DNA Physical Parameters Modulate Nucleosome Positioning in the *Saccharomyces cerevisiae* Genome

Wei Chen[*,1], Hao Lin[*,2] and Pengmian Feng[3]

[1]*Department of Physics, School of Sciences, and Center for Genomics and Computational Biology, Hebei United University, Tangshan 063000, China*

[2]*Key Laboratory for NeuroInformation of Ministry of Education, Center of Bioinformatics, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China*

[3]*School of Public Health, Hebei United University, Tangshan 063000, China*

Abstract: Nucleosome positioning plays essential roles in various cellular processes. Although many efforts have been made in this area, the rules defining nucleosome positioning is still elusive. In the present study, DNA physical parameters derived from atomistic molecular dynamic simulations were introduced to analyze nucleosomal and linker DNA sequences. The distinct structural patterns between nucleosomal and linker sequences indicate that DNA physical parameters are suitable to describe nucleosomal DNA sequences and to reveal physical mechanisms of nucleosome positioning. Further analysis of DNA flexibility around regulatory regions indicates that nucleosome positioning is closely correlated with sequence flexibility. These results demonstrate that DNA physical parameters are useful for the *in silico* nucleosome positioning prediction.

Keywords: DNA flexibility, DNA physical parameters, nucleosome positioning, replication origin, transcription start site, transcription termination site.

1. INTRODUCTION

Nucleosome is the elementary structural unit of chromatin in eukaryotes, which consists of a ~147 bp DNA sequence tightly wrapped around the histone-octamer core (composed of pairs of the four core histones H2A, H2B, H3 and H4) [1]. The packaging of DNA into nucleosomes affects the accessibility of genomic regions to regulatory proteins. It has been suggested that there are close relationships between nucleosome positioning and various cellular processes, such as mRNA splicing, DNA replication and DNA repair [2-5]. Hence, a complete understanding of gene expression in eukaryotes requires revealing the mechanism involved in nucleosome positioning.

With the high-throughput techniques chromatin immunoprecipitation (CHIP) coupled with microarrays (CHIP-chip) and CHIP coupled with sequencing techniques (CHIP-Seq), nucleosomal data are now available for the genome of yeast, worms, flies and humans [6-9]. However, the genome-wide experimental approaches are cost ineffective. Conversely, computational methods to predict nucleosome positioning sequences can be applied to genome-wide analysis without these disadvantages. In the past ten years, computational models of sequence-based prediction of nucleosome positioning have been proposed [10-13]. However, the existing sequence-based methods are limited due to either accuracy or resolution, and to which extent nucleosome positioning is determined by genomic sequence is still debated [14-18].

DNA physical parameters, including three local angular parameters (twist, tilt and roll) and three translational parameters (shift, slide and rise), have essential roles in protein-DNA interactions, formation of chromosomes and higher-order organization of the genetic material in a cell nucleus [6, 19, 20]. This suite of parameters can also modulate the accessibility of DNA to regulatory proteins [21].

In this study, the relationship between DNA physical parameters and nucleosomal DNA sequences were analyzed in the *Saccharomyces cerevisiae* genome. By defining flexibility of DNA, we investigated the distribution pattern of nucleosomes around transcription start site (TSS), transcription termination site (TTS) and origin of replication (ORI) and found that nucleosome positioning around TSS, TTS and ORI is strongly dependent on DNA physical properties.

2. MATERIALS AND METHODS

2.1. Dataset

The genome sequence of *Saccharomyces cerevisiae* was extracted from the *Saccharomyces* Genome Database (http: //www.yeastgenome.org/, downloaded April 2008). The experimentally confirmed nucleosomal sequences of

*Address correspondence to these authors at the (W. Chen) Department of Physics, School of Sciences, Hebei United University, 57 South Jianshe Road, Tangshan 063000, China; Tel: +86 315 3725715;
Fax: +86 315 3725715; E-mail: chenwei_imu@yahoo.com.cn and
(H. Lin) School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054 China;
Tel: +86-28-8320-2351; Fax: +86-28-8320-8238;
E-mail: hlin@uestc.edu.cn

2212-392X/14 $58.00+.00

Saccharomyces cerevisiae were obtained from Lee *et al.* [22]. Each of the 1,206,683 DNA fragments in the dataset was assigned a score by lasso model, where high and low score indicate that a sequence is nucleosome forming or inhibiting, respectively [22]. According to our recent work [23], the 5000 fragments of 150-bp with the highest scores and 5,000 fragments of 150-bp with the lowest scores were defined respectively as nucleosomal sequences and linker sequences.

