



MULTISCALE STUDIES OF CHROMATIN: DOMAIN BOUNDARIES, TETRANUCLEOSOME MOTIFS, AND FRACTAL DIMENSION

Walter Alvarado¹, Soren Kyhl², Aron Coraor², Joshua Moller², and Juan de Pablo²

¹Biophysical Sciences, University of Chicago, Chicago, IL, USA, ²Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL, USA

Abstract

An open question in biology is the identification of a fundamental structure of chromatin inside the cell. Our conventional understanding of this structure has been modified in the past decade to include the effects of epigenetic modifications. These chemical modifications are defined as heritable changes in gene expression, without changes in the underlying nucleic acid sequence. They have been observed to profoundly affect chromatin structure and gene expression, with implications in physiological development and several major diseases. Several studies have suggested the existence of a “30 nm fiber” of chromatin; however, recent studies have suggested that this is may not be the case for chromatin *in vivo*. Although much research has been conducted on chromatin, current experimental techniques fall short in capturing individual nucleosomal structure. Through the implementation of molecular dynamics and Monte Carlo methods, we use new state-of-the-art data driven techniques to study the fundamental building blocks of chromatin and determine how epigenetic marks affect chromatin’s structure and dynamics. In one example, nucleosome-scale simulations reveal information on the development of microphase-separated chromatin domains. In a second project, we study chromatin fractal dimension and its relation to cancer. Finally, we implement non-linear dimensionality reduction algorithms, in particular diffusion mapping, to determine the slowest dynamical modes of short chains of chromatin. We show that diffusion maps are able to identify subtle features of nucleosome dynamics and show the thermodynamics of several folding motifs identified experimentally.

Chromatin Domains

Chromatin organizes on large scales into microphase regions of silent “heterochromatin” and active “euchromatin” domains. To investigate individual gene effects on chromatin domain boundaries, our group has developed a Theoretically Informed Coarse Grain Model of chromatin capable of simulating entire chromosomes at nucleosome level resolution. The model incorporates H3K9me3 epigenetic marks and HP1 proteins which thereby bind and stabilize the heterochromatic phase. By representing nucleosomes explicitly, we are able to incorporate epigenetic mark sequences taken directly from ChIP-seq data, and can investigate the effects of altering specific epigenetic modifications on chromatin structure. Using this model, we are able to turn specific genes on and off and control heterochromatin formation. These efforts allow us to precisely tune the simulated nuclear conditions and make specific epigenetic alterations on the gene and even nucleosome level.

