How does the periodicity associated with nucleosomal DNA reflect on its intrinsic curvature?

T. Murlidharan Nair
Department of Biological Sciences and Computer Science/Informatics
Indiana University South Bend
1700 Mishawaka Ave, South Bend, IN-46634 USA

***Abstract*- Understanding signals that determine the positioning of nucleosomes along the genomic DNA is fundamental to comprehending gene regulation. Sequence dependent structural properties of DNA, like curvature and flexibility are important in DNA-protein interaction and have been shown to play an important role in nucleosome positioning. The preference of nucleosomes for specific sequences delineates the potential of those sequences to be intrinsically curved. Using theoretical models for determining the DNA curvature I have analyzed the 146 base nucleosome core DNA sequences from *C. elegans*. The analysis reveals a wide distribution of the molecular bend locus over the nucleosome core region. The results obtained using the theoretical model reveal that the nucleosomal DNA sequences have different degrees of curvature over the entire core region. The molecular bend loci associated with these sequences are delocalized and reflects a complex deviation of the DNA axis**

I. INTRODUCTION

Eukaryotic chromatin is formed by repeating units of nucleosomes [1-3]. The nucleosome core is made up of 146 bp of negatively charged DNA wrapped 1.65 times around highly basic proteins called histones, which neutralizes the negative charge. The histone proteins H2A, H2B, H3 and H4 make up the histone core. The formation of the chromatin facilitates the packaging of DNA into chromosomes by compacting it several thousand folds. While compaction facilitates easy packaging of DNA, it hinders the macromolecular machinery from reading the genetic code. Chemical and compositional modification of nucleosomes and nucleosome positioning plays an important role in genome regulation. For long it was thought that histones bound DNA randomly and were simply assigned the role of packaging proteins. Recent studies have thrown light into the basic organization of nucleosomes on chromosomes and their role in regulating genomic function (reviewed in [4]).

New technologies have paved the way towards genome-wide mapping of nucleosome positions, and several maps have now been published [5-9]. These maps have been used in understanding nucleosome organization and the underlying hidden signals for nucleosome positioning [10-13]. Briefly, some the signals that could potentially play a role in nucleosome positioning include signals for rotational and/or translational positioning [11]. The signals may be specific or degenerate, periodically dispersed or localized. In the dispersed category there are short stretches of sequences whose effects are magnified because of their repetitive appearance in a periodical manner [14]. There has been two schools of thoughts have been put forth to explain the nucleosome code, viz. the counter-phase school and the in-phase school. According to the counter-phase school the RR and YY dinucleotides dispersed along the nucleosome, are not in the same phase when they repeat (i.e. they are in alternating RR/YY pattern) [15-17]. The in-phase school argues that RR and YY dinucleotides are in the same phase when they repeat [12, 18-20]. While it is important to understand the nucleosomal DNA signals in terms of the sequence patterns embedded in them, it is equally important to understand the structures that these repeats impart to the free nucleosome DNA.

It is now well established that the structure of DNA is a function of its sequence [15, 21] and certain short stretches of sequences have preference for a specific DNA structure. For instance, occurrence of AA/TT is known to intrinsically curve the DNA axis, while (CA)n or (CG)n form Z-DNA structures. [11]. Since DNA has to wrap around the histone octamer for nucleosome formation, having sequences that have the ability to naturally curve would facilitate the wrapping process. Curved DNAs have thus been considered as signals that could be involved in nucleosome positioning [22, 23]. Recent reports have revealed a periodicity of AA and TT dinucleotides at an interval of 10.4bp within the nucleosomes which could also potentially contribute to DNA curvature [24-26]. There is also a good agreement between the intrinsically curved DNA and model based prediction of nucleosome positioning [27].