2.2. Genomic Sequence Around Functional Sites

The 5015 well-defined transcripts of the *Saccharomyces cerevisiae* genome were taken from Lee and his colleague's work [22]. The 1000-bp long region from -500 bp to +500 bp flanking TSS and TTS were obtained, respectively.

The 322 experimentally confirmed ORIs were extracted from the OriDB database (downloaded September 2009) [24]. The 1000-bp long region from -500 bp to +500 bp flanking each ORI were extracted.

2.3. DNA Physical Parameters

The six DNA physical parameters (shift, slide, rise, tilt, roll and twist), which describe the deformability of naked DNA, were derived from long atomistic molecular dynamics (MD) simulations by Goñi JR *et al.* [21] and have been used for promoter prediction [21, 25]. In order to have enough equilibrium samples for all the ten unique steps of DNA, Goñi JR *et al.* performed MD simulations of four duplexes containing several replicas of every type of base step dimer (d(GG), d(GA), d(GC), d(GT), d(AA), d(AG), d(AT), d(TA), d(TG) and d(CG)): d(GCCTATAAACGCCTATAA), d(CTAGGTGGATGACTCATT), d(CACGGAACCGGTTC CGTG) and d(GGCGCGCACCACGCGCGG) [21]. All duplexes were created in the standard B-type conformation, hydrated with around 10,600 water molecules, and neutralized by adding a suitable number of Na^+ ions [21]. The ten different nearest neighbor interaction values of the six parameters in any Watson-Crick DNA duplex structures were shown in Table **1**.

2.4. DNA Flexibility

To obtain a complete picture of the base-pair step deformability, conformational volume (*Vol*) was introduced by Olson *et al.* [26]. The *Vol* describes the flexibility of DNA dimers, and detailed descriptions about *Vol* can be referred to the previous literature [26]. By using the newly MD data [21], we calculated the *Vol* for each of the ten unique dinucleotide steps and found that the *Vol* value of CG is the largest and AT is the smallest, which is consistent with previous study [26].

On the basis of *Vol*, the flexibility of DNA (*FD*) for an *l*-bp long DNA sequence was defined by using the following formula,

$$FD = \frac{\sum_{i=1}^{10} vol_i N_i}{l-1} \qquad \sum_{i=1}^{10} N_i = l-1 \tag{1}$$

$$vol_i = \frac{Vol_i}{Vol_{CG}} \tag{2}$$

where Vol_i is the conformation volume of the *i-th* (i=1, 2,…, 10) dinucleotide, N_i is the total number of the *i-th* dinucleotide in the *l*-bp long sequence. The weighting vol_i is the normalized conformation volume relative to that of CG dinucleotide. *FD* is an index of DNA flexibility. The higher the *FD* value is, the more the flexible the sequence is.

3. RESULTS AND DISCUSSION

3.1. Structure Profiles of Nucleosomal and Linker Sequences

To investigate the structural properties of nucleosomal DNA sequences, we analyzed the six physical structure parameters (twist, tilt, roll, shift, slide, and rise) for both nucleosomal and linker sequences in the *Saccharomyces cerevisiae* genome. According to the parameters reported in Table 1, we can calculate the structure profile of any DNA double helix from its primary sequence. For a given

