

If we connect the NW endpoint to the NE endpoint and the SW endpoint to the SE endpoint of a tangle as shown in Figure 2.6, we get the numerator closure of a 2-string tangle. We denote the numerator closure of a tangle T as $N(T)$. This can be a knot or a link.

2.3 Low dimensional topology

In this thesis, all surfaces and 3-manifolds are assumed orientable and compact and surfaces in 3-manifolds are assumed properly embedded.

Definition 2.16. [17] We regard an *n-manifold* M^n to be a metric space which may

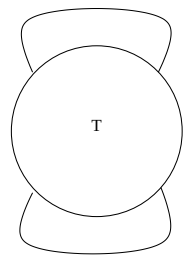

Figure 2.6: Numerator closure of 2-string tangle

be covered by open sets, each of which is homeomorphic with R^n or the half-space R^n_+.

Definition 2.17. A surface F in a 3-manifold M is *compressible* if there is a simple close curve C on F which does not bound a disk in F, but bounds a disk in M; otherwise we say F is *incompressible*.

Definition 2.18. A 3-manifold M is *irreducible* if any 2-dimensional sphere bounds a 3-ball in M; otherwise we say M is *reducible*.

Definition 2.19. A 3-manifold M is *∂-irreducible* if ∂M is incompressible in M.

Definition 2.20. A 3-manifold M is *atoroidal* if M contains no properly embedded incompressible torus.

Theorem 2.1. *[17]* ***Jordan Curve Theorem***

If Γ is a simple closed curve in $\mathbb{R}^2$ then $\mathbb{R}^2 - \Gamma$ has precisely two components, and Γ is the boundary of each.

Corollary 2.2. *[17] If Γ is a simple closed curve in the sphere S^2 then $S^2 - \Gamma$ consists of precisely two components, and Γ is the boundary of each.*

CHAPTER 3
BACKGROUND

3.1 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 broken and rejoined in a different order. 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 3.1(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 3.1(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

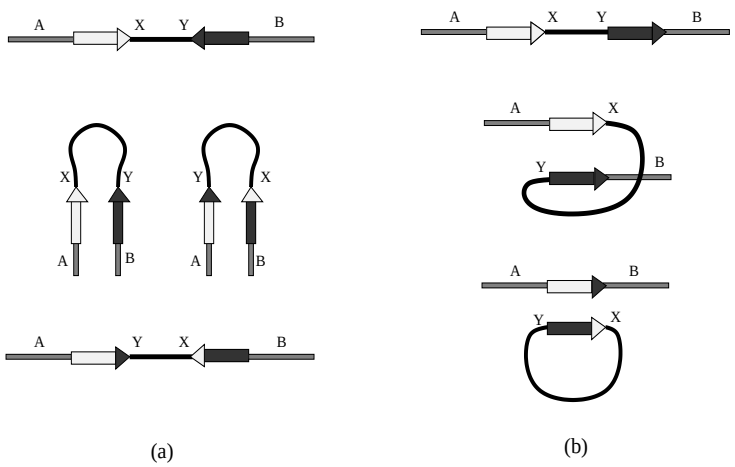

Figure 3.1: Site-specific recombinations

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

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 3.2).

3.2 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 [9]. This model assumes the protein is a three dimensional ball

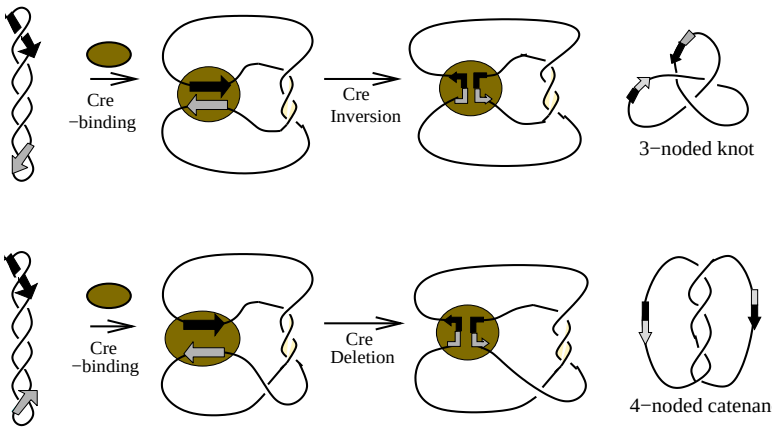

Figure 3.2: Cre recombination

and the protein-bound DNA are strings embedded inside the ball. See Figure 3.3.

Examples of 3-string tangles are given in Figure 3.4. 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 3.1. Figure 3.4 (a)∼(e) give examples of 3-string tangles. Among those, (a), (c) and (d) are examples of rational 3-string tangles.

The original tangle model was applied to proteins which bind two segments of

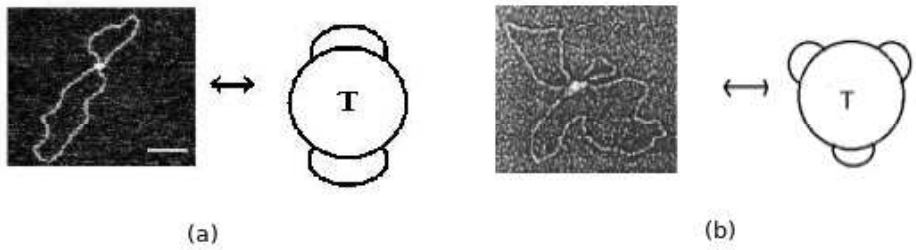

Figure 3.3: Tangle models of DNA-protein complexes

(a)AFM image of a Cre synaptic complex with circular DNA [22] and a corresponding 2-string tangle model; (b) Electron microscope image of Hin invertasome formed with circular DNA [11] and a corresponding 3-string tangle model.

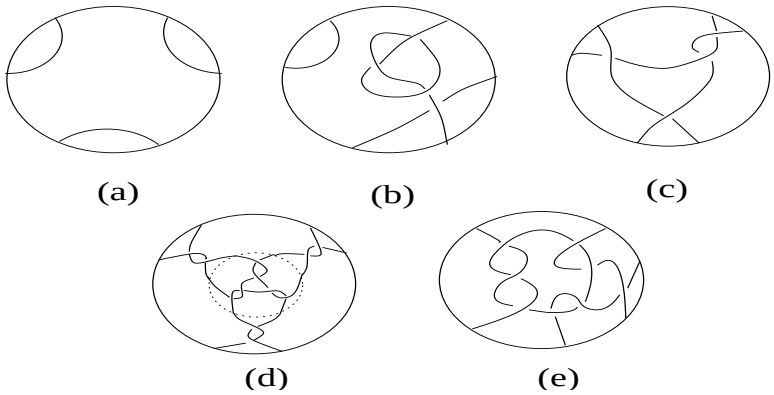

Figure 3.4: Examples of 3-string tangles

(a) Zero tangle; (b) Locally knotted tangle; (c) and (d) Rational tangles; (e) Prime tangle.

DNA and which will break and rejoin segments of DNA creating knotted DNA. For a review of 2-string tangle analysis, see for example [9, 10, 21, 4]. Software has been developed to solve 2-string tangle equations [18, 7] and 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].

3.3 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 [14]. 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 3.5). 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.

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

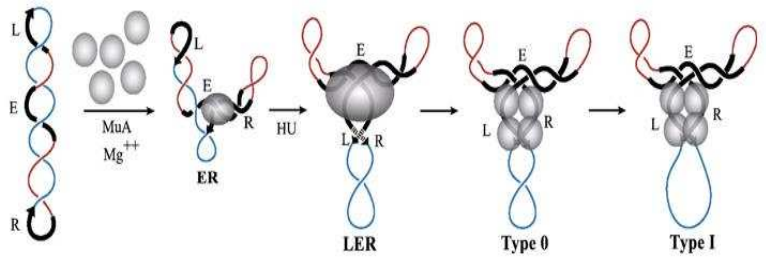

Figure 3.5: Mu transposition. Courtesy: Pathania et al.[27]

unknotted DNA not bound by any proteins except for Cre, then the main products of Cre inversion and deletion 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 [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 3.6 (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.

If the *loxP* sites are placed on the loops indicated in Figure 3.6, but directly

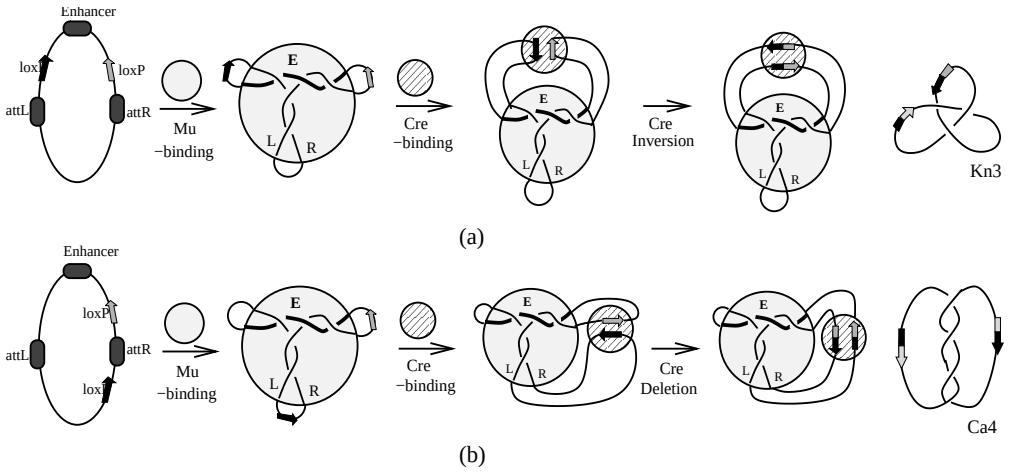

Figure 3.6: Cre recombination on the DNA-Mu protein complex

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 3.7a, 3.8a. In Figure 3.7a, the solution found in [15] is shown while Figure 3.8a 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 3.8a 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 3.7 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 3.8a 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 3.8b [15].

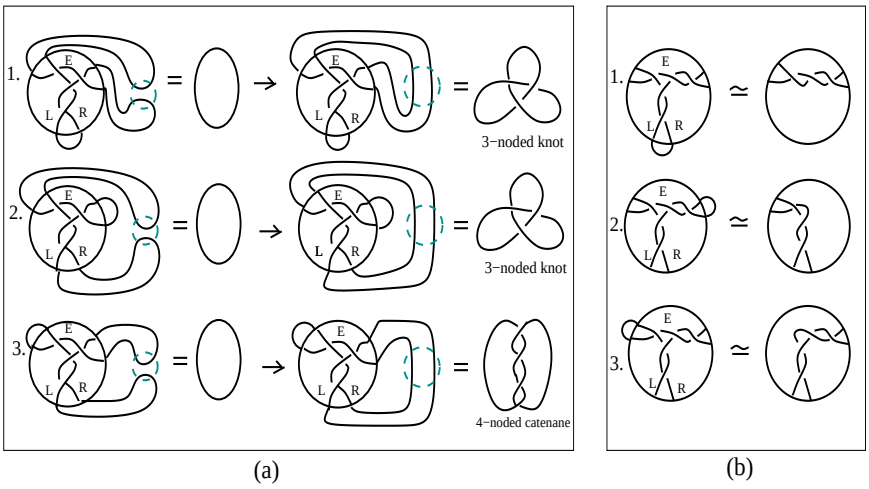

Figure 3.7: Tangle model of Mu transpososome

3.4 3-String tangle analysis

Mathematically, the Mu transpososome 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 3.8a in which the DNA bound by Mu consists of supercoiled DNA with 3 branches and 5 crossings ([15], see section 3.3). Pathania *et al.*'s experimental data [15] was mathematically analyzed by using 3-string tangle analysis [6] without the assumption that the tangle T 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 3.1. *Let T be a 3-string tangle which satisfies the system of tangle equations in Figure 3.8 (a). If T can be freely isotopic to a projection with less than*

8 crossings, then T is the tangle in the Figure 3.8 (b).

