博士論文

DNAの組み換えにおけるヌクレオソームの位相お
よびDNA湾曲部位の解析

Analysis of DNA recombination with respect to
nucleosome phasing and DNA bending

東京大学大学院 理学系研究科 生物化学 専攻

論文提出者　大木理恵子
指導教官　坂野　仁　教授

博士論文

DNAの組み換えにおけるヌクレオソームの位相および DNA湾曲部位の解析

Analysis of DNA recombination with respect to nucleosome phasing and DNA bending

論文提出者　大木理恵子

指導教官　坂野　仁　教授

Analysis of DNA recombination with respect to nucleosome phasing and DNA bending

1. Introduction

Recombination of genomic DNA is of profound significance not only for the generation of diversity of immunoglobulins but also for the maintenance of genome diversity and molecular evolution of genes. Rare and illegitimate recombination events could lead to cellular transformation, as seen in the lymphoma cells in which c-*myc* oncogene is juxtaposed to one of the immunoglobulin loci, causing deregulation of transcription of the c-*myc* gene (1-3). Mechanisms of homologous recombination, site-specific recombination, transpositional recombination and illegitimate recombination have been the subjects of extensive studies in the last decades (4, 5). These studies have been mainly carried out by employing genetic engineering methodologies to elucidate molecular mechanisms of recombination *in vivo* as well as *in vitro*. Recently, mechanisms of recombination have been discussed with respect to dynamic behavior of the chromatin structure (6, 7). As the studies on transcription have been focusing upon the chromatin structure (8, 9), it is reasonable to assume that studies on the molecular mechanism of recombination may lead to new insights if we take the dynamic behavior of the chromatin into account. In this study, I have investigated possible involvement of the chromatin structure, particularly nucleosomal structure in the process of genetic recombination in human cells.

1-1. Recombination in higher eukaryotes

Recombination in higher eukaryotes have been studied in a number of experimental systems. Among these, somatic recombination has been described in detail for the immunoglobulin and T cell receptor loci (10-

1

13). There exist organized and elaborate mechanisms in order to ensure cell type specific and developmentally regulated recombination. The rearrangement of immunoglobulin genes starts with recombination in the variable region, known as V-(D)-J recombination, where the nonamer and heptamer sequences act as recombination signal sequences. In this process, it is known that the recombination activating genes, RAG-1 and RAG-2, are involved and the regulated expression of these genes correlates perfectly with the V-D-J recombination activity (14). It has also been reported that a cell-type specific chromatin structure is required for the V-D-J recombination (15). For example, accessibility to the recombination signal sequence within the chromatin structure changes during developmental stage of B cells, which is followed by the activation of the V-D-J locus for recombination. After V-(D)-J recombination is completed, switch recombination of the constant region occurs by rearranging S regions (sequence features of S region is shown in Table 1). Despite these extensive studies, however, several questions, for example, what is the molecular mechanism causing the change in the chromatin structure and how allelic exclusion proceeds, are still open for discussion.

Another type of recombination in higher eukaryotes that has been extensively studied is homologous recombination. RAD51, a eukaryotic homologue of RecA, and meiosis-specific DCM1 have been reported to be responsible for homologous recombination in higher eukaryotes (16, 17). They were first found in yeast as proteins that share similarities both in structure and function with RecA. Although it has been established that these proteins are involved in the search for homologous sequences in the early stage of recombination process and are subsequently responsible for DNA double-strand exchange, the entire process of homologous recombination in the higher eukaryotes still

Table 1. Nucleotide sequences of repeating units in switch regions.

Switch sequences located on the 5' region of the CH genes consist of long stretches (ranging from 1 kb to 9 kb) of simple repetitive sequences.

S region	Prevalent unit sequence[a]
Sμ	*GAGCTGAGCTGGGGTGAGCT*
Sγ3	*P*GG*ACCAGGCTGGGACAGCTCTGGG* *GAGCTGGGGTGGGTGGG*AGTGTGG
Sγ1	*G*PPTCCAGG*CTGAGCCAGCT*ACAGG*G* *GAGCTGGGG*YAPP*TGGG*APTP*TPG*
Sγ2b	GGGACCAG(T/A)CCTAG*CAGCTP*T*GGGG* *GAGCTGGGG*A(A/T)GG*TPGG*APTPTGA
Sγ2a	*GGG*ACCAGGCAGTA*CAGCTCTGGG*T *P*GGG*P*NCAGGCAGTA*CAGCTCTGNGTG*
Sε	*GGGCTGGGCTGAGCTGP*GCTGA *GCTGP*GCTGAGCTGPPNT
Sα	AT*GAGCTGGG*AT*GAGCTGAGCTA*GGCT GGAATAGG*CTGGGCTGGG*CTGGTGTGA *GCTGGG*TT*AGGCTGAGCTGAGCTGG*A
common sequences	*(G)AGCT(G), TGGG(G)*

[a]Common sequences are in italic. P; purine, T; pyrimidine, N; any nucleotide.

remains to be elucidated.

Much less is known for other types of recombinations in higher eukaryotes. It has been supposed that translocation involving c-*myc* (located on chromosome 8q24) and one of the immunoglobulin loci (Igκ gene on chromosome 2q12, IgH on 14q32 and Igλ on 22q11) depends upon the B cell-specific recombination mechanisms (18, 19). The breakpoints on chromosome 8 are located upstream, downstream or within the c-*myc* gene, and those on chromosome 14 are in the switch regions, J regions or in the upstream of J regions. The mechanism of how the breakpoints are scattered over such different chromosomal regions was not well documented. Furthermore, more than one mechanism seem to take place in this process. Many events have also been advanced for the deregulation of the expression of the c-*myc* gene, which occur in all the cases so far studied (20). Involvement of the immunoglobulin heavy chain enhancer located 5' to the Sμ region or a locus control region located 3' to the Cα gene, have been reported to be responsible for the translocation (21, 22).

Translocations and chromosomal aberrations (including amplifications, deletions and LOH) observed in many locations on the chromosomes have also been studied with respect to tumorigenesis (23-26). Chromosomal gaps or breaks, known as the fragile sites, have been reported to be involved in the generation of human malignancy such as leukemias, lymphomas and malignant solid tumors (27, 28). These sites tend to be located at the junctions of Giemsa-negative and positive bands where specific T-rich sequences are often located. The molecular mechanism leading to the formation of the fragile site seems to be closely related to the chromatin structure.

Extrachromosomal (EC) DNA have also been studied as a model for more frequent recombinational events as they exist in a considerable

copy number in cells (29-32). The recombination junctions for the formation of EC DNA have been studied in detail (33-35).