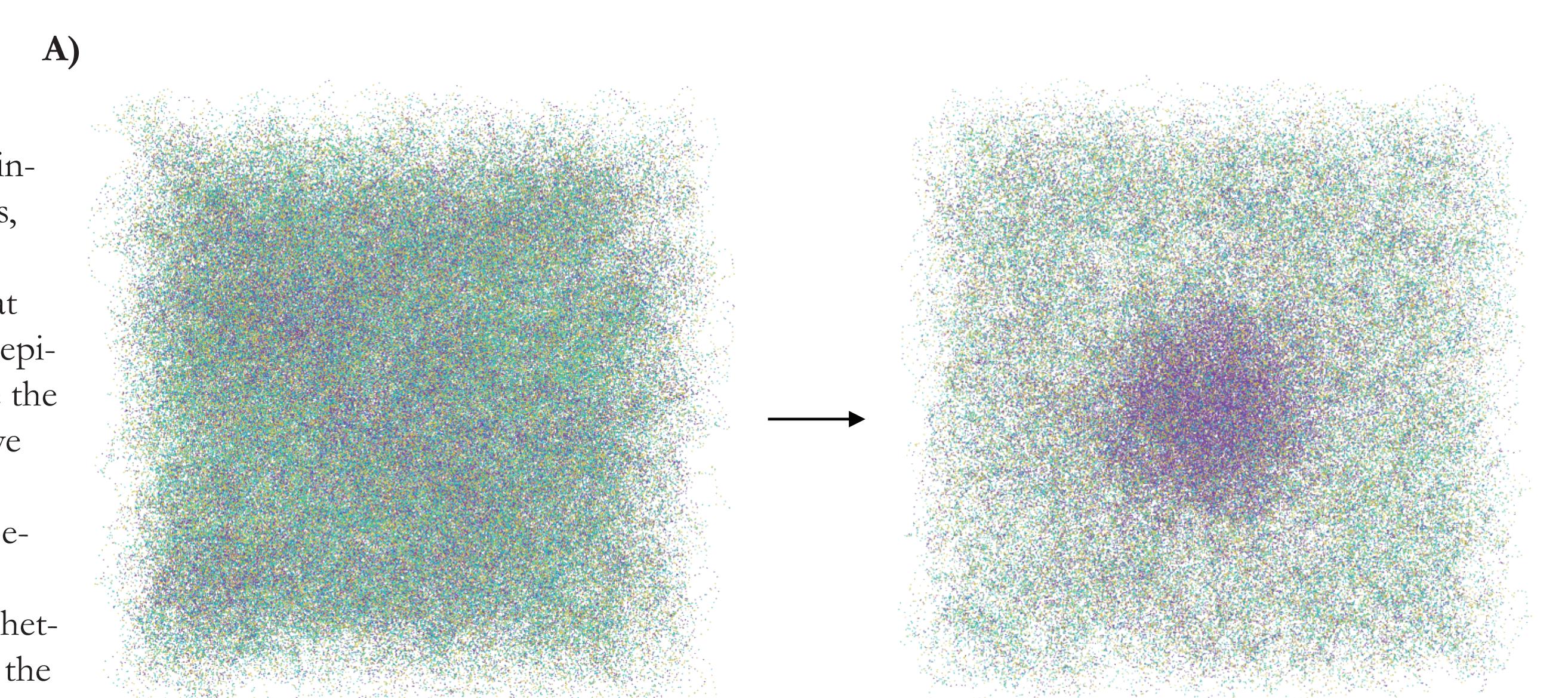


Figure 1. A) Simulation snapshot of entire-chromosome simulation of Chr. 16 with epigenetic mark H3K9me3. Nucleosomes are represented as (colored) beads, with twice-methylated (purple), singly-methylated (yellow), and not methylated (cyan) represented accordingly. These methylations are taken from Chromatin Immunoprecipitation experiments. The addition of heterochromatin protein 1 causes condensation of a heterochromatin domain. B) The resulting chromatin structure, as visualized by a contact map. Regions of high contact frequency are shaded in red.

Nucleosome Dynamics Via Diffusion Maps

Current studies have suggested two types of basic secondary structural motifs associated with epigenetic status, an *a*-tetrahedron and *b*-rhombus *in vivo* for yeast chromatin. Using Hi-CO method revealed 3D nucleosome position and orientations in chromosomes. There's coupling between nucleosome folding structure and epigenetic features at each locus which agrees with regular and irregular chromatin fiber models.

Assuming that the dynamics can be described as a diffusion process, and the similarity parameter is a good descriptor, there are two reasons. 1. It can capture diffusion distances which means capturing how easy it is to go from one snapshot to another and 2. it captures the slowest modes of the system. Meaning that we can capture paths that describe the evolution of our system in its fundamental dynamical motions. Even if they don't, the CVs we get from them are good variables with which we can use to parameterize the evolution of our system from one state to another.