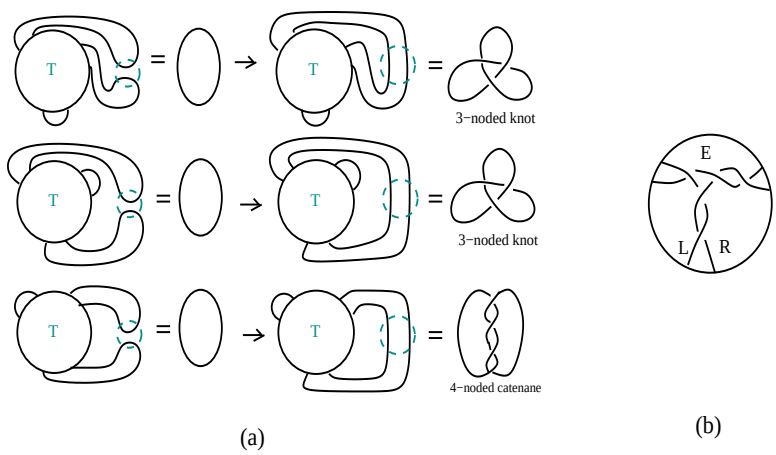

Figure 3.8: Tangle model of Mu transpososome

And this can be generalized to the following Corollary:

Corollary 3.2. *[6] Assume T is a 3-string tangle which can be freely isotoped to a projection with at most seven crossings. If T satisfies the system of tangle equations in Figure 3.9 (a), then T is the tangle in the Figure 3.9 (b).*

Two tangles are freely isotopic to each other if they are ambient isotopic allowing the boundary to move. For example, a rational tangle is freely isotopic to a tangle with no crossings. Thus Proposition 3.1 implies that the only rational tangle solution to the Figure 3.8 (a) equations is that given in Figure 3.8 (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

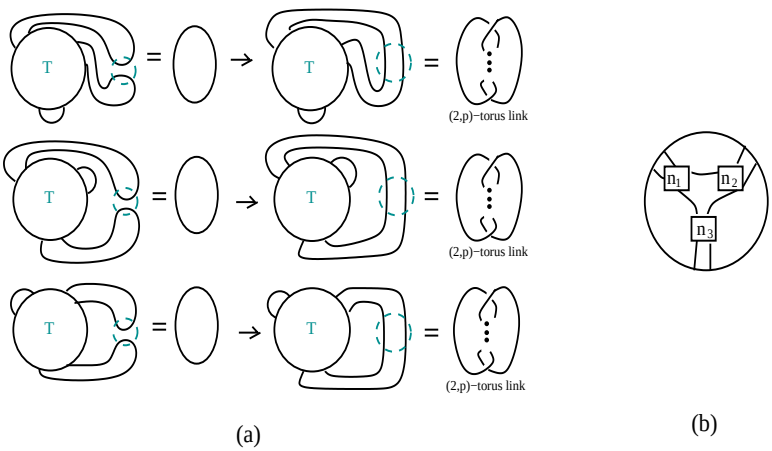

Figure 3.9: 3-string tangle solutions

Generalized 3-string tangle solutions of difference topology experiment. In (b), n_i represents n_i many left handed half twists.

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.

CHAPTER 4
4-STRING TANGLE ANALYSIS ON DNA-PROTEIN COMPLEXES

4.1 Introduction

We do not currently have experimental data for a protein-DNA complex which binds four segments of DNA. 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 4.1 (a)).

In the cell, DNA is usually circular or the ends are constrained. For both cases, DNA is usually negatively supercoiled [1]. Figure 4.1 (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 4.1 (c) shows a biologically natural tangle model of a 4-branched supercoiled DNA-protein complex, where the n_i's are the number of left-handed half twists and c_i's represent four outside loops of DNA.

In this chapter, we would like to extend 3-string tangle analysis (section 3.4) 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,

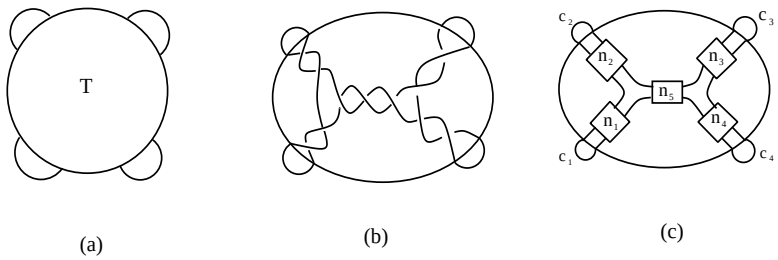

Figure 4.1: 4-string tangle model of DNA-protein complexes

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

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). With our 4-string tangle model of a branched supercoiled DNA-protein complex, the prediction of difference topology (section 3.3) indicates that 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 4.1 (c) is a biologically relevant 4-string tangle model. Assume two *loxP* sites are located on loops c_1 and c_2 of Figure 4.1 (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 on directly repeated sites assuming the tangle model Figure 4.1 (b) results in $(2,4)$-torus link. See Figure 4.2 (a). Similarly, if two *loxP* sites are located on loops c_2 and c_3, Cre recombination on inversely repeated sites results in the $(2,7)$-torus knot as shown in Figure 4.2 (b). In Figure 4.2, the dotted circle represents Cre recombinase and the two thickened arrows represent *loxP* sites. For convenience, Cre is placed on the left side and the 4-string tangle is rotated 90° counterclockwise.

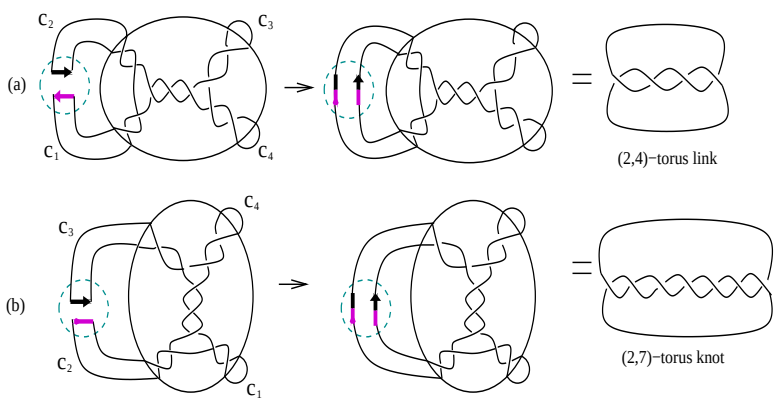

Figure 4.2: Examples of Cre recombination at *loxP* sites on two outside loops (a) Cre deletion at *loxP* sites on the loops c_1 and c_2 of this protein-bound DNA conformation results in a $(2,4)$-torus link; (b) Cre inversion at *loxP* sites on the loops c_2 and c_3 of this protein-bound DNA conformation results in a $(2,7)$-torus knot.

Hence if T is a tangle of the form shown in Figure 4.1 (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. However, the difference topology predicted that the crossing number between inversion knot and deletion link differ by one (see section 3.3). This implies that one extra crossing is introduced by either Cre inversion or deletion on a pair of outside loops. This extra crossing does not affect the DNA conformation within the complex. Hence the six deletion experiments result are enough to decide a 4-string tangle solution of our model of DNA-protein complex. Mathematically this is also a convenient assumption (see section 4.2.1, Theorem 4.2). Therefore, we assume the products of Cre recombination are $(2, p)$-torus links.

This process can be modeled by Figure 4.3. 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 summarize all these assumptions in Figure 4.3 and define a tangle satisfying these conditions a *solution tangle*.

Definition 4.1. A *solution tangle* is a tangle T which is a solution to the system of 6 difference topology equations (Figure 4.3) where the products are $(2, p_i)$ torus links.

We will first discuss branched supercoiled DNA solutions such as a tangle of the form shown in Figure 4.1 (b). A 4-string tangle model of a branched supercoiled

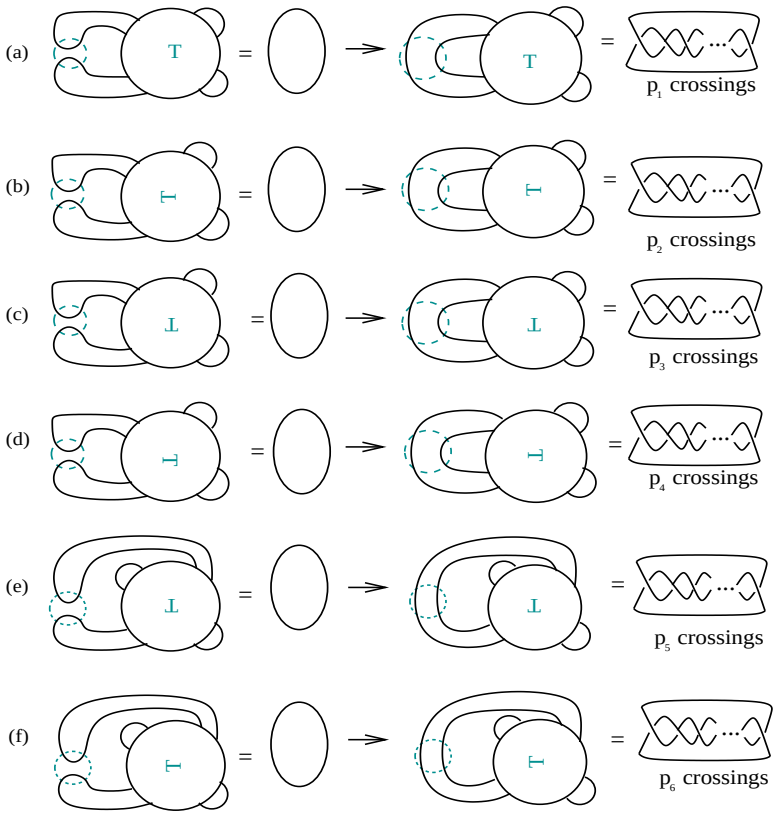

Figure 4.3: 4-string solution tangle

The equations which define a 4-string solution tangle. In (b)∼(d) and (f), T is rotated. The dotted circle represents a Cre recombinases.

DNA-complex can be represented by a weighted graph as in the following definition.

Definition 4.2. A tangle of the form shown in Figure 4.4 (a) will be called *standard*, where n_i is the number of left-handed half twists. Note that a 4-string standard tangle T can be represented by a weighted graph G_s, where G_s is as in Figure 4.4(b). Call this graph G_s a *standard graph*.

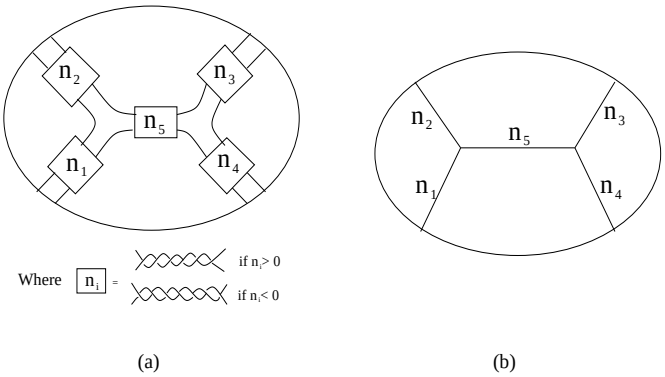

Figure 4.4: A standard tnagle and a standard graph

(a) Standard tangle; (b) A weighted graph G_s 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 4.2) to move?

Definition 4.3. A weighted graph G_R is an *R-standard graph* if it is isotopic to a

standard graph G_s allowing the boundary of G_s to move. A tangle T is *R-standard* if it corresponds to an R-standard graph G_R.

Example 4.1. Examples of 4-string standard tangles are shown in Figure 4.5 (a), (b) and an example of a 4-string R-standard tangle is shown in (b).

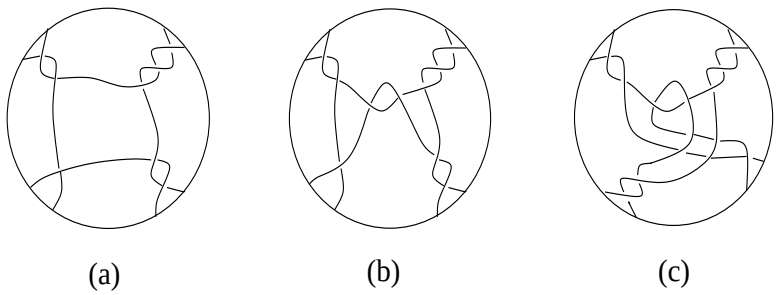

Figure 4.5: Examples of standard tangle and R-standard tangle

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

Example 4.2. Figure 4.6 (a) shows an example of a weighted graph G_R which represents the R-standard tangle T in Figure 4.6 (b).

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 4.3 must be R-standard:

Theorem 4.15 *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 4.3

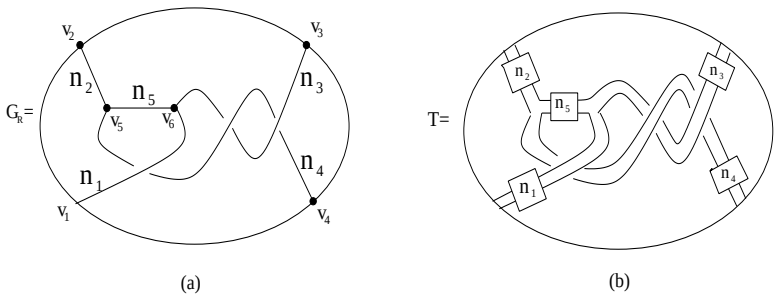

Figure 4.6: An example of a weighted graph G_R for an R-standard tangle T.

and has less than 8 crossings up to free isotopy, T can be represented 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.