Table 1. DNA Physical Parameters for the Ten Unique Dinucleotide Steps

Stiffness Constants Associated to Helical Deformations						
Step	**Twist**	**Tilt**	**Roll**	**Shift**	**Slide**	**Rise**
AA/TT	0.026	0.038	0.020	1.69	2.26	7.65
AC/GT	0.036	0.038	0.023	1.32	3.03	8.93
AG/CT	0.031	0.037	0.019	1.46	2.03	7.08
AT	0.033	0.036	0.022	1.03	3.83	9.07
CA/TG	0.016	0.025	0.017	1.07	1.78	6.38
CC/GG	0.026	0.042	0.019	1.43	1.65	8.04
CG	0.014	0.026	0.016	1.08	2.00	6.23
GA/TC	0.025	0.038	0.020	1.32	1.93	8.56
GC	0.025	0.036	0.026	1.20	2.61	9.53
TA	0.017	0.018	0.016	0.72	1.20	6.23

Constants related to rotational parameters are in kcal/mol degree^2, while those related to translations are in kcal/mol Å^2 [21].

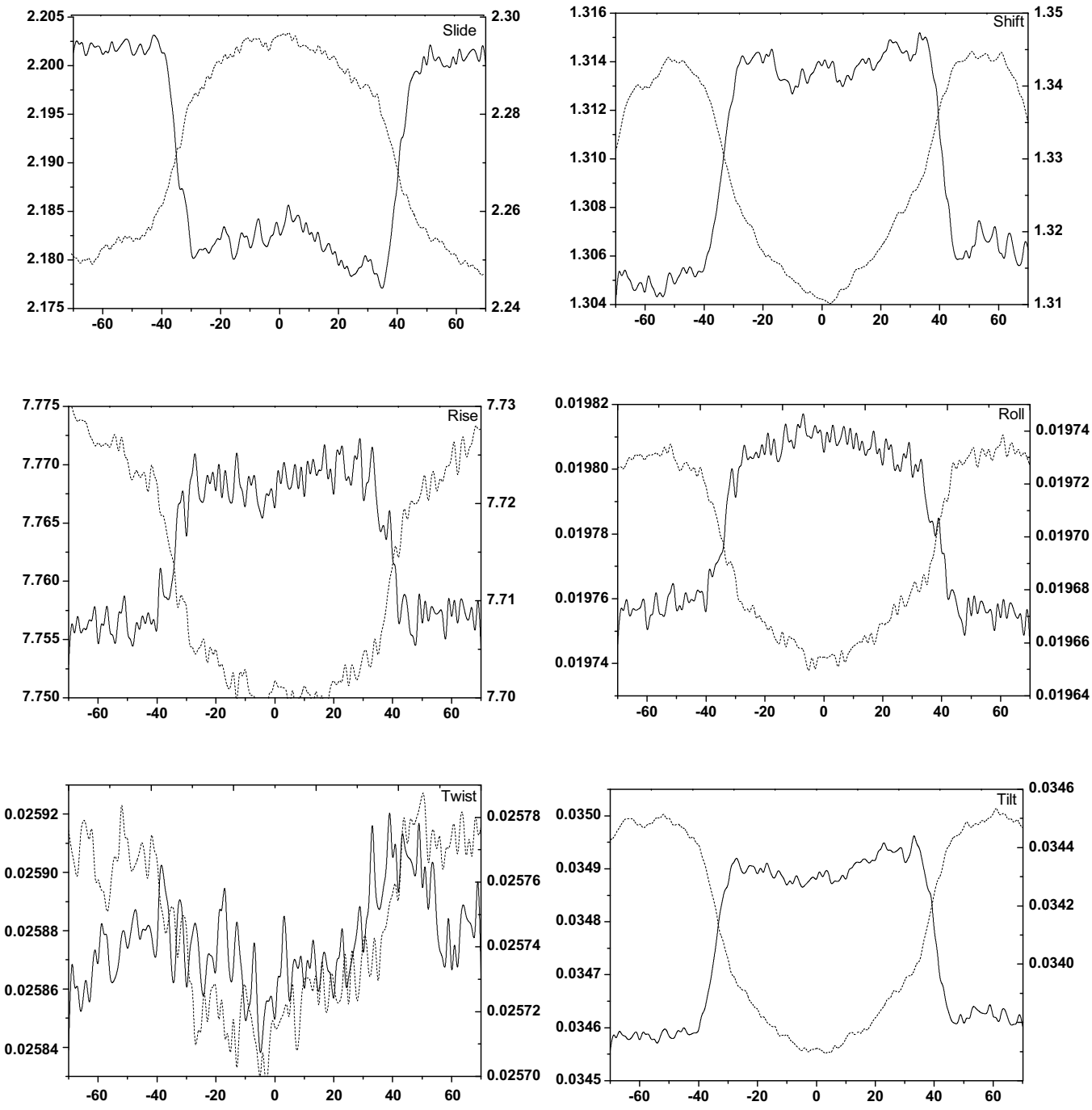

Fig. (1). DNA structural (twist, tilt, roll, shift, slide, and rise) profiles for nucleosomal (solid line) and linker (dash line) sequences. For a given deformation, the structural profile was smoothed with a 10-bp window in 1-bp increment. The horizontal axis gives the position of the sliding window labeled by its center relative to the center of nucleosomal (or linker) sequences and the vertical axis represents the profile score of each deformation.

deformation, we sum the values associated with every dinuecleotide step in the *l*-bp long sequence and divide the total by *l*-1. By using a 10-bp window with 1-bp displacement, the structural profiles for nucleosomal and linker sequences were plotted in Fig. (**1**).

Dramatic differences were found between the structural property of nucleosomal sequences and that of linker sequences (Fig. **1**). For shift, rise, tilt and roll, the scores in the central regions (-40 to +40 relative to the dyad) of nucleosomal sequences were significantly higher than those of linker sequences ($p<10^{-15}$; *t*-test). Moreover, their values in nucleosomal sequences gradually increased toward the central position of the nucleosome, which is consistent with the symmetric structure of the histone protein [1]. But the slide scores in the central regions of nucleosomal sequences