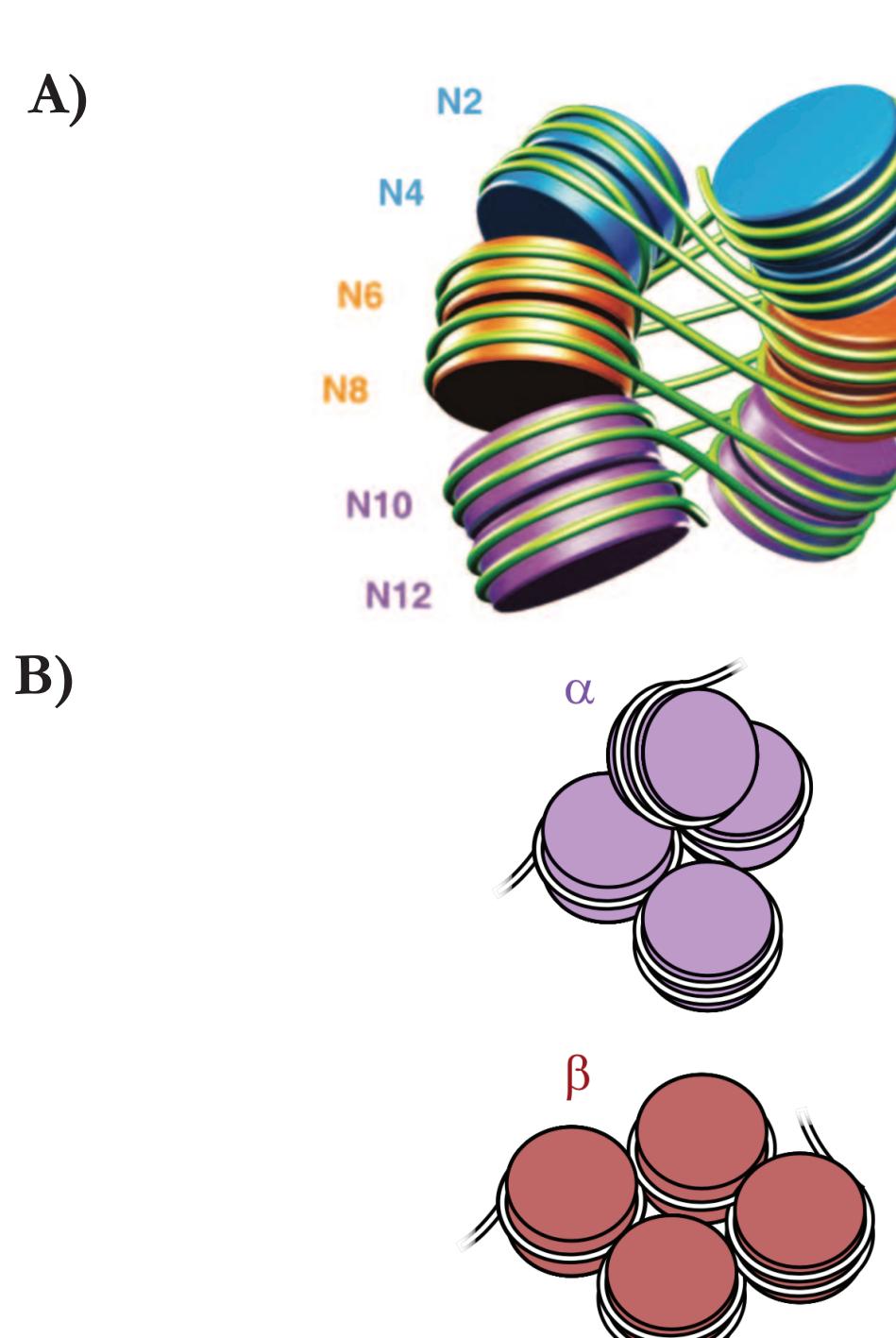


Figure 2. (A) X-ray crystallography by Song et al. suggest chromatin forms a zig-zag structure which condenses into a “30-nm fiber”. (B) Using Hi-CO, Ohno et al. suggest two possible folding motifs for yeast DNA *in vivo*.

$$\mathbf{A} \quad \mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix} \quad \text{n observations in ambient } m\text{-dimensional space}$$

$$\mathbf{B} \quad \tilde{\mathbf{Y}} = \begin{pmatrix} y_{11} & y_{12} & \dots & y_{1n} \\ y_{21} & y_{22} & \dots & y_{2n} \\ \vdots & \vdots & & \vdots \\ y_{k1} & y_{k2} & \dots & y_{kn} \end{pmatrix} \quad \text{n observations in reduced } k\text{-dimensional space}$$