4.2 4-string tangle analysis

The main goal of this section is proving our main result, Theorem 4.15. In section 4.2.1, we defind a wagon wheel graph G in B^3 which corresponds to a 4-string solution tangle T. By proving that G can lie in a properly embedded disk in the 3-ball, i.e. G is planar, we conclude that the solution tangle T is R-standard (See section 4.2.5). To prove the planarity of G, we mainly applied M. Scharlemann and A. Thompson's theorem (Theorem 4.4):

Theorem 4.4 [19] *Let $\hat{G}$ be an abstractly planar graph embedded in S^3. $\hat{G}$ is planar if and only if:*

1. *Every proper subgraph of $\hat{G}$ is planar, and*

2. $X(\hat{G})$=$S^3 - int(N(\hat{G}))$ *has compressible boundary.*

Since this theorem is about a graph which lies in S^3, we also define a new graph $\hat{\mathrm{G}}$ in S^3 which acts as same as G in B^3, see section 4.2.2. In the process of proving the planarity of every subgraph of $\hat{\mathrm{G}}$, we have to use Theorem 4.4 several times and its very complicate to show the compressibility of its exterior. To prove this, we used 4-valent graph (section 4.2.3), handle addition lemma and a property of montesinos tangles (section 4.2.4).

4.2.1 Wagon wheel graph G

We start with the following definitions:

Definition 4.4. The *abstract wagon wheel* is the graph with four vertices $\{v_1,\ v_2,\ v_3,\ v_4\}$ and eight edges $\{e_1,\ e_2,\ e_3,\ e_4,\ b_{12},\ b_{23},\ b_{34},\ b_{41}\}$ which are connected in the way of Figure 4.7(a). A graph G is a *wagon wheel graph* if G is a proper embedding of the abstract wagon wheel into a three dimensional ball, B^3 (*e.g.* Figure 4.7(b)). If there is an isotopy of B^3 which takes one wagon wheel graph to another with fixing all end points of G on ∂B^3, we say the two wagon wheel graphs are the *same* [6].

Definition 4.5. [6] If a properly embedded graph lies in a properly embedded disk in the 3-ball, we call it *planar*.

Since a wagon wheel graph G lies in B^3, we can think about a regular neighborhood of it in B^3. From eight edges of G, we get eight meridian disks in the neighborhood of G. The following definition summarizes this.

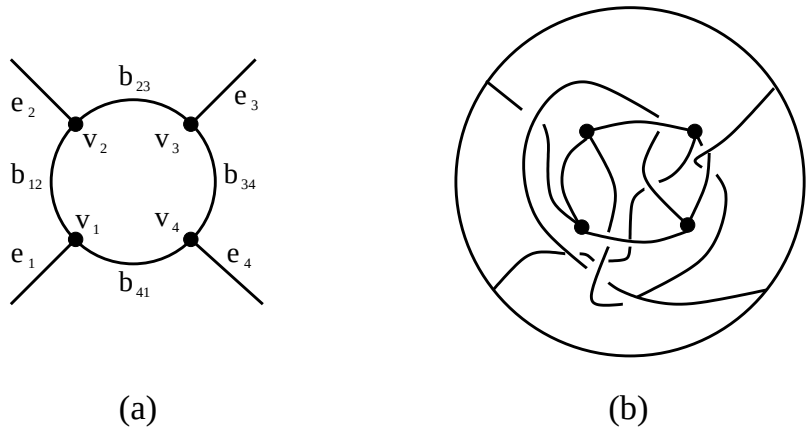

Figure 4.7: Abstract wagon wheel graph and a wagon wheel graph

(a)Abstract wagon wheel graph; (b) A wagon wheel graph.

Definition 4.6. Let G be a wagon wheel graph as in Figure 4.4, and $N(G)$ be a *regular neighborhood* of G in the 3-ball. Define $E_1,\ E_2,\ E_3,\ E_4$ to be meridian disks of $N(G)$ corresponding to edges $e_1,\ e_2,\ e_3,\ e_4$, and let $\varepsilon_i = \partial E_i$. Define $D_{12},\ D_{23},\ D_{34},\ D_{41}$ to be meridian disks of $N(G)$ corresponding to edges $b_{12},\ b_{23},\ b_{34},\ b_{41}$. Let $\delta_{ij} = \partial D_{ij}$. See Figure 4.8.

Definition 4.7. Let $X(G)$ be the *exterior of* G. Then $X(G) = B^3 - (N(G) \cup N(\partial B^3))$. Note that $\partial X(G)$ is a surface of genus 4.

We now think about a tangle T which lies on $\partial N(G)$ in some special way as follows :

Definition 4.8. Let G be a wagon wheel graph. $T(G)$ is a 4-string tangle *carried by* G if the 4-strings $s_{ij},\ s_{jk},\ s_{kl},\ s_{li}$ ($1 \leq i, j, k, l \leq 4$ and each number represents a direction as in Figure 4.9) can be isotoped to lie in $\partial N(G)$ such that

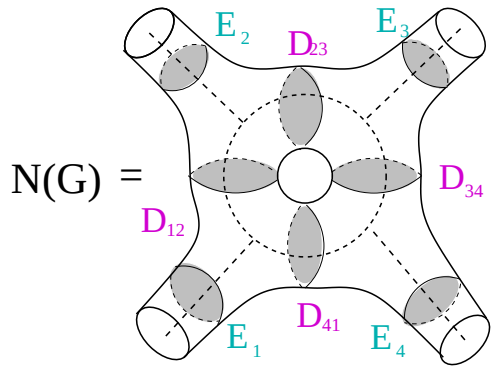

Figure 4.8: $N(G)$ is a regular neighborhood of G in the 3-ball.

- s_{ij} intersects each of $\varepsilon_i,\ \varepsilon_j$ exactly once.
- s_{ij} intersects δ_{ij} exactly once and is disjoint from the remaining δ_{st}.

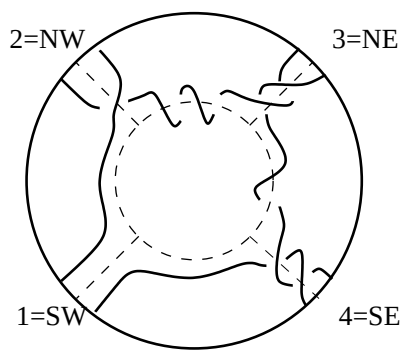

Figure 4.9: An example of a tangle carried by a wagon wheel graph.

As one can see from Definition 4.8 and the example in Figure 4.9, a string s_{ij} in a diagram of $T(G)$ is ambient isotopic to an arc which is the union of three arcs:

an arc wrapping around e_i, a parallel copy of b_{ij} and an arc wrapping around e_j. In fact, s_{ij} can wrap around b_{ij}. However, it is always isotopic to a parallel copy of b_{ij} since s_{ij} only intersects δ_{ij} once and is disjoint from the remaining δ_{st}. Hence we can conclude that $X(G)$ and $X(T(G))$ are homeomorphic. This gives us the following lemma 4.1.

Lemma 4.1. *If $T(G)$ is a tangle carried by a wagon wheel graph G, $\partial X(T(G))$ is compressible in $X(T(G))$ if and only if $\partial X(G)$ is compressible in $X(G)$.*

How are these graphs and the tangles carried by these graphs related to the solution tangles? Let T be a solution tangle. Define c_i, $1 \leq i \leq 4$, as the outside loops of a tangle in Figure 4.3 (a). Then let $C = c_i \cup s_{ij} \cup c_j \cup s_{jk} \cup c_k \cup s_{kl} \cup c_l \cup s_{li}$. Isotope C into the interior of B^3 by pushing each c_i along e_i where e_i is an arc from a point on the boundary of B^3 to C (See Figure 4.10 (a)). Then G is a wagon wheel graph and T is carried by G. Cre recombination on two outside loops of T is related to deleting two edges e_i, e_j of G. For example, Cre recombination on c_1 and c_2 corresponds to pushing c_3 and c_4 into the tangle. Thus this action corresponds to deleting e_3 and e_4 from G. See Figure 4.10 (b).

As we can see from Figure 4.10 (b), a difference topology tangle equation can be expressed by using numerator closure of a sum of 2-string tangles: $N(\frac{0}{1} + T) = N(\frac{1}{0}) \rightarrow N(\frac{1}{0} + T) = N(\frac{2p}{1})$ where T is a solution tangle and $N(\frac{2p}{1})$ is a $(2, p)$-torus link. This implies that T is a rational tangle with only vertical crossings by M.Hirasawa and K.Shimokawa's Theorem [12] and 2-string tangle calculus [9]:

Theorem 4.2. *[12, 9] Let T be a 2-string tangle. If $N(\frac{0}{1}+T) = N(\frac{1}{0})$ and $N(\frac{1}{0}+T) =$*

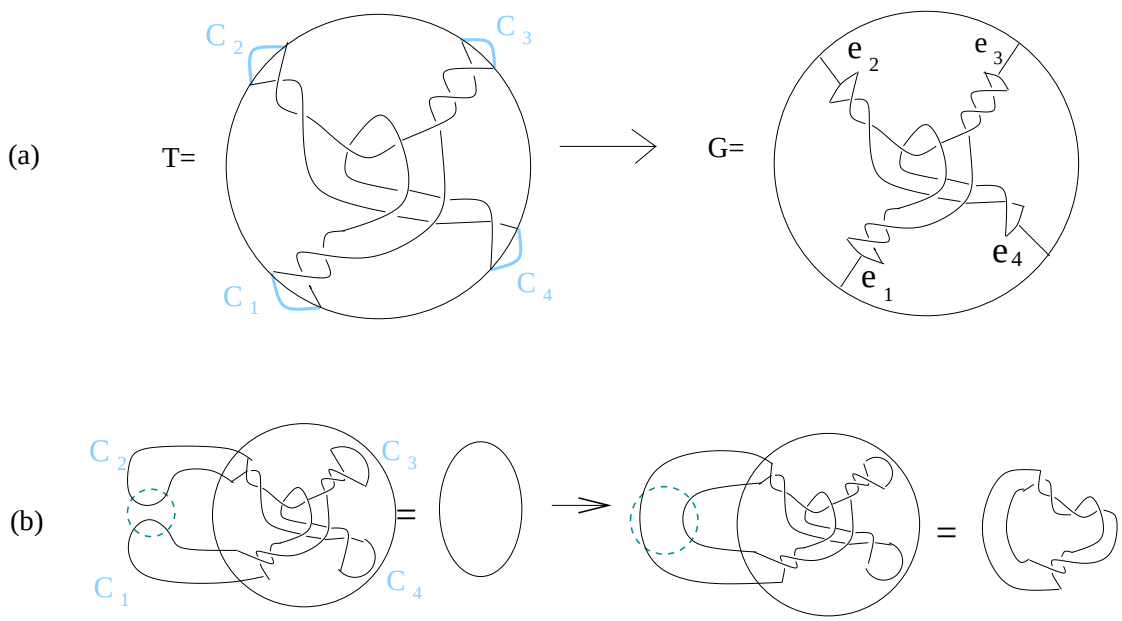

Figure 4.10: A solution tangle carried bya wagon wheel graph and a Cre recombination on a solution tangle

(a) T is a tangle carried by a wagon wheel graph G; (b) Cre recombination on c_1 and c_2 corresponds to capping off c_3 and c_4.

$N(\frac{2p}{1})$, then T is $\frac{1}{n}$ for some integer n.

Hence we get the following lemma:

Lemma 4.3. *Let T be a solution tangle and G be a wagon wheel graph constructed as in Figure 4.10 (a). Then T is carried by G. Furthermore $G - e_i - e_j$ is a planar graph for any $1 \leq i, j \leq 4$ and $i \neq j$.*

The same argument is true for a $(2, 3)$-torus knot product by [13], but false for a $(2, 5)$-torus knot product. These facts show us that it is mathematically preferable to assume that a product topology of a difference topology experiment is a $(2, p)$-torus link.

4.2.2 Pentahedral graph $\hat{\mathrm{G}}$

As we mentioned at the beginning of this section, we will apply Scharlemann and Thompson's theorem about planarity of a graph in S^3:

Theorem 4.4. *[19] Let $\hat{G}$ be an abstractly planar graph embedded in S^3. $\hat{G}$ is planar if and only if:*

1. *Every proper subgraph of $\hat{G}$ is planar, and*

2. *$X(\hat{G}) = S^3 - int(N(\hat{G}))$ has compressible boundary.*

To apply this to our theory, we define a new graph $\hat{\mathrm{G}}$ in S^3 by gluing a 3-dimensional ball to a ball in which a wagon wheel graph lies.

Definition 4.9. Let G be a wagon wheel graph which lies in a 3-ball B^3. Attach another 3-ball $B^{'}$ to B^3. Let $\hat{\mathrm{G}}$ be a new graph in S^3 given by collapsing $B^{'}$ to a

point. $\hat{\mathrm{G}}$ is a graph abstractly homeomorphic to the 1-skeleton of a pentahedron (See Figure 4.11). The edges of $\hat{\mathrm{G}}$ are $e_1,\ e_2,\ e_3,\ e_4,\ b_{12},\ b_{23},\ b_{34},\ b_{41}$ which are also the edges of G.

$\hat{\mathrm{G}}$ is planar if and only if G is planar. Also $X(G)$ and $X(\hat{\mathrm{G}})$ are homeomorphic.