1-2. Recent studies on the chromatin structure

Owing to the recent progress in the research of the chromatin structure, it has become more feasible to study biological events associated with a complex chromatin structure under more natural environment. Chromatin comprises multiple levels of compaction of nuclear structures. The most basic unit of compaction that leads to the macroscopic nuclear architecture is nucleosomes (36-38). DNA of 146 bp long is wrapped in 1.75 left-handed turns around an octamer of core histones (consisting of a tetramer of H3 and H4, and two dimers of H2A/H2B) to form a nucleosome core. Upon formation of a nucleosome, the double helical structure of DNA is known to change (39-43). Usually, DNA has a pitch of 10.5 bp per helical turn in solution, but within a nucleosome, the pitch is 10.2 bp on average. Furthermore, DNA within a nucleosome is divided into two distinct regions that have pitches of 10.0 bp or 10.7 bp per turn, respectively. The junction between the two different structures is about 60 bp away from the core-linker junction and has the most severe distortion within the nucleosome. The molecular structure of DNA within a nucleosome differs where on the nucleosome DNA is associated to.

The mechanism of nucleosome phasing (or positioning) has been drawing a considerable attention among molecular biologists with an expectation that it may play an important and critical role in gene expression and recombination. Nucleosome phasing can be induced by specific DNA sequences (44-46). For example, in the 5S ribosomal RNA gene, a periodic sequence exists that confers DNA an intrinsic curvature, and this curved DNA is favored to be incorporated into

nucleosomes (47, 48). On the other hand, the rigid structure of poly(dA)·poly(dT) sequences and left handed Z-DNA are not favored for incorporation into the nucleosomes (49, 50). The intrinsic DNA structure formed by these DNA sequences affects nucleosome phasing which are subsequently recognized by H3/H4 tetramer (51).

The understanding of chromatin compaction relies mostly on the studies on the nucleosomal level and not much is known for the higher levels of the structures. Nucleosome arrays are constrained by linker histones (H1 or H5) (52, 53) and further compaction occurs, by an irregular packing of the nucleosome arrays, to form a chromatin fiber (30-nm fibre), finally leading to formation of a chromosome (54). Several specific nuclear structures are involved in this process. These include MARs (matrix-attachment regions) (55) or SARs (scaffold-attachment regions) (57), which divide chromosomes into distinct regions, and periodically-appearing DNA bend sites as a possible signal for nucleosome phasing (58-60).

1-3. DNA bend sites and the chromatin structure

One of the candidates that are responsible for the compaction of chromosomes in the nucleosomal level is DNA bending. DNA bending is classified into two types. One is the bending caused by protein-DNA interaction (61) and the other is the intrinsic sequence-directed bending. One of the well-known examples for DNA bending caused by protein binding is the one caused by the TATA-box binding protein (62). This protein generally binds 30 bp 5' to the transcription start site of genes in mammals. Binding of the protein causes bending in DNA molecules by an angle of approximately about 90°, which presumably facilitate interactions with other proteins to form a multi-protein complex. The intrinsic DNA bending may also have the same function, as it is

suggested by that some of the proteins which cause DNA bending could be functionally replaced by specific sequences which cause intrinsic DNA bending (63). Intrinsic DNA bending is also important for the chromatin structure. DNA bend sites are known to have a high affinity to histone cores, and furthermore, chromatin assembly proteins such as high mobility group (HMG) proteins (64) and SAR binding proteins bind to the bend sites. On the other hand, there are several reports that intrinsic bending is associated with recombination. For example, DNA bend sites are found within 1 kb of recombination junctions in many illegitimate recombinations (65, 66), and HIV and MLV retroviruses favor to integrate at bend sites (67).

Although intrinsic DNA bending is caused by specific DNA sequences, the specific sequences which cause DNA bending are not yet fully understood. Sequence including short poly(dA) tracts at four to ten base intervals exhibits extensive bending but there are bend sites without such sequences (68). Thus, it is difficult to predict the location of DNA bend sites by DNA sequences alone. Kiyama *et al.* examined the human β-globin gene locus for DNA bend sites by the circular permutation assay developed by Wu and Crothers (58), and found that the bend sites appeared periodically at an interval of 680 bp on average, which is roughly four nucleosomes long (59). These bend sites were also conserved during molecular evolution of the globin genes and seem to play an important role in nucleosome phasing or chromatin folding (60).

1-4. Experimental approaches

The studies described above suggest that recombination is greatly affected by the chromatin structure like other biological reactions. To further examine and explore this possibility, I have investigated two

types of recombination, which differ each other in their aspects as shown below.

One is the recombination of the *Sau*3A family DNA which occurs at the excision of the DNA to form an extrachromosomal circular DNA (70). EC DNA is a good source for naturally occurring recombination products as described previously. Particularly, the *Sau*3A family DNA is quite useful because it is highly concentrated in the EC DNA fraction of HeLa cells (approximately 1% of this family DNA is present extrachromosomally as circular DNA). The *Sau*3A family has an 849 bp unit sequence which is divided into five homologous subunits of ~170 bp in length. From the previous study of seven recombination junctions including two extrachromosomal and five chromosomal recombinants, it has been suggested that recombination for circularization occurs between any of the two subunits with a homology of 68.4 to 77.2 % and does not require a long stretch of homology (71).

The other recombination that I describe here is the translocation between the c-*myc* and the immunoglobulin loci in lymphoma cells (1-3, 20). This type of recombination is different from that in EC DNA in the stability of the recombination products. The translocated chromosomes are usually stably maintained in the cells after cellular transformation. Approximately 90% of the translocation events observed in Burkitt's lymphoma take place between chromosomes 8 and 14 (20). In Fig.1, the translocation sites studied at the DNA sequence level in the c-*myc* gene and the immunoglobulin J3 to Sμ regions are summarized (72-76).

In this thesis which aims to elucidate the possible relationship between the chromatin structure and recombination, I especially emphasized nucleosome phasing because it is a very essential step of chromatin assembly. Furthermore, I examined DNA bend sites in the

*Sau*3A family, the c-*myc* and the immunoglobulin loci, which would provide basic information of nucleosomal phasing.

Fig. 1. Transcription view of mammalian cell lines at the c-*myc* gene region and the immunoglobulin κ locus...

Fig. 1. Translocation sites of lymphoma cell lines in the c-*myc* gene region and the immunoglobulin μ locus.

The translocation sites are indicated by arrows and their positions (relative to the P2 cap site in the c-*myc* gene region and, for the immunoglobulin locus, the positions are arranged so that the J6 coding regions are from 1482 to 1541) are shown next to their names. The names of the cell lines are indicated only for those described in this thesis.

2. Materials and Methods

2-1. Reagents

Restriction enzymes were purchased from Takara or New England Biolabs. All other reagents without reference were the reagents grade and purchased from Wako Pure Chemical.

2-2. Oligonucleotides

Oligonucleotides were synthesized using a Milligen Cyclone plus DNA synthesizer by the beta-cyanoethyl-phosphoramidite method and purified with Milligen Oligo-Pak columns.

