Nucleosome [Mis]Positioning, Chromatin Folding and a Computational Karyotype

Rod Abcede[1], James Liman[3], Victoria Bamburg[3], John Bell[2], William Johnston[3], Thomas C. Bishop[3]
[1]Wossman High School, Monroe City School District;
[2]Department of Computer Science, Southern University – Shreveport;
[3]College of Engineering and Science, Louisiana Tech University

Abstract: Given a set of experimentally or theoretically determined nucleosome positions, it is possible to rapidly construct and interactively display 3D models of entire chromosomes. Our Interactive Chromatin Modeling Web (ICM-Web) server can fold and display megabase segments of chromatin in real time (1). These models are first order approximations which assume each nucleosome is a canonical octasome and that the linker DNA assumes a sequence specific conformation similar to free DNA. Thermal fluctuations may be included in the model that alters the nucleosome wrapping (i.e. entry/exit angle) and linker conformation. The models provide valuable insights, e. g. visualization of DNA rotational phasing on individual nucleosomes and how DNA defects alter global structure even if the chromatin folds are not necessarily accurate. The DNA defects arise from known sequence specific conformations and thermal fluctuations of DNA. Nucleosome positioning or mispositioning alters chromatin topology by exposing or hiding DNA defects. We utilized the CHA1, HIS3, PHO5 and MMTV promoter complexes to illustrate these ideas. The six positioned nucleosomes in the MMTV promoter complex yield a much more extended structure than is typically represented in literature. A compact chromatin structure for the MMTV requires more than six nucleosomes and thus mispositioning of some nucleosomes. Nucleosome [mis]positoning can hide or expose DNA defects in the MMTV promoter that may regulate chromatin looping associated with this promoter. Using whole genome nucleosome positioning data, we generate chromatin folds for each chromosome of *Saccharomyces cerevisiae* to produce a computational karyotype. Thermal motion associated with linker DNA plays a critical role in chromatin flexibility, and reduces the spatial range of each fiber, and allowing it to be compacted into the nucleus. ICM is available on the Chromatin Folding tab at http://www.latech.edu/~bishop

Keywords: Chromatin, nucleosome, mispositioning, MMTV, modeling, ICM, DNA

1. Introduction

Chromatin is a complex of DNA and proteins (histones) that condenses into highly compact chromosomes inside the nucleus during eukaryotic cell division. Chromatin's primary function is to package DNA such that it fits in the cell nucleus. Packaging also impacts all genetic mechanism including gene regulation, replication, repair and transcription.

A nucleosome is a subunit of chromatin composed of a short length of DNA (147bp) wrapped around a core of histone protein molecules (an octamer). Each nucleosome is about 11nm in diameter. For more than 2 decades, chromatin structure has been extensively studied and a number of models have been proposed as to how the nucleosomes and a linker (a short sequence of DNA that connects two adjoining nucleosomes) are organized within chromatin fibers. The two-angle model characterizes the angle (α) between adjacent segments of linker

DNA and the twisting angle (β) between successive nucleosomes. Varying the values of these 2 angles determine the different conformations of polynucleosome chains. Various examples of two-angle chromatin fibers were described in the initial work by Woodcock (3). Computational (4) and mathematical analysis (5) of two-angle fibers led to the elucidation of a complete map of chromatin topology as a two-dimensional phase diagram. Coarse-grained models for simulating conformation and dynamics of chromatin fiber represent a different approach (6).

Genome-scale identification of nucleosome positions was first reported for yeast (7) and statistical positioning of nucleosomes throughout the yeast genome was identified using a barrier nucleosome model (8). A number of whole genome nucleosome maps of the yeast genome have been curated to determine consensus nucleosome positions (2). In this paper, we report findings of our current investigations in nucleosome [mis]positioning, chromatin folding, and computational karyotype utilizing our ICM-Web server.

2. Methods

ICM utilizes the algorithm for assembling 3D structures of double stranded DNA from dinucleotide stacking data presented by El Hassan (9). The ICM web server is an interactive tool that allows users to rapidly assess nucleosome stability and fold sequences of DNA into putative chromatin templates. It takes DNA sequences as input and generates a nucleosome energy level diagram, coarse-grained representations of free DNA and chromatin, and plots of helical parameters (tilt, roll, twist, shift, slide and rise) as a function of position (1). The DNA in ICM properly accounts for known sequence specific conformation and thermal fluctuations. ICM allows the user to choose these properties from several published sets. We used the DNA properties obtained from a comprehensive molecular dynamics study of B-form DNA (10). The user can select from several different energy models, nucleosome structures and methods for placing nucleosomes in the energy landscape. Alternatively, if nucleosome footprints are known from experiment, ICM web can use these positions to create a nucleosome array.

Typical cartoon representations, as shown in Fig. 1, are often regular and compact. Here we utilize experimentally determined nucleosome positioning data for CHA1 (chrIII:15937..18937), PHO5 (chrII:429991..432991), HIS3 (chrXV:720917..723917) (2) to fold these promoters into nucleosome arrays.