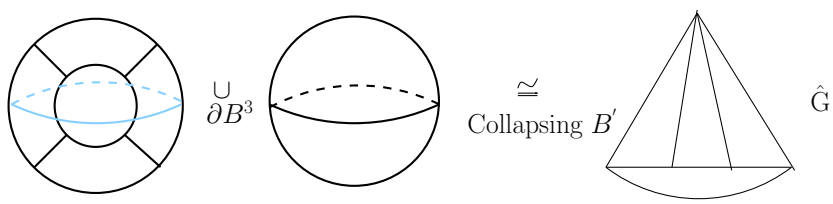

Figure 4.11: $\hat{\mathrm{G}}$ is a graph abstractly homeomorphic to the 1-skeleton of a pentahedron

4.2.3 4-valent graph $\Gamma(T)$

We will use that a 3-string tangle with small number of crossings up to free isotopy is split or has two parallel strands (see Theorem 4.5):

Theorem 4.5. *[6] If T is a 3-string tangle which can be freely isotoped to a projection with at most seven crossings, then either T is split or has two parallel strands.*

This is nice property since if a tangle is split or has two parallel strand, its exterior has compressible boundary (See Figure 2.3). We generalized this property of 3-string tangle to n-string tangle by using 4-valent graphs. Let me start with the definition of a 4-valent graph related to an n-string tangle.

Definition 4.10. [6] To a projection of an n-tangle T, we associate the 4-valent graph

$\Gamma(T)$, that is obtained by placing a vertex at each crossing. If T is not split, we label in sequence $e_1, ..., e_{2n}$ the distinct edges which are incident to the tangle circle. Let $v_1, ..., v_{2n}$ be the vertices of $\Gamma(T)$ which are endpoints of $e_1, ..., e_{2n}$. Define f_i be the face of $\Gamma(T)$ containing $e_i,\ e_{i+1}$.

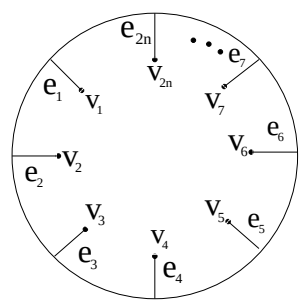

Figure 4.12: 4-valent graph.

We use the following properties of 4-valent graph:

Lemma 4.6. *[6] Assume T is not split. If $v_i = v_j$ for some $i = j$, then the crossing number of its projection can be reduced by free isotopy.*

Lemma 4.7. *[6] No two edges of f_i correspond to the same edge of $\Gamma(T)$. If two vertices of f_i correspond to the same vertex of $\Gamma(T)$, then the crossing number of the projection can be reduced by an isotopy fixed on the boundary. Finally, if v_j is incident to f_i, then $j \in i,\ i+1$ or a crossing can be reduced by a free isotopy.*

The following theorem is the generalization of Theorem 4.5 to an n-string tangle.

Theorem 4.8. *If T is a n-string tangle with at most $2n+1$ crossings up to free isotopy, then either T is split or has two parallel strands.*

Proof. Assume T is freely isotoped to a minimal crossing projection. Suppose T is not split. Then $\Gamma(T)$ is as in Figure 4.12. If there are two edges e_i and e_j such that $v_i = v_j$, then the crossing number of T can be reduced by free isotopy. This is a contradiction to the assumption that T is a minimal crossing projection. Therefore, there should be at least $2n$ vertices in $\Gamma(T)$. If $v_1, ..., v_{2n}$ are the only vertices, then by lemma 4.7, $\Gamma(T)$ must be as in Figure 4.13. This implies that there is a closed circle and this cannot happen in T by the definition of n-string tangle. Hence there will exist at least one more vertex v. From the fact that the crossing number of T is at most $2n+1$, v is the only vertex other than $v_1, ..., v_{2n}$ (Figure 4.14 (a)).

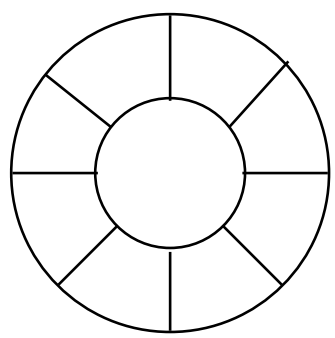

Figure 4.13: If T is not split and has only $2n$ vertices, $\Gamma(T)$ has closed circle.

Then v will be connected to v_i, v_j for some $1 \leq i, j \leq 2n$ (Figure 4.14(b)). By the assumption of minimal crossing projection, we can connect all vertices without

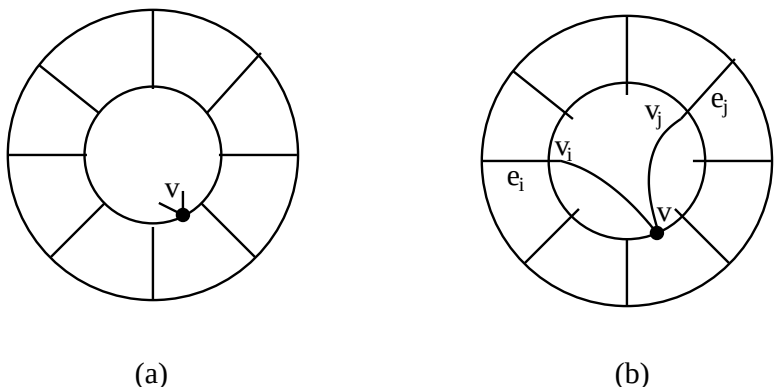

Figure 4.14: $\Gamma(T)$ with $2n+1$ vertices

(a) Since $\Gamma(T)$ has at least $2n$ and at most $2n+1$ vertices, there is the only vertex v other than $v_1, ..., v_{2n}$; (b) v is connected to v_i and v_j for some $1 \leq i, j \leq 2n$

making extra crossings. After connecting all vertices by edges, we get a 4-valent graph up to free isotopy as in Figure 4.15.

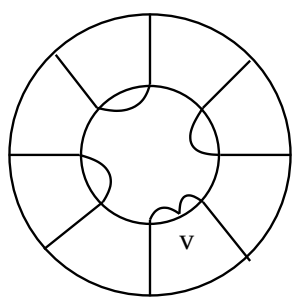

Figure 4.15: There is only one type of $\Gamma(T)$ up to free isotopy after connecting v to v_i, v_j for some $1 \leq i, j \leq 2n$

Note that we see two parallel strands in this case. See Figure 4.16.

$\square$

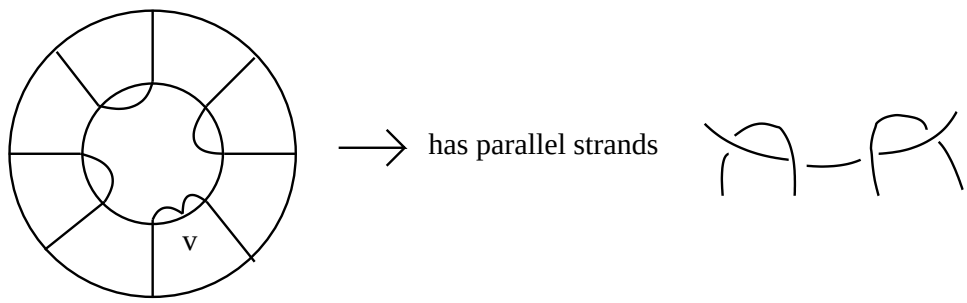

Figure 4.16: Example of two parallel strands

From the proof of Theorem 4.5 [6], we find an interesting fact that a 3-string tangle with at most seven crossings up to free isotopy has a certain type of conformation:

Corollary 4.9. *Assume T is a 3-string tangle which can be freely isotoped to a projection with at most seven crossings. If T is not split, then T is freely isotopic to the tangle in Figure 4.20 (c) or its mirror image.*

Proof. (This proof is motivated by the proof of Theorem 4.5 in [6].) Assume T is not split and is freely isotoped to a minimal crossing projection. By lemma 4.6, T should have at least 6 vertices $v_1 \cdots v_6$. From lemma 4.7, $\Gamma(T)$ must be as in Figure 4.17(a) if $v_1 \cdots v_6$ are the only vertices. Then there would be a closed curve in T. This is a contradiction. Hence there must be at least one more vertex, say v, besides $v_1 \cdots v_6$ as in Figure 4.17 (b).

We assumed that T has at most seven crossings up to free isotopy, so $v, v_1, \cdots, v_6$ are the only vertices on $\Gamma(T)$. Considering all possibilities where there is no closed curve, we get 9 cases. Six cases are as in Figure 4.18 while the other three cases can

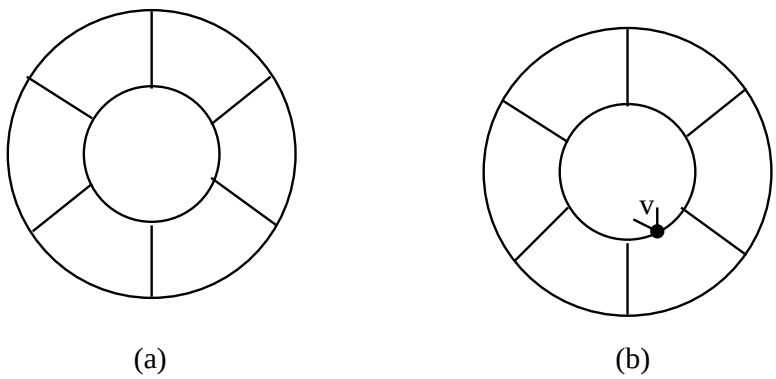

Figure 4.17: $\Gamma(T)$ with 6 or 7 vertices

(a) When $\Gamma(T)$ has exactly 6 vertices, there would be a closed curve in T; (b) v is another vertex besides $v_1 \cdots v_6$. Figure from [6]

be obtained from the top three via reflection.

Note that if T is freely isotopic to a tangle with a 4-valent graph shown in Figure 4.18, then it is freely isotopic to a tangle with any of the 4-valent graphs in Figure 4.18. We also notice that T has 2 strings (say $s_1,\ s_2$) with 2 crossings and one string (say s_3) with 3 crossings. If the crossings of s_3 are not alternating, then T will be a tangle with no crossing up to free isotopy. One example is shown in Figure 4.19. This contradicts the minimal crossing projection assumption. Hence the crossings of s_3 should be alternating and thus s_3 will be a knotted component of T.

If the crossings of s_1 or s_2 are not alternating, then T is freely isotopic to one of the tangles in Figure 4.20 (a), (b), or their mirror images. This also contradicts the minimal crossing projection assumption. Hence T is isotopic to the tangle in Figure 4.20 (c) or its mirror image.

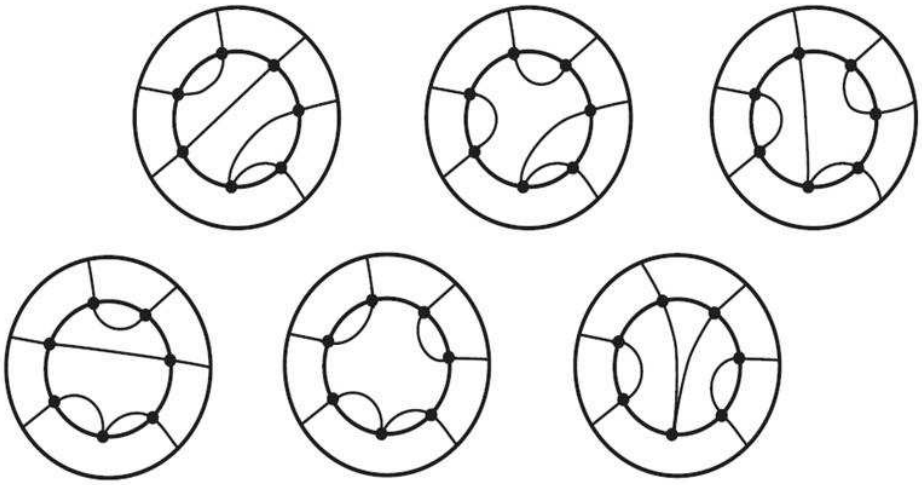

Figure 4.18: Six cases of possible $\Gamma(T)$. Figure from [6]

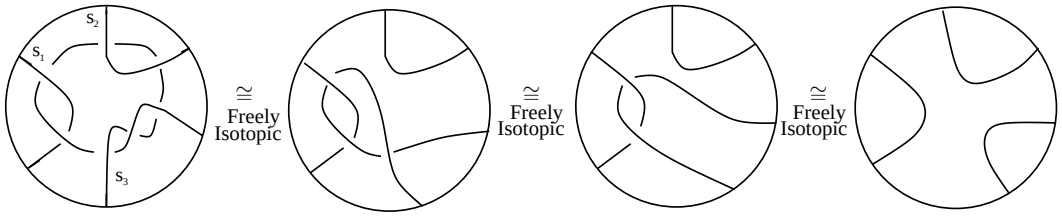

Figure 4.19: If s_3 has non-alternating crossings, then T is freely isotopic to the zero tangle.