The oligonucleotides used for amplification of chromosomal and EC DNA are as follows:

S-T, GATCCGCGGCCGCCCGAT

S-G, ATCGGGCGGCCGCG

B-C, GATCCGCGGCCGCCCGTC

B-G, GACGGGCGGCCGCG

The *Sau*3AI adaptor, S-TG, was made by annealing S-T and S-G, and the *Bfa*I adaptor, B-GC, by annealing B-G and B-C.

The *Sau*3A family-specific primers used to clone recombination junctions are as follows:

TTCCTCTTGACAGAGCAGCTCT and

AGTTCAATTCTTGAAGTGGAAC for PCR1,

GAAAGTGGGTTTGGAAACTGCG and

AACCTTGCTTTCATAGTTCAGC for PCR2, and

GTGGTGGAAAAGGAAAATCC and TTCAAAACTGCTCCATCAGC

for PCR3.

11

Oligonucleotides used for inverse PCR to determine the *in vivo* nucleosomal phases are as follows:

c-*myc* sense-1; 5' -TCG CCT GTG TGA GCC AGA TCG- 3'
c-*myc* sense-2; 5' -AGC CAG ATC GCT CCG CAG CC- 3'
c-*myc* reverse-1; 5' -TAT GCG GTC CCT ACT CCA AG- 3'
c-*myc* reverse-2; 5' -CCA AGG AGC TCA GGA TGC AAG- 3'
Igmu sense-1; 5' -TGA TGG AGT AAC TGA GCC TG- 3'
Igmu sense-2; 5' -CTT GGG GAG CCA CAT TTG GAC- 3'
Igmu reverse-1; 5' -CAA CCT GAG TCT CCA TTT TC- 3'
Igmu reverse-2; 5' -CTC CAT TTT CCA AAG GCA TCG- 3'

Double-stranded oligonucleotides used for bending assay are as follows: Thrity one-nucleotide long oligonucleotides SAUB-1s, SAUB-2s, SAUB-3s, SAUB-4s, and SAUB-5s were annealed with the complimentary oligonucleotides SAUB-1a, SAUB-2a, SAUB-3a, SAUB-4a, and SAUB-5a, respectively, which were designed to create two-base 5'-protruding ends. Thrity-nucleotide long oligonucleotides LA and RA were also annealed with the complimentary oligonucleotides LT and RT, respectively, which were designed to create two-base 5'-protruding ends. $(dA)_{31}$ or $(dA)_{30}$ were annealed with $(dT)_{30}$ or $(dT)_{31}$, respectively.

SAUB-1s; CTGTCCTCCAAAAGGAAAATCCTTCACAAAA
SAUB-2s; CTACTGTAGTAAAGGAAATAACTTCATCTAA
SAUB-3s; TTTCGTTGCAAACGGGATAAACTTCCCAGAA
SAUB-4s; CTTCCTTCGAAACGGGTATATCTTCACATCA
SAUB-5s; TTTCGTTGGAAACGGGTTCATCTTCACAGAA
SAUB-1a; GATTTTGTGAAGGATTTTCCTTTTGGAGGAC
SAUB-2a; GATTAGATGAAGTTATTTCCTTTACTACAGT

SAUB-3a; AATTCTGGGAAGTTTATCCCGTTTGCAACGA

SAUB-4a; GATGATGTGAAGATATACCCGTTTCGAAGGA

SAUB-5a; AATTCTGTGAAGATGAACCCGTTTCCAACGA

LA;GGAGGGAGGGATCGCGCTGAGTATAAAAGC

LT; CCGCTTTTATACTCAGCGCGATCCCTCCCT

RA; GTATAAAAGCCGGTTTTCGGGGCTTTATCT

RT; ACAGATAAAGCCCCGAAAACCGGCTTTTAT

2-3. DNA Isolation and Purification

EC DNA was prepared from human (HeLa) cells by the method described by Hirt. Briefly, cells (6 x 10^9) were washed 3 times with phosphate-buffered saline and suspended in 120 ml of 10 mM Tris-HCl, 10 mM EDTA (pH 8.0). After the cells were lysed with 3.6 ml of 10 % SDS and 32 ml of 5 M NaCl, the sample was incubated overnight at 4 °C, and centrifuged at 30 krpm for 2 hr at 4 °C. The supernatant was treated with proteinase K (final concentration 100 µg/ml) for 2 hr at 37 °C, followed by phenol extraction, chloroform/isoamylalcohol extraction and ethanol precipitation. The samples were further purified by three successive CsCl/ethidium bromide centrifugations at 55 krpm for 19 hr at 22 °C, to completely remove chromosomal DNA from the covalently closed circular DNA fraction. Circular mitochondrial DNA was removed by linearization with *Pac*I, followed by treatment with *Micrococcus luteus* ATP-dependent deoxyribonuclease (0.0175 units/µl) for 1 hr at 37 °C. Chromosomal DNA was prepared from the pellets obtained by Hirt extraction by the conventional phenol method.

2-4. PCR Amplification of EC DNA

EC DNA was first digested with *Sau*3AI or *Bfa*I. Then, their 3'-ends were filled with dNTPs by the Klenow fragment, and *Sau*3AI

13

adaptor S-TG or *Bfa*I adaptor B-GC was ligated. The DNA fragments were amplified by PCR in 50 mM KCl, 10 mM Tris-HCl (pH 9.0), 0.1% (w/v) Triton X-100, 1.5 mM $MgCl_2$, 0.2 mM dNTPs and 0.2 mM of the primer S-T or B-C in a total volume of 100 µl with 0.05 units/µl of *Taq* DNA polymerase. The PCR program used was: 1 min at 92 ˚C for denaturation, 2 min at 50 ˚C for annealing and 3 min at 72 ˚C for extension in each amplification cycle, with a final extension at 72 ˚C for 10 min.

2-5. Cloning of Recombination Junctions from the Extrachromosomal *Sau*3A Family DNA

The *Sau*3A family DNA was amplified by PCR (PCR1, 2 and 3) with specific primers from EC DNA or chromosomal DNA digested with *Sau*3AI (PCR1 and 3) or *Bfa*I (PCR2) in advance. Therefore, only the fragments lacking the respective restriction site can be amplified, removing most of the canonical PCR products. The positions of the PCR primers and the canonical PCR products are illustrated in Fig. 2. The PCR conditions were the same as those described above. The fragments from EC DNA with sizes different from those of chromosomal DNA were recovered from a 1.5% agarose gel and cloned into the *Eco*RV site of the pBluescript SK(-) vector (Stratagene). The nucleotide sequences of the recombination junctions were determined by the dideoxynucleotide chain termination method using 24mer reverse sequencing primer #1233 (New England Biolabs) or 17mer sequencing primer KS (Stratagene).

2-6. Southern Blot Analysis

DNA samples were electrophoresed in a 1.5% agarose gel and transferred onto Hybond-N+ nylon membranes (Amersham).

