Periodicity of DNA bend sites in eukaryotic genomes

Ryoiti Kiyama

Institute of Molecular and Cellular Biosciences, University of Tokyo, Yayoi, Bunkyo-ku, Tokyo 113, and National Institute of Bioscience and Human-Technology, Tsukuba-shi, Ibaraki 305, Japan

Correspondence to: Ryoiti Kiyama, Tel: 81-3-3812-2111, ext. 7835, Fax: 81-3-3818-9437, E-mail: kiyamar@hgc.ims.u-tokyo.ac.jp

Summary

We found that DNA bend sites are distributed regularly and periodically in the genomic DNA of eukaryotes. Their locations were conserved during molecular evolution in otherwise unstable intergenic regions of genomic DNA at intervals of approximately 700 bp, which corresponds to a length of four nucleosomes, suggesting their active role in chromatin organization. By further examination of these sites with respect to chromatin structure, we obtained evidence that these sites may act as signals for nucleosome phasing. Here, we summarize our results regarding periodic bent DNA in the human β-globin, c-*myc*, and immunoglobulin heavy chain μ loci and discuss their biological functions.

I. Bent DNA in biological reactions

Genomic DNA is a source of genetic and functional information in the form of nucleotide sequences. DNA has a relatively simple composition of purine or pyrimidine bases attached to a common phosphate backbone which would not give rise much local structural variation. However, recent studies have revealed that non-B DNA or "unusual" DNA structures are actively involved in biologically important reactions as functional elements (Crothers *et al.*, 1990). For example, Z-DNA is known to activate transcription and recombination presumably by exposing bases on the outside of the phosphate backbone, thereby increasing the chance of interaction with proteins or other bases (Rich *et al.*, 1984; Blaho and Wells, 1989). Other structures such as triplex DNA and unisomorphic DNA have been discussed in studies of transcriptional activation and recombination mechanisms (Crothers *et al.*, 1990; Soyfer and Potaman, 1996). Although their mechanisms of action are quite different, these non-B DNA structures seem to act as signals for recognition by protein factors in a way different from searching nucleotide bases.

Bent DNA was first discovered as anomalous migration of DNA fragments in gels, and has been extensively studied due to its potential involvement as a transcriptional modulator (Travers, 1989; Hagerman, 1990; Crothers *et al.*, 1990; Trifonov, 1991; Ioshikhes *et al.*, 1996; Werner *et al.*, 1996). Such structures also play important roles in activation of recombination (Nash, 1990). Binding of proteins to DNA can cause DNA bending, which would further enhance the recognition by other proteins of the site of the protein-DNA complex (Khan and Crothers, 1992). Therefore, DNA bending

formed by the intrinsic nature of the DNA or by protein binding, as well as sequence information, would be a good signal for structural recognition. Note that the non-B DNA structures themselves are the result of nucleotide sequence information, although the sequence-structure relationship is not simple.