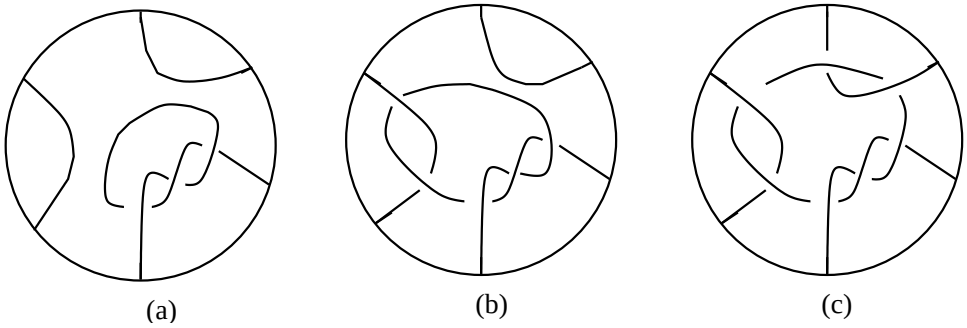

Figure 4.20: Possible 3-string tangles which corresponds $\Gamma(T)$ in Figure 4.18

(a) Neither of $s_1,\ s_2$ are linked with s_3; (b) One of $s_1,\ s_2$ is linked with s_3; (c) Both of $s_1,\ s_2$ are linked with s_3.

Hence we can conclude that if T is not split, then it is freely isotopic to the tangle in Figure 4.20 (c) or its mirror image. $\square$

4.2.4 Handle addition lemma and montesinos tangles

In this section, we use the handle addition lemma (Lemma 4.10) to find a compressing disc in $X(G)$. By using that disc and and a property of montesinos tangles (Theorem 4.13), we proved the compressibility of the exterior of subgraphs of G.

Definition 4.11. [25] Let M be an orientable, irreducible 3-manifold and F a surface in ∂M. Given a simple closed curve J on F, let $\tau(M; J)$ be the manifold obtained by attaching a 2-handle $D^2 \times I$ to M so that $\partial D^2 \times I$ is identified with a regular neighborhood of J in F. Let $\sigma(F; J)$ be the surface $(F - \partial D^2 \times I) \cup (D^2 \times \partial I)$ on the boundary of $\tau(M; J)$.

The following is the handle addition lemma:

Lemma 4.10. *[25] [Handle Addition Lemma] Let M be an orientable, irreducible 3-manifold and F be a surface in ∂M. Let J be a simple, closed curve in F. Assume $\sigma(F; J)$ is not a 2-sphere. If F is compressible in M but $F - J$ is incompressible, then $\sigma(F; J)$ is incompressible in $\tau(M; J)$.*

From the handle addition lemma, we can find a compressing disc of $X(G)$ which does not intersect with every ε_i:

Lemma 4.11. *Let G be a wagon wheel graph and $X(G)$ its exterior. Let ε_i be the meridian curves on $\partial X(G)$ as in Definition 4.6. Suppose G carries a 4-string tangle*

T which has less than 8 crossings up to free isotopy. If $T(G)$ is a solution tangle, then $\partial X(G) - (\varepsilon_1 \cup \ \varepsilon_2 \cup \varepsilon_3 \cup \varepsilon_4)$ is compressible in $X(G)$.

Proof. Let M be $X(G) = B^3 - (N(G) \cup N(\partial B^3))$. Then M is an orientable, irreducible 3-manifold.

Step 1 : Let $F = \partial X(G)$ and $J_1 = \varepsilon_1$.

Then F is surface in ∂M and J_1 is a simple, closed curve in F. Since $T(G)$ has less than 8 crossings up to free isotopy, $T(G)$ is split or has two parallel strands by Theorem 4.8. Hence $\partial X(T(G))$ is compressible in $X(T(G))$. Thus $\partial X(G)$ is compressible in $X(G)$ by Lemma 4.1. I.e., F is compressible in M. Since $\sigma(F; J_1) = \sigma(\partial X(G); \varepsilon_1)$ is homeomorphic to the exterior of a 3-string tangle with less than 8 crossings up to free isotopy which is thus split or has two parallel strands, $\sigma(F; J_1)$ is compressible in $\tau(M; J_1)$. By Lemma 4.10, $F - J_1 = F - \varepsilon_1$ is compressible in M.

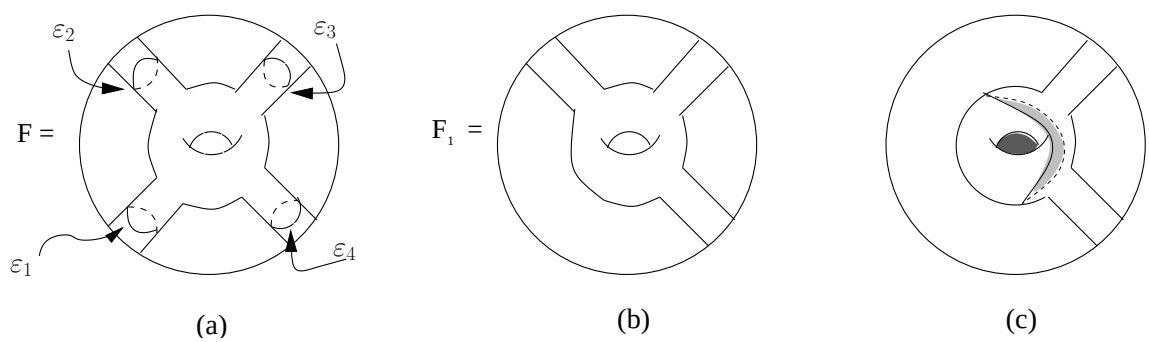

Figure 4.21: F, F_1 and $\sigma(F_1; \varepsilon_2)$

(a) $F = \partial X(G)$; (b)$F_1 = \sigma(F; \varepsilon_1)$; (c) $\sigma(F_1; \varepsilon_2)$. Shaded two disks are examples of compressing disks on $\partial X(T(G - e_1 - e_2))$.

Step 2 : Let $F_1 = \sigma(F; \varepsilon_1)$ and $J_2 = \varepsilon_2$.

$\sigma(F_1; J_2)$ is homeomorphic to $\partial X(T(G - e_1 - e_2))$. By Lemma 4.3 $G - e_1 - e_2$ is planar. Hence there is a compressing disk D on $\partial X(T(G - e_1 - e_2))$ (see Figure 4.2.4 (c)). Thus $\sigma(F_1; J_2)$ is compressible in $\tau(M, J_2)$. By Lemma 4.10, $F_1 - J_2$ is compressible in M. From the figure 4.22 (a), we can see the following relation: $F_1 - J_2$ $=\sigma(\partial X(G); \varepsilon_1) - \varepsilon_2 = \sigma(\partial X(G) - \varepsilon_2; \varepsilon_1) = \sigma(F - J_2; J_1)$.

In step 1, we proved $F - J_1$ is compressible in M. By a similar argument, we can prove that $F - J_2$ is compressible in M. Now, we know $F - J_2$ is compressible in M and $\sigma(F - J_2; J_1)$ is compressible in $\tau(M; J_1)$. Hence $F - J_1 - J_2$ is compressible in M.

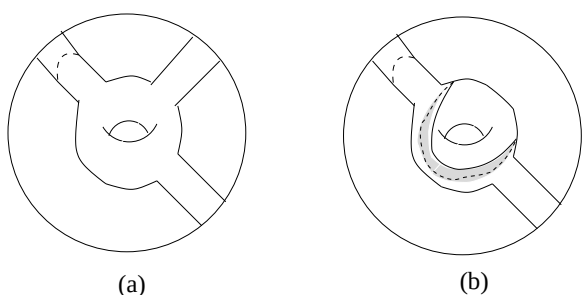

Figure 4.22: F_2 and $\sigma(F_2; J_3)$

(a) $F_2 = \sigma(\partial X(G); \varepsilon_1) - \varepsilon_2 = \sigma(\partial X(G) - \varepsilon_2; \varepsilon_1)$; (b) $\sigma(F_2, J_3)$. The shaded disk is an example of a compressing disk on $\partial X(T(G - e_1 - e_3)) - \varepsilon_2$.

Step 3 : Let $F_2 = \sigma(\partial X(G) - \varepsilon_2; \varepsilon_1)$ and $J_3 = \varepsilon_3$.

We showed that F_2 is compressible in step 2 and J_3 is a simple, closed curve. $\sigma(F_2, J_3)$

is a surface which is homeomorphic to $\partial X(T(G - e_1 - e_3)) - \varepsilon_2$. Since $G - e_1 - e_3$ is planar by the definition 4.1, $\sigma(F_2, J_3)$ is planar and thus compressible with a compressing disk D as in Figure 4.22 (b). By the Lemma 4.10, $F_2 - J_3$ is compressible. Similar to step 2, $F_2 - J_3 = \sigma(\partial X(G) - \varepsilon_2; \varepsilon_1) - \varepsilon_3 = \sigma(\partial X(G) - \varepsilon_2 - \varepsilon_3; \varepsilon_1)$ and hence $\partial X(G) - \varepsilon_1 - \varepsilon_2 - \varepsilon_3 = F - J_1 - J_2 - J_3$ is compressible in M.

Step 4 :Let $F_3 = \sigma(\partial X(G) - \varepsilon_2 - \varepsilon_3; \varepsilon_1)$ and $J_4 = \varepsilon_4$.

By step 3, F_3 is compressible and J_4 is a simple closed curve. $\sigma(F_3; J_4)$ is a subsurface of $\sigma(\sigma(F; \varepsilon_1); \varepsilon_4)$ which is planar and hence compressible. Hence $F_3 - J_4 = \sigma(F - \varepsilon_2 - \varepsilon_3 - \varepsilon_4; \varepsilon_1)$ is compressible. By step 3 and the Lemme 4.10, $F - J_1 - J_2 - J_3 - J_4$ is compressible in M. In other words, $\partial X(G) - (\varepsilon_1 \cup \varepsilon_2 \cup \varepsilon_3 \cup \varepsilon_4)$ is compressible in $X(G)$. □

By surgering along the disc in Lemma 4.11, we get another compressing disc in $X(G)$ (G is carrying a solution tangle) which intersect each δ_{ij} algebraically once and is still disjoint from all ε_i's.

Lemma 4.12. *Let ε_i and δ_{ij} be the meridian curves on $\partial X(G)$ in Definition 4.6. Suppose G carries 4-string solution tangle T which has at most 7 crossings up to free isotopy. If $T(G)$ is a solution tangle, then there is a properly embedded disk in $X(G)$ that intersects each δ_{ij} algebraically once in $\partial X(G)$ and is disjoint from the ε_i.*

Proof. (The proof of this lemma is similar to the proof of Lemma 4.11 of [DLV]) Let D' be a disk from Lemma 4.11 lying in $X(G)$ such that $\partial D' \subset \partial X(G) - \cup_{i=1}^{4} E_i$ and D' is compressible in $X(G)$. Let M be a regular neighborhood of $G \cup \partial B^3$. Then B^3

can be written by $X(G) \cup_{\partial X(G)} M$. The closure of $M - \cup_{i=1}^{4} E_i$ has two components, say M_1 and M_2. Then $M_1 = S^2 \times I$ and M_2=solid torus.

Suppose $\partial D^{'}$ lies in ∂M_1, then $\partial D^{'}$ will bound a disk $D^{''}$ on ∂M_1. We can form a sphere $S = D^{'} \cup_{\partial} D^{''}$. After slightly pushing the interior of $D^{''}$ into M_1, we can say $\partial D^{'} = \partial X(G) \cap S$. Because $D^{'}$ is a compressing disc on $\partial X(G) - \varepsilon_1 - \varepsilon_2 - \varepsilon_3 - \varepsilon_4$, there is an arc α on ∂M_1 which connects E_i and E_j for some $i,\ j$ $(1 \leq i, j \leq 4)$ such that it intersects $\partial D^{'}$ algebraically once. On the other hand, there is an arc β on M_2 which connects E_i and E_j, but is disjoint from $\partial D^{'}$, hence missing S. Thus we can get a simple closed curve $\alpha \cup \beta$ which intersects S once, i.e., S is non-separating. This is a contradiction to a Corollary of the Jordan Curve Theorem (see Theorem 2.2). Hence $\partial D^{'}$ must lie in ∂M_2.

Now, $\partial D^{'}$ is a simple closed curve on a boundary of a solid torus M_2. Suppose $\partial D^{'}$ does not bound a disk in ∂M_2. Then $\partial D^{'}$ can be expressed by $a \cdot m + b \cdot l$ where m is a meridian, l is a longitude of M_2 and $a,\ b$ are relatively prime integers. If $|b| \neq 1$, then B^3 contains a lens space summand. In other words, $\partial D^{'}$ must intersect any meridian of M_2 algebraically once. In this case, $D^{'}$ is the desired disk.

