

Student: Yi Lin (ylin7@emory.edu)

Emory University, Atlanta GA

Mentors: Dr. Wilma Olson and Dr. Irwin Tobias

Department of Chemistry, Rutgers University

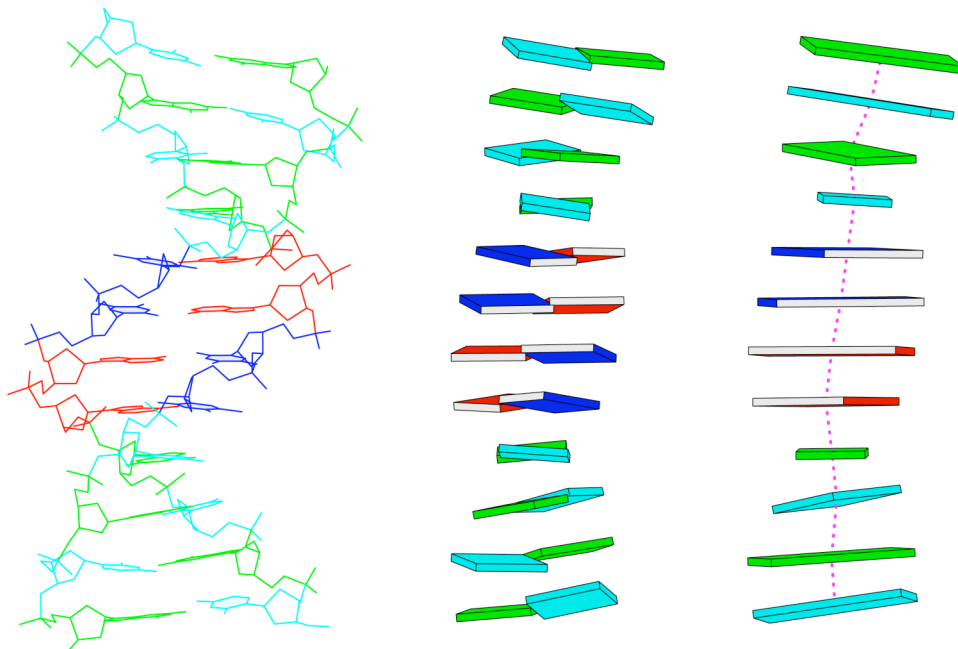
Title: A Discrete Representation of DNA Base Pair Steps

Introduction:

DNA carries genetic information and is usually 2 meters long if one were able to stretch it out. How does it fit into one single cell, which is normally 10^{-5} meters in diameter. The simple answer is it must FOLD tightly. DNA wraps around a protein called Histone forming a primary folding unit nucleosome. However, prior to being transcribed into RNA, the segment of DNA being transcribed must unwind in order for the RNA polymerase to have access to the segment.

Studies have also shown that only a small percent of DNA is expressed in any person's lifetime. Thus, how a particular DNA segment folds or how tightly it binds to Histone might dictate how likely it will be expressed.

There has been a great deal of theoretical and computational research. The traditional approach has been to regard the DNA double helix as smooth curves. However, DNA base pair steps are discrete. Thus, the new approach is to understand DNA folding by considering the base pair steps as discrete units.



Project Description:

We are most interested in finding new methods for calculating the twist number for a discrete DNA sequence.

Twist, in a continuous case, measures the magnitude of one curve around the other. However, it is a bit different in the discrete case. For example, if you only have two DNA base pair steps, you will first need to find a reference vector and find out how the vectors normal to each of the base pair steps are in relation to the reference vector.

Dr. Tobias has proposed a reference vector, r , which is defined as

$$r = \frac{v \times t}{|v \times t|}$$

where v is the vector that gives an arbitrary viewing direction and t is the vector normal to each of the base pair steps.

After working out some derivations, it is found that for a segment of any closed DNA with N base pair steps, the twist number for the DNA for the reference vector, r , is given by:

$$Tw = n + \frac{1}{2\pi} \sum_{i=1}^{n_B} \int_0^{\cos^{-1}(t_i \cdot t_{i+1})} \frac{(v \cdot b(i))(v \cdot t(i))}{1 - (v \cdot t(i))^2} d\alpha$$

where

$$t(i) = t_i \cos \alpha - (t_i \times b(i)) \sin \alpha$$

Since most of the variables above except α are constants, it is suspected that the integral can be solved explicitly. An explicit solution will save a lot of computing time as well as minimizing the computing error.

Results:

We simplified the integrand and found it to have the form:

$$\frac{\vec{v} \cdot b(\vec{1}) \int_0^{\cos^{-1}(\vec{t}_1 \cdot \vec{t}_2)} \frac{a \cos(\alpha) - b \sin(\alpha)}{1 - (a \cos(\alpha) - b \sin(\alpha))^2} d\alpha}{}$$

I then used Maple to evaluate the integral and found an explicit solution.

Next, we wanted to test the solution on a few shapes of which the twists numbers were known.

Test 1: Circle (Twist=0)

The twist calculated from our method was found to be very close to 0, which is what we expected.

Test 2: Figure8 (Twist=1 or -1)

The twist calculated based on the method was found to be close to .95. Even though it deviates a bit from 1, I suspect that it is mainly due to the fact that the figure8 that I feed into the system is not flat enough.

Test 3: Nucleosome (Twist=13.896 based on experimental data)

The twist number calculated was 13.894, which was very close to our expected number.

This method has shown to be a valid method for calculating Twist for a closed DNA sequences.

References Cited:

Representations of base sequence-dependent DNA structure (bd1084; Shui *et al.*, 1998)