II. Assays for DNA bend sites

The presence and the location of DNA bend sites can be analyzed by several assays. Among these, the circular permutation assay (Wu and Crothers, 1984) has been commonly utilized for mapping the bend sites in DNA fragments of several hundred bp to 1 kb in length. The assay procedure is schematically illustrated in **Figure 1**. Plasmids containing the tandem dimers of the fragment of interest are cloned, and after digestion of the plasmid DNA with the restriction enzymes that cut the fragment once the plasmid DNA samples are resolved by electrophoresis. We routinely use 8% polyacrylamide (mono: bis = 29: 1) gels, which can resolve bands up to 1 kb in size (Wada-Kiyama and Kiyama, 1994). Electrophoresis should be performed at 4˚C twice or three times overnight to obtain better resolution of the bands. Cloning of the tandem dimers can be performed by direct cloning of two identical fragments into the multiple cloning site of the vector, or cloning them into two different sites one after another. Most of the clones could be obtained by the former method under conditions where the fragment (0.1 to 1 μg) is present in excess over the amount of vector DNA (ten times or more) in a small-volume reaction mixture (5 to 10 μl). After transformation of *E. coli*, only direct repeats of the fragments, but not
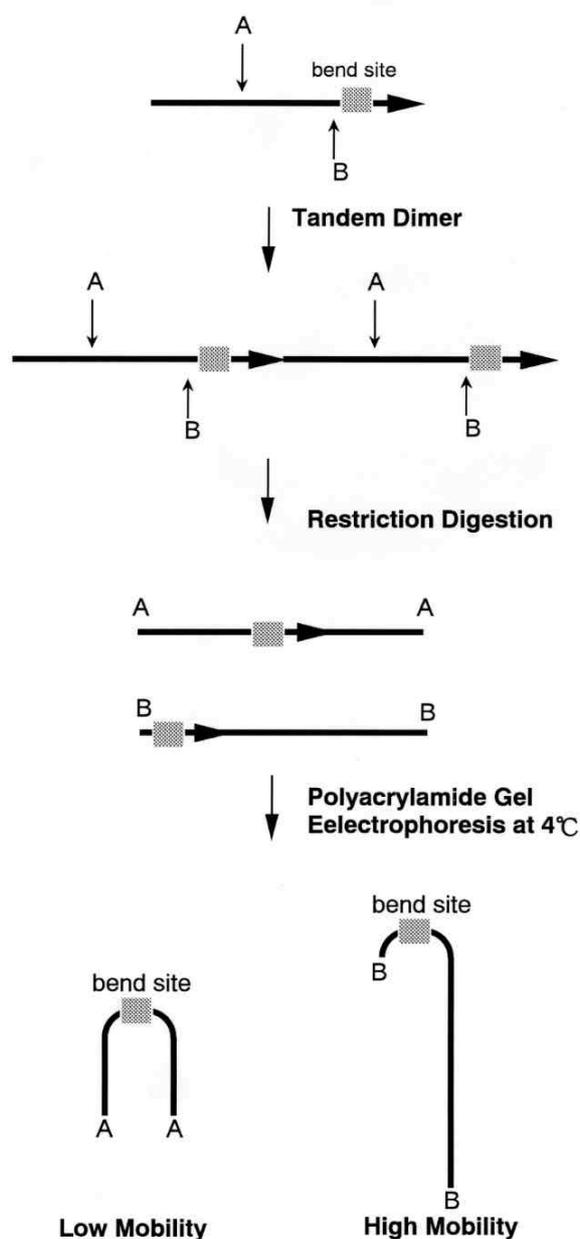
Fig. 1. Assay for bent DNA.

inverted repeats, can be obtained in dimeric form.

The circular permutation assay is a very simple and reliable method to identify and roughly map DNA bend sites. The results of mapping are reproducible under identical electrophoresis conditions and generally reproducible among different subclones containing the same site, and the patterns can be interpreted without

complicate calculations. However, this method does have some technical limitations. Firstly, the assay is totally dependent upon the availability of suitable restriction sites. If there are no appropriate restriction sites, the bend sites cannot be localized to a small region of DNA. Secondly, the DNA structure of the other part of the same fragment could influence the mobility. As a result of this effect, mapping a bend site to a very small region by this method would not always give a precise location. Although rare, we observed a slight difference in the location of a site in the -globin gene region between clones of different sizes used for mapping. Therefore, the lower limit of resolution would be 50 to 100 bp.

The more precise location of the bend sites could be achieved by several other methods. For a relatively large region, sites can be examined with deletion constructs. When the bend center is completely deleted from the construct, all restricted fragments have the same mobility. Meanwhile, mapping the site in regions of approximately 100 bp or less would be achieved by using concatenated oligonucleotides of 20 or 30 bp (Wada-Kiyama and Kiyama, 1995). When the oligonucleotide contains a bend site, the concatemers exhibit retardation on polyacrylamide gel electrophoresis. The effect of bending is greater as the length of the oligonucleotide increases. The bend angle can be estimated by comigration of standards such as A_3N_7 (0.63°/ base) (Calladine *et al.*, 1988). The bend angle could also be determined by the assay based on ring closure of concatenated oligonucleotides (Zahn and Blattner, 1987).

III. The human β-globin locus

Using the circular permutation assay, we mapped the DNA bend sites in the human -globin locus which is located on chromosome 11 and contains five (-, ^{G} -, ^{A} -, - and -) active genes and one (-) pseudogene (**Figure 2**). This locus is ideal for mapping the sites because the nucleotide sequence of over 70 kb has been reported. Furthermore, since most of the locus is intergenic, the influence of the coding region could be excluded. The similarity of the exon-intron structure and the sequences of the flanking region would be ideal for evolutionary study of the sites. The chromatin structure in this locus has been extensively characterized in that switching of globin gene expression is paralleled by alterations of chromatin structure as revealed by DNase I-hypersensitivity (reviewed by Stamatoyannopoulos and Nienhuis, 1993; Evans *et al.*, 1990).