Intrinsically curved DNA has been extensively investigated experimentally and theoretically [21, 28-31] and has been linked to nucleosome stability [32, 33]. Two classes of models have been proposed to explain the sequence-depended structure of DNA, the wedge model which is based on the assumption that the hypothetical wedges that are formed as a result of non-coplanar base planes, when repeated in phase with DNA helix repeat (10.5 bp) produces macroscopic curvature [15, 31, 34]. The junction bending model attributes DNA curvature to the distortions at the junction between different DNA structural forms [35-37]. Both models agree that the overall curvature is additive over the individual bending elements and require the phasing of (A)n tracts. DNA curvature has also been demonstrated in DNA fragments lacking poly-A tracts [38]. Experimentally DNA curvature is detected by the anomalous reptation of curved DNA during polyacrylamide gel electrophoresis [29, 37]. Mobility of DNA in gel is directly related to the mean

978-1-4244-5121-0/09/$25.00 ©2009 IEEE

square end to end distance [39]. Wu and Crothers have designed an elegant gel electrophoretic permutation assay to localize the bending locus of an intrinsically curved DNA fragment [40]. De Santis *et al*. [41-43] have proposed a theoretical model for DNA curvature, and have shown that curvature dispersion is linearly correlated with gel

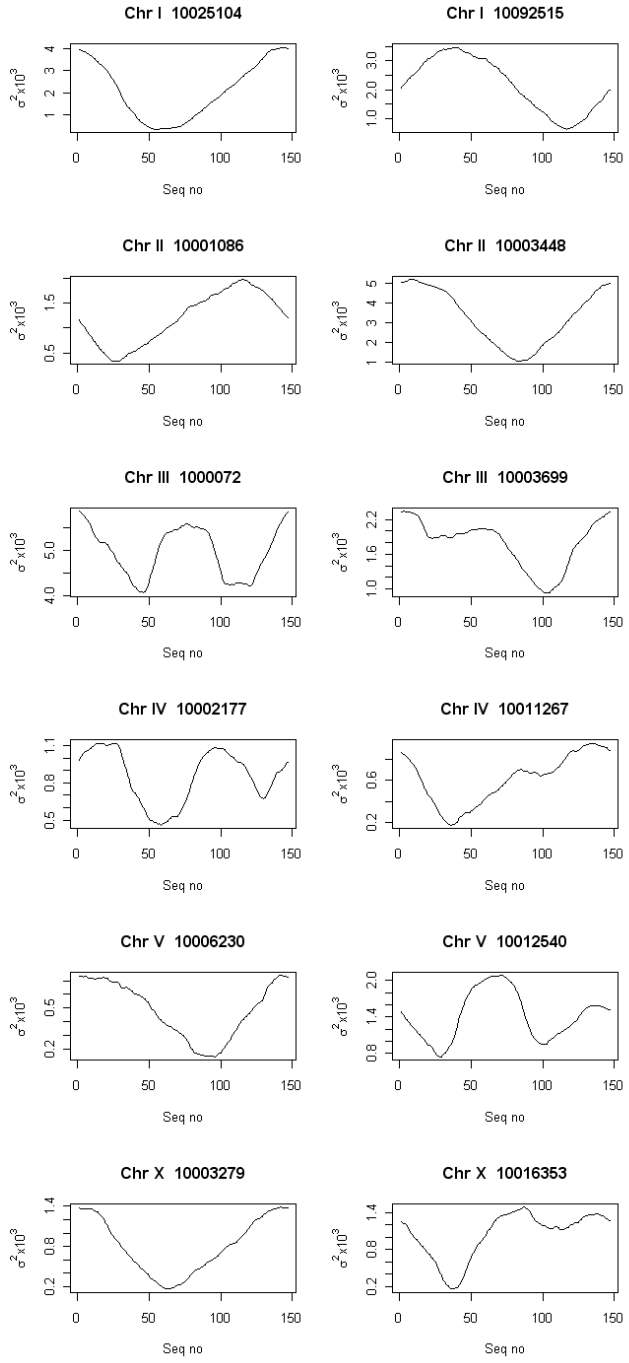

Figure 1 Curvature dispersion calculated by cyclic permutation of the sequence

electrophoretic retardation. The model has been experimentally verified and has been applied to analyze several systems [29, 44].