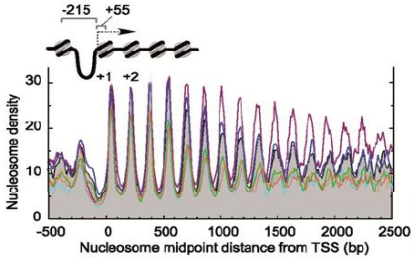


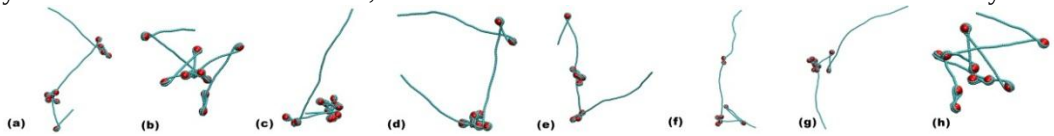
Figure 1. Shows a cartoon of chromatin above the consensus distribution of nucleosomes around the transcription start site (TSS) as obtained in reference 2. Each peak in the graph corresponds to a nucleosome position.

For nucleosome [mis]positioning, we used five tandem repeats of MMTV promoter sequence and observed chromatin bending in the 3-D image generated by ICM. We utilized RMSD fitting of every mono-, di-, and tri-nucleosomes to analyze local nucleosome packing. To produce a computational karyotype, we used whole genome nucleosome positioning data compiled by Jiang and Pugh (2) to generate chromatin folds for each chromosome of *Saccharomyces cerevisiae.*

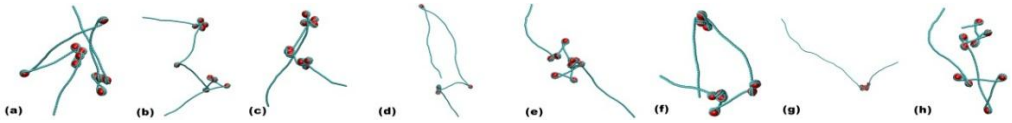
3. Results and Discussion

The folding of promoter sequences into chromatin images are shown in figure 2. These models were generated by ICM (www.latech.edu/~bishop) using promoter sequences from *S. cerevisiae.* The first six models utilize experimentally determined positioning data as follows: (a) Mavrich(11); (b) Field(12); (c) Jiang and Pugh(2); (d) Shivaswamy(13); (e) Whitehouse(14); and (f) Lee(15). The seventh model (g) utilized consensus

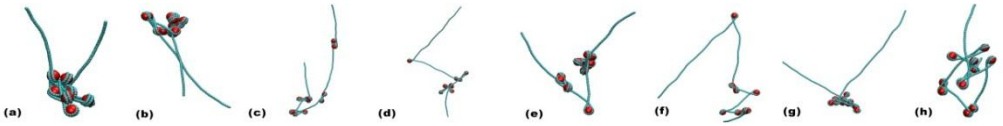
positioning reported in Jiang and Pugh (2). The last model (h) used positioning automatically determined by ICM with default energy parameters and minimum linker length of 19bp. In each image red spheres represent histone octamers and smaller cyan spheres represent 5bp long segments of DNA. In all cases there are unoccupied segments of DNA resulting in an extended conformation for the chromatin fold. In some instances there is a tendency to form clusters of nucleosomes, but in no instance can a fiber-like structure be definitively identified.



CHA1 folding, S. cerevisiae, http://www.yeastgenome.org, chrIII:15937...18937



HIS3folding, S. cerevisiae, http://www.yeastgenome.org, chrXV:720917...72391



PHO5 folding, S. cerevisiae, http://www.yeastgenome.org, chrII:429991...432991

Figure 2.Promoter folding as indicated by text.*(See results and discussion for details)*

Figure 3 shows a fiber like chromatin model generated by ICM using a 19bp linker and tandem repeats (5x) of the MMTV sequence. The fiber is clearly bent over a local region (see insert in a). Since all nucleosomes in this model are identical, the bend can only arise due to sequence specific conformations of linker DNA revealed by changes in histone positions. Since the sequence is repeated five times, in some instance(s) positioning reveals a bent linker. In other instances, the bent linker is hidden. Figures 3b to 3d seek to identify the origin of the fiber bending by illustrating RMSD fitting of every mono- di- and tri- nucleosome, respectively. The images indicate variability in linker conformations mono-, di-, and tri-stacking of nucleosome. However, images do not clearly identify a single linker or local stacking that can be associated with the bend in Figure 3a. We conclude that the bend may extend over more than three nucleosomes or result from additive effects of linker deformations that are far separated in sequence space rather than additive bending by defects in successive linkers.