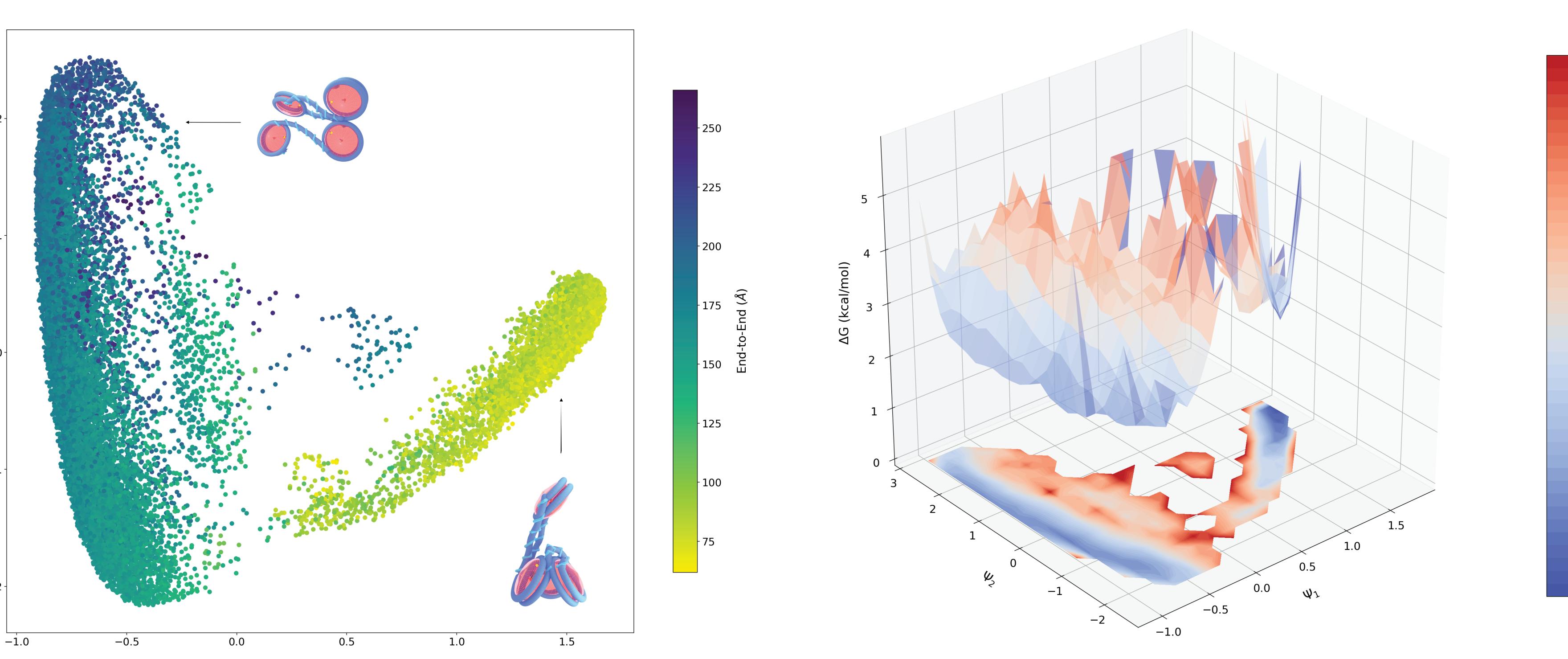


Figure 3. (A) Diffusion maps involves mapping a high-dimensional data-set, such as coordinates, to a low dimensional manifold to extract essential collective variables. (B) Using the diffusion map approach we observed two key states for an unbiased molecular dynamics run of four nucleosome with 187 nucleotide repeat length. By sampling the set of points from diffusion maps for the lowest diffusion modes, the free energy landscape can be mapped.

Fractal Dimension

New electron microscopy methods have also revealed new connections between local chromatin structure and global gene expression. The Backman Lab has used new Partial Wave Spectroscopy (PWS) methods to measure the local fractal dimension of chromatin, and have revealed statistically significant changes in the fractal dimension in pancreatic cells which develop cancer. It would be highly desirable to understand the local rearrangements which occur in the cancerous cells.

These data provide subnucleosome-level resolution which may be used to compare against simulations of a mesoscale model of DNA developed in our group. Using machine learning techniques, we can identify the chromatin configuration most consistent with this experimental data and conduct coarse-grain simulations of chromatin to interpret experimental changes in fractal dimension.

