DNA Bend Sites in the Human β-Globin Locus: Evidence for a Basic and Universal Structural Component of Genomic DNA

Yuko Wada-Kiyama,* Kiyoko Suzuki,* and Ryoiti Kiyama†

*Department of Physiology, Nippon Medical School, Tokyo, Japan; and †National Institute of Bioscience and Human Technology, Ministry of International Trade and Industry, Ibaraki, Japan

Here we summarize the DNA bend sites in a 66-kb region of the human β-globin locus. A total of 98 sites were mapped by circular permutation assay along the locus with an average interval of 679.2 ± 229.6 bp between them. The distribution of the bend sites indicated that although the most frequent distance was about 650–700 bp, there appeared to be preferences at 300–400, 500–550, 800–850, 1,000–1,050, and 1,150–1,200 bp, indicating that these distances are multimers of a 170-bp basic unit. DNA bend sites in the globin-encoding regions indicated that most of their locations relative to the cap sites were conserved during evolution. Insertion of Alu and L1 sequences that occurred at various times and changed the distances of the sites was corrected for the e-, gβ-, and δ-globin genes.

The only exception of the conservation was observed at the duplication junctions of the two γ-globin genes, which occurred 25–35 MYA. Among the 75 A/A/A (A2N1A2N1A2) sequences found in the 51 bend sites, 59 sequences from 47 sites showed bending profiles by oligonucleotide-based assay. All of these sites were included in the sites predicted by computer analysis based on the distribution of AA and TT dinucleotides. These lines of evidence suggest that these DNA bend sites are one of the basic structural components universally present in genomic DNA.

Introduction

Methods for DNA structural analyses including X-ray crystallography, NMR spectroscopy, and other techniques have improved the accuracy of structural information of longer DNA (Olson 1996; Widom 1997; Schurr et al. 1997). DNA structures formed by specific DNA sequences appearing in functional regions such as the promoter regions of genes and origins of replication are important in vivo for recognition by trans-acting factors. However, long-range coordinations between functional regions, between the enhancer elements and the cognate promoters for example, require more experimental evidence to determine the mechanisms of their interactions. For this purpose, the structural bases of such coordinations and not just the factors involved should be more extensively investigated. Signals at one site can be transmitted to the other sites by allosteric transitions in DNA upon binding of proteins (Schurr et al. 1997) or by deformations of DNA at specific secondary structures (Olson 1996). Another line of evidence to support this argument is that the sequences between separated elements adjust the translational and rotational positions (Drew and Calladine 1987; Pina et al. 1990). Nucleosome phasing is involved in this long-range coordination through positioning of regulatory elements (McPherson et al. 1996; Wong et al. 1997), although much of the mechanism still remains unclear. To obtain structural evidence, we searched for basic structural components universally present in genomic DNA. We previously proposed that DNA bend sites that appear periodically (periodic bent DNA) are associated with long-range coordinations (Ohki et al. 1998).

Periodicity of DNA bend sites was reported first in the human e-globin gene region and subsequently in other regions of the same locus and in the mouse β\textsuperscript{maj}-globin gene, as well as in human c-myc, immunoglobulin heavy chain μ, and estrogen receptor genes (reviewed in Kiyama 1998). The average interval between the neighboring bend sites is approximately 680 bp, suggesting that a novel repeating unit is present in the genomic DNA. To address the biological significance of periodic bent DNA, we focused on the interaction between bent DNA and the core histones. Since the core histones have affinity to curved DNA and there is a rotational preference for AA and TT dinucleotides in the nucleosomal phase which would cause DNA bending, it is natural to suspect that periodic bent DNA should be involved in this nucleosomal phase as a signal. Evidence obtained in the human g\textsuperscript{α}-\textsuperscript{γ}-\textsuperscript{β}-globin gene region suggested that some of the bending sequences in periodic bent DNA have higher affinity to the core histones than do nonbending sequences with similar base compositions (Wada-Kiyama and Kiyama 1996b; Onishi, Wada-Kiyama, and Kiyama 1998).

