Computational analysis suggests a highly bendable, fragile structure for nucleosomal DNA

Tadasu Nozaki a,b, Nozomu Yachie c, Ryu Ogawa a,d, Rintaro Saito a,b,k, Masaru Tomita a,b,d

a Institute for Advanced Biosciences, Keio University, Tsuruoka, 997–0017, Japan
b Systems Biology Program, Department of Environment and Information Studies, Keio University, Fujisawa, 252–8520, Japan
c Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02445, USA
d Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa, 252–8520, Japan

Article history:
Accepted 14 February 2011
Available online 19 February 2011

Keywords:
Bendability
Cleavage intensity
Nucleosome
DNA structure
Histone
Epigenetics

1. Introduction

Eukaryotic chromosomal DNA coils around histones to form nucleosomes. Although histone affinity for DNA depends on DNA sequence patterns, how nucleosome positioning is determined by them remains unknown. Here, we show relationships between nucleosome positioning and two structural characteristics of DNA conferred by DNA sequence. Analysis of bendability and hydroxyl radical cleavage intensity of nucleosomal DNA sequences indicated that nucleosomal DNA is bendable and fragile and that nucleosome positional stability was correlated with characteristics of DNA. This result explains how histone positioning is partially determined by nucleosomal DNA structure, illuminating the optimization of chromosomal DNA packaging that controls cellular dynamics.

The efficiency of the coiling of DNA around histones to form nucleosomes depends strongly on DNA sequence (Yuan et al., 2005; Ogawa et al., 2010). In vitro studies have shown that the range of affinities between histones and DNA is at least 1000-fold and that DNA sequence pattern is an important factor for accurate nucleosome positioning (Thastrom et al., 1999). Furthermore, the results of several other studies suggest that the DNA coiled around the histones is bendable, and these studies have shed light on the importance of DNA bendability and the optimization of chromosomal DNA packaging that controls cellular dynamics.

Along with the progress in computational methods for analyzing nucleosome data (Zhang et al., 2008; Tolstorukov et al., 2009), recent remarkable progress in chromatin immunoprecipitation (ChIP) techniques, including ChIP–chip (ChIP combined with microarray technology) and ChIP–Seq (ChIP combined with deep sequencing), has permitted detection and prediction of nucleosome positioning at the whole-genome level (Yuan et al., 2005; Barski et al., 2007; Park, 2009). The increase in the number of large-scale nucleosome data sets has permitted systematic analysis of nucleosome data in conjunction with other “omic” data (e.g., genome-wide relationships between nucleosome positioning and transcription regulation (Schones et al., 2008; Lister et al., 2009)).
In this study, we used a computational analysis of a large-scale nucleosomal data set to evaluate two overall characteristics of chromosomal DNA structure: DNA bendability and DNA cleavage intensity. DNA bendability is calculated from parameters for a set of trinucleotide patterns (Brukner et al., 1995) or tetranucleotide patterns (Packer et al., 2000) in a given sequence. The parameters for each trinucleotide pattern were obtained by Brukner et al. (1995) using DNA binding proteins which generally bind to DNA that bends toward the major groove. The parameters were derived from 709 DNA binding regions. Bendability is frequently used to analyze DNA structure. For example, the location of transcription initiation promoters is associated with areas of bendable DNA (Abeel et al., 2008; Akan and Deloukas, 2008; Choi and Kim, 2009).

Cleavage intensity indicates the likelihood of DNA cleavage by hydroxyl radicals and provides a map of local variation in the shape of the DNA backbone. Cleavage intensity is calculated from parameters for a set of tetranucleotide patterns in a given sequence. The parameters were derived by Greenbaum et al. (2007) from experiments in which DNA sequences were exposed to hydroxyl radicals; they analyzed how cleavage of DNA sequences depends on their sequence patterns. Cleavage intensity also reflects the solvent-accessible surface area of DNA. This index of DNA fragility has also been used to characterize structural features of DNA (Tullius and Greenbaum, 2005; Greenbaum et al., 2007).

These two structural parameters cover all combinations of trinucleotide or tetranucleotide patterns and are sufficiently reliable to analyze DNA structure (Brukner et al., 1995; Greenbaum et al., 2007). We conducted genome-wide computational analysis of the bendability and cleavage intensity of nucleosome and linker DNA regions and found that nucleosomal DNA was overall highly bendable and fragile. We also determined whether the positional stability of nucleosomes under various cellular conditions depended on these two structural properties of DNA, and we found that stable nucleosomal DNA was more bendable and cleavable than unstable nucleosomal DNA. We discuss which types of DNA structure confer the optimal DNA packaging in chromosomes and which types are optimal for dynamic cellular processes.

