Understanding Nucleosome Dynamics using Diffusion Maps

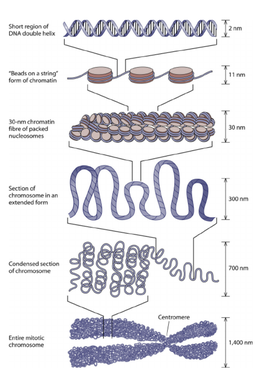
Ashley Z. Guo, Joshua Lequieu, Juan J. de Pablo
Institute for Molecular Engineering, The University of Chicago

Abstract

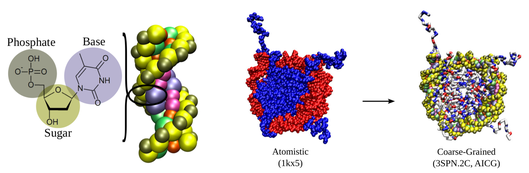
The identification of effective collective variables remains a challenge in molecular simulations of complex systems. Here, we use a nonlinear manifold learning technique known as the diffusion map to extract key dynamical motions from a complex biomolecular system, namely the nucleosome: a DNA-protein complex consisting of 147 base pairs of DNA wrapped around a disc-shaped group of eight histone proteins. We show that diffusion maps are effective at extracting collective variables previously found through a detailed free energy analysis in addition to revealing more subtle features involving looping conformations, in which DNA bulges away from the histone complex. This work demonstrates that diffusion maps can be a promising tool for analyzing very large molecular systems and its characteristic slow modes.

Introduction & Motivation

Eukaryotic genomes undergo packing into an array of DNA-protein complexes called chromatin that is folded into successively higher order structures to form the mitotic chromosome (below left). Chromatin is a dynamic material that may simultaneously expand and compact as different genes are expressed. The nucleosome is the basic building block of chromatin and is made up of 147 base pairs of DNA wrapped around a disc-shaped complex of eight histone proteins. Understanding the key dynamics exhibited by this basic unit of chromatin is of utmost importance.



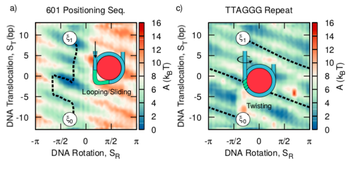
Jansen & Verstrepen, 2011.

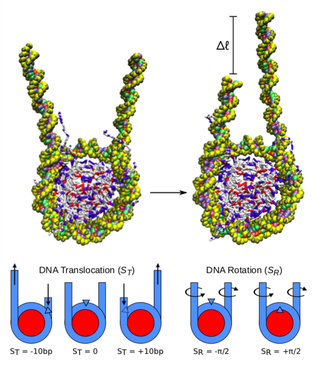


To model the nucleosome, we use a coarse-grained model of DNA developed by our group called 3SPN (3-Sites-Per-Nucleotide), which captures many properties of DNA including major and minor groove widths, excluded volume, and melting behavior. Histone proteins are modeled using the AICG model.

3SPN.2C: Freeman, Hinckley, Lequieu, Whitmer, de Pablo. J. Chem. Phys. 141:165103 (2014)
AICG: Li, Wolynes, Takada. PNAS 108:3504 (2011)
3SPN.2C + AICG: Freeman, Lequieu, Hinckley, Whitmer, de Pablo. Phys. Rev. Lett. 113:168101 (2014)

Recent work from our group examines the sequence-dependence of nucleosome dynamics. By performing molecular dynamics (MD) simulations of the nucleosome using nine DNA sequences with varying binding affinities and performing a free energy analysis, we show strongly binding sequences (e.g. 601) tend to reposition on the nucleosome through a sliding motion (S_T at right), while weakly binding sequences (e.g. TTAGGG) tend to reposition by a twisting motion (S_R at right).





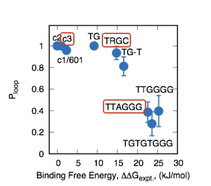
Lequieu, Schwartz & de Pablo, 2017.

Applying the Diffusion Map to Nucleosome MD Data

We invested a huge amount of computational resources into simulating 9 DNA sequences, 100 simulations per sequence, for a total of approximately 5 microseconds simuated time. **Can we make the most of our own MD data and mine it for additional information about the nucleosome?**

Two key questions: Can we extract underlying dynamics that
(1) are consistent with previous free energy analysis?
(2) were difficult to identify through free energy analysis and can provide some fresh perspective?

We use the diffusion map to reduce our high-dimensional molecular dynamics data for the nucleosome to fewer dimensions which capture the key dynamics of the system. **For ease of analysis, we choose three representative sequences out of the total nine simulated: strongly binding sequence c3, moderately binding sequence TRGC, and weakly binding sequence TTAGGG.** Eigenvalue spectra suggest that c3 and TRGC exhibit a hierarchy of three dominant modes and three moderate modes, while TTAGGG is primarily dominated by a single slow mode.