The human β-globin locus contains five active genes (e-, g\textsuperscript{α}-, g\textsuperscript{β}-, δ-, and β-globins) and β-globin pseudogene (ψβ-globin) in a region larger than 44 kb (Stamatoyannopoulos and Nienhuis 1993). Each gene has three exons but occupies only 1.5 kb. In the non-coding duplicated regions, similarity has been mostly lost, and insertions of Alu and L1 repetitive sequences have occurred since their separation, resulting in a total disruption of the original distances between the conserved patches of the nucleotide sequences. However, these genes show marked coordination when they switch expression during development and differentiation. This requires strict control of the transcription machinery, including RNA polymerases and the interactions between the promoters located as far apart as 42 kb. The locus control region that governs this coordination is located from 6 to more than 20 kb upstream of the first gene, the e-globin gene (reviewed in Stamatoyannopoulos and Nienhuis 1993; Hardison et al. 1997). Therefore, the regulation of expression of these genes needs some coordination or interaction between their regulatory re-
regions and between each gene and the locus control region. We show here a conservation of bent DNA structure among globin gene members other than the conserved motifs of transcription factors.

Materials and Methods

Circular Permutation Assay

DNA bend sites were determined by the circular permutation assay, as described previously (Wada-Kiyama and Kiyama 1994). DNA fragments of 500 bp to 1 kb were cloned into the pBluescript vector as tandem duplicates. Each clone was digested with the restriction enzymes that cut only once in the fragment, thus producing the same size fragments but in the permuted orders of the sequence, and the fragments were then resolved by 8% polyacrylamide gel (29:1 mono:bis-acrylamide) electrophoresis at 4°C for 2–3 days at 1 V/cm. M13mp18 digested with PvuII or AvaI was used as an internal migration marker. The results of the assay were analyzed by calculation of relative mobilities of each fragment to the fastest-migrating fragment, summarized in figure 1. Bent DNA should be located close to restriction site, where the relative mobility is highest, based on the phenomena that when the bent DNA is located close to either end of the fragment, its migration within the gel matrix is unaffected, otherwise showing retardation. The degrees of retardation depend on the relative distance from the bent DNA. When two bend sites were located within a cloned fragment, the resultant graph showed two peaks at different positions or an apparent deformation of the graph. In such cases, the regions were further divided to separate the sites. Although most of the bend sites were used for statistical analysis, only two sites, δB-1′ and βB-2′, were excluded because they appeared to be exceptional in the alignments of the bend sites among the globin genes. However, this exclusion does not affect the conclusions of our analysis. Complete maps of the clones and the positions of the bend sites can be provided on request.

Oligonucleotide-Based Bending Assay

Oligonucleotides were unidirectionally ligated and electrophoresed as previously described (Wada-Kiyama and Kiyama 1995). The A/A/A or complementary T/T/T sequences found in the bend sites and used in the assays were 22-nt sequences starting from nucleotides 402, 505, 3523, 6836, 9920, 10884, 10972, 11607, 12058, 12194, 12204, 12573, 13571, 13611, 14451, 14551, 15959, 17858, 20761, 21425, 21524, 22657, 23841, 24326, 25552, 25639, 25661, 25706, 25726, 26714, 26756, 27339, 27400, 28057, 28153, 28499, 29799, 32104, 32133, 33316, 33342, 33840, 36472, 37249, 37955, 37970, 38267, 38374, 42659, 44676, 44729, 46561, 49827, 50539, 50594, 53589, 53614, 55326, 55366, 56678, 57494, 57888, 57991, 58824, 59068, 59611, 59679, 59730, 61109, 61836, 63033, 63187, 65157, 65184, 65520, and 66291 in the DDBJ/GenBank/EMBL database (entry name HSHBB). The first 20-nt sequences on both strands were used to make the oligonucleotides, allowing unidirectional ligation of the oligonucleotides. The last AA or TT dinucleotides of the 22-nt sequences were supplied by the first AA or TT after ligation, reproducing the complete 22-nt sequences. Fragments of 100 bp, 5mers of the oligonucleotide unit, were used to calculate migration relative to the control (A_20\cdot T_20)_n (Koo et al. 1986).

Prediction of Bend Sites by TRIF Analysis

DNA curvature was analyzed by using TRIF 1.00 software, which calculates the angle of duplex axis and DNA curvature in the curvature unit based on the distribution of dinucleotides (Shigelman, Trifonov, and Bolshoy 1993). One curvature unit is the value obtained for the mean DNA curvature in the crystallized nucleosome. For globin genes, a smoothing window of 50 bp in the scanning window of 200 bp was applied.