In the present study higher order DNA structures associated with 146 base nucleosome core DNA sequence from *C. elegans* [9] has been analyzed theoretically. Curvature dispersion associated with the 146 base nucleosome core DNA sequence has been calculated by cyclically permuting the sequence and the distribution of the molecular bend locus of the nucleosome core regions determined. The results indicate a wide distribution of the bend locus, having delocalized curvature throughout the nucleosome core region.

II. Materials and Methods

A. Data

The data for the current study was obtained from *C. elegans* UUPc (Unique unambiguous pyrocore) database. The database contained 28,230 sequences from chromosome I, 30,310 sequences from chromosome II, 26,111 sequences from chromosome III, 30,177 sequences from chromosome IV, 39,547 sequences from chromosome V, and 33, 488 sequences from chromosome X [9].

B. Curvature dispersion calculation

Curvature dispersion has been calculated following the model proposed by De Santis *et al* [41, 42]. The model uses conformational energy calculations to approximate the local deviations of the 16 different dinucleotide steps from the standard B-DNA structure. Deviations from the canonical B-DNA structure are integrated and represented as a curvature vector $C(n,\nu)$, which represents the directional change of the double helical axis between sequence number n and $n+\nu$. The dispersion of curvature σ^2 is calculated as the second moment of the curvature vector, and is shown to be linearly correlated with electrophoretic retardation [42]. Calculating σ^2 by cyclically permuting the sequence is a theoretical alternative for localizing the molecular bend locus. For details refer to De Santis *et al* [41].

C. DNA path calculation

DNA path was calculated following the model developed by Shpigelman et. al. [45]. The overall DNA path is calculated using the local helix parameters viz. helix twist angle, wedge angle and the direction of deflection angle. The coordinates of the successive base pair stacks are calculated by applying (i) translation by half the average rise per residue (average rise per residue = 3.39) along the Z axis (ii) half the helical twist rotation about Z axis (iii) rotation by the wedge angle in the XY-plane (iv) rotation by another half helical twist about the Z axis (v) translation by another half of the average rise per residue. These transformations can be described in the following equation:

$$M_n = T\left(-\frac{3.39}{2}\right)\times R_z\left(\frac{\Omega}{2}\right)\times Q(\sigma,\delta-90^\circ)\times R_z\left(\frac{\Omega}{2}\right)\times T\left(-\frac{3.39}{2}\right) \quad \text{where}$$

$$Q(\sigma,\delta-90^\circ) = R_z\left(-(\delta-90^\circ)\right)\times R_x(\sigma)\times R_z\left(\delta-90^\circ\right) \text{ and } R_n \text{ is the rotation about n axis}$$

The programs for computing the coordinates were developed in R (www.r-project.org). The angles of Twist (Ω), Wedge (σ) and Direction (δ) were taken from those determined by Bolshoy *et al* [38, 45] experimentally as well as those

determined by De Santis *et al* [42]. Both angles essentially predicted the same structure.

III. Results and Discussion

Sequence directed structures associated with the nucleosome DNA of *C. elegans* has been analyzed using the theoretical models for DNA curvature. Recent analysis by measuring the distance between YY, YR, RR and RY dinucleotides of nucleosome DNA fragments from *C. elegans* revealed a consensus sequence structure of the nucleosome DNA repeat to be (YYYYYRRRRR)n [14]. Phase shifts of between various dinucleotides within ~10 base nucleosome sequence repeat have been reported earlier [25, 46]. A bendability matrix has been used to represent these phase preferences, and it has been noted that AA and TT dinucleotides counter-phase one another, may reflect the periodical pattern of the nucleosome DNA [25]. Nucleosome DNA bendability matrix that was recently determined from nucleosome core DNA sequences of *C. elegans* revealed a consensus repeat of A(TTTCCGGAAA)T [47]. With a view to understand how the periodicity affects the overall structure of free nucleosomal DNA, the UUPc database was analyzed using the theoretical models for DNA curvature. The curvature dispersion calculated as the second moment of the curvature vector by cyclically permuting the sequences revealed the molecular bend locus of the nucleosomal DNA sequence. In the interest of brevity curvature dispersion for two sequences from each of the chromosomes is represented in Figure 1. Curvature dispersion calculations were done for all the sequences in the database. Curvature dispersion retains all the characteristic of the curvature profile, but has the added advantage that it improves the signal to noise ratio. Since curvature dispersion is linearly correlated with gel electrophoretic retardation, calculating curvature dispersion by cyclically permutation of the sequence is equivalent to performing a cyclic permutation assay theoretically [30, 44]. The minima of the curve corresponds the bend locus of the fragment. This is equivalent to the experimental cyclic permutation assay in which a linear faster reptating fragment is obtained if its bend locus is destroyed by restriction digestion [40]. Delineating the bend loci associated the nucleosome DNA sequence helps understand the regions where the curvature is concentrated which in turn helps describe the wrapping of the DNA. Figure 2 shows the distribution of the bend locus for each chromosome as obtained using the theoretical permutation assay. The graphs correspond to the distribution of the minima. Results show a rather even distribution of the loci with relatively fewer loci concentrated at position 140 and beyond. The distribution points to the fact that that nucleosome core region has curved regions throughout the entire stretch, and may depend on how it is being packaged. While the histograms in Figure 2 correspond to the minima, it is noteworthy to point out that several nucleosome core sequences had local minimas. Presence of these local minima reveals a much more complicated deviation of the DNA axis associated with the nucleosome DNA.

