Developing the Interactive Chromatin Modeling Program

John Bell[1], Rod Abcede[2], William Johnston[3], James Liman[4], Victoria Bamburg[5],Thomas C. Bishop[3]

[1]Department of Computer Science, Southern University at Shreveport-Bossier
[2]Wossman High School, Monroe City School District
[3]Department of Physics and Chemistry, Louisiana Tech University
[4]Department of Biomedical Engineering, Louisiana Tech University
[5]Department of Chemical Engineering, Louisiana Tech University

Abstract: The Interactive Chromatin Model web server (ICM-Web, www.latech.edu/~bishop) is a tool to model our present understanding ofhow chromatin folds experimentally. ICM-Web integrates both bioinformatics and computational biology allowing the user to input a sequence of DNA and choose from several different energy models and nucleosome placements. ICM-Web then generates a nucleosome energy diagram, 3D representation of the molecule and plots of helical parameters. The program is being changed from FORTRAN 77/90 to an object-oriented design format in C++, additionally the program will be expanded to allow for greater selection of genomic data and moved into a free source code hosting platform

Keywords:DNA, nucleosome, chromatin modeling, bioinformatics, computational biology

1. Introduction

The nucleosome hypothesis was unveiled in 1974and states that a nucleosome consists of DNA tightly wrapped around a of core histones. In 1993 Woodcock introduced a two-angle model that describesthe folding of nucleosomes into chromatin[1]. Computational chromatin modeling and imaging has advanced[2] greatly in the last two decades. However the structure of chromatin is still subject to debate [3].

ICM is based on El Hassan's algorithm for reconstructing a 3D model of DNA from a set of DNA helical parameters[4]. Our program generates a 3D model of DNA by utilizing base pair specific helical parameters values. To make a model of chromatin values for free DNA are replaced with helical parameter values describing the nucleosome super helix where ever the histones contact the DNA, i.e. the nucleosome footprint. ICM automatically locates the footprints at minimum energy locations in the nucleosome energy [5]. Alternatively, the footprints may be obtained from experimental data (Jiang and Pugh's 2009 [6]) or other prediction algorithms. Using this approach, ICM can generate an all-atom model or coarse-grain representation at various resolutions.

The current version of ICM evolved rapidly from its original design. The rapid evolution produced an unstructured mix of FORTRAN 77 and FORTRAN 90. The primary limitation is that the program allows only one type of nucleosome footprint.Addressing this limitation requires a program that is both dynamic and better structured. This program will be written in an object-oriented design (OOD) structure. There will be a provision in ICM to take into account several types of nucleosomes in the molecule at the same time. For sharing and

collaborative development the program will be uploaded into a free source hosting platform, Bitbucket.

2. Methods

ICMis designed to be controlled by a configuration file. The web interface (ICM-Web) generates this configuration file by requiring the user to specify the following: a DNA sequence, nucleosome placement options, nucleosomal and free DNA helical parameters, and the amount of thermal motion The user has a choice of how to input a DNA sequence either by choosing from a predefined list, inputting in their sequence or uploading a sequence from PubMed. When the user clicks the 'Go' button, ICM initiates three calculations and generates the results page. The first calculation determines the energy landscape. The second determines nucleosome footprints via an iterative search criteriaand checks the validity of the user defined start positions before placing nucleosomes at the indicated positions. Two sets of helical parameters are then generated: one describes free DNA; the other describes the putative chromatin template. The third calculation converts these two sets of DNA helical parameters to Cartesian coordinate representations using the algorithm in [5].

The underlying algorithm for folding DNA and chromatin will remain the same, but the capabilities of the program will fundamentally differ in several ways. The issue ofthere being only one type of nucleosome for the whole molecule will be addressed.To achieve thisthe user must select a list of nucleosomes, rather than just one type and a set of non-overlapping footprints must be determined. The footprints can be determined automatically or specified by the user.For example, generating the chromatin fold in Figure 2, requires only knowledge of the DNA sequence and footprint locations as depicted in Figure 1. Here we explicitly allow for different types of nucleosome footprints. Sequence and occupancy work closely together. For a given sequence of DNA the type of footprint will determine the occupancy value for each set of base pair. Each base pair corresponds to an array of sixteen helical parameters based on the footprint. El Hassan's algorithm is then used to calculate the path of DNA generated and histones docked onto the DNA as each footprint location.

Chromatin
[footprint type 0][footprint type 1] [footprint type 0][footprint type 2] [footprint type 0]
Sequence: T G C C T A G G T C T T C T T T A A A A A G T A A A A A G G G G G A A A T G C C
Occupancy: 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 0 0 0 0 0 0 0 0

Figure 1.Chromatin Structure showing DNA sequence and Occupancy code according to footprints.

