Geometric Analysis of DNA in Molecular Dynamic Simulations of Nucleosomes

Ronald Brown[1] , Victoria Bamburg[2], James Liman[2], James Solow[2], Thomas C. Bishop
[1]Radford University
[2]Louisiana Tech University

LA-SiGMA — Louisiana Alliance for Simulation-Guided Materials Applications

NSF

Abstract

DNA sequence has a tendency to either include or exclude nucleosomes. The nature of such sequence specificity is not yet clear. However, since nucleosomes alter access to DNA, a proper understanding of sequence specificity is needed in order to fully comprehend genomic function, e.g., gene regulation. By analyzing the geometric properties and possible peculiarities that specific DNA sequences contain, we hope to gain insight into why nucleosomes position as they do on the genome. Here, our goal is to analyze DNA helical parameter data obtained from molecular dynamic simulations of nucleosomes. From this analysis, we can determine if the helical parameters are conserved throughout the simulations or whether the parameters are influenced by sequence. We seek to establish metrics for analysis that can be used in future simulations to identify and classify DNA sequence properties related to nucleosome positioning.
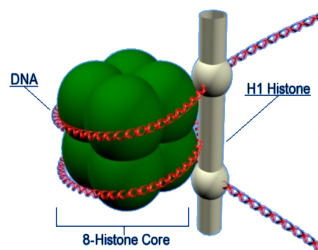
• An exaggerated visual representation of a nucleosome. Note that the 147 base pair of DNA wraps nearly two times around the eight histones.
Ref: http://www.molecularstation.com/molecular-biology-images/502-dna-pictures/60-nucleosome.html

Introduction

A nucleosome core particle (NCP) is a biomolecular complex of eight histone proteins around which is wrapped 147-base pair(bp) of DNA. Thus, nucleosomes fold long lengths of DNA in to a highly compact superhelix. This folding or packing influences genetic functions such as transcription, replication, regulation and repair. Nucleosome formation requires the 147bp of DNA to assume a specific super-helical conformation. Our goal is twofold: 1) determine if there is only one super-helical conformation or multiple separate conformational substates for the superhelix, and 2) determine if DNA sequence alters these findings. For this purpose we have analyzed a collection of all atom molecular dynamic simulations (Bishop2005) of nucleosomes containing different sequences of DNA. DNA helical parameter data is extracted from these simulations which we then use to analyze DNA conformation during the simulations.

There are twelve DNA helical parameters. They consist of two types: base pair parameters and dimer step parameters. The base pair parameters- Shear, Stretch, Stagger, Buckle, Propeller, and Opening- are used to define the relative position and orientation of two bases in a pair with respect to each other, while the dimer step parameters- Shift, Slide, Rise, Tilt, Roll, and Twist- define the relative position and orientation of two base pairs with respect to each other. In each case the relative orientation includes three rotations and three translations. In total, the helical parameters embody a complete description of DNA conformation that is equivalent to a Cartesian coordinate description. By analyzing the helical parameters, we can quantify the influence of DNA sequence on the geometric properties of super-helical DNA in nucleosomes.
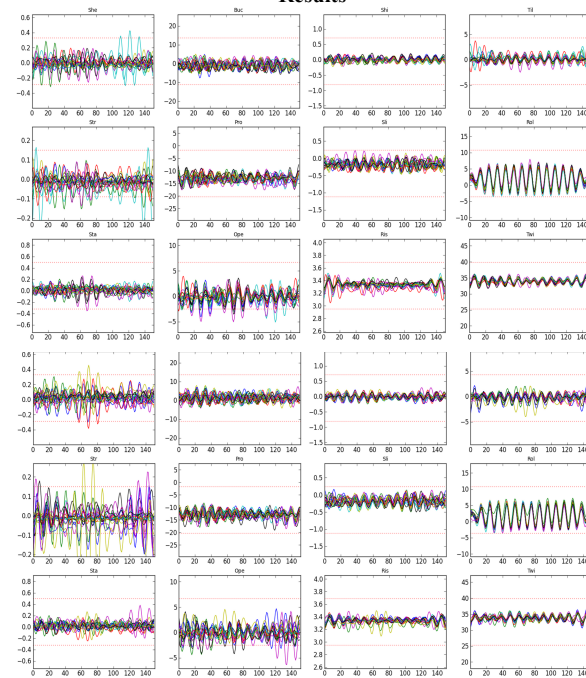
Acknowledgments

