A Wavelet Based Method to Predict Nucleosome Positions

Laura Riva [(1)(2)], Giorgia Menozzi [(1)], Manuela Sironi[(1)], Sergio Cerutti[(2)] and Uberto Pozzoli [(1)]

[(1)]IRCCS E. Medea, Associazione La Nostra Famiglia, 23842 Bosisio Parini (LC) -Italy
laurariva@bp.lnf.it
[(2)]Department of Biomedical Engineering, Polytechnic University, Milan -Italy

Keywords: Nucleosome, Bendability, Wavelet Transform.

Introduction

The nucleosome core particle is the fundamental repeating subunit of chromatin. It consists of two molecules each of the four 'core histone ' proteins, H2A, H2B, H3 and H4, and a 147 bp stretch of DNA. A better knowledge of the chromatin nucleosomal organization is crucial to understand many important phenomena occurring in chromosomes. Regulatory mechanisms of gene expression are partially influenced by nucleosome positioning and regions with exposed chromatin (i.e. where nucleosomes are more distant) can be more prone than others to double strand breaks. Analysis of nucleosomal DNA has demonstrated the existence of a weak sequence-dependent signal for nucleosome positioning, this makes classical computational biology methods, like alignment and consensus sequences, poorly applicable here. The ability of DNA to assume certain conformation in certain positions can considerably enhance its binding potential to nucleosomes. According to recent X-ray structure studies, the 147 bp nucleosomal DNA has detectable bends [1] symmetrically displaced around the central position, this suggests the presence of localized periodicities in DNA bendability. Wavelet transform can be used to locally evaluate periodicities allowing to detect positions with a bend distribution similar to known nucleosomal DNA.

Methods

Nucleosomal DNA sequences were retrieved from PDB [2] and from nucleosomal database [3]. G+C content profile (window 10bp, step 1 bp) and bendability [4] have been compared in terms of frequency coherence. The coherence between the two signals showed high values (>0.8) in a wide range between 3 bp and 1000 bp periodicity. Thus we used G+C content to evaluate the periodicity profiles using a Morlet wavelet transform. We calculated the wavelet transform of well studied nucleosomal DNAs (146bp) using their positive envelopes to obtain models of the nucleosomal periodicity pattern. To predict nucleosomes in generic DNA sequences we calculate the wavelet envelope of the sequence, then, we assign to each position the correlation coefficient between the 146 long envelope fragment around that position and the model pattern. The obtained correlation profile is used to predict nucleosome positions using a threshold.

Results and Discussion

In figure 1 there is the confrontation between recon profile prediction [5] [6] (dotted line) and wavelet profile prediction (continuous line) for Xenopus TFIIIA gene that contains two positive *cis*-elements positioned very close to each other separated from a TATA element by approximately 190 nucleotides. This 190-nucleotide spacing is critical for efficient transcription of the gene and a nucleosome is positioned in this region [5]. In figure 1 the TATA box is located within 394-400 region relative to sequence start and the centre of nucleosome site is placed approximately at 290 (hatched vertical line). Both our and Recon prediction seem to localize the dyad axis of nucleosome, nevertheless our profile is more allow for easier identification of nucleosomal positions. Fig 2 shows the wavelet profile prediction (A) and the recon profile (B) for SV40 whole genome. Shading shows regions of 100 bp centered around each experimentally detected nucleosome (vertical black lines). Our method seems to better localize nucleosome dyad axis. It behaves more efficiently at the end of the sequence in accordance with experimental results where the late SV40 region seems to have more strong nucleosome localization sites than the early one [7]. 21 nucleosomes are predicted (crosses); 13 of them (indentified with arrows in the figure) are located within 50bp from the nearest experimental position (shaded vertical bands) and considered correct predictions. The hits number is 27% higher than expectation if prediction was at random.

The physical properties of nucleosomes depend on environmental conditions such as ionic strength, divalent ion concentration, histon modification state: there are intrinsic experimental difficulties to determine nucleosomes. Not all nucleosomes experimentally identified are sequence dependent and our analysis may give an indication of the most probable sites of nucleosome center sequence driven positioning, moreover an already formed nucleosome may not be entirely fixed in position. Results seem to be encouraging but a direct validation would require more experimental identified nucleosomal DNA sequences. The algorithm herein presented is symmetric and enhances intrinsic rotational signals proper of a nucleosome.