With a view to understand the deviation of the DNA axis, the DNA paths of the nucleosome core DNA sequences were computed. In the interest of brevity, DNA paths for only two

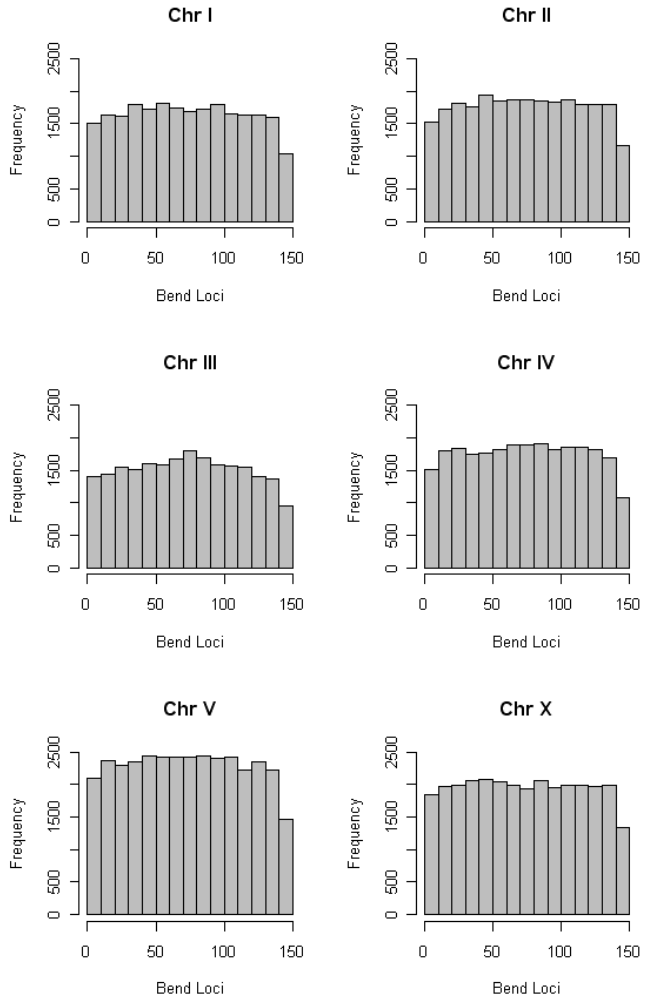

Figure 2: Distribution of the bend locus as calculated using the theoretical model

of the nucleosome core sequence for each of the chromosomes is shown in Figure 3. The paths reveal the complex trajectories assumed by nucleosome DNA. It is important to point out that these are theoretically computed results using well accepted models that have been experimentally tested on other systems. In trying to understand the sequence directed curvatures associated with the nucleosomal DNA, it is important to recap the well established fact that DNA is anisotropic. The anisotropy may be a result of the helical structure of the DNA itself or it may be function of its sequence [48]. In either case it makes it more bendable towards the groove even for unperturbed DNA [49, 50]. For the nucleosomal DNA to conveniently wrap around the histone octamer, the sequence repeats should be such that, it facilitates this process. Every dinucleotide is capable of deflecting the DNA axis depending on the wedge angles associated with it [38, 42]. Periodicity in