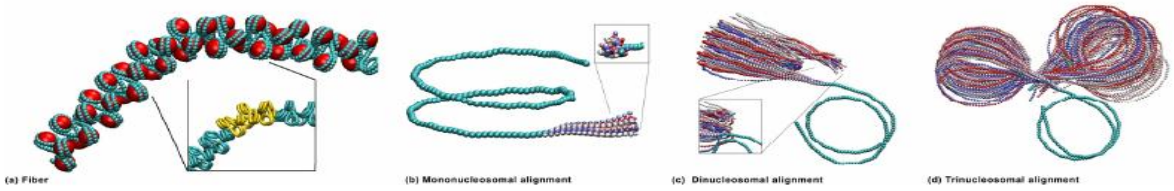


Figure 3. [Mis]positioning and Chromatin Bending. (a) Tandem repeats of the MMTV sequence (5x). Root Mean Square Deviation (RMSD) fitting of every (b) mono-, (c) di-, (d) tri-nucleosome of the MMTV.

Figure 4 provides images for the folding of entire chromosomes by ICM. Figure 4a shows the folding of chromosome I as a function of nucleosome occupancy. The model is extended and irregular when only few nucleosomes are present. As nucleosome occupancy increases, crowding requires the nucleosomes to assume regular spacing producing a fiber like structure that can be used to investigate gross features of chromatin folding in figure 4b. Assembling fiber like structures for all sixteen chromosomes of *S. cerevisiae* produces a so-called computational karyotype (Bar indicates ~5 μm). On a workstation, ICM is fast enough to fold the entire yeast genome into a karyotype in under 5 minutes and a netbook grade computer is sufficiently powerful to display this entire karyotype as an interactive molecular graphic.

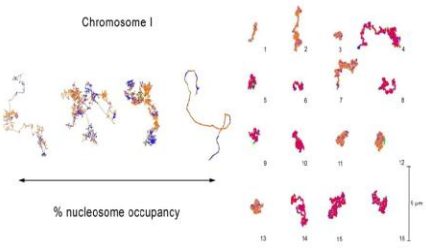


Figure 4.a. Chromosome I occupancy; b. Karyotype of 16 chromosomes of *S. cerevisiae*

4. Conclusion

The model of nucleosome folding utilized by ICM is presently overly simplistic. A more complete description of chromatin should account for all intra- and extra-nucleosomal interactions. A number of efforts are presently underway to upgrade the features of our ICM web tools. Nevertheless, as shown above, there is no similar conformation for both consensus and specific experiment data for CHA1, HIS3, and PHO5 folding. Any single set of positions may differ significantly from the consensus. The nucleosome [mis]positioning data do not clearly identify a single linker or local stacking that can be associated with the bend in MMTV tandem repeats. We were unable to identify any single linker or set of adjacent linkers that causes the MMTV to bend in our ICM models. More careful analysis is required to identify the source of bending. Finally, our computational karyotypes can be utilized in a manner similar to traditional karyotype studies, namely, to identify gross physical properties rather than detailed features. Identification of relationships between structural features and bioinformatics data, e.g. nucleosome free regions and sequence labels, can be easily highlighted in a computational karyotype. ICM Web is a crucial tool for investigating molecular mechanisms and for rapidly assembling models of chromatin that can be employed to rationalize biophysical data, especially spatial relations. To our knowledge, there is no similar tool available to the scientific community.

5. Acknowledgments

The current work is funded by the NSF EPSCoRCyberTools project under award #EPS-0701491.

6. References

(1) R. Stolz & T. Bishop, Nucleic Acids Research, 2010, Vol. 38, doi:10.1093/nar/gkq496.
(2) C. Jiang & B. F. Pugh, Nature Review Genetics 10, 161-172 March 2009, doi:10.1038/nrg2522
(3) Woodcock et al. Proc. Natl. Acad. Sci. USA, Oct. 1993, 90: 9021-9025
(4) P. M. Diesinger & D.W. Heermann, Biophysical Journal, Oct. 2009, Vol. 97, 2146-2153
(5) H. Schiessel J. Phys.: Condens. Matter, 2003, 15: R699-R774
(6) N. Korolev et al. Current Opinion in Structural Biol, 2012, 22:151-159
(7) G.C. Yuan, Science 309, 626 (2005), doi: 10.1126/science.1112178
(8) Mavrich et al. Genome Res. June 2008, doi: 10.1101/gr.078261.108
(9) M.A. El Hassan & C.R. Calladine, J. Mol. Biol. 1995, 251:648-664
(10) Lavery et al. Nucleic Acids Research, 2010, Vol.38, No. 1, 299-313
(11) Mavrich et al. Nature 2007, 446:572-576
(12) Field et al. PLoS Comput Biol. 2008, 4:e1000216
(13) Shivaswamy et al. PLoS Biol 2008 6:e65
(14) Whitehouse et al. Nature 2007, 450:1031-1035
(15) Lee et al. Nat Genet 2007, 39:1235-1244