were significantly lower than that in linker sequences ($p<10^{-11}$; *t*-test). These results suggest that the unusual physical properties might control nucleosome positioning, which in turn would affect the DNA accessibility to regulatory proteins and ultimately impact gene regulation.

3.2. Nucleosomal Sequences with Higher Flexibility

To investigate the flexibility of nucleosomal sequences and validate if the flexibility observed in nucleosomal DNA sequences are unique, we compared the flexibility between the 5,000 nucleosomal and 5,000 linker sequences by calculating *FD* value. We found that the *FD* value of nucleosomal sequence was significantly higher than that of linker sequence ($p<10^{-10}$; *t*-test), indicating that we can predict nucleosome positions according to *FD*. As no training is performed, *FD* will be fully an *ab initio* descriptor for nucleosome positioning prediction.

3.3. Nucleosome Positioning Near Regulatory Sites

Nucleosomes are known to play key roles in biological processes by controlling the accessibility of regulatory protein to DNA and modulating the high-order structure of chromatin. Recently, several genome-scale experimental maps and computational works have demonstrated that TSS and TTS regions are strongly nucleosome depleted [22, 27]. To demonstrate that *FD* can accurately predict nucleosome positioning, we analyzed nucleosome positing around several functional sites by calculating *FD* scores.

In a 50-bp sliding window with a step size of 10-bp, we calculated the *FD* score for genomic regions -500 bp to +500 bp relative to TSS. The average *FD* profile around TSS was plotted in Fig. (**2**). An *FD* score trough locating at approximately 50 bp upstream of the TSS (Fig. **2**), indicating that this region is too rigid to wrap around the histone core. While the *FD* scores in flanking regions are higher, suggesting that these regions are flexible and biased to form nucleosomes. These results are in accordance with previous work that a pronounced nucleosome depleted region was found upstream of the TSS [22]. The nucleosome depleted region upstream of the TSS may allow the binding of the pre-initiation complex to this region.