Fig. 2. **Map of the *Sau*3A family repetitive sequence.** *Sau*3AI, *Xba*I, *Bfa*I and *Taq*I sites are indicated by vertical arrows. PCR primers used to amplify the *Sau*3A family sequence are shown by horizontal arrows with the expected canonical PCR products. The positions of the primers were as follows: PCR1, 77-98 on subunit 5 and 99-120 on subunit 5; PCR2, 112-133 on subunit 2 and 95-117 on subunit 3; PCR3, 4-23 on subunit 1 and 104-123 on subunit 4. The position of the *Sau*3A family probe (*Stu*I-*Sph*I probe) used for Southern hybridization is also indicated.

Hybridization was performed overnight at 65 °C with probes prepared with a random priming DNA labeling kit (Boehringer Mannheim) and, after hybridization, the membranes were washed 3 times in 2 x SSC (SSC; 15 mM sodium citrate, 0.15 M NaCl) , 0.1% SDS for 30 min at 65 °C and twice in 0.2 x SSC, 0.1% SDS for 20 min at 65 °C, and autoradiographed.

2-7. Clones for Bending Assay

DNA fragments containing the human *Sau*3A family DNA were obtained from a plasmid clone, pXR-4, containing 4 units of the repeat , and those containing the human immunoglobulin μ locus gene from a phage clone Ch. H. Ig myu-24, containing 12 kb of immunoglobulin heavy chain Cμ region. The human c-*myc* gene was obatained by PCR amplification from HeLa cell genomic DNA. The DNA fragments to be duplicated were recovered from low-melting point agarose gels after digestion with restriction enzymes to obtain the regions of interest and cloned into the vector (pBluescript SK(-), Stratagene). The clones with tandem duplication of the inserts were screened after transformation of *E. coli* DH5α. Restriction sites were obtained from the GenBank or EMBL databases (entryname: HSRSSAU for the *Sau*3A family DNA, HSMYCC for the c-*myc* locus, HSJHCMU, HUMIGJEN, HSIGHJ6, and HUMIGHBP for the Ig μ locus) and also by sequence analysis of the phage clone Ch. H. Ig myu-24.

2-8. Assay for DNA Bend Sites

The circular permutation assay for DNA bend sites was originally described by Wu and Crothers (77). The priciple is shown in Fig. 3. Plasmid DNAs that contain regions of interest were linerlized with a series of restriction enzymes which cut at a single site and produce fragments of the same size but with the

Fig. 3. **Principle of the circular permutation assay.**

Plasmid DNAs containing duplicates of the regions of interest were linearlized with a series of restriction enzymes which cut the regions at a single site and produce the fragments of the same size but with the permuted order of nucleotide sequences. Each sample was then electrophoresed at 4°C on 8% polyacrylamide gels (mono : bis = 29 : 1). DNA fragments containing a DNA bend site in the middle of the fragments show a low mobility while those containing the site close to either end show a high mobility.

permuted order of nucleotide sequences. DNA fragments that have bend sites in the middle will show a low mobility and the ones that have them at their ends will show a high mobility. One μg of plasmid DNAs that contained duplicates of the regions of interest were linearized with the restriction enzymes that cut at a single site within the regions. After mixing with an internal calibration marker (*Pvu*II digests of M13mp18) the DNAs were electrophoresed at 4 °C on 8% polyacrylamide gels (mono: bis = 19:1) in 50 mM Tris-borate, 5 mM EDTA (TBE) buffer under the following conditions: 1 V/cm for 60 to 70 hrs for fragments of over 600 bp or 1.5 V/cm for 30 to 40 hrs for those smaller than 600 bp. Electrophoresis at 55 °C (3V/cm, for 7 hrs) was performed with at least one clone for each bend site to confirm that the bending was abolished at a high temperature.

2-9. Oligonucleotide Bending Assay

The duplex oligonucleotides (0.2 mg/ml) of 30 bp or 31 bp long (detailed in 1-2) were incubated with T4 polynucleotide kinase in the presence of 1 mM ATP at 37°C for 30 min, followed by the reaction with T4 DNA ligase at 4°C overnight. The ligation products were resolved by an 8% polyacrylamide (29: 1= mono: bis) gel in 45 mM Tris-borate, 1 mM EDTA buffer at 4°C for 23 hr.

2-10. Preparation of Nucleosome Core Particles

Nucleosome core particles were prepared as follows. Nuclei were prepared from 1×10^{10} HeLa cells and suspended in 0.34 M sucrose, 60 mM KCl, 15 mM NaCl, 15 mM 2-mercaptoethanol, 0.5 mM spermidine, 0.15 mM spermine, 15 mM Tris-HCl (pH 7.5) and 1 mM $CaCl_2$. They were then treated with 40 units/ml micrococcal nuclease for 2 min at 37°C. Core particles were recovered from nuclei by

centrifugation for 2 min at 8,000 x g as a supernatant after elution with 0.2 mM EDTA (pH 8.0). Core particles were depleted of histone H1 with 0.6 M NaCl, 0.2 mM 2-mercaptoethanol and 5 mM Tris-HCl (pH 7.5) and purified by gel filtration with Sepharose CL-6B in the same buffer, which was followed by dialysis against 20 mM ammonium acetate, 0.2 mM EDTA, 2 mM 2-mercaptoethanol and 5 mM Tris-HCl (pH 7.5). Core particles were stored at 4°C.

2-11. *In vitro* Reconstitution of Nucleosomes

Nucleosome core particles were mixed with DNA in 100 μl of TEP (10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA, 0.1 mM PMSF) supplemented with 2 M NaCl. The samples were first dialyzed against 1 liter of TEP with 0.4 M NaCl for 3 hr at 4°C and then against 1 liter of TEP with 16 mM NaCl for overnight at 4°C. For micrococcal nuclease digestion to recover 146 bp fragments, 10 μg of core particles were mixed with 20 μg of plasmid DNA (pBluescript and its derivatives) linearized with *Xmn*I, *Bam*HI (pHmBAB31) or *Sac*I (pHμD16).