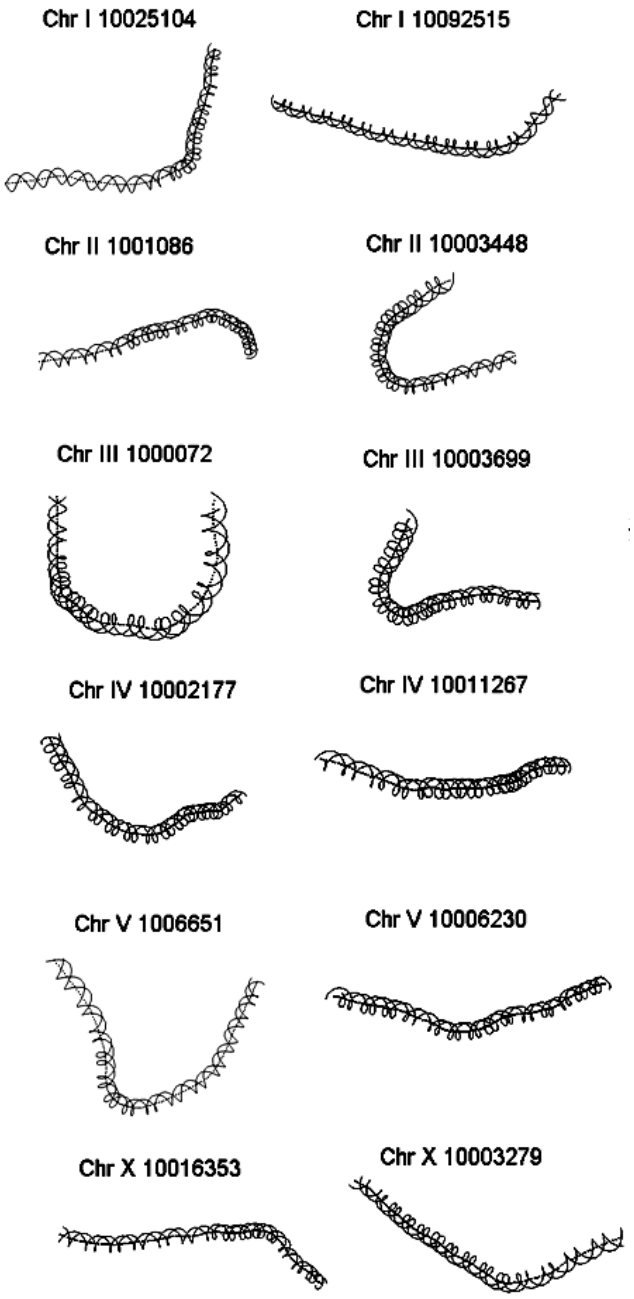

Figure 3 DNA path calculated for the nucleosome DNA whose bend loci is represented in figure 1

the distribution of any particular dinucleotide will produce deflection in the DNA axis that will be additive over their individual wedge angle contributions. Towards describing the nucleosome sequence patterns, with the two major competing schools of thought, the "counter-phase" school that claims the RR and YY dinucleotides are distributed in alternating RR/TT fashion and the "in-phase" school that claims the RR and YY dinucleotides are in the same phase within the repeat unit, it is important to understand how these repeats translate into structure and to decipher other messages that nucleosome DNA carry. Further, there are other components that should not be ignored, which include the histone induced bending component and the role of polarization interactions in the wrapping/unwrapping of nucleosomal DNA [51]. The results presented here lends credence to the recent report by Gabdank *et al* [47], wherein they infer from their analysis that bendability is not the sole reason for positional preference of dinucleotides.