2. Materials and methods

2.1. Genomic DNA and nucleosome positioning data

The reference genome sequence of Homo sapiens (hg18) was obtained from the University of California at Santa Cruz genome database (http://genome.ucsc.edu/). To determine genomic histone positions, we used the previously reported ChIP–Seq data set for histone H3 of human CD4+ T-cells obtained under two cellular conditions, with and without the T-cell receptor signal, which are designated as the activated and resting conditions, respectively (Schones et al., 2008). We modified a previously developed method (Tolstorukov et al., 2009) to define nucleosomal positions from the ChIP–Seq of both resting and activated cells (for details of the method, see Supplementary methods). We assessed the positional stability of nucleosomes under the two cellular conditions by computing the distance of the genomic position of a given nucleosome in the resting cell from its nearest nucleosomal position in the activated cell. For all the nucleosomes identified under the resting conditions, we defined stable nucleosomes as those for which the distance between the nucleosome under resting conditions and the nearest nucleosome under activated conditions was within 12 bps, and unstable nucleosomes as those for which that distance ranged from 13 to 73 bps (half the size of one nucleosomal DNA unit). Nucleosomes that did not fit either definition were not used for the positional stability analysis. The boundary between the two types of nucleosomes (i.e., the 12–13-bp boundary) was defined from the inflection point in the distance distribution (Supplementary methods).

2.2. DNA bendability calculation

To assess nucleosomal DNA bendability, we used the trinucleotide bendability parameter profile obtained by Brukner et al. (1995). For each of the nucleosomes identified under the resting conditions (including both stable and unstable nucleosomes), DNA bendability was calculated for every trinucleotide in the region from −400 to +400 bps from the center position of the nucleosomal DNA region. For each position relative to the central position of the nucleosome, the average DNA bendability was calculated.

2.3. DNA cleavage intensity calculation

We used the OH radical cleavage intensity database (ORChID; http://dna.bu.edu/orchid/) program, which is used to estimate the hydroxyl radical cleavage strength and solvent-accessible surface area using a tetranucleotide DNA cleavage intensity profile (Greenbaum et al., 2007), to calculate the DNA cleavage intensity. (The source code of ORChID was kindly provided by its authors; ORChID is a web-based application.) For each of all the nucleosomes identified under the resting conditions, we computed tetranucleotide DNA cleavage intensities for all the DNA positions from −400 to +400 bps from the center position by querying to ORChID. For each position relative to the central position of the nucleosome, the average DNA cleavage intensity was calculated.

2.4. Guanine–cytosine content and statistical analysis

For a control analysis with respect to guanine–cytosine (GC) content, we randomly sampled 10 000 out of the ~5 000 000 nucleosomes identified under the resting conditions as a positive reference set (PRS) to reduce the calculation time. For each nucleosome in the PRS, sliding window averages of DNA bendability, cleavage intensity, and GC content for the region surrounding the nucleosome (from −400 to +400 bps) were calculated using a 30-bp window with 1-bp displacement. To prepare a random reference set (RRS), we randomly selected 1 000 000 genomic positions and used their surrounding sequences (from −400 to +400 bps) to calculate DNA bendability, cleavage intensity, and GC content using a 30-bp sliding window average. Additionally, the DNA bendability and cleavage intensity expected from each GC content value were calculated from the above data, and we set them as RRS with GC content. Then, we randomly correlated the sliding window for each of the 10 000 nucleosomes in the PRS with the sliding windows of the RRS having the same GC content, so that the DNA bendabilities and cleavage intensities of the same relative sliding window positions could be compared with the DNA bendabilities and cleavage intensities expected from the GC content and the genomic context. For each sliding window position relative to the center of the nucleosomal region, differences in DNA bendability and cleavage intensity between the PRS and the RRS were evaluated by t-test.

3. Results

3.1. Bendability of nucleosomal DNA

To investigate the structural properties of nucleosomal DNA, we first analyzed the bendability of the DNA sequences surrounding the nucleosomal region. Using all the nucleosomal positions of human CD4+ T-cells identified under the resting conditions, we calculated average DNA bendabilities of the genomic positions relative to the central position of the nucleosome (from −400 to +400 bps) (Fig. 1A). Bendability scores within the nucleosomal region were significantly higher than those within the linker regions (P < 10^-15; t-test). Moreover, the bendabilities in the nucleosomal region gradually increased toward the central position of the nucleosome,
3.2. Fragility of nucleosomal DNA

