Geometric Analysis of DNA in Molecular Dynamics Simulations of Nucleosomes

Rocky Brown[1], Victoria Bamburg[2], James Liman[2], James Solow[2],
Thomas C. Bishop[2]

[1]Department of Physics, Radford University
[2] Department of Chemistry and Physics, Louisiana Tech University

Abstract: There are specific DNA sequences that have a tendency to allow geometric conformations that either include or exclude nucleosomes. Since nucleosomes alter access to DNA, they are essential in regulating gene expressions. It is necessary that nucleosomes be properly understood in order to fully comprehend genomic function. 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 DNA. Here, our goal is to analyze DNA helical parameter data obtained from molecular dynamic simulations of nucleosomes. From this analysis, we will determine if the helical parameters are conserved throughout the simulations or if the parameters are influenced by sequence. Furthermore, this will provide a metric for analysis methods that can be used in future simulations to identify and classify DNA sequence properties related to nucleosome positioning.

Keywords: Nucleosomes, Simulations, Helical parameters

1. Introduction

A nucleosome core particle (NCP) is a biomolecular complex of eight histone proteins around which is wrapped 147-base pair of DNA. Nucleosomes pack long lengths of DNA. This packing or folding influences genetic functions such as transcription, replication, regulation and repair. Nucleosome formation requires the 147bp of DNA to assume a specific superhelical conformation. Our goal is two fold: 1) determine if there is only one superhelical conformation or many separate conformational substates, 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. These simulations provided DNA helical parameter data which we then used as a measure of conformation.

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 using the helical parameters, we can quantify the influence of DNA sequences on the geometric properties of super-helical DNA in nucleosomes.

2. Methods of Analysis

2. 1 Tools

In order to analysis the massive amount of the data that has been collected, it was necessary to generate computer programming software 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 nucleosome simulation workflow more efficient, we developed a set of tools in Python, called HPTools, that can be used to analyze helical parameter data in near real time. This will allow for data processing to occur at nearly the same time as the simulations, and will convert the generated data into a reduced set of observations that are biologically relevant.

2. 2 Dataset

The data to be analyzed was previously generated by running simulations of sequences found in the yeast genome. The simulations were executed in the molecular dynamics program NAMD using the same methods as previous experiments (Bishop2005). The simulations utilized 16 sequences (one from each chromosome of *S. cerevisiae*). The sequences correspond to the most highly occupied and least variable nucleosome footprint observed for each chromosome. In the molecular dynamics simulations, a 20bp window about the chosen position that included 10bp upstream and 10bp downstream was scanned. This gave us a 167bp sequence and 21 nucleosomes to simulate. In total, there were 336 systems that 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 conformations is independent of the sequence.

2. 3 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 will filter the simulation data for these wavenumbers to determine if the necessary and sufficient result is valid for our simulations with different sequences of DNA, or if DNA sequence somehow alters the Fourier spectra associated with super-helical DNA conformation.

3. Results and Discussions

The following graphs represent results from 21 separate simulations. Each line on the plots corresponds to one of the 21 positions associated with one of the 16 sequences that was “threaded” around the nucleosome. Below we

display only the mean values in a figure generated with our HPTools. The tools are sufficiently general and powerful to allow us to rapidly load, analyze and display virtually any data set associated with the 336 simulations, including routines to achieve the required Fourier analysis. From the images below it is clear that there remains a data formatting issue, note the spikes.