The results of the analysis demonstrate that variable degrees of curvature are associated with nucleosomal DNA. The results also reveal that nucleosomal DNAs do not conform to the same exact sequence dependent structure, and have bend loci localized at different positions along the sequence. Nucleosomal DNA has been attributed to carrying more messages than just the chromatin code, and is considered the most degenerate code [17, 52]. From the biological functional perspective, the non-optimal positions of the dinucleotides may actually be an advantage, facilitating important biological processes of replication and transcription. Nature has optimized the chromatin code for multiple functions, making it one of the most difficult feature extraction problems.

ACKNOWLEDGMENT

I would like to thanks Prof. E. N. Trifonov for useful discussion on nucleosomal DNA. I also thank Prof. Arthur Lim for help in understanding the change-of-basis matrix calculation. This work is supported by IUSB summer research grant.

REFERENCES

[1] R. D. Kornberg, and Y. Lorch, "Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome," Cell, vol. 98, no. 3, pp. 285-94, Aug 6, 1999.
[2] R. D. Kornberg, and A. Klug, "The nucleosome," Sci Am, vol. 244, no. 2, pp. 52-64, Feb, 1981.
[3] K. Luger, A. W. Mader, R. K. Richmond et al., "Crystal structure of the nucleosome core particle at 2.8 A resolution," Nature, vol. 389, no. 6648, pp. 251-60, Sep 18, 1997.
[4] C. Jiang, and B. F. Pugh, "Nucleosome positioning and gene regulation: advances through genomics," Nat Rev Genet, vol. 10, no. 3, pp. 161-72, Mar, 2009.
[5] A. Barski, S. Cuddapah, K. Cui et al., "High-resolution profiling of histone methylations in the human genome," Cell, vol. 129, no. 4, pp. 823-37, May 18, 2007.
[6] T. N. Mavrich, C. Jiang, I. P. Ioshikhes et al., "Nucleosome organization in the Drosophila genome," Nature, vol. 453, no. 7193, pp. 358-62, May 15, 2008.
[7] D. E. Schones, K. Cui, S. Cuddapah et al., "Dynamic regulation of nucleosome positioning in the human genome," Cell, vol. 132, no. 5, pp. 887-98, Mar 7, 2008.
[8] W. Lee, D. Tillo, N. Bray et al., "A high-resolution atlas of nucleosome occupancy in yeast," Nat Genet, vol. 39, no. 10, pp. 1235-44, Oct, 2007.
[9] S. M. Johnson, F. J. Tan, H. L. McCullough et al., "Flexibility and constraint in the nucleosome core landscape of Caenorhabditis elegans chromatin," Genome Res, vol. 16, no. 12, pp. 1505-16, Dec, 2006.
[10] E. Segal, and J. Widom, "Poly(dA:dT) tracts: major determinants of nucleosome organization," Curr Opin Struct Biol, vol. 19, no. 1, pp. 65-71, Feb, 2009.
[11] R. Kiyama, and E. N. Trifonov, "What positions nucleosomes?--A model," FEBS Lett, vol. 523, no. 1-3, pp. 7-11, Jul 17, 2002.
[12] E. Segal, Y. Fondufe-Mittendorf, L. Chen et al., "A genomic code for nucleosome positioning," Nature, vol. 442, no. 7104, pp. 772-8, Aug 17, 2006.