Results

Figure 1| Above are a set of twelve plots from sequences in chromosome 1 and 2 respectively. Each set represents the twelve helical parameters as a function of the base pair vs the Fourier filtered mean data. Each plot contains 21 lines representing the 21 separate simulations for each chromosome.
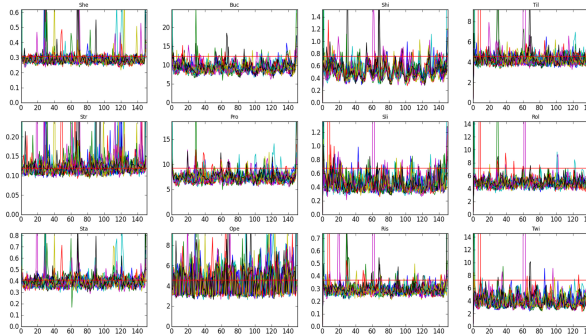

Figure 2| This is a plot of the standard deviation of the mean values observed during the 21 simulations of chromosome 1. The solid red line is the corresponding value of standard deviation of DNA free in solution.

Methods of Analysis

In order to analyze the tens of terabytes of simulation data that has been collected, it was necessary to develop automated software tools to assist in the task. Python was the language of choice due to its portability (operating system to operating system), its ability to execute high level functions such as Fourier Transforms, and its support of graphing and plotting utilities. To make the entire nucleosome simulation workflow more efficient, we developed a set of tools in Python, called HPTools. This project has demonstrated that our HPTools are sufficiently fast and powerful to allow us to analyze helical parameter data in near real time. Thus, we can incorporate analysis into our simulation workflow. HPTools includes a number of utilities for data manipulation and display including: multiple plotting commands, Fourier filtering, and file collection utilities. The goal is to convert the simulation trajectory data into a reduced set of observations that are biologically relevant.

Dataset

The sample data to be analyzed was generated previously from simulations of nucleosomes containing nucleosome positioning sequences found in the yeast genome. The simulations were executed in the molecular dynamics program NAMD using the methods described in Bishop2005. The simulations modeled 16 sequences. The 16 sequences correspond to the most highly occupied and least variable nucleosome footprint observed for each of the 16 chromosomes in *S. cerevisiae*. In the molecular dynamics simulations, a 20bp window about the chosen position-10bp upstream and 10bp downstream- was scanned. This gave us a 167bp sequence and 21 nucleosomes to simulate. In total, there were 336 nucleosomes simulated. Every system was simulated for 20ns. Here, we only consider the last nanosecond and a subset of the 336 systems. Statistical analysis - mean, range, standard deviation and normality – were used to quantify the differences in conformation and determine whether or not conformation is independent of sequence.

16 x   = 336 simulations

threading 167bp onto a nucleosome yields 21 nucleosomes
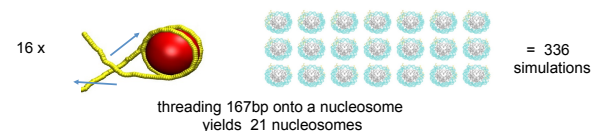
Process

It was previously demonstrated that twelve Fourier wavenumbers were both necessary and sufficient to model super-helical DNA in atomic resolution (Bishop2008) using the DNA helical parameters. The necessary and sufficient wavenumbers are: 1, 2, 3, 10, 11, 12, 13, 14, 15, 16. We filtered the simulation data for these wavenumbers to determine if DNA sequence altered the DNA super-helical conformation, i.e., altered the associated Fourier spectra data.

Conclusion

The HPTools that we have developed have proven to be sufficiently fast, powerful, and generalizable to enable us to incorporate analysis directly into our high throughput simulation workflow. All of the figures shown here were created in real time, and each represents a complete conformational analysis of 21 separate simulations. HPTools can be readily utilized to analyze our entire set of simulation data.

From the graphs of helical parameter data in Figure 1, it is clear that the values for Roll, Twist, and Slide are highly conserved in 21 separate simulations. This result is consistent with our previous analysis of all available x-ray structures (Bishop2008). Our findings support the conclusion that the patterns of Roll, Twist, and Slide that are necessary and sufficient for nucleosome formation are not affected by DNA sequence. Based on the simulations that we have analyzed, there is only one conformation of DNA superhelix. However, it is also clear that the values of all helical parameters are not conserved, suggesting there may be conformations unique to each sequence. These conformational substates do not alter the DNA superhelical path, only the local geometry of the DNA. Analysis of DNA fluctuations- as indicated by values of standard deviations (Figure 2)- demonstrate that nucleosomal DNA is less flexible than free DNA. This finding extends our previous results (Bishop 2005) which were based on only one simulation.