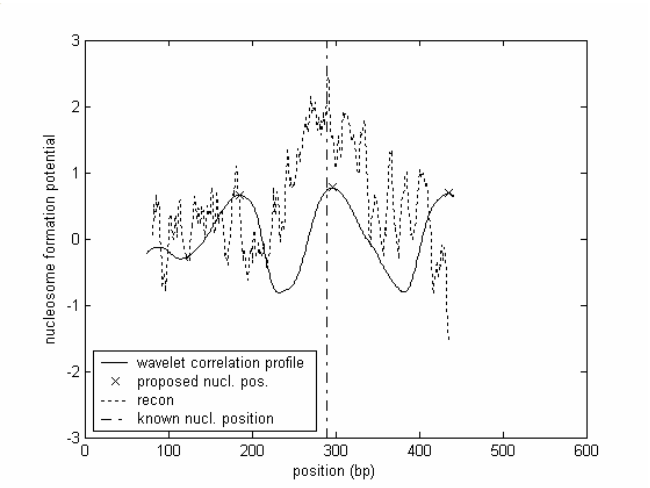

Fig. 1 recon profile prediction and wavelet profile prediction for Xenopus TFIIIA

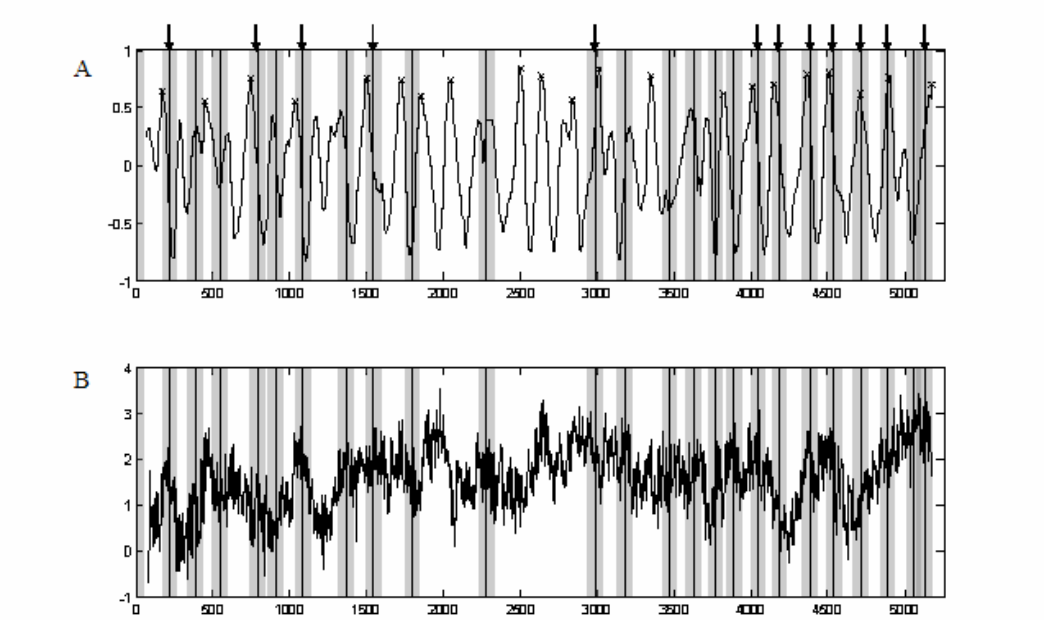

Fig. 2 Comparison between wavelet profile in A e recon profile in B for nucleosome prediction

References

[1] K. Luger, A.W. Mader, R.K. Richmond, D.F. Sargent, T.J. Richmond, Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, 389(6648):251-260, 1997

[2] http://www.rcsb.org/pdb/

[3] I. Ioshikhes, E.N. Trifonov, Nucleosomal DNA sequence database. *Nucleic Acids Res*., 21(21):4857-9, 1993

[4] I. Brukner, R. Sanchez, D. Suck, S. Pongor., Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides. *EMBO J*., 14(8):1812-8, 1995

[5] http://wwwmgs.bionet.nsc.ru/mgs/programs/recon/

[6] V.G. Levitsky, O.A. Podkolodnaya, N.A. Kolchanov, N.L. Podkolodny, Nucleosome formation potential of eukaryotic DNA: calculation and promoters analysis. *Bioinformatics*, 17(11):998-1010, 2001

[7] C. Ambrose, H. Lowman, A. Rajadhyaksha, V. Blasquez, M. Bina, Location of nucleosomes in simian virus 40 chromatin. *J Mol Biol.,* 214(4):875-84, 1990