[13] N. Kaplan, I. K. Moore, Y. Fondufe-Mittendorf et al., "The DNA-encoded nucleosome organization of a eukaryotic genome," Nature, vol. 458, no. 7236, pp. 362-6, Mar 19, 2009.
[14] F. Salih, B. Salih, and E. N. Trifonov, "Sequence structure of hidden 10.4-base repeat in the nucleosomes of C. elegans," J Biomol Struct Dyn, vol. 26, no. 3, pp. 273-82, Dec, 2008.
[15] E. N. Trifonov, "Sequence-dependent deformational anisotropy of chromatin DNA," Nucleic Acids Res, vol. 8, no. 17, pp. 4041-53, Sep 11, 1980.
[16] A. B. Cohanim, Y. Kashi, and E. N. Trifonov, "Yeast nucleosome DNA pattern: deconvolution from genome sequences of S. cerevisiae," J Biomol Struct Dyn, vol. 22, no. 6, pp. 687-94, Jun, 2005.
[17] S. Kogan, and E. N. Trifonov, "Gene splice sites correlate with nucleosome positions," Gene, vol. 352, pp. 57-62, Jun 6, 2005.
[18] S. C. Satchwell, H. R. Drew, and A. A. Travers, "Sequence periodicities in chicken nucleosome core DNA," J Mol Biol, vol. 191, no. 4, pp. 659-75, Oct 20, 1986.
[19] J. Widom, "Role of DNA sequence in nucleosome stability and dynamics," Q Rev Biophys, vol. 34, no. 3, pp. 269-324, Aug, 2001.
[20] J. Widom, "A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells," Proc Natl Acad Sci U S A, vol. 89, no. 3, pp. 1095-9, Feb 1, 1992.
[21] E. N. Trifonov, "Curved DNA," CRC Crit Rev Biochem, vol. 19, no. 2, pp. 89-106, 1985.
[22] S. B. Kogan, M. Kato, R. Kiyama et al., "Sequence structure of human nucleosome DNA," J Biomol Struct Dyn, vol. 24, no. 1, pp. 43-8, Aug, 2006.
[23] B. Pina, U. Bruggemeier, and M. Beato, "Nucleosome positioning modulates accessibility of regulatory proteins to the mouse mammary tumor virus promoter," Cell, vol. 60, no. 5, pp. 719-31, Mar 9, 1990.
[24] I. Ioshikhes, A. Bolshoy, and E. N. Trifonov, "Preferred positions of AA and TT dinucleotides in aligned nucleosomal DNA sequences," J Biomol Struct Dyn, vol. 9, no. 6, pp. 1111-7, Jun, 1992.
[25] G. Mengeritsky, and E. N. Trifonov, "Nucleotide sequence-directed mapping of the nucleosomes," Nucleic Acids Res, vol. 11, no. 11, pp. 3833-51, Jun 11, 1983.
[26] F. Salih, B. Salih, and E. N. Trifonov, "Sequence-directed mapping of nucleosome positions," J Biomol Struct Dyn, vol. 24, no. 5, pp. 489-93, Apr, 2007.
[27] Y. Wada-Kiyama, K. Kuwabara, Y. Sakuma et al., "Localization of curved DNA and its association with nucleosome phasing in the promoter region of the human estrogen receptor alpha gene," FEBS Lett, vol. 444, no. 1, pp. 117-24, Feb 5, 1999.
[28] P. J. Hagerman, "Sequence-directed curvature of DNA," Annu Rev Biochem, vol. 59, pp. 755-81, 1990.
[29] T. M. Nair, K. Madhusudan, V. Nagaraja et al., "On the mobility behavior of a curved DNA fragment located in circular permutation," FEBS Lett, vol. 351, no. 3, pp. 321-4, Sep 12, 1994.
[30] T. M. Nair, K. Madhusudan, V. Nagaraja et al., "Theoretical permutation gel electrophoretic analysis of a curved DNA fragment located in circular permutation," Electrophoresis, vol. 17, no. 4, pp. 633-41, Apr, 1996.
[31] L. E. Ulanovsky, and E. N. Trifonov, "Estimation of wedge components in curved DNA," Nature, vol. 326, no. 6114, pp. 720-2, Apr 16-22, 1987.
[32] C. Anselmi, G. Bocchinfuso, P. De Santis et al., "Dual role of DNA intrinsic curvature and flexibility in determining nucleosome stability," J Mol Biol, vol. 286, no. 5, pp. 1293-301, Mar 12, 1999.