2-12. Mapping of Nucleosomal Boundaries with Restriction Enzymes

Reconstituted core particles were treated with 20 units/ml micrococcal nuclease for 40 min on ice followed by incubation at 37 °C for 3 min. The 146 bp fragments protected by core particles were recovered from a 6% polyacrylamide gel (19:1= mono: bis) and treated with bacterial alkaline phosphatase. The fragments were then end-labeled with $[\gamma\text{-}^{32}P]$ ATP by T4 polynucleotide kinase, digested with restriction enzymes and resolved by 6% polyacrylamide-7M urea gels under denaturing conditions. The scheme for mapping the *in vitro* phases is shown in Fig. 4.
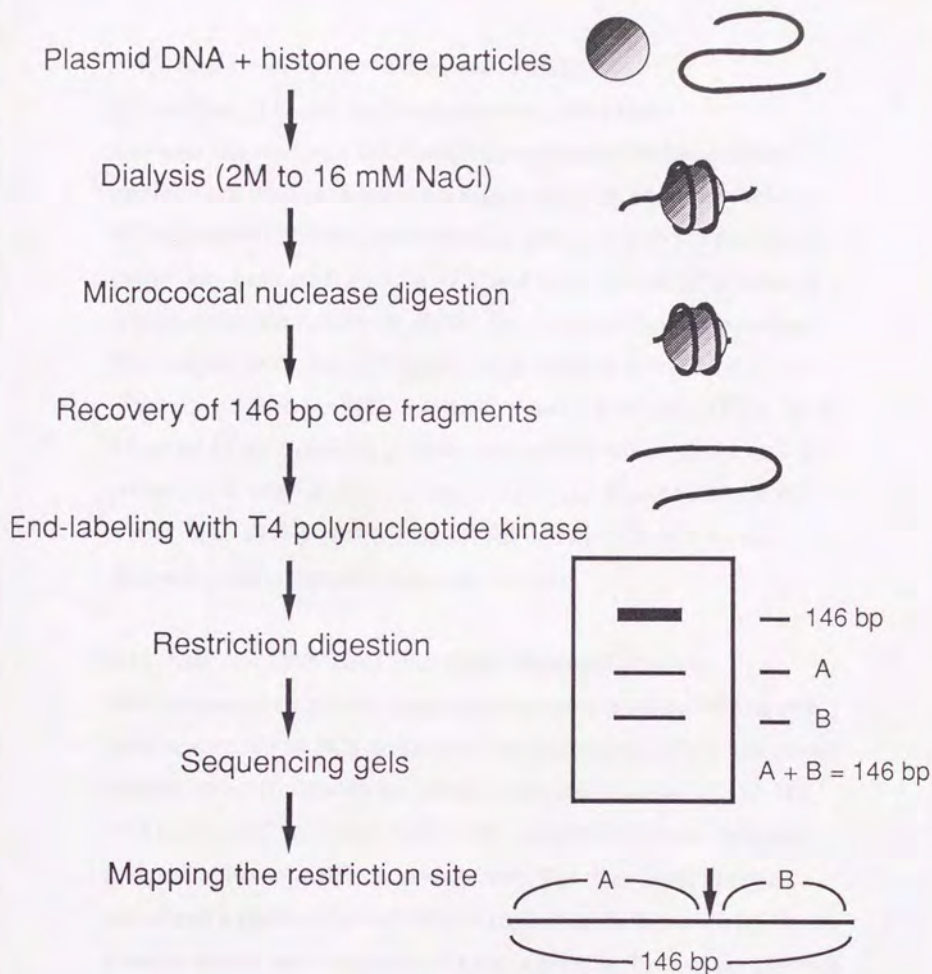
Fig. 4. **Analysis of nucleosomal phases by restriction digestion of reconstituted chromatin.**

The analysis of *in vitro* nucleosomal phases is schematically illustrated. The nucleosomal phases were determined according to the length of the restriction fragments of the nucleosomal core DNA.

2-13. Analysis of *in vivo* Nucleosomal Phases

For analysis of *in vivo* nucleosomal phases, the 146 bp mononucleosomal core DNA was recovered from the histone core particles and used as template for inverse PCR. In advance to inverse PCR, the core DNA was bunt-ended by treatment with 1.5 units/µl of mung-bean nuclease for 1 hr at 37°C and with 0.4 units/µl of Klenow fragment supplemented with dNTPs for 30 min at room temperature. The samples were then self-ligated using Ligation Kit version 2 (Takara) and used for PCR. For the first and second round PCR, 35 or 15 cycles of the following program was applied using primer set I or primer set II, respectively; 0.5 min at 92°C, 0.5 min at 50°C and 0.5 min at 72°C. PCR products were cloned into the pGEM T-vector (Promega) and subjected to sequence analysis.

2-14. Assay for DNA Bend Sites Using Deletion Constructs

DNA fragments containing serial deletions encompassing -452 to +49 were constructed by PCR and cloned into pBluescript SK(-). The clones pHmNa, pHmSb, pHmMc and pHmLd contained regions -452 to -102, -452 to -52, -452 to -2 and -452 to +49, respectively. Each clone was designed to have *Bam*HI site at their ends. The three latter plasmids containeed a portion (Sb and Mc), or all (Ld) of the bend site (cMB+1), whereas the site was completely deleted in pHmNa. Each of the plasmids was digested with *Bam*HI or *Sma*I and resolved by electrophoresis at 4°C to detect the region responssible for DNA bending.

2-15. Assay for DNA Bend Sites within the Sµ Region

Sµ region contains a highly repetitive sequence, and therefore, lacks restriction sites that can be used for the circular permutaion assay. After roughly determining the bend sites by the cicular permutation assay, the

potential regions of DNA bending were cloned into vector pBlVHD8 (shown in Fig. 5), which contained duplicated 466 bp *Eco*RV-*Hinc*II fragment of pBR322 at both *Hinc*II and *Sma*I sites of pBluescript. The fragments to be assayed were blunt-ended by T4 DNA polymerase and cloned into the *Eco*RV site of the vector located between the duplicated fragments. Each clone was digested with *Nhe*I, *Bam*HI or *Eco*NI, which cut uniquely within the duplicated 466-bp fragments, and electrophoresed on an 8% polyacrylamide (mono: bis= 29: 1) gel at 4°C.

pBIVHD8

3890 bp

*Eco*RV- *Hin*cII
(pBR322)

*Eco*RV- *Hin*cII
(pBR322)

*Nhe*I 759

*Bam*HI 905

*Sph*I 1092
*Eco*NI 1152

*Eco*RV

*Nhe*I 1264
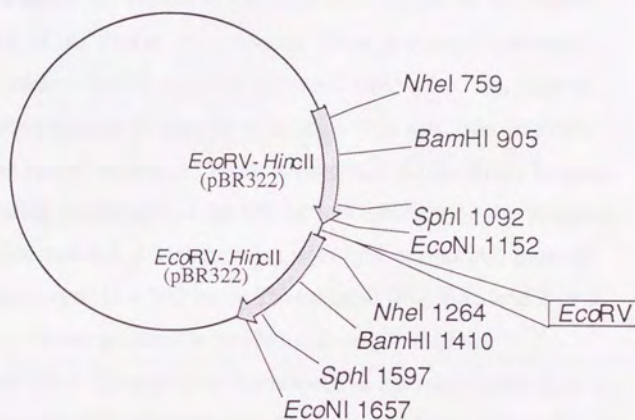*Bam*HI 1410

*Sph*I 1597
*Eco*NI 1657

Fig. 5. **Plasmid construct of vector pBIVHD8**
The 466 bp fragment from pBR322 was inserted into pBluescript at the *Sma*I
and *Hin*cII sites. The fragments for the bending assay were cloned into the
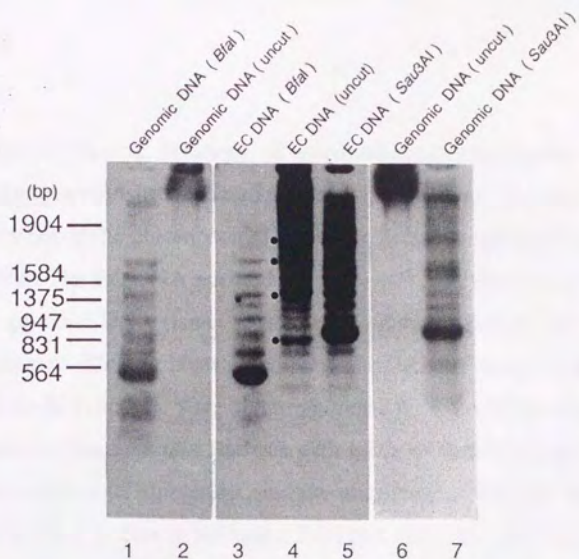*Eco*RV site which is located between the inserted 466 bp fragments.

