ANALYSIS OF PERIODICITIES IN THE DINUCLEOTIDE CONTEXT OF NUCLEOSOMAL DNA USING THE METHOD PHASE

Levitsky V.G.*, Katokhin A.V., Furman D.P.

Institute of Cytology & Genetics SB RAS, Novosibirsk, Russia
* Corresponding author: e-mail: levitsky@bionet.nsc.ru

Keywords: nucleosome positioning, dinucleotide context periodicity, nucleosomal DNA

Summary

Motivation: Accumulation of the data on the role of chromatin nucleosomal organization in transcription, replication, and other basic biological processes makes ever more important the development of software tools for detecting the patterns of nucleosome location in the genomic DNA.

Results: A method PHASE for assessing the density of phased dinucleotides in arbitrary DNA sequences was designed. The tool PHASE was used to analyze periodic characteristics of dinucleotide context of nucleosome formation sites (NFS). It was demonstrated that the density of periodic AA and TT dinucleotides along NFS was maximal in the regions encompassing positions ±41 relative to the site center, reflecting the highest capability of DNA regular bending there.

Availability: The software package realizing the tool PHASE is available at http://wwwmgs.bionet.nsc.ru/mgs/programs/phase/.

Introduction

Numerous data of experiments and computer studies suggest that trans-interactions of protein factors with their cognate sites and cis-interactions of the latter with nucleosome positioning sites determine an ordered location of nucleosomes in genomic DNA. The totality of these interactions forms the so-called code of nucleosomal chromatin organization (Lower, Widom, 1997; Trifonov, 1997; Kiyama, Trifonov, 2002).

Periodic location of di- and trinucleotides is among the most important characteristics of this code (Ioshikhes et al., 1996; Stein, Bina, 1999). Earlier, we demonstrated that particular characteristics of dinucleotide composition of nucleosomal DNA allowing the latter to attain the conformation necessary for DNA–histone interactions and form nucleosomes (Levitsky et al., 1999; 2001). It is also known that a periodic arrangement of AT-containing dinucleotides, in particular, phased location of AA and TT dinucleotides, make a most weighty contribution to the contextual code of nucleosome positioning (Ioshikhes et al., 1996; Kiyama, Trifonov, 2002). Thus, in this work, we focused our attention on analyzing these dinucleotide types and quantifying the periodicity in question by the tool PHASE, designed for this particular purpose.

Methods and Algorithms

A sample of 142 nucleotide sequences with a length of 300 bp carrying nucleosome formation site (NFS) in the center was used in the work. These sequences were extracted from EMBL databank entries according to the codes and positions indicated by Ioshikhes and Trifonov (1993). The tool PHASE is based on estimation of the measure of similarity between the frequencies of phased dinucleotides in actual DNA sequences and the anticipated frequencies.

The dinucleotides located at a distance of one or two DNA helix turns (one turn on the average takes 10.5 bp) from each position considered were taken into account. Let us consider a nucleotide sequence $S$ and a set of $N = 10$ periods $\{D_j\} = \{±10, ±11, ±20, ±21, ±22 \text{ bp}\}$ (five periods in both directions from any position of the sequence that is located at least 22 bp from the sequence borders).
Let us calculate the complete set of dinucleotide frequencies in the sequence $S$ and generate a sample of randomized sequences $\{R\}$ with the same dinucleotide composition. Let the dinucleotide of $j$th type ($1 \leq j \leq 16$) be present at the $i$th position of a sequence from the sample $\{R\}$. Let the number of dinucleotides of the same type as in this position that are located at positions $\{i+D_n\}$ be $k$ ($0 \leq k \leq N$). Let us designate the probability of observing $k$ periodically located dinucleotides of $j$th type in all the sequences of sample $\{R\}$ as $p_j(k)$. Then the frequency of dinucleotides of $j$th type, $f_j$, in the sample $\{R\}$ is determined as the following sum:

$$f_j = \sum_{k=0}^{N} p_j(k).$$

(1)

Note that the probability $p_j(0)$ means the absence of periodic dinucleotides or the presence of exclusively “nonperiodic” dinucleotides of $j$th type, whereas the probability $p_j(N)$ corresponds to the highest periodicity of dinucleotides permitted by the set $\{D_n\}$.