Assume $\partial D^{'}$ bounds a disk $D^{''}$ in ∂M_2. Then the intersection of a neighborhood of $\cup_{i=1}^{4} E_i$ and ∂M_2 must lie in $D^{''}$ as $S = D^{'} \cup D^{''}$ is separating. Since M_2 is an unknotted solid torus (our solution tangle with 4 outside loops is a model of an unknotted circular DNA-protein complex), there exists a disc D properly embedded in $B^3 - int M_2$ such that ∂D= longitude of ∂M_2 and hence essential. Since intersection of D and $D^{''}$ is a set of disks, we can arrange that D is disjoint from $D^{''}$. Then $D \cap S$

is a set of trivial circles in D' and D. Hence after surgeries which take off the trivial circles from D without changing ∂D, D is a disk properly embedded in $X(G)$ since all ε_i's lie on opposite side of S from D. Hence D is the desired disk.

□

To determine the compressibility of the exterior of subgraphs of $\hat{\mathrm{G}}$, we have to determine the compressibility of $\hat{\mathrm{G}}-b_{ij}$ and $\hat{\mathrm{G}}-b_{ij}-b_{kl}$. This follows from showing the compressibility of the exterior of $T-s_{ij}$ and $T-s_{ij}-s_{kl}$ where T is a tangle carried by G. We use Lemma 4.12 and the following Theorem 4.13 about montesinos tangles developed by Y. Wu:

Theorem 4.13. *[26] Suppose $T = T_1 + T_2$ is a nontrivial sum of atoroidal tangles. Then T is a nontrivial tangle, and it is ∂-reducible if and only if, up to relabeling, T_1 is a 2-twist tangle, and T_2 is a rational tangle, in which case exterior of T is a handlebody.*

The following Lemma 4.14 summarizes the compressibility of the exteriors of $T,\ T-s_{ij}$ and $T-sij-s_{kl}$ ($sij,\ s_{kl}$ are the strings of G).

Lemma 4.14. *Let T be a 4-string solution tangle which can be freely isotoped to a projection with less than eight crossings. Then $X(T), X(T-s_{ij})$ and $X(T-s_{ij}-s_{kl})$ have compressible boundary where $s_{ij},\ s_{jk},\ s_{kl},\ s_{li}$ are the 4 strings of T, $1 \leq i,j,k,l \leq 4$.*

Proof. We assume T is freely isotoped to a minimal crossing projection. By Theorem 4.8, T is split or has two parallel strands. Hence $X(T)$ has a compressing disk, see

Figure 2.3 (b).

Since T is a 4-string tangle with less than 8 crossing up to free isotopy, $T - s_{ij}$ is a 3-string tangle which has also less than 8 crossings up to free isotopy. By Theorem 4.5, $T - s_{ij}$ is split or has two parallel strands. In either case, $X(T - s_{ij})$ has a compressible boundary.

To prove the compressibility of $T - sij - s_{kl}$, we consider two cases:

Case 1. $T - s_{ij}$ is split. Let D be a compressing disc in the exterior of $T - s_{ij}$ which splits 3 strings into two and one on each side of D. All cases can be drawn as in Figure 4.2.4.

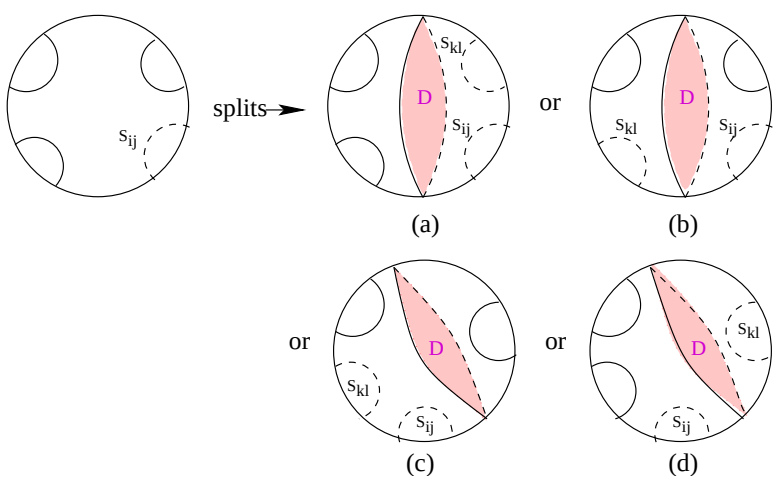

Figure 4.23: 4 different cases of the position of D when $T - s_{ij}$ is split

For the cases (b) and (c) of the Figure 4.2.4, $T - s_{ij} - s_{kl}$ still split. Thus, the compressing disk D of $X(T - s_{ij})$ is still a compressing disc of $X(T - s_{ij} - s_{kl})$.

To prove the cases of Figure 4.2.4 (a) and (d), we define a wagon wheel graph

as in Figure 4.10 (a). Note that an edge b_{ij} corresponds to a string s_{ij} and $X(G)$ and $X(T)$ is homeomorphic.

For the case Figure 4.2.4 (a), by Lemma 4.12, there is a properly embedded disk $D^{'}$ in $X(T)$ that intersects each d_{ij} which is a meridian algebraically once in $\partial X(G)$ and is disjoint from the ε_i . After surgering along D , we may assume that $D^{'}$ is disjoint from D, and hence from d_{ij} and d_{kl}. But then $D^{'}$ is a compressing disk in $X(G - b_{ij} - b_{kl})$. Hence $T - s_{ij} - s_{kl}$ also has compressible boundary.

For the case Figure 4.2.4 (d), let the boundary of the ball of the tangle T be the union of two disks B_1 and B_2 where $B_1 \cap B_2 = \partial D$. Then $T_1 = closure(B_1 \cup D)$ is isotopic to a 3-string tangle including the string s_{ij}. We can consider two cases again:

(i) If T_1 is split, then there is a compressing disk $D^{'}$ in $X(T_1)$ which separates 3 strings into two and one on each side of $D^{'}$. If s_{ij} is in the side which has two strings, then a disk $D^{'}$ after surgering along D is the desired disk. If s_{ij} is in the side which has one string, by the similar argument of case Figure 4.2.4 (a), there is a disk $D^{''}$ in $X(T_1)$ which intersects each δ_{ij} algebraically once in $\partial X(G)$ and is disjoint from the ε_i. Then a disk $D^{''}$ after surgering along $D^{'}$ and D is the desired disk.

(ii) If T_1 is not split, by Corollary 4.9, it is isotopic to the tangle in Figure 4.20 (c) or its mirror image. If s_{ij} is the string with 3 self-crossings, then $T_1 - s_{ij}$ is isotopic to a two string tangle with two parallel strands. Hence it has exterior has a compressing disk $D^{'}$ which is the desired disc after surgering along D, If s_{ij} is not the string with 3 self-crossings, then $T_1 - s_{ij}$ will be a montesinos tangle with the form $M(\pm 1/2, \pm 1/3)$.

By Theorem 4.13, $X(T_1 - s_{ij})$ is handlebody thus it has a compressing disc D' which is the desired disc after possibly surgering along D.

Case 2. $T - s_{ij}$ is not split. If $T - s_{ij}$ is not split, then it is isotopic to the tangle in Figure 4.20 (c) or its mirror image by Corollary 4.9. If s_{kl} is the string with 3 self-crossings, then $T - s_{ij} - s_{kl}$ is isotopic to a two string tangle with two parallel strands. Hence it has exterior with compressible boundary. If s_{kl} is not the string with 3 self-crossings, then $T - s_{ij} - s_{kl}$ will be a montesinos tangle with the form $M(\pm 1/2, \pm 1/3)$. By Theorem 4.13, $X(T - s_{ij} - s_{kl})$ is handlebody thus its boundary is compressible.

Hence we can conclude that $\partial X(T - s_{ij} - s_{kl})$ is compressible.

□

4.2.5 Proof of main theorem

Theorem 4.15. *Suppose T is 4-string tangle which has less than 8 crossings up to free isotopy. If T is a solution tangle, then T is R-standard.*

Proof. Let T be a solution tangle with strands $s_{ij},\ s_{jk},\ s_{kl},\ s_{li}$ for $1 \leq i, j, k, l \leq 4$. Define G as in Figure 4.10 (a). Then G is a wagon wheel graph and T is carried by G.

If G is planar,then there is an isotopy from G to a very simple wagon wheel graph (see Figure 4.24 (a)) after allowing boundary points to move. A tangle carried by a wagon wheel graph in Figure 4.24 (a) has strands s_{12}, s_{23}, s_{34} and s_{41} where s_{ij} is a union of three arcs: an arc wrapping around e_i, a parallel copy of b_{ij} and an arc

wrapping around e_j. See an example in Figure 4.24 (b). This tangle corresponds to a graph in Figure 4.24 (c) which is a standard graph. Since an R-standard graph is a graph which is isotopic to a standard graph allowing the boundary to move, we can conclude that $T(G)$ is R-standard. Hence it is sufficient to show that G is planar to finish the proof of this theorem.

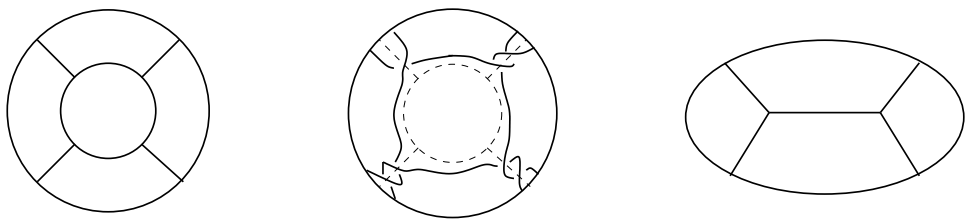

Figure 4.24: A simple wagon wheel graph and a standard graph

Since $\hat{\mathrm{G}}$ is planar if and only if G is planar, we will show that $\hat{\mathrm{G}}$ is a planar graph. To use Theorem 4.4, we have to show that every proper subgraph of $\hat{\mathrm{G}}$ is planar, and $X(\hat{\mathrm{G}})$ has compressible boundary.

• Claim I : $X(\hat{\mathrm{G}})$ has compressible boundary.

Since T is a 4-string tangle with at most 8 crossings, T is split or has parallel strands by Theorem 4.8, and hence $X(G)$ has compressible boundary. Because $X(\hat{\mathrm{G}})$ is homeomorphic to $X(G)$, $X(\hat{\mathrm{G}})$ has also compressible boundary.

• Claim II : Every proper subgraph of $\hat{\mathrm{G}}$ is planar.

It is sufficient to show that $\hat{\mathrm{G}}$-e_n and $\hat{\mathrm{G}}$-b_{ij} are planar for $1 \leq i, j, n \leq 4$ and $i \neq j$.

- Subclaim II-1 : $\hat{\text{G}}$-e_n is planar for $1 \leq i \leq 4$.

  $G - e_n$ carries 3-string tangle $T(G - e_n)$ with at most 7 crossings. By theorem 4.5, $T\,(G - e_n)$ (a tangle carried by $G - e_n$) is split or has two parallel strands. Then by cororally 3.2, $T(G - e_n)$ is planar and hence $G - e_n$ is planar. Thus $\hat{\text{G}}$-e_n is also planar.

- Subclaim II-2 : $\hat{\text{G}}$-b_{ij} are planar for $1 \leq i, j \leq 4$ and $i \neq j$.

  Unlike $\hat{\text{G}}$-e_n, we cannot sure that $\hat{\text{G}}$-b_{ij} carries a 3-string solution tangle. Hence we would like to use theorem 4.4 for $\hat{\text{G}}$-b_{ij}. In other words, we need to show that every proper subgraph of $\hat{\text{G}}$-b_{ij} is planar, and $X(\hat{\text{G}}$-$b_{ij})$ has compressible boundary.

- Subclaim II-2-(1) : $X(\hat{\text{G}}$-$b_{ij})$ has compressible boundary.

  An edge b_{ij} of the graph $\hat{\text{G}}$ corresponds to a string s_{ij} of the tangle T, and $\hat{\text{G}}$-b_{ij} corresponds to a 3-string tangle $T - s_{ij}$. By the hypothesis, T is 4-string tangle having less than 8 crossings up to free isotopy. Then $T - s_{ij}$ is 3-string tangle having at most 7 crossings up to free isotopy, and hence $T - s_{ij}$ is split or has parallel strands by Theorem 4.5. From Lemma 4.14, $X(\hat{\text{G}}$-$b_{ij})$ has compressible boundary.

- Subclaim II-2-(2) : Every proper subgraph of $\hat{\text{G}}$-b_{ij} is planar.

  Since there are 4 different b_{ij}, there are 4 different cases of $\hat{\text{G}}$-b_{ij} as shown in Figure 4.25 (a)$\sim$(d). It is sufficient to show that graph in Figure 4.25 (d) is planar since other cases are similar. For the rest of the proof, we assume $\hat{\text{G}}$-b_{ij}

is the graph in Figure 4.25 (d).