The periodicity of the bend sites at intervals of 680 bp on average was first identified in the -globin region (Wada-Kiyama and Kiyama, 1994). Further studies of the sites in the regions of other globin genes revealed that relative locations of the sites to their cap sites were conserved among most of the members of this family which were separated as much as 200 million years ago.
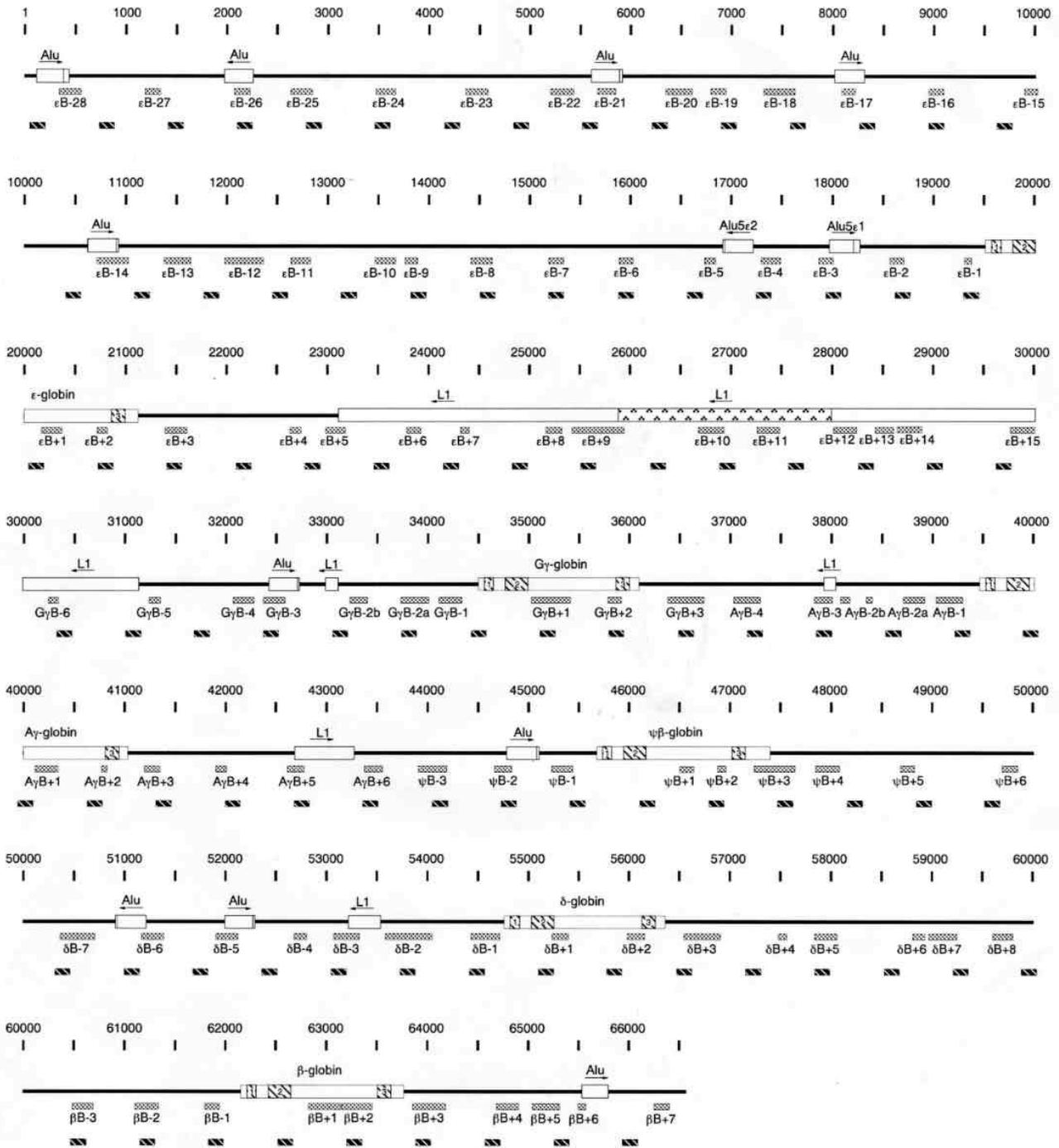
Fig. 2. Periodic bent DNA in the human -globin locus. Mapped DNA bend sites are shown as shadowed boxes. Hatched boxes indicate putative 150 bp sites aligned at 680 bp intervals as a reference for periodicity.