Results

Mapping of DNA Bend Sites in the Human β-Globin Locus

We used the circular permutation assay (Wu and Crothers 1984) for mapping DNA bend sites in the two regions, nucleotides 22000–33000 and 45000–60000 (fig. 1A and B). This assay is based on the physical interaction between DNA and the gel matrix which affects the migration of DNA fragments, and when the bend site is located at the center, DNA fragments show the highest degree of retardation. Therefore, restriction enzymes that cut close to the bend center would generate the fastest-migrating bands. Figure 1C summarizes the results of mapping of the DNA bend sites spanning 66 kb in the human β-globin locus. A total of 98 major and several minor or additional bend sites were mapped. The average interval of DNA bend sites in the locus was 679.2 ± 229.6 bp, which is close to the values previously obtained for subregions of the globin locus and other loci (Kiyama 1998), supporting the suggestion that this is the common feature of genomic DNA. We observed strict periodicity of the sites in the 5’ regions of the ε- and ψβ-globin genes (14–20 kb and 41–46 kb, see fig. 1C). In contrast, there were several regions that showed disruption of periodicity. These included the locus control region (first 14 kb), regions containing the β-like globin genes (19.5–21 kb for ε, 34.5–36 kb for γ, 39.5–41 kb for δγ, 45.5–47 kb for ψβ, 55–56.5 kb for δ, and 62–63.5 kb for β), the region containing L1 repetitive sequences (23–31 kb), regions containing duplication junctions for the γγ- and δγ-globin genes (33–34 kb and 38–39 kb), and the replication origin region between the δ- and ε-globin gene (57.5–59 kb).

The distribution of the distances between the neighboring sites (fig. 2A) indicated that, in addition to the highest peak at 650–700 bp, there were several other peaks, at 300–400, 500–550, 800–850, 1000–1050 and 1150–1200 bp. This and the 680-bp average suggested the presence of a unit length of 170 bp, which theoretically should have peaks at 340, 510, 680, 850, 1,020, and 1,150–1,200 bp. This and the 680-bp average indicated that, in addition to the highest peak at 650–700 bp, there were several other peaks, at 300–400, 500–550, 800–850, 1000–1050 and 1150–1200 bp. This and the 680-bp average suggested the presence of a unit length of 170 bp, which theoretically should have peaks at 340, 510, 680, 850, 1,020, and 1,190 bp. This was again observed in the distribution of the distances between any two of the bend sites (fig. 2B), which were generally observed at every fourth
Conservation of the DNA Bend Sites Among the β-like Globin Genes

We demonstrated that the DNA bend sites are conserved in the promoter regions of the β-like globin genes (Wada-Kiyama and Kiyama 1996b). Figure 3 shows that they are conserved even in the 3′ regions and introns among these genes. Since these globin genes separated as early as 150–200 MYA, it is surprising that their DNA structure, but not the nucleotide sequences, is conserved. Based on this observation and the structural characteristics, it is reasonable to suspect that these genes are a structural component with no coding function. An exception to this conservation was observed in the 5′ region of the δγ- and γγ-globin genes (GγB-2a, GγB-2b, AγB-2a, and AγB-2b). These globin genes were duplicated much later than the other members, and the recombination junctions are located in the LI sequences 1.5 kb upstream of the cap sites (33- and 38-kb regions). This genomic rearrangement seems to have caused this irregularity.

Sequence Features of the Bend Sites

To understand the mechanism of bending, we adopted two approaches: oligonucleotide-based assay and computer analysis. First, we examined potential bend core sequences with the consensus A/A/A (A2-N8-A2-N8-A2) by oligonucleotide-based bending assay (fig. 4). Oligonucleotides of 20mer sequences containing A/A/A were unidirectionally ligated to form multimers of each 20-bp unit. When bent DNA was included, deviation from the standard (A20-T20) became significant as the lengths of the ligation products increased. For the standard oligonucleotides showing a bend angle of approximately 40°/20 bp (A3N7), 7.2% retardation was observed with 5mer oligonucleotides (100 bp in length). Among the total of 75 A/A/A sequences found in the 51 bend sites, 19 sequences from 17 sites showed at least 7.2% retardation and 59 sequences from 47 sites showed at least 2.7% retardation, shown by a standard A2N8, which corresponded to approximately 28°/20 bp.

170-bp division (indicated by the arrows in the figure), i.e., every 680 bp in total. This indicated that, on average, every fourth 170-bp division in genomic DNA is marked with DNA bending. It should be noted that the 680-bp repeat was disturbed when the lengths reached approximately 3,000 bp (or four repeats). However, the periodicity again became significant from 8 to 11 kb, suggesting that there is a long-range coordination (see Discussion).
(data not shown). There were four sites, \(\epsilon B - 24\), \(\epsilon B - 15\), \(\epsilon B + 15\), and \(\delta B + 4\), for which the A/A/A sequences did not show the bending profile (less than 2.7% retardation).