The average *FD* profiles around TTS regions were also examined in the same way and distinctive flexibility profile was observed as shown in Fig. (**3**). A deep *FD* valley was found in the intergenic border downstream of TTS, indicating strong nucleosome depletion in this region. The nucleosome depleted region near TTS may contribute to the assembly of anti-sense pre-initiation complexes, disassembly of polymerase machinery and recycling of RNA polymerase to the promoter by DNA looping [27].

DNA replication is thought to be one of the most highly regulated processes referring to interactions between regulatory proteins and DNA sequences. The initiation of DNA replication is also regulated by chromatin structure. In a 50-bp sliding window with a step size of 10-bp, we calculated the *FD* score of the 322 experimentally confirmed ORIs. Fig. (**4**) shows that the average *FD* value in the core replication region (0~+250 bp) was lower than those within

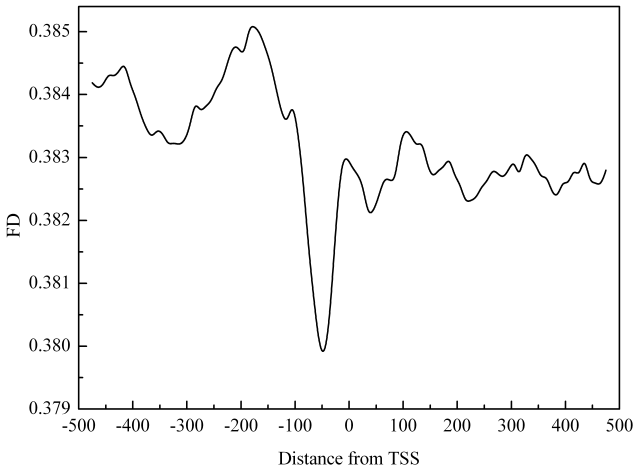

Fig. (2). DNA flexibility around transcription start site (TSS) of the *Saccharomyces cerevisiae* genome. The data were smoothed with a 50-bp sliding window in 10-bp increments from -500 bp to 500 bp relative to TSS. The horizontal axis gives the position of the sliding window labeled by its center relative to TSS and the vertical axis represents DNA flexibility (*FD*).

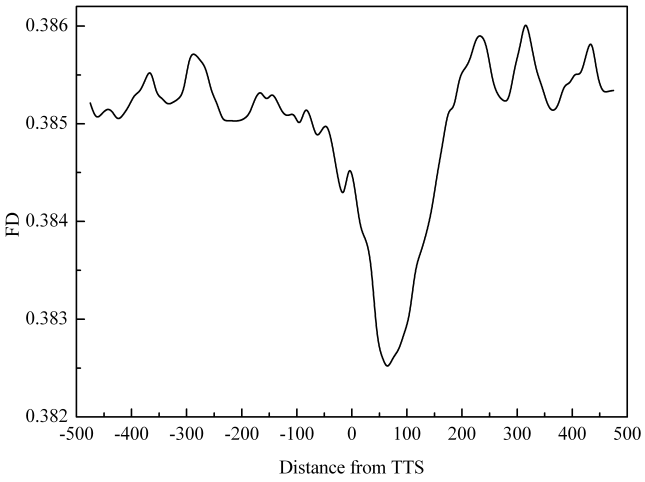

Fig. (3). DNA flexibility around transcription termination site (TTS) of the *Saccharomyces cerevisiae* genome. The data were smoothed with a 50-bp sliding window in 10-bp increments from -500bp to 500bp relative to TTS. The horizontal axis gives the position of the sliding window labeled by its center relative to TTS and the vertical axis represents DNA flexibility (*FD*).

surrounding regions. And a *FD* valley was found at the core replication region (Fig. **4**). These results demonstrate that the core replication region shows a significantly lower flexibility than flanking regions. It has been reported that nucleosomes are depleted in core replication region but are well positioned in the flanking regions of ORI [5, 28]. Thus, the different flexibility distribution around ORI may be the signal of nucleosome positioning and facilitate DNA unwinding, protein binding and then replication fork progression during the process of genome replication.