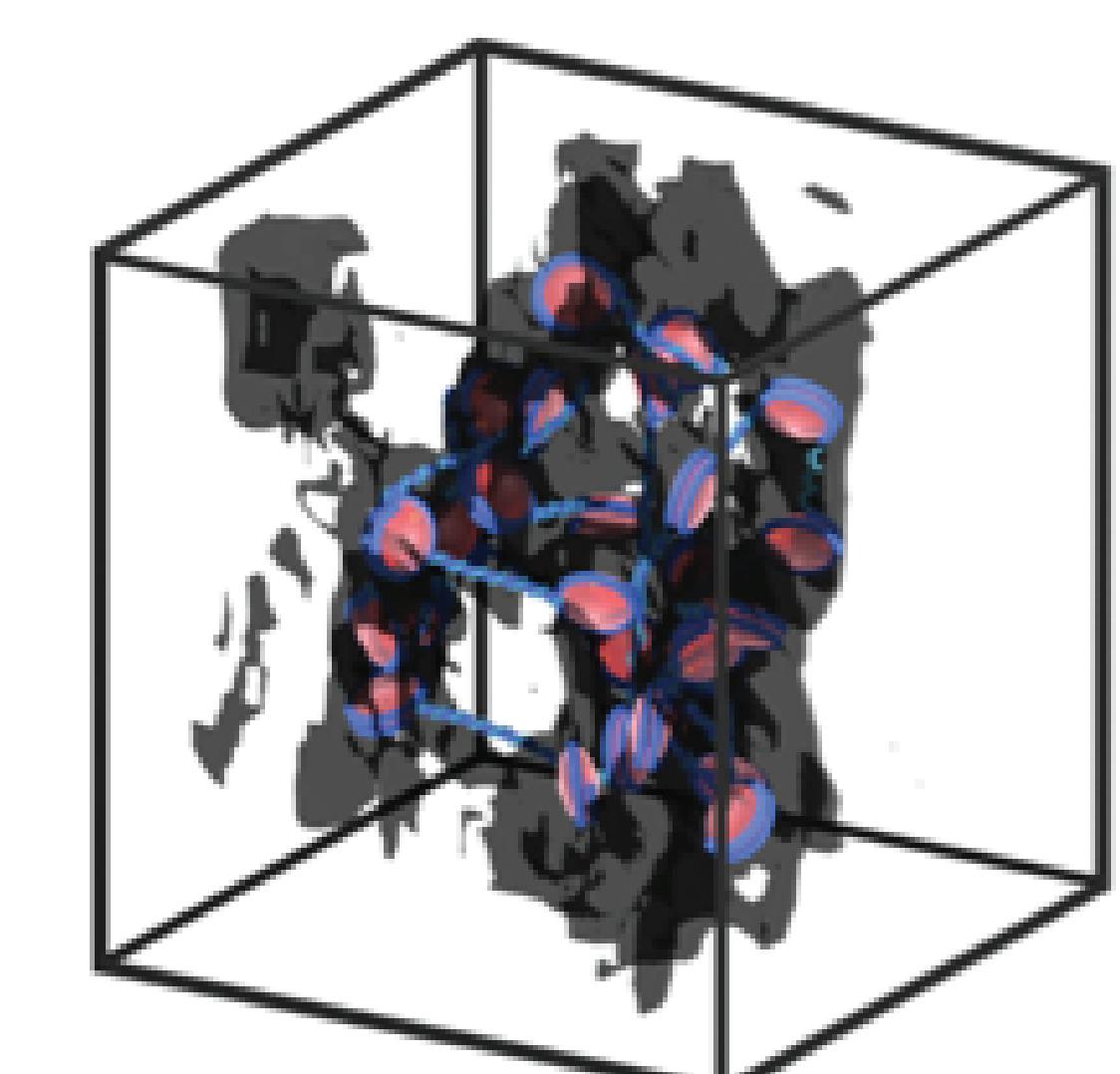


Figure 4. (grey) chromSTEM data representing electron density of dyes bound to DNA. (blue, red) 1 cylinder-per-nucleotide coarse-grain model of chromatin. Red represents histones, connected by blue linker DNA.

References

- Berger, Shelley L., et al. "An operational definition of epigenetics." *Genes & development* 23.7 (2009): 781-783.
- Moosavi, Azam, and Ali Motalevazadeh Ardekani. "Role of epigenetics in biology and human diseases." *Iranian biomedical journal* 20.5 (2016): 246.
- Imai, Ryosuke, et al. "Density imaging of heterochromatin in live cells using orientation-independent-DIC microscopy." *Molecular biology of the cell* 28.23 (2017): 3349-3359.
- Machida, Shinichi, et al. "Structural basis of heterochromatin formation by human HP1." *Molecular cell* 69.3 (2018): 385-397.
- Ferguson, Andrew L., et al. "Nonlinear dimensionality reduction in molecular simulation: The diffusion map approach." *Chemical Physics Letters* 509.1-3 (2011): 1-11.