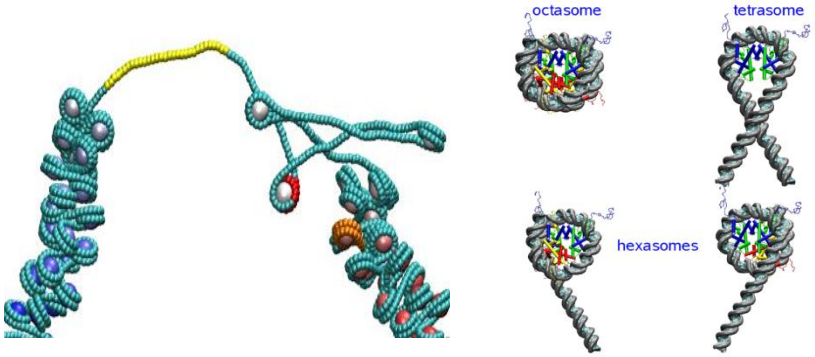

Figure 2.The figure on the left shows how the current program only allows for one type of nucleosome. Thefigure on the right displays four different types of nucleosomes.

ICM will implement an OOD data structure. The various functions of the program will be treated as objects. Each object will consist of data and operations on that data. The program will consist of a collection of interacting objects. C++ is designed especially to implement OOD.

Uploading the program into Bitbucket will give the program a greater audience because Bitbucket is a free source code hosting platform. With Bitbucket's editing features anyone can make recommendations, give feedback on the program or write additional code which this group can accept or reject. Because Bitbucket uses distributed version control systems it makes it easy to manage the program because development is tracked and logged. The program is open to the whole scientific community and updates can occur almost instantaneously. Documentation on ICM can be found when logging into Bitbucket in the README file, which explains the program including the web address for the web interface.

3. Results and Discussions

The resulting program will have several key features: Open source distribution, highly portable C++, dynamic memory, user defined diagnostics anda graphical user interface. Molecular images can be displayed by Jmol via a web interface or saved and uploaded in VMD. The program will maintain its present functionality and be extended to generate structures for any type of footprint on the DNA. Upload the program onto Bitbucket provides a platform for scientists and programmers to use ICM-Web and leave feedback which is a great boost in its availability, Figure 3.

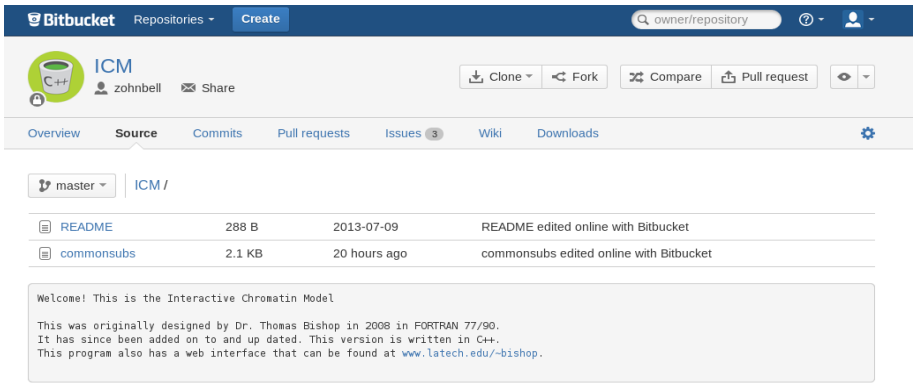

Figure 3. Bitbucket page showing ICM

4. Conclusion

As we progress in our knowledge of chromatin structure and the influences that determine that structure, we plan to modify ICM to more accurately reflect the true structure of chromatin and how it is folded. Additionally the webpage update is set for a future goal. This will require further integration of bioinformatics and computational biology.These are all necessary steps because tools like ICM-Web are critical for the rational design of chromatin studies in addressing questions and for structural analysis of proposed mechanisms of action for chromatin.

5. Acknowledgments

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

6. References

[1] Woodcock, Grigoryev; Horowitz, Whitaker.(1993). A Chromatin Folding Model That Incorporates Linker Variability Generates Fibers Resembling Native Structures. Proc. *Natl. Acad. Sci. USA.*Vol. 90, pp. 9021-9025 Oct. 1993

[2] Korolev, Nikolay; et al. (2012). Modeling chromatin structure and dynamics: status and prospects.*Current Opinion in Structural Biology*.Vol 22 Issue 2. April 2012. pp 151-159.

[3] Bishop, T C. (2011). Enhancing ICM-Web: A Prototype Computational Core Activity.*LBRB*. September 2011.

[4] El Hassan, M. A., Calladine, C. R.(1995).The Assessment of the Geometry of Dinucleotide Steps in Double-Helical DNA; a New Local Calculation Scheme. *Journal of Molecular Biology*, 251, 648-664.

[5]Stolz, Richard; Bishop, Thomas. (2010). ICM-Web: The interactive chromatin modeling webserver. *Nucleic Acids Research*.Vol 38 2010. Webserver issue W254-W261.

[6] Jiang, Cizhong; Pugh, Franklin. (2009). A compiled and systematic reference map of nucleosome positions across the *Saccharomyctescervisiae genome*.Genome Biology 2009.R109.1 – R109.1.