We next computed DNA cleavage intensity, which is an estimate of hydroxyl radical cleavage strength and the solvent-accessible area, for the DNA sequence surrounding the nucleosome using ORChID. Like the DNA bendability pattern, the DNA cleavage intensity pattern for the nucleosomal DNA region was symmetrical. The cleavage intensity in the nucleosomal region was higher than that of the linker regions on both sides ($P<10^{-15}$, t-test), and a $-10$-bp periodic fluctuation of DNA cleavage intensity was also observed, especially in the nucleosomal DNA region (Fig. 2A). The cleavage intensity of the nucleosomal region was statistically higher than the intensities of the RRS expected from the GC content, especially at the central position of the nucleosome (Fig. 2B). Moreover, interestingly, the cleavage intensities in the regions immediately inside the nucleosome–linker junctions were markedly higher than average, whereas the intensities in the regions immediately outside the junctions were markedly lower than average (that is, the distribution dropped sharply from the inside to the outside of the nucleosome region). However, the difference from the RRS was statistically significant only for the downstream junction ($P<10^{-5}$).

Fig. 2. Cleavage intensity of nucleosomal DNA and the surrounding regions. (A) DNA cleavage intensity. The black oval denotes the nucleosome region (from $-73$ to $+73$ bps), and the white diamonds denote linker regions (from $-123$ to $-74$ bps and from $+74$ to $+123$ bps). The polygonal line and the error bars represent average DNA cleavage intensity and s.d. (Note that s.d. values are shown only for every 5th bp, to improve viewability.) (B) Statistical significance ($P$-values) of differences in cleavage intensity between PRS and RRS as determined by t-test (see Materials and methods). The polygonal line represents $P$-values for each genomic position.

3.3. Structural features of nucleosomal DNA and positional stability of the nucleosome

If nucleosomal DNA structure determines nucleosome positioning, the positional stability of the nucleosome should correlate with structural characteristics of nucleosomal DNA. To determine whether this correlation was in operation, we compared nucleosome positions obtained from the resting state of CD4+ T-cells with those expected from the RRS (see Materials and methods). For each group of stable and unstable nucleosomes, we analyzed DNA bendability and cleavage intensities of the region surrounding the nucleosome (from $-400$ to $+400$ bps) and found that the overall nucleosomal DNA bendability and the overall cleavage intensity of stable nucleosomes were markedly higher than those of unstable nucleosomes (Fig. 3).
more, although a previous study suggested that DNA bendability is higher, higher, and lower, respectively, than the average. Further bendability scores (0.182, 0.068, and frequency) (Peckham et al., 2007; Ogawa et al., 2010), but their formation (neither of the papers offered any reasons for this low example is presented in Supplementary Figure 2). We note that our we found many local regions that did show such tendencies (an local region of DNA showed exactly the same tendencies. However, the linker regions. These observations were true overall, but not every bendability score and a higher DNA cleavage intensity score than did The sequence around the nucleosome overall had a higher DNA bendability score of 0.019. ATA/TAT, TAA/TTA, and AAA/TTT have

structural features of DNA around stable and unstable nucleosomes. (A) DNA bendability and (B) DNA cleavage intensity. The black oval denotes the nucleosome region (from −73 to +73 bps), and the white diamonds denote linker regions (from −123 to −74 bps and from +74 to +123 bps).

4. Discussion

In this study, we computationally analyzed two structural features of the DNA sequences around nucleosomes and found several significant differences between nucleosomal and linker DNA regions. The sequence around the nucleosome overall had a higher DNA bendability score and a higher DNA cleavage intensity score than did the linker regions. These observations were true overall, but not every local region of DNA showed exactly the same tendencies. However, we found many local regions that did show such tendencies (an example is presented in Supplementary Figure 2). We note that our preliminary results indicated that these structural features of DNA sequences associated with nucleosomal and linker regions were also found in Arabidopsis thaliana (Supplementary Figure 3), which implies that these structural features may be conserved in other species.

Previous studies have reported the existence of specific DNA sequence patterns in the nucleosomal DNA region (Segal et al., 2006; Peckham et al., 2007; Ogawa et al., 2010). For instance, a previous support vector machine-based learning approach was used to derive short DNA fragments that frequently appear in the nucleosomal region (Ogawa et al., 2010). However, the correlation between these short DNA fragments and nucleosome formation and DNA bendability scores is far from perfect. Although the trinucleotides CCA/TGG and CAG/CTG have been reported to have the highest nucleosome-formation potentials (Peckham et al., 2007; Ogawa et al., 2010), the bendability scores of CCA/TGG and CAG/CTG, −0.246 and 0.175, respectively, are lower and higher, respectively, than the average bendability score of −0.019. ATA/TAT, TAA/TTA, and AAA/TTT have been suggested as having the potential to inhibit nucleosome formation (neither of the papers offered any reasons for this low frequency) (Peckham et al., 2007; Ogawa et al., 2010), but their bendability scores (0.182, 0.068, and −0.274, respectively) are higher, higher, and lower, respectively, than the average. Furthermore, although a previous study suggested that DNA bendability is strongly correlated with GC content (Peckham et al., 2007), we showed that bendable DNA structures around the center positions of nucleosomes and in the regions around the nucleosome–linker junctions cannot be explained by GC content alone. Accordingly, the high bendability score of the nucleosomal DNA region is not merely a consequence of coincidence between DNA bendability parameters and previously known sequence features of nucleosomal DNA, and this fact supports the plausibility that nucleosomal DNA is bendable.

