Genomics Proteomics Bioinformatics 12 (2014) 249–253

HOSTED BY

ELSEVIER

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com

APPLICATION NOTE

nuMap: A Web Platform for Accurate Prediction of Nucleosome Positioning

Bader A. Alharbi [1,#], Thamir H. Alshammari [2,#], Nathan L. Felton [3], Victor B. Zhurkin [4], Feng Cui [1,*]

[1] *Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, NY 14623, USA*
[2] *B. Thomas Golisano College of Computing & Information Sciences, Rochester Institute of Technology, Rochester, NY 14623, USA*
[3] *Information Technology Services, Rochester Institute of Technology, Rochester, NY 14623, USA*
[4] *Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA*

Received 5 July 2014; revised 3 August 2014; accepted 5 August 2014
Available online 16 September 2014

Handled by Zhihua Zhang

KEYWORDS

Nucleosome rotational positioning;
Sequence-dependent DNA anisotropy;
Prediction of nucleosome positioning;
Sequence patterns;
Web server

Abstract Nucleosome positioning is critical for gene expression and of major biological interest. The high cost of experimentally mapping nucleosomal arrangement signifies the need for computational approaches to predict nucleosome positions at high resolution. Here, we present a web-based application to fulfill this need by implementing two models, YR and W/S schemes, for the translational and rotational positioning of nucleosomes, respectively. Our methods are based on sequence-dependent anisotropic bending that dictates how DNA is wrapped around a histone octamer. This application allows users to specify a number of options such as schemes and parameters for threading calculation and provides multiple layout formats. The nuMap is implemented in Java/Perl/MySQL and is freely available for public use at http://numap.rit.edu. The user manual, implementation notes, description of the methodology and examples are available at the site.

Introduction

The basic repeating unit of eukaryotic chromatin is the nucleosome, which contains a 145–147 bp DNA fragment wound around a histone core. The positioning of nucleosomes on a DNA fragment is critical for gene expression [1]. Along with other factors such as DNA methylation and histone modifications, DNA sequence is one of the most important factors guiding nucleosome positioning *in vitro* and *in vivo* [2]. Accurate determination of nucleosome positions is extremely important to study gene regulatory mechanisms. Unfortunately, micrococcal nuclease, which is commonly used for mapping nucleosome positions experimentally, exhibits high AT-specificity [3,4] and lacks the resolution to identify exact nucleosome positions, especially for nucleosomes with the GC-rich boundaries [5]. Therefore, prediction of nucleosome positioning on genomic sequence at high resolution is of great biological interest.

Nucleosome positioning is usually characterized by two parameters: rotational positioning, which describes the side of the DNA helix that faces the histones, and translational

* Corresponding author.
E-mail: fxcsbi@rit.edu (Cui F).
# Equal contribution.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

http://dx.doi.org/10.1016/j.gpb.2014.08.001

positioning, which determines the nucleosome midpoint (or dyad) with regard to the DNA sequence [6]. We have developed two simple models, the YR scheme [7] and W/S scheme [8], for the prediction of translational and rotational positioning of nucleosomes, respectively. Both methods successfully predict ~80% of nucleosome positions *in vitro* with 2-bp precision (see Supplementary materials at http://numap.rit.edu/app/dna/suppMaterial.xhtml). Here, we present a web-based application that implements both models for prediction of nucleosome positioning patterns.