[33] A. Scipioni, S. Pisano, C. Anselmi et al., "Dual role of sequence-dependent DNA curvature in nucleosome stability: the critical test of highly bent Crithidia fasciculata DNA tract," Biophys Chem, vol. 107, no. 1, pp. 7-17, Jan 1, 2004.
[34] E. N. Trifonov, and J. L. Sussman, "The pitch of chromatin DNA is reflected in its nucleotide sequence," Proc Natl Acad Sci U S A, vol. 77, no. 7, pp. 3816-20, Jul, 1980.
[35] S. D. Levene, H. M. Wu, and D. M. Crothers, "Bending and flexibility of kinetoplast DNA," Biochemistry, vol. 25, no. 14, pp. 3988-95, Jul 15, 1986.
[36] D. M. Crothers, J. Drak, J. D. Kahn et al., "DNA bending, flexibility, and helical repeat by cyclization kinetics," Methods Enzymol, vol. 212, pp. 3-29, 1992.
[37] J. C. Marini, S. D. Levene, D. M. Crothers et al., "Bent helical structure in kinetoplast DNA," Proc Natl Acad Sci U S A, vol. 79, no. 24, pp. 7664-7668, Dec, 1982.
[38] A. Bolshoy, P. McNamara, R. E. Harrington et al., "Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles," Proc Natl Acad Sci U S A, vol. 88, no. 6, pp. 2312-6, Mar 15, 1991.
[39] P. G. de Gennes, "Passive entry of a DNA molecule into a small pore," Proc Natl Acad Sci U S A, vol. 96, no. 13, pp. 7262-4, Jun 22, 1999.
[40] H. M. Wu, and D. M. Crothers, "The locus of sequence-directed and protein-induced DNA bending," Nature, vol. 308, no. 5959, pp. 509-13, Apr 5-11, 1984.
[41] P. De Santis, A. Palleschi, M. Savino et al., "A theoretical model of DNA curvature," Biophys Chem, vol. 32, no. 2-3, pp. 305-17, Dec, 1988.
[42] P. De Santis, A. Palleschi, M. Savino et al., "Validity of the nearest-neighbor approximation in the evaluation of the electrophoretic manifestations of DNA curvature," Biochemistry, vol. 29, no. 39, pp. 9269-73, Oct 2, 1990.
[43] G. Zuccheri, A. Scipioni, V. Cavaliere et al., "Mapping the intrinsic curvature and flexibility along the DNA chain," Proc Natl Acad Sci U S A, vol. 98, no. 6, pp. 3074-9, Mar 13, 2001.
[44] T. M. Nair, "Evidence for intrinsic DNA bends within the human cdc2 promoter," FEBS Lett, vol. 422, no. 1, pp. 94-8, Jan 23, 1998.
[45] E. S. Shpigelman, E. N. Trifonov, and A. Bolshoy, "CURVATURE: software for the analysis of curved DNA," Comput Appl Biosci, vol. 9, no. 4, pp. 435-40, Aug, 1993.
[46] I. Ioshikhes, A. Bolshoy, K. Derenshteyn et al., "Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences," J Mol Biol, vol. 262, no. 2, pp. 129-39, Sep 20, 1996.
[47] I. Gabdank, D. Barash, and E. N. Trifonov, "Nucleosome DNA bendability matrix (C. elegans)," J Biomol Struct Dyn, vol. 26, no. 4, pp. 403-11, Feb, 2009.
[48] J. A. Schellman, "Flexibility of DNA," Biopolymers, vol. 13, no. 1, pp. 217-26, Jan, 1974.
[49] N. B. Ulyanov, and V. B. Zhurkin, "Sequence-dependent anisotropic flexibility of B-DNA. A conformational study," J Biomol Struct Dyn, vol. 2, no. 2, pp. 361-85, Oct, 1984.
[50] V. B. Zhurkin, "[Local mobility of the DNA double helix. Comparison of conformational analysis with experiments]," Mol Biol (Mosk), vol. 17, no. 3, pp. 622-38, May-Jun, 1983.
[51] A. M. Ababneh, "The role of polarization interactions in the wrapping/unwrapping of nucleosomal DNA around the histone octamer: Implications to gene regulation," J Theor Biol, vol. 258, no. 2, pp. 229-39, May 21, 2009.
[52] A. B. Cohanim, Y. Kashi, and E. N. Trifonov, "Three sequence rules for chromatin," J Biomol Struct Dyn, vol. 23, no. 5, pp. 559-66, Apr, 2006.