Bendability was examined using TRIF 1.00 software, which was developed to predict DNA curvature based on the distribution of AA and TT dinucleotides as major determinants (Bolshoy et al. 1991; Shpigelman, Trifonov, and Bolshoy 1993). These dinucleotides provide the sharpest wedge angles and contribute most to bent DNA when they appear on one side of the double helix, and therefore they are spaced at intervals of multiples of 10.5 bp. As shown in figure 5, the peaks indicated by the software were generally included in those determined experimentally. Almost all of the 98 sites (marked in the figure) were distinguished from the nearby regions by these analyses.

**Discussion**

Periodicity of DNA Bend Sites in the Human \(\beta\)-Globin Locus

We summarized here the DNA bend sites appearing in the human \(\beta\)-globin locus. A total of 98 sites were mapped according to the circular permutation assay, which showed 5%–20% retardation of the DNA fragments containing the sites upon electrophoresis at 4°C. When these fragments were electrophoresed at 55°C (data not shown), or fragments without the bend sites (pBMH23, p\(\gamma\)EBY8, and p\(\beta\)R65 for example; fig. 1) were used, almost no retardation was observed. However, as predicted by computer analysis (fig. 5), there are many sites with various degrees of bending. Therefore, local DNA bending could be canceled by major bent DNA or by other local bending, and here we observed the average of the bending profile and/or the major bend sites that were not canceled. However, these experimentally determined bend sites should have some biological significance because (1) the distances between the neighboring sites showed a preference at several lengths, particularly at 680 bp (fig. 2A); (2) these preferential lengths were multiples of a 170-bp unit; and (3) arrays of the sites with intervals of 680 bp are disrupted at functionally important regions, including the coding regions. As suggested above, we believe that this function is related to nucleosomes for the following rea-
FIG. 2.—Distribution of DNA bend sites. A, Distances between the neighboring sites. The distances calculated between the nearest restriction sites in each bend site are shown in the graph. The highest peak and the minor peaks (including the shoulder at 300–400 bp) are indicated by solid and hatched arrows, respectively. B, Distances between any two of the bend sites. Peaks corresponding to multiples of 700 bp are indicated by arrows. The lower peaks in the region between 3 and 8 kb are indicated by hatched arrows.

FIG. 3.—Conservation of bend sites among the β-like globin genes. Five active genes (ε, ξγ, ηγ, δ, and β) and one pseudogene (εβ) were aligned to examine the conservation of the bend sites. The regions showing conservations are shaded, and the recombination junctions in the ξγ- and ηγ-globin genes are indicated by arrows.

Sequence Features of the Bend Sites

As observed previously, the major sequence elements that appear in the experimental bend sites are adenine or thymine tracts longer than two nucleotides (Wada-Kiyama and Kiyama 1994, 1996a). We observed here that A/A/A sequences appeared in the 51 bend sites and the A/A/A sequences from 47 sites showed at least 2.7% retardation for 5mer oligonucleotides (fig. 4). Furthermore, 79 of the 98 sites contained sequences of AₙNₘAₙNₘAₙ (n and m = 7, 8, or 9) (data not shown), which would cause bending. Therefore, the major sequence features that contributed to bending here were the dinucleotides AA and TT. These dinucleotides provide higher wedge angles, and when they are localized on one side of the double helix by appearing nearly every 10.5 bp, or multiples of this value, a higher degree of DNA curvature is observed. This should be the case for most, if not all, of the bend sites shown in the β-globin locus. TRIF 1.00, used here, or the previous version, CURVATURE, have been used for detection of bent DNA in the promoter region of the human c-myc and cdc2 genes (Nair 1998; Ohki et al. 1998). In both regions, bent DNA was located within the regions of a gradual curvature.