Methods

Both YR and W/S schemes are based on the sequence-dependent DNA anisotropy [9,10], which dictates how DNA is wrapped around a histone octamer. One of the best established sequence patterns consistent with this anisotropy is the periodic occurrence of AT-containing dinucleotides (WW) and GC-containing dinucleotides (SS) in the nucleosomal locations where DNA is bent in the minor and major grooves, respectively [11]. The minor and major groove bending sites are defined by the base-pair step Roll values observed in nucleosome crystal structures, in which a DNA fragment of 147 bp or 146 bp in length is co-crystallized with histones [12]. Based on the Roll values, 14 minor-groove DNA bending sites and 12 major-groove DNA bending sites were selected from the 147-bp or 146-bp nucleosomal DNA template; each site is 4 bp in length (see the Methods section on the server website for details). The W/S scheme implements the aforementioned periodic WW and SS patterns. The W/S score $S_{W/S}(n)$ of the threaded sequence with the center at position n can be calculated as

$$S_{W/S}(n) = \sum_{\text{minor site}=1}^{14} C_{WW} + \sum_{\text{major site}=1}^{12} C_{SS} - \sum_{\text{minor site}=1}^{14} C_{SS} - \sum_{\text{major site}=1}^{12} C_{WW} \quad (1)$$

where C_{WW} and C_{SS} are the total occurrences of WW and SS dinucleotides at a given minor-groove or major-groove DNA bending site.

In addition to the WW and SS motifs, the YR scheme incorporates GC, pyrimidine–purine (YR), YYRR and RYRY motifs to take into account their anisotropic bending into DNA grooves (see Methods at the site for details). The YR score $S(n)$ of the threaded sequence with the center at position n can be calculated as

$$S(n) = \sum_{\text{pattern}=1}^{28} \left(w_x \sum_{\text{site}=1}^{12 \text{ or } 14} C_x \right) \quad (2)$$

where C_x is the occurrence of a designated motif x at a given site and w_x is the weight of the motif at this site. A total of 28 DNA sequence patterns are used in the YR scheme [7]. If a motif occurs at the site where DNA is severely bent, its occurrence is given a higher weight than at other sites (see Methods section on the website for details).

Implementation

The nuMap web application takes a DNA sequence as input and returns both numeric scores and corresponding profiles. We use the model-view-controller (MVC) design, in which the communication between the client and the database is mediated by the web application server (**Figure 1**). The server is implemented using a combination of extensible hypertext markup language pages (XHTML), and JavaServer Faces (JSF) as a rich component-based user interface. Nucleosome positioning scores can be computed by the YR and W/S schemes implemented at the backend of the server as Perl scripts. An open-source reporting engine, JasperReports library, is used with the combination of Java codes for reporting the results in graphical output in multiple formats including PDF, Excel and CSV (see Implementation Notes at the server site for details).

Prediction options and features

Figure 2 shows the general layout for nuMap user interface. The user can select from four different output layouts. The "Single Layout" is used to generate single graph for each of the input sequences, whereas the "Superimpose Layout" overlays more than one graph, if multiple sequences are used as input. The "Average Layout (Asymmetric)" is used to calculate the average scores for multiple sequences, while the "Average Layout (Symmetric)" produces a symmetrical graph by generating the reverse complement of all the input sequences and then calculating the average scores at each base pair position. The setting panel allows the user to choose from two prediction schemes, YR and W/S. By default, a chosen scheme will calculate a score for the first nucleosome starting at the position 1 of the input sequence, and assign that score to the dyad position 74. The user can also give a specific value to the start of the input sequence (if not 1) by changing the value in the "Starting Position" box. The "Weight Set" option is specific to the YR scheme, the user can choose from three different sets used initially to establish this scheme with each set evaluating the occurrence of specific pattern at specific site differently (see Table SIII in [7]).

The nuMap provides numerous reporting formats such as PDF (**Figure 3**), Word and Excel. Moreover, the server offers many features that allow the user to interact with output data in different ways such as zooming in a large graph, which is useful to investigate a certain range of values, and exploring dynamic data features, which show the corresponding score/base pair position when a user places the mouse over any point in the graph (**Figure 4**).

Future developments

Other models [13,14] work well in predicting nucleosome occupancy *in vivo* by introducing a position-independent component, P_L, to represent sequences that are generally favored or disfavored regardless of their position within the nucleosome (most notably, poly(dA:dT) tracts, which are strongly disfavored by nucleosomes) [14]. Detailed comparisons in prediction accuracy between these models and ours have been made and will be published separately. The incorporation of this component into the YR or W/S scheme has a potential to improve both the rotational and translational positioning predictions of nucleosomes *in vivo*.