3. Results

3-1. Preference of the Recombination Sites involved in Excision of the Human Alphoid *Sau*3A Repeat Family

3-1-1. Recombination Junctions in the Extrachromosomal *Sau*3A Family DNA

Fig. 6A shows the results of Southern blot analysis of the *Sau*3A family in the EC as well as chromosomal DNA. A typical ladder-like pattern was observed with purified uncut EC DNA (Fig. 6A, lane 4), indicating the presence of circular molecules with one, two, or more unit(s) of the family sequence. There existed four minor bands between the major bands (multimers of the 849 bp unit), indicating the presence of a composite subunit. After digestion with *Bfa*I or *Sau*3AI, most of the bands converged to a 500 bp or 850 bp band (Fig. 6A, lane 3 or 5, respectively), whose patterns were identical with that of the chromosomal DNA digested with the respective enzymes (lanes 1 or 7). To ensure that the PCR procedure could be applied for amplification of the repetitive sequence, we first examined the *Sau*3A family DNA after PCR amplification from chromosomal and EC DNA to see whether there is any difference of the DNA fragments before and after PCR. As shown in Fig. 7, lanes 1, 2, 4 and 5, the pattern of the restriction fragments of the genomic DNA (between the lengths of 300 bp and 1.5 kb) was identical before and after PCR amplification. On the other hand, the pattern obtained with the PCR products of the EC DNA indicated the presence of the fragments that were not present in the chromosomal DNA (Fig. 7, lanes 3 and 6). Since these fragments should contain recombination junctions, I cloned these fragments and analyzed their nucleotide sequences.

A

Lane labels (top):
Genomic DNA (*BfaI*)
Genomic DNA (uncut)
EC DNA (*BfaI*)
EC DNA (uncut)
EC DNA (*Sau3AI*)
Genomic DNA (uncut)
Genomic DNA (*Sau3AI*)

(bp)
1904 —
1584 —
1375 —
947 —
831 —
564 —

1 2 3 4 5 6 7

B

Subunit Number

5   1   2   3   4   5   1   2   3   4   5

B(X) B   S   B   B(X) B   S   B

PCR1
PCR2
PCR3

C

Cut   PCR

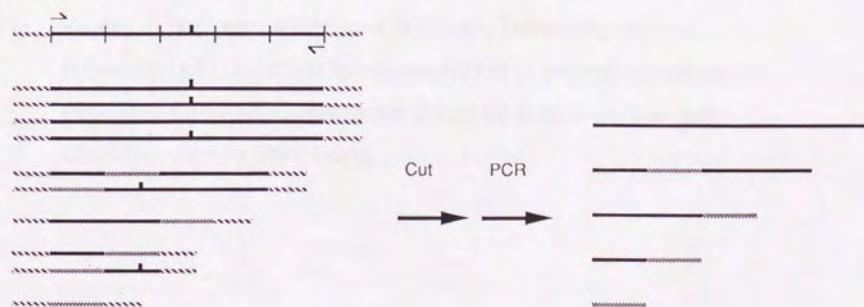Fig. 6. **Cloning strategy of recombination junctions from the extrachromosomal *Sau*3A family DNA.** (A) Southern blot patterns of the chromosomal and extrachromosomal *Sau*3A family DNA. The EC DNA purified from 3×10^8 cells (lanes 3 to 5) and 2 μg of genomic DNA (lanes 1, 2, 6 and 7), either digested with restriction enzymes (*Bfa*I for lanes 1 and 3 and *Sau*3AI for lanes 5 and 7) or uncut (lanes 2, 4, and 6), were electrophoresed on a 1.5 % agarose gel and used for Southern blot analysis with a *Sau*3A family probe containing the whole unit. Monomers, and the multimers of the unit, exhibiting a ladder-like pattern in the uncut EC DNA (lane 4), were indicated by dots. **(B)** Map of the *Sau*3A family repetitive sequence. *Sau*3AI (S), *Bfa*I (B) and *Xba*I (X) sites are indicated by vertical arrows. PCR primers used to amplify the *Sau*3A family sequence are shown by horizontal arrows with the expected canonical PCR products. The positions of the primers were as follows: PCR1, 77-98 on subunit 5 and 99-120 on subunit 5; PCR2, 112-133 on subunit 2 and 95-117 on subunit 3; PCR3, 4-23 on subunit 1 and 104-123 on subunit 4. **(C)** Cloning strategy. Circular EC DNA was first treated with the restriction enzymes which cut the unit. The products were then subjected to PCR amplification with the primers (horizontal arrows) located in both sites of the restriction site. Therefore, only the fragments lacking the site by recombination or point mutation can be amplified. Composite subunits are shown by shadowed lines and restriction sites by filled boxes.

Fig. 7. **PCR amplification of the extrachromosomal *Sau*3A family DNA.**

(A) One μg of the PCR products amplified from the EC DNA (lane 3) and those from chromosomal DNA (lane 2), both digested with *Sau*3AI, and *Sau*3AI-digested chromosomal DNA (lane 1) were electrophoresed on a 1.5 % agarose gel and hybridized with the *Stu*I-*Sph*I probe (indicated in Fig. 2). **(B)***Bfa*I-digested PCR products amplified from the EC DNA (lane 6) and those from chromosomal DNA (lane 5), *Bfa*I-digested chromosomal DNA (lane 4) were electrophoresed and hybridized with the *Stu*I-*Sph*I probe as in A.

3-1-2. Molecular Cloning of the Recombination Junctions

   To clone the fragments containing the recombination junctions, I
amplified the *Sau*3A family DNA with three different pairs of PCR
primers (detailed in Fig. 6B) and obtained different size classes of
fragments (PCR1 to 3). I first employed PCR amplification of EC DNA
using adaptors (S-TG or B-GC, detailed in Materials and Methods) in
order to supply enough materials for the analyses by Southern
hybridization and others, and eventually for the cloning procedure. I
then amplified the PCR products again, but this time, using *Sau*3A
specific primers (PCR1-3). The PCR products in each scheme contained
the fragments of the canonical length (shown in Fig. 6B) as well as
fragments containing a composite subunit. As shown in Fig. 6C, circular
EC DNA was digested with the restriction enzyme (*Sau*3AI, *Bfa*I or
*Xba*I) before PCR amplification, which prevents amplification of most
of the canonical fragments. Note that the canonical fragments without
the restriction sites could be amplified, although they are minor species
(see Fig. 6A).