4. CONCLUSIONS

In this study, we investigated six DNA physical structure parameters in nucleosomal and linker sequences. We found

that the nucleosomal sequences had higher shift, rise, tilt and roll scores than linker sequences, while the slide score was lower in nucleosomal sequences than that in linker sequences. Since the structural properties of nucleosomal DNA sequences show significant difference from that of linker sequences, we defined the flexibility of DNA (*FD*) to predict nucleosome positions. By calculating *FD*, nucleosome positions around TSS, TTS and ORI were systematically analyzed and nucleosome depleted regions around these functional sites were observed (Figs. **2-4**). These results are in accordance with previous studies [22, 26, 27] and suggest the importance of DNA physical parameters (twist, tilt, roll, shift, slide and rise) in regulating nucleosome positioning.

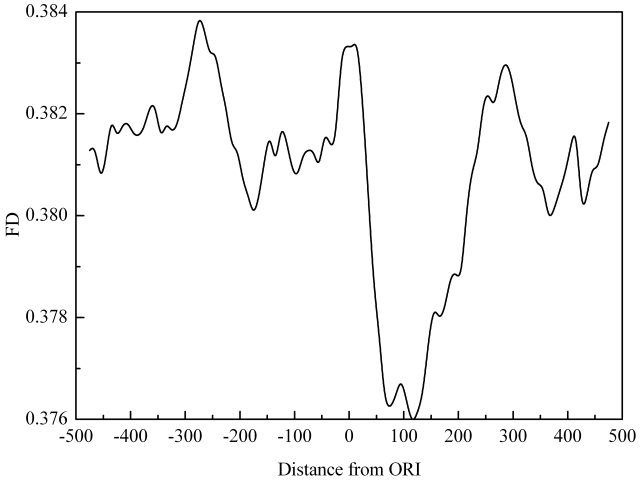

Fig. (4). DNA flexibility around replication origin (ORI) of the *Saccharomyces cerevisiae* genome. The data were smoothed with a 50-bp sliding window in 10-bp increments from -500bp to 500bp relative to ORI. The horizontal axis gives the position of the sliding window labeled by its center relative to ORI and the vertical axis represents DNA flexibility (*FD*).

Due to the training data may not be representative of direct histone-DNA binding, the accuracy and resolution of previous sequence-based predictions of nucleosome positioning were still far from satisfactory. Unlike the previous computational approaches, our present model is not trained on genomic data and only depended on the DNA physical parameters of base-pair steps. The above findings have demonstrated that nucleosomal and linker sequences are signaled by unusual physical properties. Therefore, we expect that this suite of parameters will be useful for further elucidating nucleosome positioning mechanisms in eukaryotic genomes.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

We wish to express our gratitude to the editor and three anonymous reviewers whose constructive comments were very helpful in strengthening the presentation of this article. This work was supported by Grants from The National Nature Scientific Foundation of China (No. 61100092 and 61202256), Nature Scientific Foundation of Hebei Province (No. C2013209105).

REFERENCES

[1] Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 1997; 389: 251-60.

[2] Yasuda T, Sugasawa K, Shimizu Y, *et al*. Nucleosomal structure of undamaged DNA regions suppresses the non-specific DNA binding of the XPC complex. DNA Repair (Amst) 2005; 4: 389-95.

[3] Schwartz S, Meshorer E, Ast G. Chromatin organization marks exon-intron structure. Nat Struct Mol Biol 2009; 16: 990-95.

[4] Chen W, Luo LF, Zhang LR. The organization of nucleosomes around splice sites. Nucleic Acids Res 2010; 38: 2788-98.

[5] Berbenetz NM, Nislow C, Brown GW. Diversity of eukaryotic DNA replication origins revealed by genome-wide analysis of chromatin structure. PLoS One 2010; 6: e1001092.

[6] Miele V, Vaillant C, d'Aubenton-Carafa Y, Thermes C, Grange T. DNA physical properties determine nucleosome occupancy from yeast to fly. Nucleic Acids Res 2008; 36: 3746-56.