Various models have been developed for nucleosome positioning predictions [15–19] and for gene regulatory

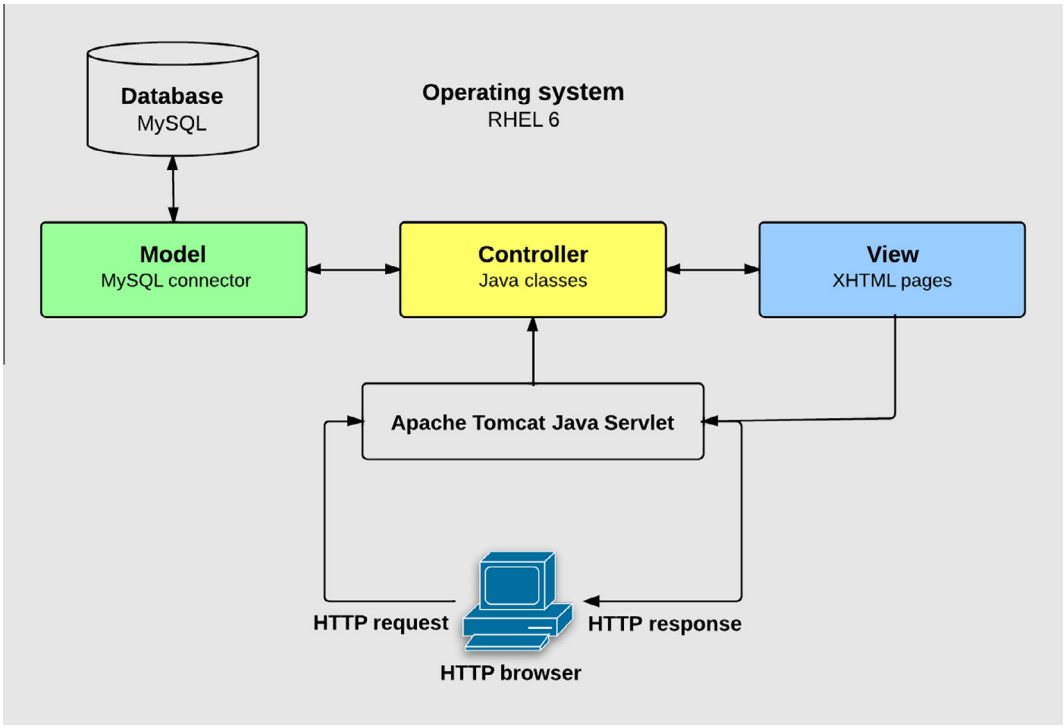

Figure 1 System infrastructure of the nuMap server
The MVC model describes an application in terms of three different design modules, the Model, the View and the Controller. The Model describes how the application represents the data model, including data files and a database management system (*e.g.*, MySQL). The View describes what the user interface looks like and how the information is displayed at the user end (*e.g.*, XHTML). The Controller part represents how the user interface is managed and controlled by the application code. MVC, model-view-controller.