All the sites contained regions that showed curvature units of more than 0.04 in the scanning window of 20 bp (data not shown). Furthermore, under appropriate smoothing conditions, higher peaks can be observed in the subkilobase-level window that corresponds to the experimental bend sites (fig. 5). However, all peaks did not correspond to the experimentally determined sites.
Assay with the oligonucleotides containing A/A/A sequences. A total of 75 20mer oligonucleotides containing A/A/A sequences were assayed by 8% polyacrylamide gel electrophoresis at 4°C for 20–23 h. The oligonucleotides showing 7.2% retardation, as shown with A3N7 (Calladine, Drew, and McCall 1988), or 5.0% or 2.7% retardation for 5mers (100 bp) were labeled “+”, “+,” and “-,” respectively. Those showing less than 2.7% are labeled “...” The nucleotide sequences of the markers are: A6N4, (AAAAACG GCC); A5N5, (AAAAACG GCC); A4N6, (AAAAACG GCC); A3N7, (AAACCGGGCC); A3N7A10, AAACCGGGCCAAAGAAAA; A2N8, (AACCGGGCC).
Fig. 5.—Prediction of the bend sites by computer analysis. DNA bend sites were predicted by TRIF software with 200-bp scanning and 50-bp smoothing windows. Corresponding peaks in the curvature mapping are marked by dots.
Apparently, some of the predicted sites were not revealed by the gel assay. This indicated that mathematical smoothing of DNA curvature is different from experimental smoothing, which should include the interaction of DNA with the gel matrix. For example, changing smoothing conditions sometimes reverses the peak heights (data not shown), although the same peak positions are retained. Effective lengths of DNA fragments on retardation in the gel should be dependent on the size and density of the gel matrix, thereby changing the effect of DNA curvature below certain lengths, while the mathematical smoothing consists of just averaging all minor and major curvatures. Such two-phase behaviors of DNA within the gel matrix have been reported by Calladine et al. (1991). Another explanation of the differences might be provided by the superhelicity formed by the actual DNA fragments, which might add a rigidity to DNA in the gels, which is neglected in the mathematical smoothing.

Conservation and Maintenance of DNA Bend Sites

We observed here that dramatic reorganization events of the locus, insertion of Alu or L1 sequences and duplication of the γ-globin region, did not change many of the locations of DNA bend sites (fig. 3). The insertion of Alu or L1 sequences should have added at least 300 bp to the original distances between εB-3 and εB-2, between δB-2 and δB-1, and between δB-3 and δB-2 after generation of the cognate genes, but these distances were adjusted later so that the periodicity maintained. The mechanism of conservation would be relocation of the sites either by deletion of the elongated distances, likely in the case of L1 sequences, or by mutations accumulated at the old and new sites, removing or creating the sites, but not likely by gene conversion because of lack of sequence homology among the sites. The only exception of conservation was in the case of the sites in the upstream regions of the γ-globin genes. Since the duplication of the γ-globin genes occurred recently (25–35 MYA; Fitch et al. 1990), more time would be needed to adjust their distances. Another possibility is that there is some reason for disruption only for the γ-globin gene, although this is less likely. So far, we have no idea as to the cause of this disruption, which was not located at the direct duplication break points (arrows in fig. 3).

Long-Range Coordination on the Genomic DNA

Distances between the neighboring bend sites indicated that sites as far apart as 8–10 kb showed preferred distances, while those separated by 3–8 kb showed less preference (fig. 2B). This indicated that a local periodicity has an influence on or has a relationship with the periodicity of the other part of the locus. We observed conservation of the relative locations of the bend sites among the six globin gene regions, which are separated by as much as 40 kb (fig. 3). However, the periodicity was disturbed at or within the various locations: the locus control region (Onishi, Wada-Kiyama, and Kiyama 1998), the first and second exons of the globin genes, the long L1 sequence, duplication junctions of the γ-globin genes, and the replication origin of the locus (see Results for details). As previously discussed (Wada-Kiyama and Kiyama 1996b), we believe that the disturbance in these regions is caused by the structural hierarchy of these functions over bent DNA or is the consequence of recent genomic rearrangements. Meanwhile, there are a total of 11 regions that contain at least three consecutive bend sites of the 680-bp periodicity (fig. 1C). If these 11 regions were placed independently, we would not see such a long-range preference. These lines of evidence suggest that there is a long-range coordination based on the DNA bend sites in genomic DNA. Although whether this structure is the basis for the long-range coordination of DNA–protein complexes on the genomic DNA is still under investigation, we believe that DNA bending is a basic and universal structural component of genomic DNA and has a direct or indirect influence on biological phenomena.

Acknowledgments

We thank Dr. Ed Trifonov for critical comments and Dr. T. Tsukahara for computer analysis. This work was supported by a grant from the Ministry of International Trade and Industry (to R.K.), a grant-in-aid from the Ministry of Education, Science, Sports and Culture of Japan (to Y.W.-K.), and grants from Asahi Foundation (to Y.W.-K. and R.K.).

Literature Cited


ANTONY DEAN, reviewing editor

Accepted March 4, 1999