Combinatorics and Knot Theory for RNA-DNA Complexes II

Pacific Institute for the Mathematical Sciences

Margherita Maria Ferrari [1]; Chris Soteros [2] Matthew Schmirler [2] Olivia Numedahl [2]; Shruti Karen Kaur [2]; Mingze Sun [3]; Brynn Williams [4]

[1]University of Manitoba [2]University of Saskatchewan [3]University of Alberta [4]University of Victoria

VXML

Project Overview

A 3-d lattice polygon model is developed to study models of 3-stranded RNA-DNA complexes called R-loops. To develop the model we focus on the following aspects:

- How to use a random sample of lattice tube polygons to model RNA-DNA geometry.
- How to visualize the model geometry.
- How to relate DNA base pairs to lattice edges.

We present progress made on each of these aspects.

What is an R-loop?

Figure 1. Schematic representation of an R-loop (blue: non-template DNA; red: template DNA; green: RNA transcript; blue box: RNA. polymerase)

R-loops occur during the process of transcription; experiments indicate they can play either destructive or regulatory roles in cellular processes. Thus, it is important to determine the factors influencing R-loop formation and stability.

Transcription involves an enzyme, called RNA polymerase, acting on double-stranded DNA to create a new RNA molecule using the template strand of the DNA. R-loops occur when the newly formed RNA binds to the template DNA strand. This results in a 3-stranded RNA-DNA structure consisting of an RNA-DNA complex along with the displaced non-template DNA strand.

Experiments indicate that R-loop formation is favoured when a G-rich RNA transcript is created from the template DNA. DNA supercoiling also affects R-loop formation, with more negatively supercoiled DNA promoting R-loop formation [1]. Hence both DNA sequence and DNA topology affect R-loop formation.

Experimental studies of R-loops [1] for specific DNA sequences have yielded site-specific probabilities for R-loop formation under different topological states of the DNA. Existing models [1, 2, 3] can be used to predict where R-loops are more likely to appear, but do not include detailed representations of the R-loop geometry.

The goal of this work is to develop a 3-d lattice model of DNA-RNA complexes that incorporates information from site-specific R-loop formation probabilities. Such a model will allow us to address more detailed geometric and topological questions about R-loops.

Acknowledgments

The team acknowledges helpful discussions with M. Vazquez and J. Lusk. We also acknowledge the usefulness of the CoCalc account provided by PIMS and Overleaf accounts provided by MMF and CES. CES acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) [funding reference number RGPIN-2020-06339].

First Step 3-d Lattice Model

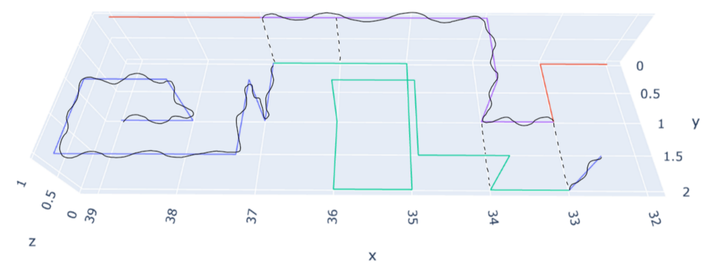


Figure 2. Portion of 2x1 lattice model of an R-loop in a tube: red: single-stranded RNA before and after R-loop; blue/black: DNA/DNA before and after R-loop; green: single-stranded DNA within an R-loop; purple/black: RNA/DNA within an R-loop.

Using a random sample of polygons from a lattice tube to model RNA-DNA geometry:

Samples of random polygons with fixed *span* m (the maximum extent in the x-direction) were available from [4]. Polygons were divided into two walks by deleting an edge at the $x = 0$ plane and an edge at the $x = m$ plane. One walk is considered double-stranded and the other single-stranded.