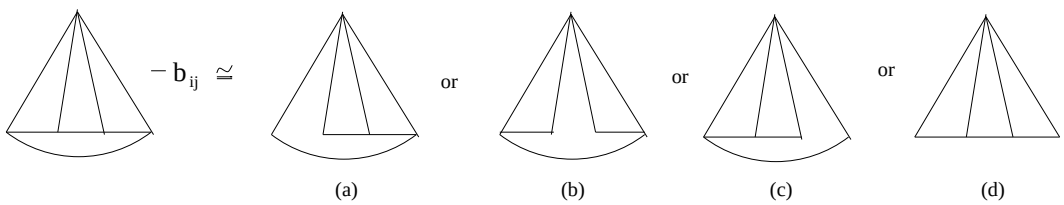

Figure 4.25: $\hat{\mathrm{G}}$-b_{ij} can be 4 different graphs

- Subclaim II-2-(2)-(i) : ($\hat{\mathrm{G}}$-b_{ij})-e_l is planar.

  ($\hat{\mathrm{G}}$-b_{ij})-e_l=($\hat{\mathrm{G}}$-e_l)-b_{ij} is a subgraph of $\hat{\mathrm{G}}$-e_l $(i, j, l \in 1, 2, 3, 4)$. Since $\hat{\mathrm{G}}$-e_l is planar, all subgraphs of $\hat{\mathrm{G}}$-e_l are also planar. Hence ($\hat{\mathrm{G}}$-b_{ij})-e_l is planar.

- Subclaim II-2-(2)-(ii) : ($\hat{\mathrm{G}}$-b_{ij})-b_{kl} is planar.

  ($\hat{\mathrm{G}}$-b_{ij})-b_{kl} is homeomorphic to Figure 4.26 (a) or (b) $(i, j, k, l \in 1, 2, 3, 4)$.

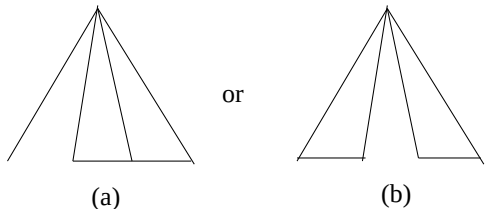

Figure 4.26: $\hat{\mathrm{G}}$-$b_{ij} - b_{kl}$ can be 2 different graphs

In the case of (a), this graph has an edge with a free end, say e_n. Since an edge with a free end can always lie on a disk properly embedded in B^3, the

planarity of this graph is the same as the planarity of a graph which we get after deformation retract of e_n to a vertex. Hence the planarity of the graph in (a) is same as the planarity of a subgraph of $\hat{\mathrm{G}}$-e_n. We already showed that $\hat{\mathrm{G}}$-e_n is planar, hence the graph in (a) is also planar.

In the case of (b), it's not simple like (a). Again, we will apply Theorem 4.4. To do that, we should check the compressibility of exterior of ($\hat{\mathrm{G}}$-b_{ij})-b_{kl} and planarity of all subgroups of it.

- Subclaim II-2-(2)-(ii)-(a) : X(($\hat{\mathrm{G}}$-b_{ij})-b_{kl}) has compressible boundary.

  Note that $X(G)$=$X(\hat{\mathrm{G}})$ is isotopic to $X(T(G))$, hence it is sufficient to show that $X(T - s_{ij} - s_{kl})$ has compressible boundary. This is true by Lemma 4.14.

- Subclaim II-2-(2)-(ii)-(b) : Every subgraph of $\hat{\mathrm{G}}-b_{ij} - b_{kl}$ are planar.

  If $\hat{\mathrm{G}}-b_{ij} - b_{kl}$ has the form of Figure 4.26 (a), then it has an edge with a free end, say e_n. Hence the planarity of this graph is the same as the planarity of $\hat{\mathrm{G}}$-e_n for some $n \in \{1, 2, 3, 4\}$. Since $\hat{\mathrm{G}}$-e_n is planar, a subgraph of it is also planar. Now, assume the graph $\hat{\mathrm{G}}-b_{ij} - b_{kl}$ has the form of Figure 4.26(b). A subgraph of $\hat{\mathrm{G}}-b_{ij} - b_{kl}$ obtained by deleting either e_n or b_{mn} has an edge with a free end, $m,\ n \in \{1, 2, 3, 4\}$. Again, after deformation retract of the edge with a free end, the planarity of this types of subgraph would be the same as the planarity of a subgraph of $\hat{\mathrm{G}}$-e_n for some $n \in \{1, 2, 3, 4\}$. Hence all subgraphs of $\hat{\mathrm{G}}-b_{ij} - b_{kl}$ are planar.

So far, we showed that every proper subgraph of $\hat{\mathrm{G}}$ is planar, and $X(\hat{\mathrm{G}}) = S^3$

- int(N($\hat{G}$)) has compressible boundary. By Theorem 4.4, we conclude that $\hat{G}$ is a planar graph. Thus G is also planar. Therefore T is R-standard.

□

4.3 Branched supercoiled DNA solutions

In this section, we discuss more about the branched supercoiled DNA solutions. We start with the following definition:

Definition 4.12. 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 4.27). For example, the graph G_R in Figure 4.6 (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 topologies of a DNA molecule when $n_5 \neq 0$ and $(i, j) \neq (k, l)$.

4.3.1 Simple branched solution tangles

We will first focus on tangles of the form shown in Figure 4.28. Suppose a tangle whose corresponding graph has the form shown in Figure 4.28 (a) is a solution

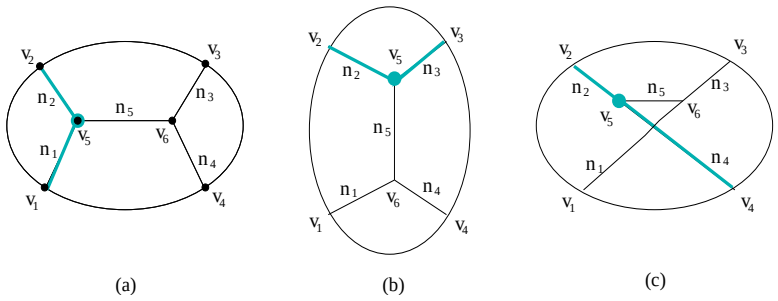

Figure 4.27: (i, j)-branched weighted graphs for R-standard tangle

(a)(1,2)-branched; (b)(2,3)-branched; (c)(2,4)-branched weighted graph for R-standard tangle (Note that the dege connecting v_4, v_5 can pass either over or under the edge connecting v_1, v_6.)

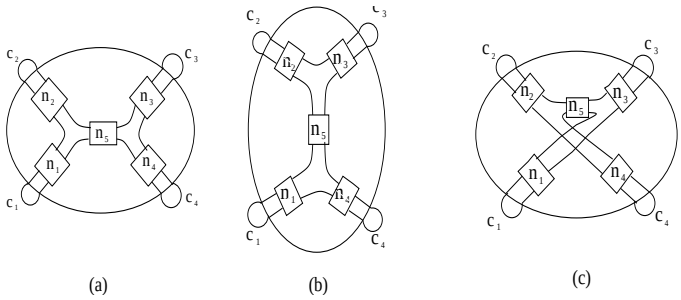

Figure 4.28: Examples of R-standard tangle model of a branched DNA-protein complex corresponding to a weighted graph

(a) (1,2)-branched; (b) (2,3)-branched; (c) (2,4)-branched (Similar to Figure 4.27 (c), there are two cases at the crossing of two branches depending on which branch is over than another.)

to the system of equations in Figure 4.3. Then we have the equations in 4.1. The values $p_1, \cdots, p_6$ in Figure 4.3 must be determined experimentally. Our goal is to find $n_1, \cdots n_6$ in terms of the p_i's.

$$\begin{aligned} n_1 + n_2 &= p_1 \\ n_2 + n_3 + n_5 &= p_2 \\ n_3 + n_4 &= p_3 \\ n_1 + n_4 + n_5 &= p_4 \\ n_1 + n_3 + n_5 &= p_5 \\ n_2 + n_4 + n_5 &= p_6. \end{aligned} \tag{4.1}$$

The solution to the system of equations 4.1 is the following:

$$\begin{aligned} n_1 = \frac{p_1 + p_2 - p_6}{2}, \quad n_2 = \frac{p_1 - p_4 + p_6}{2} \\ n_3 = \frac{p_2 + p_3 - p_6}{2}, \quad n_4 = \frac{-p_2 + p_3 + p_6}{2} \\ n_5 = \frac{-p_1 + p_2 - p_3 + p_4}{2}, \quad p_2 + p_4 = p_5 + p_6. \end{aligned} \tag{4.2}$$

We can solve similar equations for tangles corresponding to the graphs in Figure 4.28 (b) and (c). The summary of the results is the following:

- The solution to the Figure 4.3 equations is the following if the solution is of the

form Figure 4.28 (b):

$$\begin{aligned} n_1 &= \frac{-p_3+p_4+p_5}{2}, \quad n_2 = \frac{p_3+p_4-p_5}{2} \\ n_3 &= \frac{p_1+p_2-p_5}{2}, \quad n_4 = \frac{-p_1+p_2+p_5}{2} \\ n_5 &= \frac{p_1-p_2+p_3-p_4}{2}, \quad p_1+p_3 = p_5+p_6. \end{aligned} \tag{4.3}$$

- The solution to the Figure 4.3 equations is the following if the solution is of the form Figure 4.28 (c) (In this equation, ± 2 means that 2 if the branch with n_4 half twists is over and -2 if it is under.):

$$\begin{aligned} n_1 &= \frac{p_1-p_2+p_5}{2}, \quad n_2 = \frac{p_1-p_4+p_6 \pm 2}{2} \\ n_3 &= \frac{-p_1+p_2+p_5}{2}, \quad n_4 = \frac{-p_1+p_4+p_6 \pm 2}{2} \\ n_5 &= \frac{p_2+p_4-p_5-p_6 \pm 2}{2}, \quad p_1+p_3 = p_2+p_4 \pm 2. \end{aligned} \tag{4.4}$$

Note that the n_i must be integral. To have an integer solution set $n_1, \cdots, n_5$, all numerators of Equations 4.2, 4.3 and 4.4 should be even. In fact, there are eight possible cases to have an integer solution set for equation 4.2, shown in the following table. Equations 4.3 and 4.4 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 4.2, 4.3 and 4.4:

Lemma 4.16. *The graph G_R corresponding to an R-standard tangle in Figure 4.28 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.*

	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

- *If $p_1 + p_3 = p_5 + p_6$ holds, G_R is (2,3)-branched.*
- *If $p_1 + p_3 = p_2 + p_4 \pm 2$ holds, G_R is (2,4)-branched (± 2 means that 2 if the branch with n_4 half twists is over and -2 if it is under).*

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

We have only proved Lemma 4.16 for tangles corresponding to the graphs shown in Figure 4.28. However, Lemma 4.16 can be extended to R-standard tangles as discussed in the next section.

4.3.2 Discussion On complicated branched solution tangles

We will now consider a more complicate branched solution tangle like that in Figure 4.29.

Example 4.3. Let G be a graph which corresponds to the standard graph in Figure 4.29(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 4.29 (b)). Then G_R is the weighted graph which corresponds to the R-standard tangle T in Figure 4.29 (b). Since v_5 is connected to $v_2 = NW$ and $v_4 = SE$, G_R is (2,4)-branched.

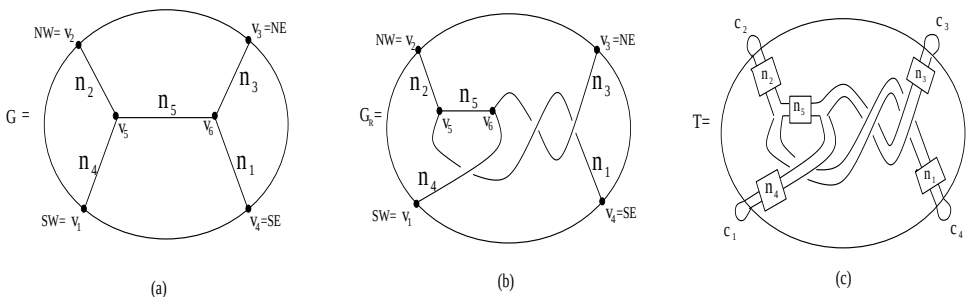

Figure 4.29: Example of R-standard tangle

Let's compare this example with the tangle in 4.28 (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 4.30 [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 4.4

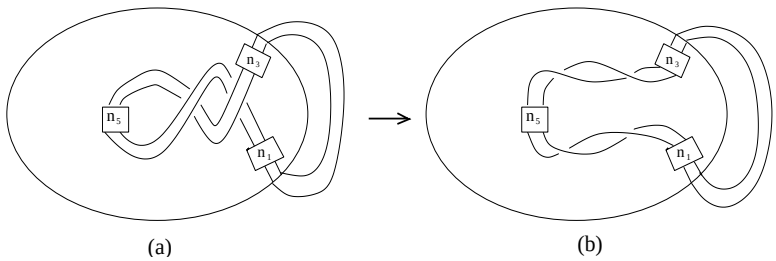

Figure 4.30: Writhe and twist

(a) A link obtained by doing Cre recombination in example 4.29 has writhe two; (b) The writhe two can be converted to 4 half twists.

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 4.2, 4.3 and 4.4 will still be even after adding or subtracting an even number. Hence the last equation in 4.4 determines if a tangle of the form T or T' can be a solution.