We also found that high DNA cleavage intensity scores, especially at the central positions of the nucleosomes, were independent of GC content. We also calculated average DNA cleavage intensities of tetrancleavetides containing CCA/TGG and CAG/CTG trinucleotides. Although CCA/TGG and CAG/CTG were the two sequence fragments with the highest potentials to form nucleosomes (Ogawa et al., 2010) (as mentioned previously), the average DNA cleavage intensities of tetrancleavetides containing CCA/TGG and CAG/CTG were 0.37 and 0.55, respectively, that is, the same as and higher than the average tetrancleavetide cleavage intensity (0.37). This result implies that, like DNA bendability, the high DNA cleavage intensity score of the nucleosomal DNA region is not due completely to the known sequence features enriched within the nucleosomal DNA region.

A recent study has suggested a relationship between the bendabilities and cleavage intensities of DNA sequences: these features commonly exist within positions of the minor groove of the DNA structure (Heddi et al., 2009). Our results indicate a similarity between the distribution patterns of DNA bendability and cleavage intensity with respect to genomic position relative to the central position of the nucleosome, and both distributions have −10-bp periodic fluctuation, which is consistent with length of one pitch of the double-stranded DNA structure (i.e., the periodicity of the minor groove positions that face the histone). To determine whether the similar distribution patterns were independent from or tightly associated with each other, we compared trinucleotide bendability scores and DNA cleavage intensity scores of tetrancleavetides. For each of the tetrancleavetides having DNA bendability scores, we calculated the average DNA cleavage intensity of all tetrancleavetides containing the corresponding tetrancleavetide in the nucleosomal DNA region. We found that the Pearson’s correlation coefficient between the tetrancleavetide bendability scores and their corresponding average tetrancleavetide cleavage intensity scores was 0.26, which suggests no direct relationship between bendability and cleavage intensity.

Taken together, our results suggest that the high bendability scores of the nucleosomal DNA region computed in this study and the symmetric bendability pattern around the nucleosomes reflect the fact that, regularly repeated, bendable DNA regions have evolved so that chromosomal DNA can sinusoidally coil around histones and form the compact chromatin structure. Moreover, the high cleavage intensity of the nucleosomal region suggests that histones confer durability to nucleosomal DNA against hydroxyl radicals and that this DNA region has not needed to gain the durability evolutionarily. Abrupt changes in DNA bendability and cleavage intensity were found at the nucleosome–linker junctions. In particular, the DNA bendability of the junction region significantly exceeded that expected from the GC content, which suggests that the DNA at nucleosome–linker junctions plays an important structural role in the stability of nucleosome positioning. The overall trends in DNA bendability and cleavage intensity for the stable nucleosomes were more remarkable than those for the unstable nucleosomes; this result supports our speculations about the biological functions of bendability and cleavage intensity. Gene expression is thought to be controlled by nucleosomal positioning and stability (Cairns, 2009). The relationship between the positional stability of nucleosomes and DNA structural preferences also suggests that nucleosomal DNA sequences are encoded to control gene expression. Although known DNA sequence features that are suggested to facilitate high-efficiency formation of nucleosomes and DNA sequences conferring DNA bendability and cleavage intensity must not be independent from each other, we believe that our results
shed light on how nucleosome positioning’s control of cellular dynamics (e.g., gene expression pattern (Schones et al., 2008; Cairns, 2009)) is regulated by DNA structural preferences.

Acknowledgments

This research was supported by research funds from the Yamagata Prefectural Government and the City of Tsuruoka, Japan, and by a grant-in-aid from the Japan Society for the Promotion of Science (JSPS) to N.Y. (grant no. 1924). During this research, N.Y. was supported by a JSPS Postdoctoral Fellowship (PD) and a JSPS Postdoctoral Fellowship for Research Abroad. The source code of ORChID was kindly provided by Dr. Steve Parker.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.gene.2011.02.004.

References