   Fig. 8 shows that the products from EC DNA (Fig. 8, lanes 2, 4
and 6) contained additional fragments that were not observed in the
products from chromosomal DNA (Fig. 8, lanes 1, 3 and 5). These
additional fragments were cloned and their nucleotide sequences were
determined. All of the 32 clones analyzed (16, 9 and 7 clones for PCR1-
A, 2-A and 3-A, respectively) contained at least one composite subunit.
Fig. 9A shows the subunit structure of three clones. Clone I-11
contained subunits 5 and 1, and a 2/5 composite. Clones II-17 and II-36
similarly contained the whole subunits and a composite, 3/5 or 1/2,
respectively. The nucleotide sequences of their recombination junctions
are shown in Fig. 9B. Recombination seems to have occurred
somewhere between positions 23 and 27 for I-11, 127 and 152 for

Fig. 8. **PCR amplification of the fragments containing recombination junctions.**

PCR-amplified EC DNA (lanes 2, 4 and 6) and chromosomal DNA (lanes 1, 3 and 5) were used for PCR amplification with specific primers shown in Fig. 6B, and the PCR products were electrophoresed in a 1.5 % agarose gel. PCR amplification was performed with *Sau*3AI fragments for PCR1 (lanes 1 and 2) and PCR3 (lanes 5 and 6), and *Bfa*I fragments for PCR2 (lanes 3 and 4) in a total volume of 100 µl containing 300 ng of DNA. Arrows indicate the PCR products with recombination junctions. The bands at 850 bp and 1700 bp (dimer) in PCR1, 680 bp in PCR2, 630 bp or 1480 bp (dimer) in PCR3 were derived from the canonical fragments.

II-17, and 12 and 19 for II-36.

3-1-3. Distribution of Recombination Junctions

I also employed direct amplification of recombination junctions from purified EC DNA digested with *Sau*3AI, *Bfa*I or *Xba*I (PCR1-B, 2-B, 3-B and 3-C) and added to the data shown above. The positions of recombination junctions observed in each PCR scheme (a total of 67 clones containing 68 junctions) are shown in Fig. 10. The results of three PCR schemes were combined and are summarized in Fig. 11A. The distribution of the recombination junctions showed a marked centering at four regions, positions 10 to 25 (A), 40 to 50 (B), 85 to 90 (C) and 135 to 160 (D). When this was compared with the degree of sequence homology between subunits (shown as a dashed line in Fig. 11A), regions B and C overlapped the regions with a high sequence homology between the subunits, while regions A and D showed less apparent association between them, indicating that some factor other than sequence homology has a dominant effect on the events (see Discussion). I included region D in the latter case because, although the region is correlated with the homologous regions, the recombination frequency is apparently higher than expected when compared with regions B and C. Table 2 and Fig. 11B is the summary of the recombination events within the unit structure. There was a strong preference for ± 2 subunit shift (87%), especially for 2/4 or 4/2 (22%), 3/5 or 5/3 (22%), or 2/5 or 5/2 (19%), while 13% of the cases occurred between the adjoining subunits (± 1 shift).

A

Subunit Number   5   1   2/5

I-11

Subunit Number   3/5   1   2

II-17

Subunit Number   3   4   5   1/2

II-36

B

```
                  10         20         30         40         50         60
Subunit 2  CTGTAGTAAAGGAAATAACTTCATCTAAAAACCAAACGGAAGCATTCACAGACAATTCTT |
                               ***      *   *              *    * * * *
I-11       CTGTAGTAAAGGAAATAACTTCACAGAAAAACTAAACAGAAGCATTCTCAGAAACTACTC |
           ** * *  * **** *
Subunit 5  CGTTGGAAACGGGTTCATCTTCACAGAAAAACTAAACAGAAGCATTCTCAGAAACTACTT
```

```
                  130        140        150        160        170 1       10
Subunit 3  TCAAACACTCTATTTGTAGAATCTGCAAGTGGATATTTGGACCACTTTGTGGCCTTCCTT
                     *                          *        ***    ** *** ****
II-17      TCAAACACTCTTTTTGTAGAATCTGCAAGTGGACATTTGGAGGGCTTTGGAGGGCTTTGA
           *   *       *
Subunit 5  TGAAACCCTCTTTTTCTAGAATCTGCAAGTGGACATTTGGAGGGCTTTGGAGGGCTTTGA
```

```
                  170 1       10         20         30         40
Subunit 1  GGGCTTTGAGGCCTGTGGTGGAA-AAGGAAAATCCTTCACAAAAAAACTAGATGGAAGCAT
                            |        *** *  ***         * * *
II-36      GGGCTTTGAGGCCTGTGGTGGAAAAAGGAAATAACTTCATCTAAAAACCAAACGGAAGCAT
           **         *** * *|
Subunit 2  CTGCTTTGAGGCCTACTGTAGTA-AAGGAAATAACTTCATCTAAAAACCAAACGGAAGCAT
```

Fig. 9. **Clones containing a composite subunit (A), and the nucleotide sequences of their recombination junction (B).** Clones containing a composite subunit (A), and the nucleotide sequences of their recombination junction (B). **(A)** Three clones, I-11, II-17 and II-36, contained composite subunits 2/5, 3/5 and 1/2, respectively. Composite subunits are indicated as striped boxes. **(B)** The sequences of the clones are aligned with those of the parental subunits and the region from which the sequence was derived are shadowed. A vertical bar represents a mismatch to both of the parental subunits and an asterisk represents a mismatch to either of them. Accession numbers of the nucleotide sequences of these recombination junctions in DDBJ/GenBank/EMBL databases are: I-11, D49597;II-17, D49591; II-36, D49595.

