Using Knot Theory to Model DNA
An Undergraduate Research Project

William Schellhorn

Department of Mathematics
SIMPSON COLLEGE

MathFest, July 31, 2008
Outline

1. Description of the Project

2. Selected Topics
   - Example 1: Tn3 Resolvase
   - Example 2: Mu Transposase

3. My Perspectives
At Simpson College, every mathematics student completes a semester-long research project.

Outstanding students can apply for the Honors in Mathematics major.

An honors student completes a two-semester research project.
The student was interested in both mathematics and biology.

She approached me after I gave a talk on knot theory and mentioned some of its applications.
The goal of the project was for my student to explore how knot theory is being used to model DNA.

We found many suitable papers and sections of books.

It was her responsibility to read the sources and then report back to me.

Weekly meetings for one hour with additional opportunities to ask questions throughout the week.
My student was not required to produce original results.
The project ended up being a survey of existing research.

I am not a geneticist nor a biologist.
The Structure of DNA

- DNA (deoxyribonucleic acid) is formed by pairs of molecular strands that form a double helix.

The Genetic Code

- The molecular strands determine a sequence of bases (adenine, thyamine, cytosine, guanine).
- This sequence is the genetic code that gives a “blueprint” for life.

DNA is Naturally Knotted

- DNA in the nucleus of a cell is a knotted mess (think of a bunch of tangled fishing line in a beach ball).
- Here is an electron microscope image of some knotted DNA.

Electron microscope image by S. Wasserman, J. Dungan, and N. Cozzarelli (1985)
Biological Functions

- DNA must be utilized to perform various biological functions that are necessary for life.
  - Replication - reproducing a given DNA molecule.
  - Transcription - copying segments of DNA.
  - Recombination - modifying DNA molecules.
Enzymes

- Nature manipulates DNA molecules via enzymes called *topoisomerases*.
- Topoisomerases modify the DNA topologically.
Biochemists want to determine how particular enzymes act on DNA. These complex biological actions can be modeled topologically. The action of the Cre enzyme is modeled below.

[Diagram of Cre enzyme action on DNA]

Graphic from http://www.callutheran.edu/Academic_Programs/Departments/BioDev/omm/jmol/cre/11_8_2.gif

William Schellhorn
Using Knot Theory to Model DNA
The actions of some enzymes can be described as *site-specific recombination*.
- The enzyme recognizes a particular sequence of base pairs.
- The enzyme attaches to two specific sites corresponding to the sequence.
- The enzyme recombines the molecular strands.

![Diagram of recombination process](image)
Procedure

- Biochemists create a circular DNA molecule that contains copies of each of the recombination sites necessary for the reaction.
- View the closed loop before and after the enzyme acts, but the action itself is unknown.
Tangles

- Depict the circular DNA molecule before the reaction using two *tangles*.
  - The *substrate tangle* $S$ is unchanged by the enzyme.
  - The *site tangle* $T$ is where the enzyme acts.
- The enzyme replaces $T$ with a new tangle $R$, called the *recombination tangle*.
Tangle Notation

- The *substrate knot* and the *product knot* are both known.
- The tangles $S$, $T$, and $R$ are variables that are unknown.

\[ N(S+T) = \text{substrate} \]
\[ N(S+R) = \text{product} \]
System of Equations

- Have two equations with three variables here, so we cannot determine $S$, $T$, and $R$ uniquely.
- Should be able to determine two of them if the other happens to be known.

\[
N(S + T) = \text{substrate} \\
N(S + R) = \text{product}
\]
Outline

1. Description of the Project
2. Selected Topics
   - Example 1: Tn3 Resolvase
   - Example 2: Mu Transposase
3. My Perspectives

William Schellhorn
Using Knot Theory to Model DNA
The enzyme Tn3 resolvase replaces a site tangle $T$ with a recombination tangle $R$ and then releases the DNA molecule. Will occasionally repeat the tangle replacement a number of times.
Example 1: Tn3 Resolvase