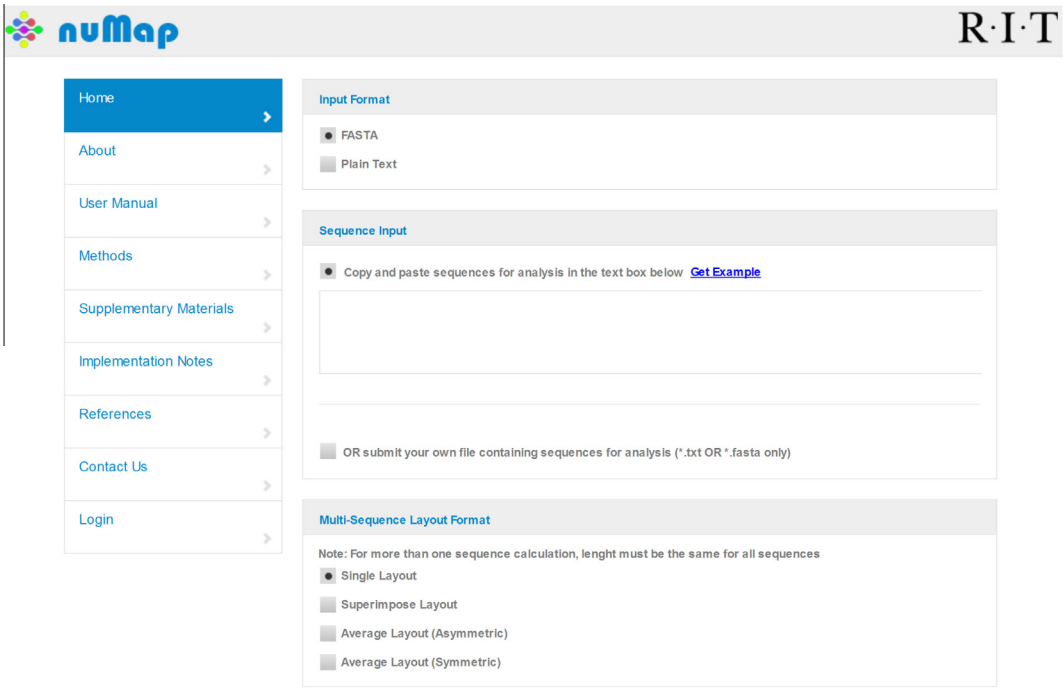

Figure 2 A screen shot of the nuMap server interface

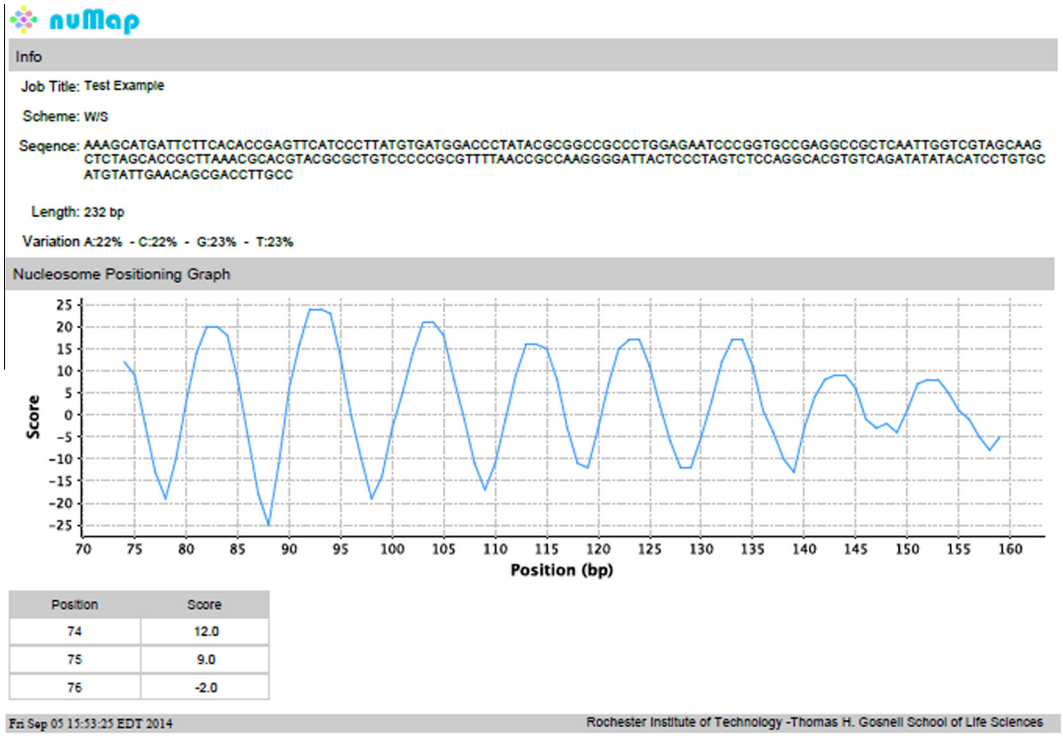

Figure 3 Example of nuMap output in PDF format

A sample out of the nuMap server is shown. The Information Panel is shown on the top, containing job-related information such as sequence name and user-selected options including prediction scheme, sequence, its length and composition. A profile of the nucleosome positioning score of the '601' sequence, in this case, predicted by the W/S scheme is shown underneath. The physical meaning of the score is straightforward – the lower the DNA deformation energy for particular sequence motifs placed at the sites of DNA distortion, the higher the score. The score is calculated for each 147-bp window and is assigned to the center position of a given window. The actual score at each position is given in a tabular form.

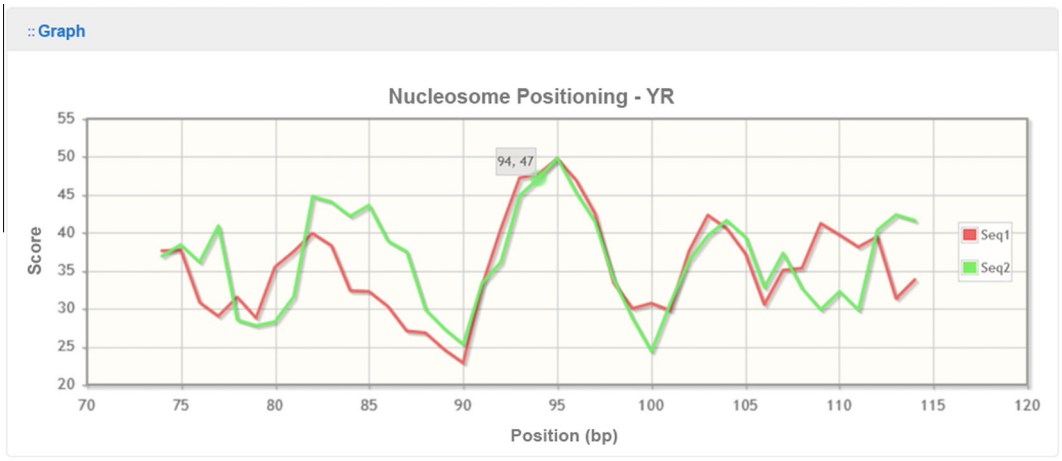

Figure 4 A screen shot of nuMap output of superimposing profiles for two sequences

A screen shot of the nuMap output for two sequences superimposed together is shown. The score value on a particular position can be shown on the screen by mouse. In this case, a mouse hovering over position 94 shows the score, 47, at this position. This example illustrates dynamic data features of the server. The YR (pyrimidine–purine) scheme was used to calculate the scores. The physical meaning of the scores is given in the legend in Figure 3.

analysis [20,21]. Nonetheless, one has to browse through different servers, which often have different formats and representations, making the comparison of the results extremely difficult. We will overcome this obstacle by reprogramming these methods, incorporating them into the nuMap server and presenting the results in the same format. Comparison of all the methods in this way will allow detailed analyses of the strengths and weaknesses of each approach, facilitat-

ing our understanding of the biophysical principles of nucleosome positioning.

Authors' contributions

BAA and THA developed the server, performed the analyses and drafted the paper. NLF set up the server. VBZ and FC contributed to data analysis and paper writing. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This study was supported by the scholarship of the King Abdullah International Medical Research Center (KAIMRC) in Saudi Arabia to BA, the start-up fund, Faculty of Development (FEAD) fund and Dean's Research Initiation Grant (D-RIG) fund of Rochester Institute of Technology in the USA awarded to FC and the Intramural Research Program of National Cancer Institute, NIH of the USA to VBZ.