4.3.3 Generalized R-standard solution tangles

We developed the concept of R-standard tangle because of the possibility that a pair of supercoiled DNA branches can be twisted. An R-standard tangle T corresponds to a weighted graph G_R as in Figure 4.6. This implies that twisting a pair of branches of T corresponds to twisting a pair of edges (excluding the one connecting v_5 and v_6)of G_R along the boundary of the 3-ball. This action of adding twists to G_R is exactly the same as the action of constructing a rational 2-string tangle (See

section 2.2) except for the fact that a tangle starts from a zero tangle versus a graph starts from a standard graph (see Definition 4.2). Hence we can get an R-standard graph by adding alternating horizontal and vertical twists to a standard graph, and thus an R-standard graph can be denoted by conway notation.

Let $[a_1, b_1, ..., a_{n-1}, b_{n-1}, a_n]$ be a conway notation for G_R, where a_i is the number of horizontal right-handed half twists and b_i is the number of vertical left-handed half twists. The $a_i^{'}$s and $b_i^{'}$s are integers. In other words, we start with a standard graph, add a_1 horizontal half twists by moving vertices at NE and SE along the boundary of 3-ball, add vertical b_1 half twists by moving vertices at SW and SE along the boundary of 3-ball, and add a_2 horizontal half twists, etc. Similar to rational 2-string tangles, a unique rational number ($\in \mathbb{Q} \cup \infty$) is decided by a continued fraction and two R-standard graphs are the same (i.e. ambient isotopic) if the two rational numbers from each graph are the same. See Figure 4.31. for an example.

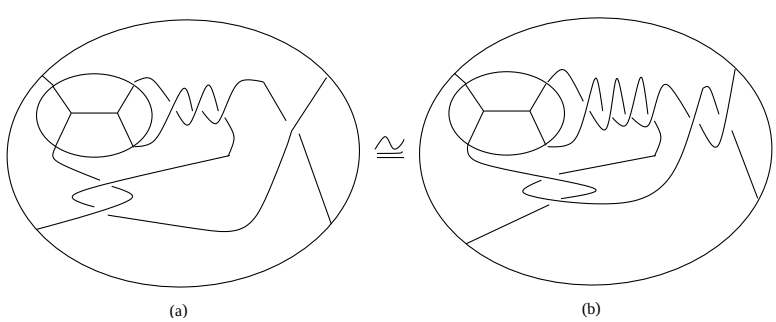

Figure 4.31: Two ambient isotopic R-standard tangles

(a)[-3,-2,-1]$\leftrightarrow -\frac{10}{7} = -1 + \frac{1}{-2+\frac{1}{-3}}$; (b)[-4,2,-2]$\leftrightarrow -\frac{10}{7} = -2 + \frac{1}{2+\frac{1}{-4}}$

An interesting property of the continued fraction is that the numbers corresponding to a conway notation can always be even numbers. Without loss of generality, let $\frac{A}{B}$ be a reduced (A, B are relatively prime) rational number which corresponds to G_R with $B = A \cdot q + r$ $(0 < r < A)$. We can assume $A < B$ since if $A > B$, then $\frac{A}{B} = q_0 \pm \frac{C}{D}$ where $C < D$ and q_0 is even. Then $\frac{A}{B} = \frac{1}{\frac{B}{A}} = \frac{1}{\frac{A \cdot q + r}{A}} = \frac{1}{q + \frac{r}{A}}$. If q is an even number, it will be an entry of the conway notation for G_R. If q is an odd number, then $q + 1$ will be an entry of the conway notation for G_R since $\frac{A}{B} = \frac{1}{\frac{B}{A}} = \frac{1}{\frac{A \cdot (q+1) + (r-A)}{A}} = \frac{1}{(q+1) + \frac{r-A}{A}}$. In this manner, we can always obtain a conway notation for an R-standard tangle with all even entries. See Equation 4.5 for an example.

$$\frac{12}{65} = \frac{1}{\frac{65}{12}} = \frac{1}{\frac{12 \cdot 6 + (5-12)}{12}} = \frac{1}{6 + \frac{-7}{12}} = \frac{1}{6 + \frac{1}{\frac{7 \cdot 2 + (5-7)}{-7}}} = \frac{1}{6 + \frac{1}{-2 + \frac{2}{7}}} = \frac{1}{6 + \frac{1}{-2 + \frac{1}{4 + \frac{1}{-2}}}} \tag{4.5}$$

In section 4.2.5, we proved that a solution tangle (with less than 8 crossings up to free isotopy) of the equations in Figure 4.3 is an R-standard tangle. Let T be a solution tangle and G_R be the weighted R-standard graph (see Figure 4.6) which corresponds to T. Then G_R has a conway notation $[a_1, b_1, ..., a_{n-1}, b_{n-1}, a_n]$ with all even integral entries (see Figure 4.32 (a)). Hence G_R can be obtained by adding alternating full horizontal twists and even full vertical twists to a weighted standard graph. Since we add full twists to a weighted graph, the vertices are still $v_1 = SW,\ v_2 = NW,\ v_3 = NE,\ v_4 = SE$. This is a very important clue to find a solution tangle.

Let $p_1, ..., p_6$ be the numbers from the tangle equations in Figure 4.3. In other words, p_i is the number of half twists on the link from Cre recombination on two

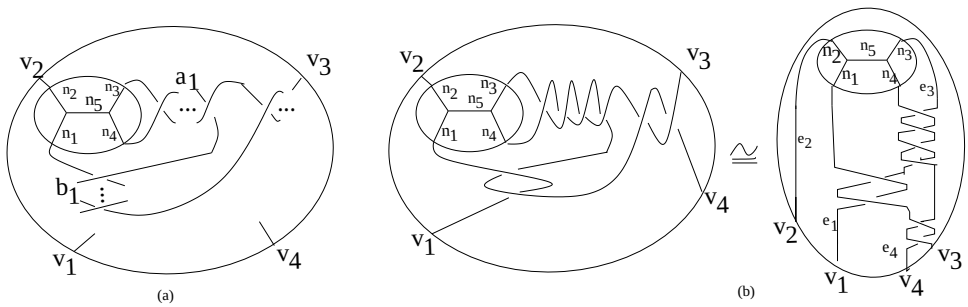

Figure 4.32: G_R

(a) G_R; (b) G_R with a conway notation [-4,2,-2] and a different view of it

outside loops of a solution tangle T which corresponds to G_R in Figure 4.32 (a). For the standard tangle, p_i values are related to the $n_i^{'}$s as in Equation 4.1. Since we didn't move v_2 at all when adding twists to a standard tangle, the equation involving the p_1, p_2, p_5, p_6 are the same as in Equation 4.1, and only p_3, p_4 are changed. This change can be easily seen in the different view of G_R in Figure 4.32 (b). Let e_i be the edge of G_R with weight n_i. Then e_2 has no crossings with any other edge. e_1 and e_4 may have crossings with only e_3 because the number of added twists are all even numbers. For example, in the graph of Figure 4.32 (b), if we walk along the edge e_3 from v_3, we cross only e_4. Same is true for e_1. This implies that we get $(a_1 + \cdots + a_n)$(sum of the numbers of all horizontal half twists added to a standard tangle) many crossings when we connect v_3 and v_4 by using an arc. Hence the link obtained from Cre recombination on c_3 and c_4 of a solution tangle T has -$(a_1+\cdots+a_n)$ writhe which can be converted to $2\cdot(a_1+\cdots+a_n)$ half twists [2] (See Figure 4.30). I.e., $p_3 = n_3+n_4+2\cdot(a_1+\cdots+a_n)$. Similarly,the link obtained from Cre recombination on

c_1 and c_4 of T has $(b_1+\cdots+b_{n-1})$ writhe which can be converted to $-2\cdot(b_1+\cdots+b_{n-1})$ half twists. Hence $p_4 = n_1 + n_4 + n_5 - 2 \cdot (b_1 + \cdots + b_{n-1})$. We can summarize all p_i values of a solution tangle T in Equation 4.6.

$$\begin{aligned} n_1 + n_2 &= p_1 \\ n_2 + n_3 + n_5 &= p_2 \\ n_3 + n_4 + 2 \cdot (a_1 + \cdots + a_n) &= p_3 \\ n_1 + n_4 + n_5 - 2 \cdot (b_1 + \cdots + b_{n-1}) &= p_4 \\ n_1 + n_3 + n_5 &= p_5 \\ n_2 + n_4 + n_5 &= p_6. \end{aligned} \tag{4.6}$$

We can solve the system of equations in Equation 4.6, and the solution is the following:

$$\begin{aligned} n_1 &= \frac{p_1 + p_2 - p_6}{2} \\ n_2 &= \frac{p_1 - p_4 + p_6 + 2 \cdot (b_1 + \cdots + b_{n-1})}{2} \\ n_3 &= \frac{p_2 + p_3 - p_6 - 2 \cdot (a_1 + \cdots + a_n)}{2} \\ n_4 &= \frac{-p_2 + p_3 + p_6 - 2 \cdot (a_1 + \cdots + a_n)}{2} \\ n_5 &= \frac{-p_1 + p_2 - p_3 + p_4 - 2 \cdot (a_1 + \cdots + a_n) + 2 \cdot (b_1 + \cdots + b_{n-1})}{2} \\ p_2 + p_4 + 2 \cdot (b_1 + \cdots + b_{n-1}) &= p_5 + p_6. \end{aligned} \tag{4.7}$$

CHAPTER 5
FUTURE DIRECTION

First of all, I am looking for a real biological data for our 4-string tangle model. Recently, I found an interesting figure (Figure 5.1) in the biology paper [16] which is about the MukB protein. MukB is a bacterial protein which is needed to compact DNA in a cell in *Escherichia coli*. In this paper [16], the difference topology technique is used with Topoisomerase II instead of Cre. MukB binds to an unknotted circular DNA and after that Topoisomerase II acts on the outside loops. The product topology is usually a trefoil. Figure 5.1 is a working model of a DNA-MukB protein complex. Even though this is not a real configuration of the complex, it shows us the possibility of a DNA-protein complex which can be modeled by a n-string tangle.

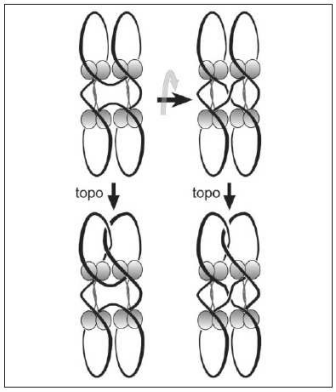

Figure 5.1: An working model of DNA-MukB protein complex with Topoisomerase II. Figure from [16]

Second, I would like to extend the number of crossings in Theorem 4.15. Although a rational tangle is the most biologically relevant tangle model of DNA-protein complexes, mathematically, extending the number of crossings in Theorem 4.15 is still an interesting problem.

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 enumaration 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*, 10:269–294, 2001.

[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. Vazquez. Tangle analysis of difference topology experiments: Applications to a Mu protein-dna complex. *Algebraic and Geometric Topology*, 9:2247–2309, 2009.

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

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

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

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

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

[12] M. Hirasawa and K. Shimokawa. Dehn surgeries on strongly invertible knots which yield lens spaces. *Proc. Amer. Math. Soc.*, 128:3445–3451, 2000.

[13] P. Kronheimer, T. Mrowka, P. Ozsvath, and Z. Szabo. Monopoles and lens space surgeries. *Ann. of Math.*, 165(2):457?546, 2007.

[14] B. Lewin. *Genes*. Oxford University Press, 5th ed., New York, 1994.

[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] Z. Petrushenko, C. Lai, R. Rai, and V. Rybenkov. DNA reshaping by MukB. *J. Biol. Chem.*, 281:4604–4615, 2006.

[17] D. Rolfsen. *Knots And Links*. AMS Chelsea Publishing, Providence, Rhode Island, 1976.

[18] Y. Saka and M. Vazquez. Tanglesolve: topological analysis of site-specific recombination. *Bioinformatics*, 18:1011–1012, 2002.

[19] M. Scharlemann and A. Thompson. Detecting unknotted graphs in s^3. *J. Diff. Geom.*, 34:539–560, 1991.

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

[21] D. W. Sumners, C. Ernst, N.R. Cozzarelli, and S.J. Spengler. Mathematical analysis of the mechanisms of DNA recombination using tangles. *Quarterly Reviews of Biophysics*, 28, 1995.

[22] Alexandre A. Vetcher, Alexander Y. Lushnikov, Junalyn Navarra-Madsen, Robert G. Scharein, Yuri L. Lyubchenko, Isabel K. Darcy, and Stephen D. Levene. DNA topology and geometry in Flp and Cre recombination. *J. Mol. Biol.*, 357:1089–1104, 2006.

[23] 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.

[24] 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.

[25] Y. Wu. A generalization of the handle addition theorem. *Proceeding Of The American Mathematical Society*, 114:237–242, 1992.

[26] Y. Wu. The classification of nonesimple algebraic tangles. *Math. Ann.*, 304:457–480, 1996.

[27] Z. Yin, A. Suzuki, Z. Lou, M. Jayaram, 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.