Selected Topics

My Perspectives

Example 1: Tn3 Resolvase

Example 2: Mu Transposase

Using Knot Theory to Model DNA
Example 1

Using Knot Theory to Model DNA

\[ N(S + T) = \text{unknot} \]
\[ N(S + R) = \text{Hopf link} \]
\[ N(S + R + R) = \text{figure-eight knot} \]
\[ N(S + R + R + R) = \text{Whitehead link} \]
Example 1

- Four equations with three variables.
- Solved by De Witt Sumners (FSU) and Claus Ernst (WKU) in 1993.
- They also proved that $N(S + R + R + R + R)$ should be the knot 6$_2$, which was later observed as a product.
Note that solutions are topological rather than actually revealing the action of the enzyme.

One illustration is recombination by the XER enzyme (Darcy 2001). *Does it create or remove a crossing?*
Outline

1. Description of the Project

2. Selected Topics
   - Example 1: Tn3 Resolvase
   - Example 2: Mu Transposase

3. My Perspectives
Mu Transposase

- Mu transposase is a protein that moves fragments of DNA along a genome.
- Certain retroviruses (like HIV) integrate into their host genome in a process that is very similar to transposition.
- Models use 3-tangles rather than 2-tangles.
Example 2

- Biochemists S. Pathania, M. Jayaram, and R. Harshey performed eight laboratory experiments (2002).
- Determined a possible shape for the DNA bound within the Mu transposase protein complex, which is referred to as the *PJH solution*. 

![Diagram of DNA and protein complex](image-url)

These other configurations involve about 25 crossings each, as opposed to the five in the PJH solution.

They showed that the PJH solution is the only “biologically reasonable” solution.

They also showed that a restrictive assumption made by Pathania, Jayaram, and Harshey (about the most biologically likely shape of the DNA) was not necessary.
Mathematical techniques used in some of the proofs include:

- knot theory (including a calculus of rational tangles and linking numbers)
- graph theory (including the planarity of tetrahedral graphs)
- topology of 3-manifolds and surfaces (including Dehn surgery)

There are many more enzymes to study!
The Topic

- **Visual**: manipulating knot diagrams helps build understanding.
- **Wide range of sources**: both introductory and more advanced articles.
- **New developments**: always able to find recent articles.
- **Context**: more advanced topics can be introduced when they are needed.
- **Reference**: valuable to have a fellow professor who has studied genetics.
Students should have at least one class on genetics.

Students should have experience with reading, understanding, and creating logical arguments (ideally a class on proofs).

Students need to be self-motivated and responsible.
When would we have been limited by our college’s technology?

Would advanced laboratory equipment be needed eventually?

Would contacts at a university or laboratory be willing to offer assistance?
I strongly encourage you to try to advise a project on a topic that you are not too familiar with.

I learned a lot!

You cannot be too proud - will frequently have to admit that you do not know something.

Interesting topic of conversation at recruiting events with prospective students.

Another direction: purely mathematical projects on tangles.
Interdisciplinary Study

- Just beginning to explore the interaction between knot theory and knotted structures in biology and the other sciences.
- Both the mathematics and biology departments at Simpson where interested in the project.
- My hope is that the project will help stimulate more interdisciplinary undergraduate research.
- Possibilities exist for joint undergraduate research projects with professors and students from multiple disciplines.
My student was accepted into the Program for Women and Mathematics at the Institute for Advanced Study at Princeton.

She met another student who studied the topic - I am interested to hear about their discussions.

I will be working on a similar two-semester project with a student over the upcoming academic year.

New student is pursuing a Biochemistry major in addition to an Honors in Mathematics major.

I am excited about exploring the applications of knot theory and I am eager to see how the new project will take shape.
For Further Reading

C. C. Adams.  
_The Knot Book_.  

I. K. Darcy, J. Luecke, and M. Vazquez.  
Tangle Analysis of Difference Topology Experiments: Applications to a Mu Protein-DNA Complex.  

I. K. Darcy.  
Biological Distances on DNA Knots and Links: Applications to XER Recombination.  