References

[1] Kornberg RD, Lorch Y. Twenty-five years of the nucleosomes, fundamental particle of the eukaryotic chromosome. Cell 1999;98:285–94.
[2] Struhl K, Segal E. Determinants of nucleosome positioning. Nat Struct Mol Biol 2013;20:267–73.
[3] Dingwall C, Lomonossoff GP, Laskey RA. High sequence specificity of micrococcal nuclease. Nucleic Acids Res 1981;9:2659–73.
[4] Hörz W, Altenburger W. Sequence specific cleavage of DNA by micrococcal nuclease. Nucleic Acids Res 1981;9:2643–58.
[5] Nikitina T, Wang D, Gomberg M, Grigoryev SA, Zhurkin VB. Combined micrococcal nuclease and exonuclease III reveals precise positions of the nucleosome core/linker junctions: implications for high-resolution nucleosome mapping. J Mol Biol 2013;425:1146–60.
[6] Travers AA, Klug A. The bending of DNA in nucleosomes and its wider implications. Philos Trans R Soc Lond B Biol Sci 1987;317:537–61.
[7] Cui F, Zhurkin VB. Structure-based analysis of DNA sequence patterns guiding nucleosome positioning *in vitro*. J Biomol Struct Dyn 2010;27:821–41.
[8] Cui F, Zhurkin VB. Rotational positioning of nucleosomes facilitates selective binding of p53 to response elements associated with cell cycle arrest. Nucleic Acids Res 2014;42:836–47.
[9] Zhurkin VB. Anisotropic flexibility of DNA and the nucleosomal structure. Nucleic Acids Res 1979;6:1081–96.
[10] Trifonov EN. Sequence-dependent deformational anisotropy of chromatin DNA. Nucleic Acids Res 1980;8:4041–53.
[11] Satchwell SC, Drew HR, Travers AA. Sequence periodicities in chicken nucleosome core DNA. J Mol Biol 1986;191:659–75.
[12] Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. J Mol Biol 2002;319:1097–113.
[13] Segal E, Fondufe-Mittendorf Y, Chen L, Thåström A, Field Y, Moore IK, et al. A genomic code for nucleosome positioning. Nature 2006;442:772–8.
[14] Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, et al. The DNA-encoded nucleosome organization of a eukaryotic genome. Nature 2009;458:362–6.
[15] Tolstorukov MY, Choudhary V, Olson WK, Zhurkin VB, Park PJ. nuScore: a web-interface for nucleosome positioning prediction. Bioinformatics 2008;24:1456–8.
[16] Guo SH, Deng EZ, Xu LQ, Ding H, Lin H, Chen W, et al. iNuc-PseKNC: a sequence-based predictor for predicting nucleosome positioning in genomes with pseudo k-tuple nucleotide composition. Bioinformatics 2014;30:1522–9.
[17] Xi L, Fondufe-Mittendor Y, Xia L, Flatow J, Widom J, Wang JP. Predicting nucleosome positioning using a duration Hidden Markov Model. BMC Bioinformatics 2010;11:346.
[18] Stolz RC, Bishop TC. ICM web: the interactive chromatin modeling web server. Nucleic Acids Res 2010;38:W254–61.
[19] Gabdank I, Barash D, Trifonov EN. FineStr: a web server for single-base resolution nucleosome positioning. Bioinformatics 2010;26:845–6.
[20] Guan D, Shao J, Zhao Z, Wang P, Qin J, Deng Y, et al. PTHGRN: unraveling post-translational hierarchical gene regulatory networks using PPI, ChIP-seq and gene expression data. Nucleic Acid Res 2014;12:W130–6.
[21] Guan D, Shao J, Deng Y, Wang P, Zhao Z, Liang Y, et al. CMGRN: a web server for constructing multilevel gene regulatory networks using ChIP-seq and gene expression data. Bioinformatics 2014;30:1190–2.