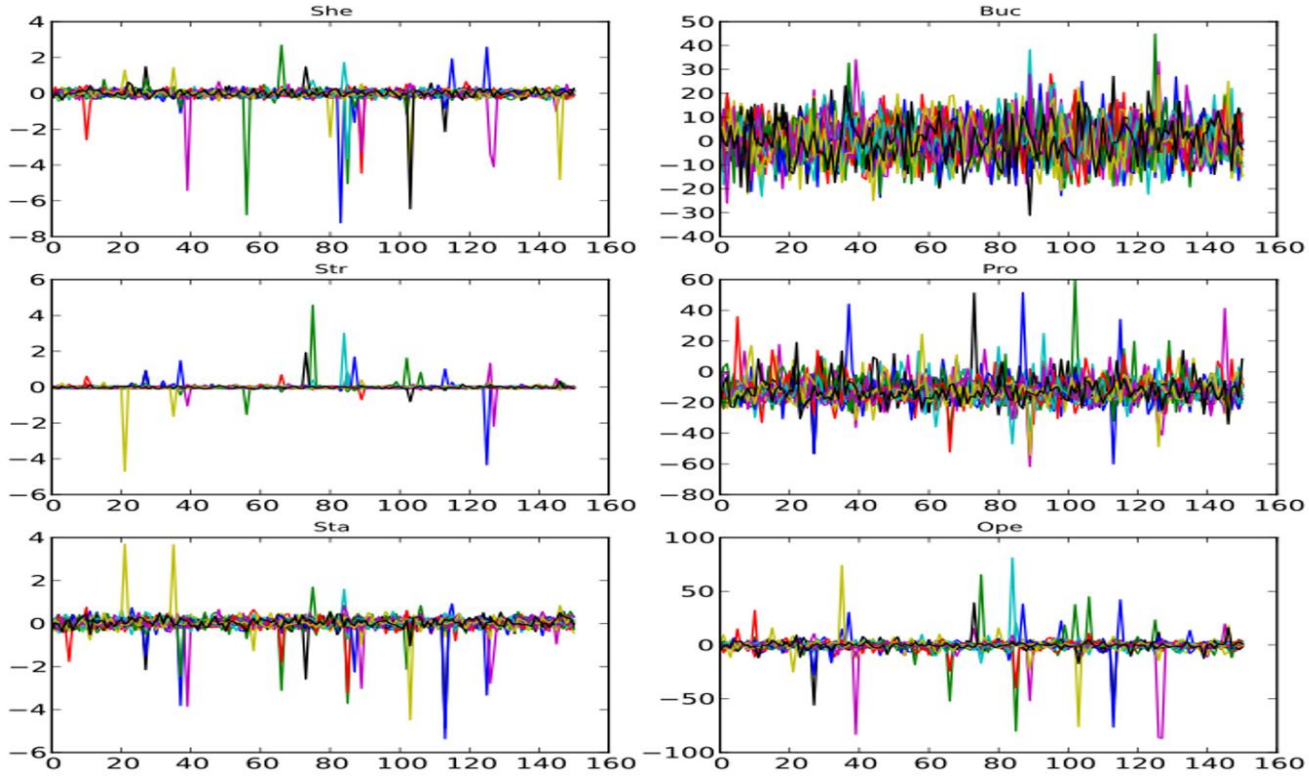

Fig. 1| Base Pair Parameters This is a graph of the mean values of the base pair parameters. As stated above, there are 21 separate systems here, each is represented by a different color. The peaks will be discussed in the conclusions.

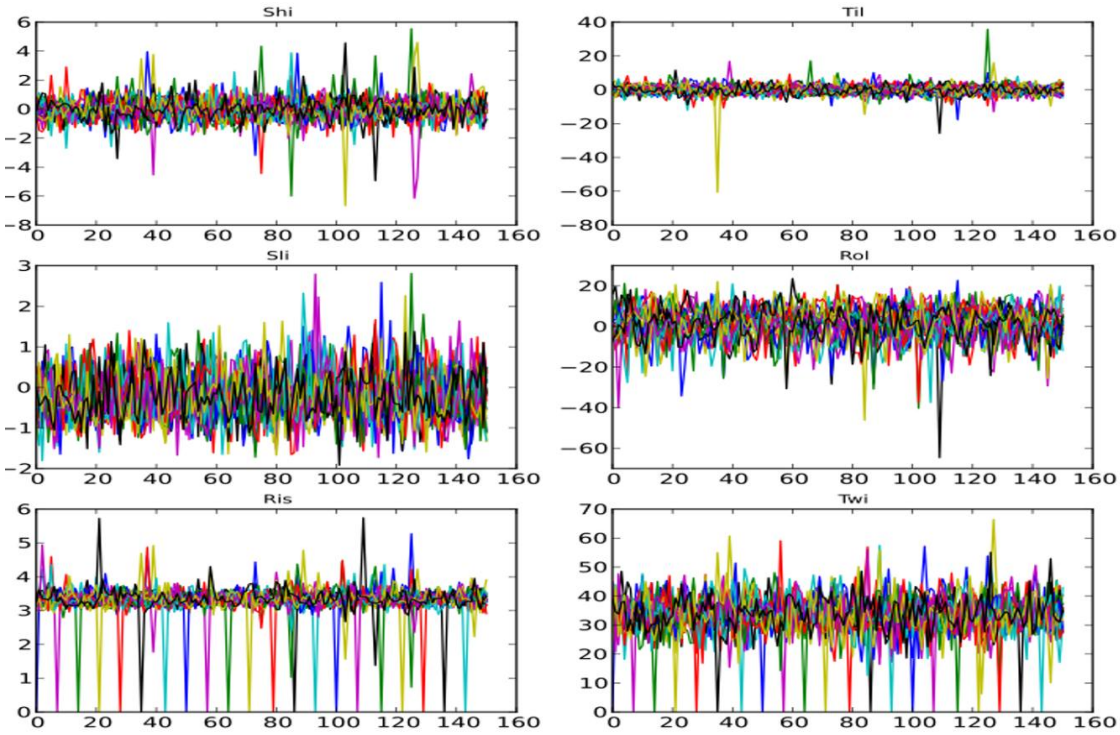

Fig. 2| Dimer Step Parameters This is a graph of the mean values of the dimer step parameters. As stated above, there are 21 separate systems here, each is represented by a different color. The peaks will be discussed in the conclusions.

4. Conclusions

The HPTools routines that we have developed have proven to be sufficiently fast, powerful, and generalizable to enable us to achieve data processing at the same time as the simulations. See our poster for applications of these tools.

5. Acknowledgments

The current work is funded by the NSF EPSCoR LA-SiGMA project under award #EPS-1003897.

6. References

[1] Bishop. T.C. (2009). Geometry of the nucleosomal DNA superhelix. *Biophys J,* 95(3), 1007-1017.

[2] Mukherjee, R., Bishop, T. C. (2011). Nucleosomal DNA: Kinked, Not Kinked, or Self-Healing Material?.

Frontiers Nucleic Acids, 5, 69-92.