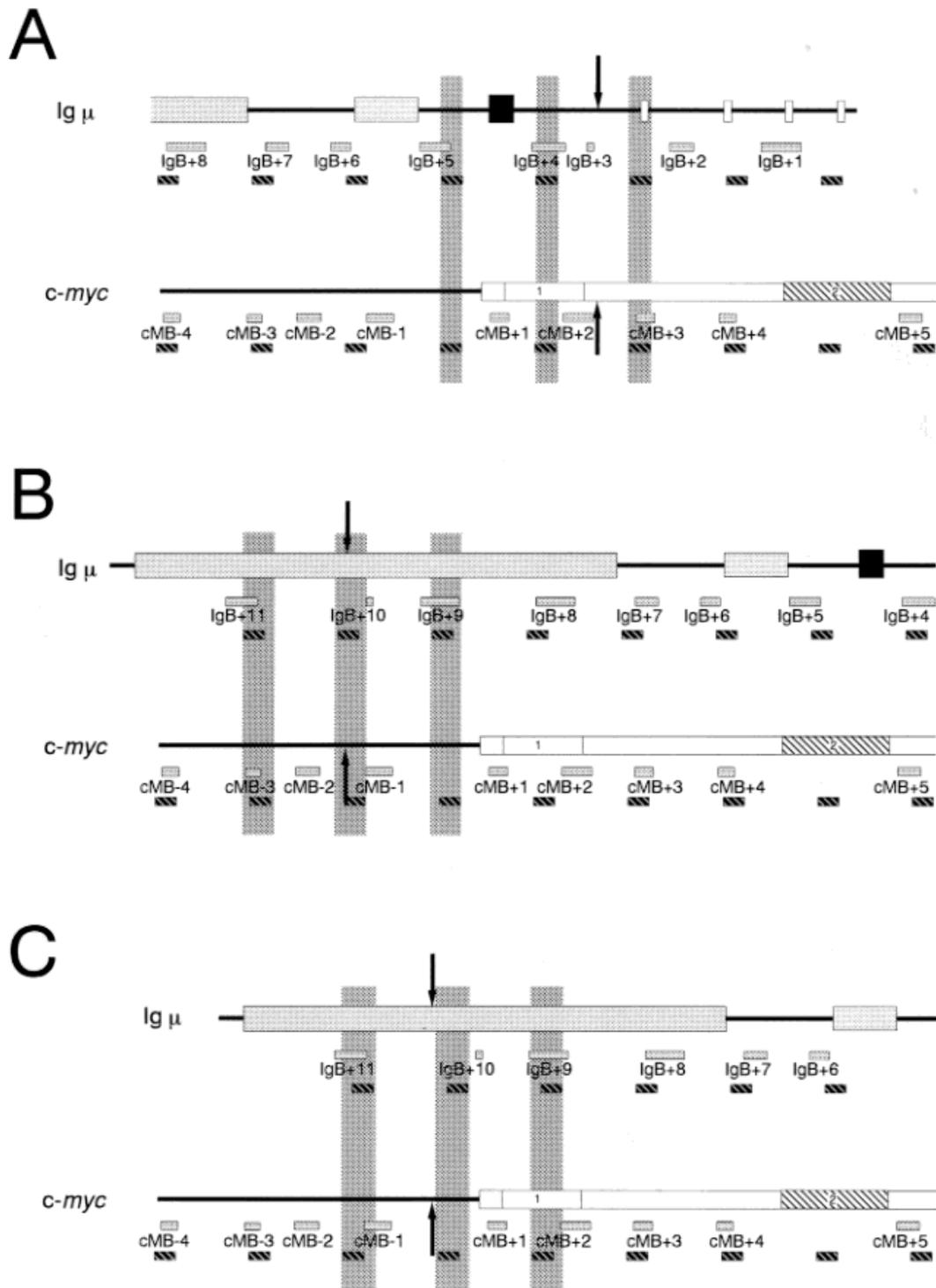
Fig. 3. Conservation of periodic bent DNA in the translocation of the c-*myc* and Igµ loci. DNA bend sites in the c-*myc* (bottom) and Igµ (top) loci are aligned to highlight the conservation of the periodicity of the hypothetical sites (shadowed columns) based on their universal periodicity, after the rearrangements observed in Manca (**A**), BL22 (**B**) and Ramos (**C**) cell lines. Only three hypothetical sites near the junctions are shadowed but they were matched throughout the loci. Reproduced from Ohki *et al.* (1997).