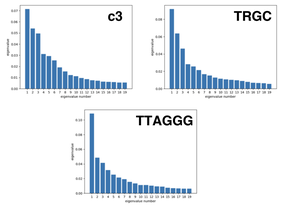
$$A_{ij} = \exp\left(\frac{-d_{ij}{}^{2\alpha}}{2\epsilon^2}\right) \qquad 0 < \alpha < 1$$

$$d_{ij} = \text{RMSD between two configs}$$

$$\mathbf{M} = \mathbf{D}^{-1}\mathbf{A}$$

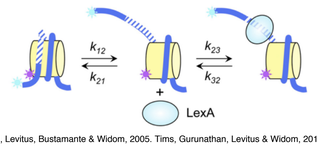
$$D_{ii} = \sum_k A_{ik}$$

$$\text{config}_i \mapsto (\psi_1(i), \psi_2(i), \ldots, \psi_k(i))$$

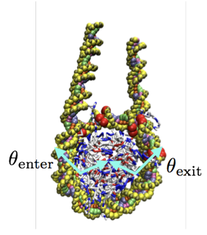


DNA Breathing Dynamics Tied to Sliding Dynamics

Single molecule FRET studies have shown that DNA can spontaneously unwrap and rewrap from the histone in a motion called DNA breathing. This spontaneous unwrapping and rewrapping of DNA allows proteins to access DNA segments that were previously buried near the histone complex.

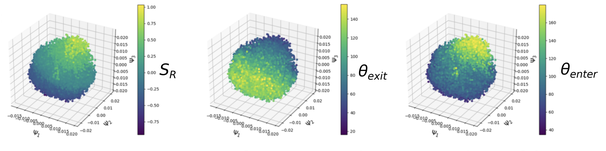


Li, Levitus, Bustamante & Widom, 2005. Tims, Gurunathan, Levitus & Widom, 2011.



As proxies for DNA breathing, we define two angles that characterize the unwrapping of DNA at either side of the nucleosome, θ_{enter} and θ_{exit}. We find that these two angles correlate well with the same eigenvector used to identify DNA translocation/sliding (S_T) and rotation (S_R). This suggests that DNA breathing and sliding are not independent motions: **the diffusion map approach suggests that breathing and sliding motions are inseparable.**

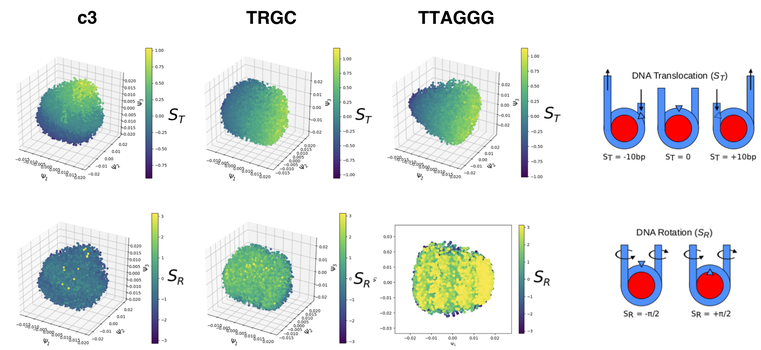
Sequence c3:



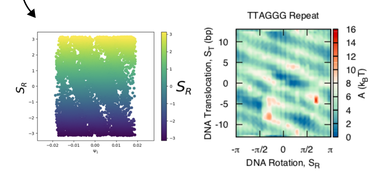
Same pattern is observed for TRGC and TTAGGG.

Extracting DNA Translocation (S_T) and Rotation (S_R)

DNA translocation emerges in one of the top two eigenvectors across all three sequences, regardless of binding strength. Rotation is found to correlate with the same eigenvector for only the weakly binding sequence TTAGGG, the only sequence that primarily repositions by rotation.



Plotting S_R vs ψ_1, which correlates with S_T, recovers the same pattern as the previously studied free energy landscape for TTAGGG.



Moderate Eigenvectors Reveal Looping Conformations

Sequences c3 and TRGC exhibited hierarchical eigenvalue spectra. The moderate eigenvectors in this case (ψ_4, ψ_5, and ψ_6) are found to isolate looping conformations, in which DNA bulges out from the histone complex (see right). To quantify this, we use the collective variable *loopiness*, which is the distance between the histone and DNA integrated along the entire circumfrence of the nucleosome. Loopiness emerges in higher eigenvectors for sequences more likely to exhibit translocation, and loopiness is not found for sequence TTAGGG, which is weakly binding and primarily repositions by rotation.