Let us consider again the sequence $S$ and assume that a dinucleotide of $j$th type is present at its $i$th position. Let $k$ dinucleotides of $j$th type ($0 \leq n \leq 10$) be observed at positions $\{I+D_n\}$. Then, we determine the next function, $w_j(k)$, as a logarithm of the relative probability of occurrence of the $j$th type periodic dinucleotides:

$$w_j(k) = -\log \frac{p_j(k)}{f_j}.$$  

(2)

The anticipated mean value of the function $w_j$ is equal to

$$w_j = -\frac{1}{N} \sum_{k=0}^{N} \left( \log \frac{p_j(k)}{f_j} \right).$$

(3)

The maximal value of the function $w_j$, as follows from definition (2), amounts to $w_j(N)$. Basing on the mean and maximal values of the function $w_j(k)$ (2), let us determine the function $PDD_j(k)$ of the periodic dinucleotide density (PDD) as follows:

$$PDD_j(k) = \frac{w_j(k) - w_j}{w_j(N) - w_j}.$$  

(4)

Let us define the integral function $PDD$ of the periodic densities for an arbitrary set of dinucleotides $\{J\}$ as follows:

$$PDD = \sum_{j \in \{J\}} PDD_j.$$  

(5)

As it follows from definition (4), the maximal value of functions (4) and (5) over the sequence positions equals unity, whereas the zero value corresponds to the mean calculated for randomized sequences, i.e., in the absence of phased dinucleotides.

**Results and Discussion**

The software package PHASE was used to analyze periodic contextual characteristics of sequences containing NFS. Fig. shows the profile of PDD function for AA and TT dinucleotides constructed for the sample of sequences containing NFS using an averaging window of 10 bp. The region $[-73; +73]$ corresponds to NFS; regions $[-110; -74]$ and $[+74; +110]$, to linker regions. Analysis of the PDD
profile in the region [–73; +73] suggested the following inferences: (i) the maximal PDD values for AA and TT dinucleotides are observed in the regions encompassing positions ±41 and (ii) the minimal PDD values for the same dinucleotides are found in a wider regions surrounding positions ±89 and ±1.

The linker regions also display low PDD values. These results comply well with earlier published data. For example, multiple alignment of nucleosome DNA sequences demonstrated that pronounced periodic characteristics of the context were absent in the central region of NFS (Ioshikhes et al., 1996). The result we obtained agrees as well with the X-ray structure analysis data showing that the terminal 10-bp sequences of nucleosomal DNA, adjacent to the linker regions, are unbend (Luger et al., 1997) and are bound least stably to the core histones (Polach, Widom, 1995).

As the high values of PDD profile obtained by PHASE indicate the presence of a phased dinucleotide at a distance of one or two DNA helix turns, while the presence of phased dinucleotide itself determines an increase in the ability of DNA to bend regularly, we may infer that nucleosome DNA contains two more pronouncedly bent regions with a length of 40–50 bp at positions [–73; –25] and [+25; +73] relative to NFS center. These two regions are separated with a less bent region of 30–50 bp located at [–25; +25]. Calculation of characteristics of nucleosomal DNA 3D trajectory (Fitzgerald et al., 1994) lead the authors to similar conclusions, as it was demonstrated that bending pattern consisted of repeating units of two 50–60 bp bending elements separated by a 20–30 bp region of a low curvature.

**Acknowledgements**

The work was supported in part by the Russian Foundation for Basic Research (grants Nos. 03-04-48555-a and 03-07-96833); Ministry of Industry, Science, and Technologies of the Russian Federation (grant No. 43.073.1.1.1501); Russian Federal Research Development Program «Research and Development in Priority Directions of Science and Technology» (contract No. 38/2004); and Program for Basis Research of the Presidium of the Russian Academy of Sciences (contract No. 10002-251/TI-25/155-270/200404-082); Project No. 10.4 of the RAS Presidium Program «Molecular and Cellular Biology».
References