Table 1. Average intervals of periodic bent DNA.

Locus	Mapped region (kb)	No. of sites	Average (bp)	S. D. [a] (bp)	Ref.
-globin	66	98	679.2	229.6	**b**
c-*myc*	8	12	694.2	281.4	Ohki *et al.* 1997
Igμ	7	11	654.5	222.7	Ohki *et al.* 1997
Erythropoietin receptor	9	13	651.2	221.0	**b**
Estrogen receptor [c]	3	5	688.1	210.9	Kuwabara *et al.* 1997

[a] Standard deviation.

[b] Unpublished results.

[c] 5'-region containing the alternative cap site, P0.

The duplication of the two -globin genes, which occurred most recently, was immediately followed by diversification of the non-coding region by as much as 70%, while all of the bend sites were conserved (Slightom *et al.*, 1980). Insertion of an *Alu* sequence might have disturbed the periodicity, although as observed in the region upstream of the -globin gene, the interval seemed to have returned to the average after a long period of molecular evolution. The positions of the bend sites were conserved even between the human - and mouse ^{maj}-globin genes (Wada-Kiyama and Kiyama, 1996b).

Mapping of over 90 bend sites in the locus revealed that the periodicity of the bend sites exists throughout the locus with an average interval of 680 bp (Wada-Kiyama and Kiyama, 1994, 1995, 1996b; unpublished results). However, we observed disturbance of the periodicity at several locations. Interestingly, all of the locations of the disturbed periodicity located upstream of the -globin genes that caused the distances of the adjacent bend sites to be longer than average were found in or close to the DNase I-hypersensitive sites, which constitute the locus control region (-LCR). The -LCR is composed of four or five developmentally-regulated DNase I-hypersensitive sites (Crossley and Orkin, 1993; Evans *et al.*, 1990; Felsenfeld, 1993). These sites are designated as open chromatin regions and act as the sites of interaction of transcription factors and the enhancer-binding protein NF-E2. This region interacts with the promoter region of each member of the -like globin gene family and controls their expression during development. One of the DNase I-hypersensitive sites, HS2 located 11 kb upstream of the cap site of the -globin gene, was located in the center of two adjacent bend sites separated by a distance of 860 bp, which is longer than average (unpublished results). It seemed as if HS2 was placed far from the bend sites to minimize the influence of the sites. This is discussed again below.

IV. The human c-*myc* and the immunoglobulin heavy chain μ loci

The human c-*myc* gene has three exons and occupies a region of approximately 5.5 kb on chromosome 8. As observed in the -globin locus, this locus contained periodic bent DNA. DNA bend sites were mapped at an average interval of 694.2 bp and were present in the 5'- and 3'- non-coding regions, introns and the non-coding exon (exon 1), but not present in the coding region (Ohki *et al.*, 1997). Interestingly, one of the bend sites corresponded to the location of TATA box of the P2 promoter, suggesting that prebending of the promoter region can facilitate transcriptional enhancement.

The c-*myc* gene is involved in the progression of Burkitt's lymphoma by translocation of the locus into one of the immunoglobulin genes located on chromosomes 2, 14 or 22. These translocation events often result in reshuffling the location of regulatory elements. Deregulation of the expression by juxtaposition of the μ enhancer to the c-*myc* promoter is one of the mechanisms of tumor progression caused by this oncogene. The mechanism of these translocation events has not been well documented except that immunoglobulin-specific recombination is somehow involved (Specer and Groudine, 1991). Translocation junctions were formed at various locations in the locus yet no specific sequences were commonly found in their immediate proximity. However, when the periodic bent DNA was mapped in the c-*myc* and the Igμ loci, at least three stable cell lines containing the translocation junctions within these regions showed conservation of the periodicity before and after the rearrangements (**Figure 3**). This would be readily explained if we assume that the periodic bent DNA is a key element for chromatin structure. It would be necessary for a stable cell line to maintain a similar chromatin structure as to that before the rearrangement. Otherwise, a secondary rearrangement could alter the sequence until a stable structure is eventually formed.