[7] Schones DE, Cui K, Cuddapah S, *et al*. Dynamic regulation of nucleosome positioning in the human genome. Cell 2008; 132: 887-98.

[8] Yuan GC, Liu JS. Genomic sequence is highly predictive of local nucleosome depletion. PLoS Comput Biol 2008; 4: e13.

[9] Valouev A, Ichikawa J, Tonthat T, *et al*. A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res 2008; 18: 1051-63.

[10] Segal E, Mittendorf YF, Chen L, *et al*. A genomic code for nucleosome positioning. Nature 2006; 442: 772-78.

[11] Ioshikhes IP, Albert I, Zanton SJ, *et al*. Nucleosome positions predicted through comparative genomics. Nat Genet 2006; 38: 1210-15.

[12] Gupta S, Dennis J, Thurman RE, *et al.* Predicting Human Nucleosome Occupancy from Primary Sequence. PLoS Comput Biol 2008; 4: e1000134.

[13] Peckham HE, Thurman RE, Fu Y, *et al*. A genomic code for nucleosome positioning. Genome Res 2007; 17: 1170-77.

[14] Trifonov EN. Nucleosome Positioning by Sequence, State of the Art and Apparent Finale. J Biomol Struct Dyn 2010; 27: 741-46.

[15] Zhang Y, Moqtaderi Z, Rattner PB, *et al*. Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions *in vivo*. Nat Struct Mol Biol 2009; 16: 847-53.

[16] Lantermann AB, Straub T, Strålfors A, Yuan GC, Ekwall K, Korber P. *Schizosaccharomyces pombe* genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. Nat Struct Mol Biol 2010; 17: 251-7.

[17] Travers A, Hiriart E, Churcher M, Caserta M, Di Mauro E. The DNA sequence dependence of nucleosome positioning *in vivo* and *in vitro*. J Biomol Struct Dyn 2010; 27: 713-24.

[18] Arya G, Maitra A, Grigoryev SA. A structural perspective on the where, how, why, and what of nucleosome positioning. J Biomol Struct Dyn 2010; 27: 803-20.

[19] Abeel T, Saeys Y, Bonnet E, *et al*. Generic eukaryotic core promoter prediction using structural features of DNA. Genome Res 2008; 18: 310-23.

[20] Goñi JR, Fenollosa C, Pérez A, *et al*. DNAlive: a tool for the physical analysis of DNA at the genomic scale. Bioinformatics 2008; 24: 1731-32.

[21] Goñi JR, Pérez A, Torrents D, *et al*. Determining promoter location based on DNA structure first-principles calculations. Genome Biol 2007; 8: R263.

[22] Lee W, Tillo D, Bray N, *et al*. A high resolution atlas of nucleosome occupancy in yeast. Nat Genet 2007; 39: 1235-44.

[23] Chen W, Lin H, Feng PM, *et al*. iNuc-PhysChem: A sequence-based predictor for identifying nucleosomes *via* physicochemical properties. Plos One 2012; 7: e47843.

[24] Nieduszynski CA, Hiraga S, Ak P, *et al*. OriDB: a DNA replication origin database. Nucleic Acids Res 2007; 35: 40-6.

[25] Zuo YC, Li QZ. The hidden physical codes for modulating the prokaryotic transcription initiation. Physica A 2010; 389: 4217-23.

[26] Olson WK, Gorin AA, Lu XJ, *et al*. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proc Natl Acad Sci USA 1998; 95: 11163-68.

[27] Mavrich TN, Ioshikhes IP, Venters BJ, *et al*. A barrier nucleosome model for statistical positioning of nucleosome throughout the yeast genome. Genome Res 2008; 18: 1073-83.

[28] Yin S, Deng W, Hu L, *et al.* The impact of nucleosome positioning on the organization of replication origins in eukaryotes. Biochem Biophys Res Commu 2009; 385: 363-8.

Received: August 29, 2012 Revised: September 30, 2012 Accepted: October 30, 2012