

A TANGLE ANALYSIS OF A DNA-PROTEIN COMPLEX WHICH BINDS FOUR DNA SEGMENTS

University of Iowa, Soojeong Kim

Biology Background

Site-Specific Recombination

(http://www.mun.ca/biochem/courses/3107/Lectures/Topics/Site_specific_Recomb.html)

- Site-specific recombination requires **short specific sequences** for the recombination. And those are the **only sites** at which recombination occurs.

Inverted repeats

If the two short specific sites are oriented oppositely to one another in the same DNA molecule, then the following illustrates the sequence of events that will take place. And a recombination at inverted repeats causes an inversion.

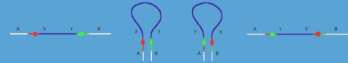


Figure 1.- from (http://www.mun.ca/biochem/courses/3107/Lectures/Topics/Site_specific_Recomb.html)

Direct repeats

If the two short specific sites are oriented in the same direction in the same DNA molecule then the following illustrates the sequence of events. And a recombination at direct repeats causes a deletion.

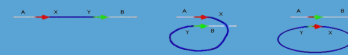


Figure 2.- from (http://www.mun.ca/biochem/courses/3107/Lectures/Topics/Site_specific_Recomb.html)

Topological Consequences of deletion and inversion:

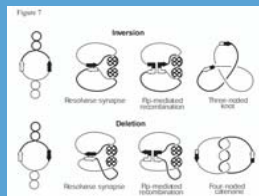


Figure 3.- from [2]

Cre and Flp, members of the tyrosine family recombinases,

- can carry out **DNA inversion and deletion equally well**, depending on the relative orientation of their target sites: **loxP** and **FRT**, resp.

Difference Topology [2]

- In 2002, Pathania, Jayaram and Harshey [2] employed a new methodology called *difference topology* in order to derive the number of DNA crossings trapped in an unknown synapse. Pathania et al revealed the topological structure within the Mu transpososome consisted of three DNA segments containing five nodes.

Phage Mu

a virus that infects bacteria.
(the most efficient transposable element known)
Mu proteins binding with three DNA sequences at the enhancer and the two Mu ends (attL & attR) is called the **'transpososome'**.

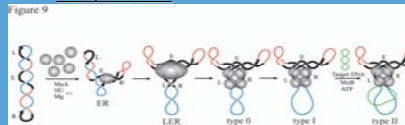


Figure 4.-from [2]

- Experimental results with the Mu transposition synapse and their interpretations:
 - $E^*[L,R]$; 3-noded knot for inversion & 4-noded catenane for deletion=> there must be 3 crossing between E and [L,R].
 - $L^*[E,R]$; same as $E^*[L,R]$.
 - $R^*[E,L]$; 4-noded catenane and 5-noded knot=> 4 crossings between R and [E,L].

$$x+z=3$$

(x=# of crossing [E,R],
y=# of crossing [R,L],
z=# of crossing [L,E].)

$$x+y=3$$

$$y+z=4$$

$$;x=1,y=2,z=2$$

Figure 5.- from [2]



Result:

- one crossing between E and R,
- two crossing between L and R,
- two crossing between L and E.

In other words, the transposition synapse has the 3-branched 5-noded configuration



figure from [3]

Tangle Analysis

These 'difference topology' experimental results were also analyzed in [3] by I. K. Darcy, J. Luecke, and M. Vazquez. They described infinitely many solutions to the same tangle equations, but argued that Pathania et al.'s model is the only biologically reasonable one.

The Mu transpososome protein complex binds three DNA segments. There exist protein complexes which bind more than three DNA segments. I would like to extend some of the 3-string tangle analysis in [3] to DNA protein complex containing four DNA segments.

•4-string tangle model

The unknotted substrate captured by a protein is modeled as the union of the two 4-string tangles $T_0 \cup T$ where T is the DNA-protein complex tangle and T_0 is the outside the complex. Let c_1, c_2, c_3, c_4 be the string of T_0 and s_1, s_2, s_3, s_4 be the strings of T .

Rational tangle solutions are generally believed to be the most likely biological models [3]. Suppose T is rational 4-string tangle, and define $O_1 = T \cup c_1 \cup c_2, O_2 = T \cup c_2 \cup c_3, O_3 = T \cup c_3 \cup c_4, O_4 = T \cup c_4 \cup c_1, O_5 = T \cup c_1 \cup c_3$ and $O_6 = T \cup c_2 \cup c_4$. Then those are 2-string tangles.

Suppose that in each experiment, the two recombination sites for deletion or inversion (ex. Cre, loxP sites) are located on two strings c_i and c_j ($i \neq j$). And after deletion or inversion experiment, assume the results are (2, p)-torus link or knot. These assumptions are motivated by [2][3]. This can introduce us the following six 2-string tangle equations;

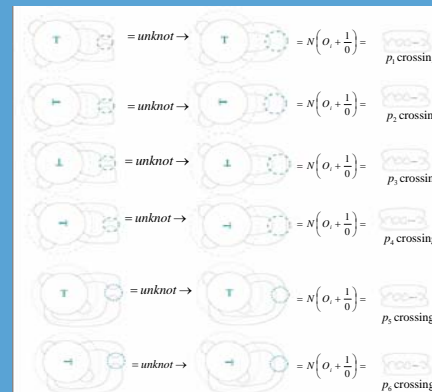


Figure 6.

•For example, let $T_0 \cup T$ be the following rational 4-string tangle;

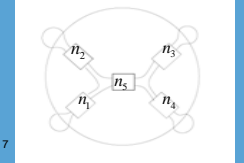


Figure 7

,where n_i is the number of right-handed twists in the box. Then by equations in Figure 6, we would get the following linear system:

$$n_3 + n_4 = p_1, \quad n_1 + n_4 + n_5 = p_2$$

$$n_1 + n_2 = p_3, \quad n_2 + n_3 + n_5 = p_4$$

$$n_2 + n_4 + n_5 = p_5, \quad n_1 + n_3 + n_5 = p_6$$

Then we can get the following results;

$$n_1 = \frac{p_1 - p_4 + p_5}{2}, \quad n_2 = \frac{p_3 + p_4 - p_5}{2}, \quad n_3 = \frac{p_1 - p_2 + p_5}{2}$$

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

•The type of tangle in above example can be expressed as a graph . Actually there are 3 possible cases as in figure 8.

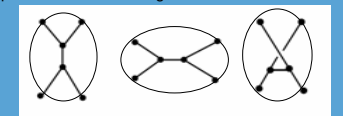


Figure 8

•Definition

A 4-string tangle is *standard* if it can be isotoped rel end points to one of figure 8. And a tangle is *R-standard* if it is isotopic to standard tangle after allowed to move boundary points of the graph.

•Theorem

Suppose T is rational tangle which has less than 8 crossings, and T is an solution tangle (i.e., a solution to the 6 equations in figure 6), then T is R-standard.

•References

- C. Ernst, D. W. Sumners, A calculus for rational tangles: applications to DNA recombination, *Math. Proc. Camb. Phil. Soc.* 108 (1990), 489-515.
- S. Pathania, M. Jayaram, and R. Harshey, Path of DNA within the Mu transpososome: Transposase interaction bridging two Mu ends and the enhancer trap five DNA supercoils, *Cell* 109 (2002), 425-436.
- I. K. Darcy, J. Luecke, and M. Vazquez, A tangle analysis of the Mu transpososome protein complex which binds three DNA segments, *Preprint*.