V. Other loci in eukaryotic genomes

We have already determined that periodic bent DNA is present in the human erythropoietin receptor and the human estrogen receptor loci (Kuwabara *et al.*, 1997; unpublished results). The intervals of the sites in these loci were 651.2 or 688.1 bp, respectively, which are close to the values for other loci (**Table 1**). In both cases, its periodicity was disturbed by exons. For the estrogen receptor gene, the alternative cap site located approximately 2 kb upstream of the canonical site caused a shift of the nearby sites. For the erythropoietin receptor gene, the 700 bp periodicity of bend sites was conserved even within the long introns (1st and 6th introns) although the sites were shifted when the length of introns was not sufficient to accommodate two sites. One of the sites in the estrogen receptor gene contained motifs of the estrogen response element, the binding site for the hormone-responsive trans-activating factors, and had the affinity to the nuclear scaffold. This site might play a role in determining the nuclear localization of this gene as well as a role in transcriptional regulation.

For other loci of eukaryotes, we examined the potential bend sites by a computer search (Wada-Kiyama, and Kiyama, 1996a). Based on the observation that physically mapped bend sites often contain A+T-rich sequences including A_n or T_n tracts at intervals of ten or multiples of ten nucleotides, we searched $A_2N_8A_2N_8A_2$ and the complementary $T_2N_8T_2N_8T_2$ for a periodicity. There was a statistically significant sequence periodicity at an interval of roughly 700 bp in eukaryotic genomic DNA. This tendency was absent in prokaryotes and in eukaryotic cDNA, suggesting that the periodicity is universal among eukaryotic genomes, especially in intergenic regions.

VI. Biological significance of the bend sites

The observations that periodic bent DNA is conserved during molecular evolution and its intervals are maintained precisely in otherwise unstable intergenic regions suggested that these sites are biologically relevant. Despite the systematic and organized patterns of chromatin folding, no specific signals have been determined as key elements for the folding mechanism. A computer search further revealed the non-random distribution of nucleotide sequences on the genomic DNA, while it failed to deduce any specific sequences in common, suggesting the presence of unidentified codes which are not apparent from sequence information alone. Therefore, judging from the periodicity and the length of their intervals, periodic bent DNA may be closely associated with chromatin structure, presumably with the formation of nucleosomes. We reported that some of the sites were indeed involved in the formation of nucleosomes by having high affinity to histone core particles (Wada-Kiyama and Kiyama, 1996b). Chromatin structure seems to be extensively stabilized when the overall periodicity is maintained before and after the rearrangement. Meanwhile, open chromatin regions,

revealed by DNase I-hypersensitivity (Gross and Garrard, 1988), could be at least partly due to disturbance of the periodicity. While the nucleosome phasing activity of these sites might be effective when the distances of the bend sites are less than or equal to the length of four nucleosomes, open chromatin regions would be more efficiently formed when their distances are more than the length of four nucleosomes. Some of the sites seem to be used as multiple sites for chromatin organization, as observed in the estrogen receptor gene. We are currently investigating chromatin structure at the replication origin based on the alignment of bend sites to examine the relationship of these sites with DNA replication. Our results indicated that periodic bent DNA is a key element of chromatin structure and plays a role in various biological reactions.

References

Blaho, J. A. and Wells, R. D. (**1989**) Left-handed Z-DNA and genetic recombination. **Prog. Nucl. Acid Res. Mol. Biol**. 37, 107-126.

Calladine, C. R., Drew, H. R. and McCall, M. J. (**1988**) The intrinsic curvature of DNA in solution. **J. Mol. Biol**. 201, 127-137.

Crossley, M. and Orkin, S. (**1993**) Regulation of the -globin locus. **Curr. Opin. Genet. Dev**. 3, 232-237.

Crothers, D. M., Haran, T. E. and Nadeau, J. G. (**1990**). Intrinsically bent DNA. **J. Biol. Chem**. 265, 7093-7096.