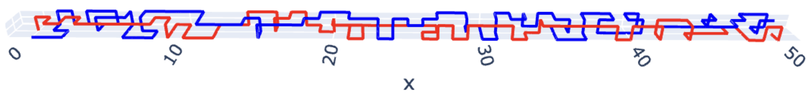
Choosing locations for the start and end of an R-loop:

An R-loop beginning and ending requires the two walks to be close. As a first model for this, we assume this can happen at 2-sections (half integer x-planes that contain only two edges of the polygon) where the edges are one edge apart in the y or z direction, called 2*-sections.

Visualization

Examining a 3-d model of a polygon of span 50 in a 2x1 tube:

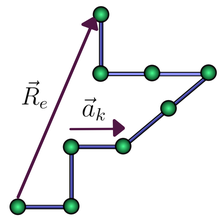
We split a span m polygon into two walks by removing one edge in each of its first ($x = 0$) and last ($x = m$) planes. The double-stranded and single-stranded pieces of the RNA-DNA complex are represented by these two walks. For example, the blue walk could represent a double-stranded molecule and the red one could represent a single-stranded molecule.



In the Figure below, the colours change at each 2*-section. For this model, we are assuming that these are possible places where an R-loop could start or end.

Relating DNA Base Pairs and Lattice Edges via Persistence Length

Persistence length is a geometric property which quantifies the bending stiffness of a polymer. The persistence length of double-stranded DNA and single-stranded DNA/RNA have been measured experimentally; if we can estimate the persistence length for walks in our model, we can use this to approximate the number of DNA base pairs in a walk.



$$c(k) = \vec{a}_k \cdot \vec{R}_e$$

Figure 3. A position dependent correlation at position k is given by $c(k) = \vec{a}_k \cdot \vec{R}_e$, where $\vec{a}_k$ is the direction of the k^{th} edge and $\vec{R}_e$ is the end-to-end vector from the first to last vertex in the walk. Average values of this correlation can be used to measure persistence length [5].

Given an independent sample of length-n walks ($n = m$ = polygon span) from our model, we computed the average value of $c(k)$ for $k = 1, \ldots, n$. We then plotted these average $c(k)$ values as a function of chain position k/n. The plateau of this plot serves as an estimate for model persistence length [5].

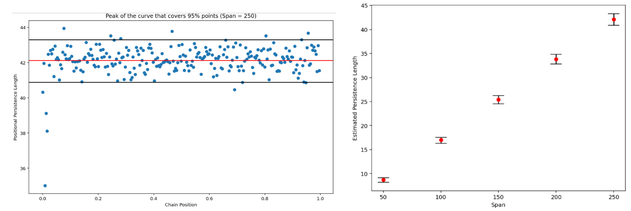


Figure 4. LEFT: Persistence length of double-stranded DNA is 50nm ≈ 150 base pairs (bp). At walk span = 250, the persistence length is 42 lattice edges. 150 / 42 = 3.57 bp per edge. RIGHT: Persistence length (in lattice edges) grows linearly with walk length = polygon span.

Future Work

Future directions will begin with finishing the 3-d lattice model and assigning probabilities to each possible model configuration. From this, we can get average geometric and topological properties of the RNA-DNA complexes. For this purpose, we can upload the lattice walks and polygons from the model into Knotplot [6] and measure entanglement complexity using the knotplot toolbox.

References

[1] R. Stolz et al. *Proceedings of the National Academy of Sciences* **116**.13 (2019), pp. 6260–6269.

[2] N. Jonoska et al. *Using Mathematics to Understand Biological Complexity: From Cells to Populations*. Ed. by R. Segal, B. Shtylla, and S. Sindi. Cham: Springer International Publishing, 2021, pp. 35–54. DOI: 10.1007/978-3-030-57129-0_3.

[3] M. M. Ferrari et al. *In preparation* (2024).

[4] J. W. Eng. PhD thesis. University of Saskatchewan, 2020.

[5] H. Hsu, W. Paul, and K. Binder. *Macromolecules* **46** (2010).

[6] R. Scharein. http://knotplot.com.