Evans, T., Felsenfeld, G. and Reitman, M. (**1990**) Control of globin gene transcription. **Ann. Rev. Cell Biol**. 6, 95-124.

Felsenfeld, G. (**1993**) Chromatin structure and the expression of globin-encoding genes. **Gene** 135, 119-124.

Gross, D. S. and Garrard, W. T. (**1988**) Nuclease hypersensitive sites in chromatin. **Ann. Rev. Biochem**. 57, 159-197.

Hagerman, P. J. (**1990**). Sequence-directed curvature of DNA. **Ann. Rev. Biochem**. 59, 755-781.

Ioshikhes, I., Bolshoy, A., Derenshteyn, K., Borodovsky, M. and Trifonov, E. N. (**1996**) Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences. **J. Mol. Biol**. 262, 129-139.

Khan, J. D. and Crothers, D. M. (**1992**). Protein-induced bending and DNA cyclization. **Proc. Natl. Acad. Sci. USA** 89, 6343-6347.

Kuwabara, K., Wada-Kiyama, Y., Sakuma, Y. and Kiyama, R. (**1997**) Multiple interactions of periodic bent DNA in the promoter region of the human estrogen receptor gene with the nuclear scaffold, core histones and nuclear factors. submitted.

Nash, H. A. (**1990**) Bending and supercoiling of DNA at the attachment site of bacteriophage lambda. **Trends Biochem. Sci.** 15, 222-227

Ohki, R., Hirota, M., Oishi, M. and Kiyama, M. (**1997**) Conservation and continuity of periodic bent DNA in genomic rearrangements between the c-*myc* and immunoglobulin heavy chain μ loci. submitted.

Rich, A., Nordheim, A. and Wang, A. H.-J. (**1984**) The chemistry and biology of left-handed Z-DNA. **Ann. Rev. Biochem.** 53, 791-846.

Slightom, J. L., Blechl, A. E. and Smithies, O. (**1980**). Human fetal ^{G}- and ^{A}-globin genes: Complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. **Cell** 21, 627-638.

Soyfer, V. N. and Potaman, V. N. (**1996**) *In* **Triple-Helical Nucleic Acids.** Springer Verlag, New York.

Specer, C. A. and Groudine, M. (**1991**). Control of c-*myc* regulation in normal and neoplastic cells. **Adv. Cancer Res.** 56, 1-48.

Stamatoyannopoulos, G. and Nienhuis, A. W. (**1993**) Hemoglobin switching. *In* **The molecular basis of blood diseases**, Stamatoyannopoulos, G., Nienhuis, A. W., Majerus, P. and Varmus, H. (eds). W. B. Saunders, Philadelphia. pp107-155.

Travers, A. A. (**1989**) DNA conformation and protein binding. **Ann. Rev. Biochem**. 58, 427-452.

Trifonov, E. D. (**1991**) DNA in profile. **Trends Biochem. Sci**. 16, 467-470.

Wada-Kiyama, Y. and Kiyama, R. (**1994**). Periodicity of DNA bend sites in the human -globin gene region: Possibility of sequence-directed nucleosome phasing. **J. Biol. Chem**. 269, 22238-22244.

Wada-Kiyama, Y. and Kiyama, R. (**1995**). Conservation and periodicity of DNA bend sites in the human -globin gene locus. **J. Biol. Chem**. 270, 12439-12445.

Wada-Kiyama, Y. and Kiyama, R. (**1996a**) Conservation and periodicity of DNA bend sites in eukaryotic genomes*.* **DNA Res.** 3, 25-30.

Wada-Kiyama, Y. and Kiyama, R. (**1996b**) An intrachromosomal repeating unit based on DNA bending. **Mol. Cell. Biol.** 16*,* 5664-5673.

Werner, M. H., Gronenborn, A. M. and Clore, G. M. (**1996**) Intercalation, DNA kinking, and the control of transcription. **Science** 271, 778-784.

Wu, H.-M. and Crothers, D. M. (**1984**). The locus of sequence-directed and protein-induced DNA bending. **Nature** 308, 509-513.

Zahn, K. and Blattner, F. R. (**1987**). Direct evidence for DNA bending at the lambda replication origin. **Science** 236, 416-422.