A

B

Recombination Frequency vs Position

Clone Number	Size (bp)
II-52 [3 / 1]	170
II-17 [3 / 5]	340
II-32 [3 / 5]	340
II-48 [3 / 2]	340
15 [1 / 2]	510
II-36 [1 / 2]	510
II-51 [1 / 2]	510
II-30 [1 / 2]	850
I-74 [3 / 1, 4 / 2]	1020
X-2 [3 / 1]	170
X- 101 [3 / 1]	170
X- 103 [4 / 2]	170
X-A [3 / 1]	170
X-C [3 / 1]	170
X-4 [4 / 2]	340
X-5 [3 / 5]	340
X-6 [5 / 2]	340
X-8 [3 / 5]	340
X-18 [5 / 2]	340
X- 106 [3 / 5]	340
X-117 [3 / 1]	340
X-118 [4 / 2]	340
X-119 [4 / 2]	340
X-120 [4 / 1]	340
X-121 [4 / 2]	340
X-122 [4 / 1]	340
X-125 [4 / 1]	340

PCR2-A
PCR2-B

C

Recombination Frequency vs Position

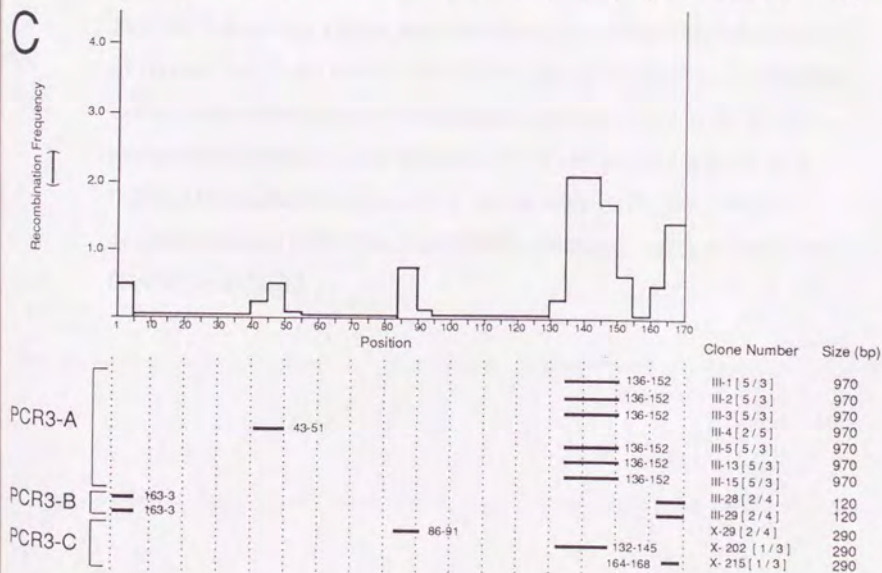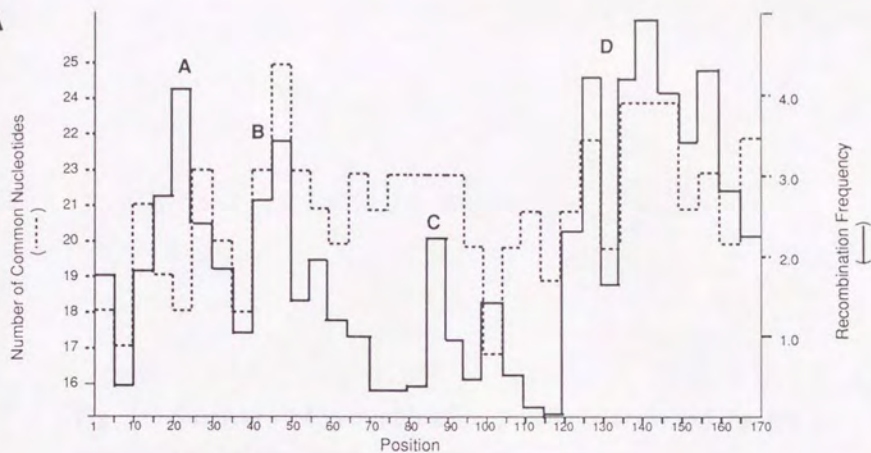| Clone Number	Size (bp)
III-1 [5 / 3]	970
III-2 [5 / 3]	970
III-3 [5 / 3]	970
III-4 [2 / 5]	970
III-5 [5 / 3]	970
III-13 [5 / 3]	970
III-15 [5 / 3]	970
III-28 [2 / 4]	120
III-29 [2 / 4]	120
X-29 [2 / 4]	290
X- 202 [1 / 3]	290
X- 215 [1 / 3]	290

PCR3-A
PCR3-B
PCR3-C

Fig. 10. **Distribution of the recombination junctions from** *Sau*3AI, *Bfa*I and *Xba*I digested EC DNA observed in PCR1-3.

The recombination junctions of the 67 clones (68 junctions) containing composite subunits are shown as horizontal bars. Clones in PCR1-A (16 clones), PCR2-A (9 clones), PCR3-A (7 clones) and PCR3-B (2 clones) were obtained by two-step amplification using primers S-T and S-G (PCR1-A and PCR3-A) or B-C and B-G (PCR 2-A and PCR3-B) for the first step after *Sau*3AI (PCR1-A and PCR3-A) or *Bfa*I (PCR2-A and PCR3-B)digestion, respectively, and primers shown in Fig. 6B for the second step. Clone I-74 contained two recombination junctions. Clones in PCR1-B (12 clones), PCR2-B (18 clones) and PCR3-C (3 clones) were obtained by direct amplification with primers for PCR1 through 3 after digestion with *Sau*3AI (PCR1) or *Xba*I (PCR2-B and 3-C). Recombination frequencies were calculated by adding the probabilities of recombination for every 5 bp of the subunits for each recombination event. If the clone has the recombination junction between 21-27, its probabilities for the regions between 21-25 and 26-30 are 0.714 and 0.286. The nucleotide sequences of the recombination junctions are available through DDBJ/GenBank/EMBL databases: accession numbers D49585 to D49652.

Fig. 11. **Summary of recombination junctions arranged by the subunit (A) or the whole unit (B) structure.**

(**A**) The frequencies of recombination within the subunits for PCR1 through PCR3 were combined and are shown by a solid line. Four recombination hotspots are indicated as A to D (A: 10 to 25, B: 40 to 50: C: 85 to 90 and D: 135 to 160). The number of common nucleotides (common for at least 3 subunits) among 5 subunits are shown as a dashed line. (**B**) The recombination events are arranged according to the type of the subunit shift: ± 1 (*upper*) or ± 2 (*lower*) subunit shift. The frequencies (%) of each type of recombination are shown above.

Table 2. Subunits involved in recombination classified by the type of subunit shift.

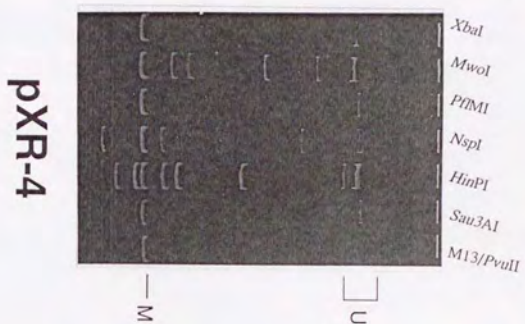| Type of recombination[a]	Type of subunit shift[b]	Number of junctions			
		Each	(%)	Subtotal	(%)
1 / 2 or 2 / 1	±1	4	(6)		
2 / 3 or 3 / 2	±1	1	(1)		
3 / 4 or 4 / 3	±1	1	(1)	9	(13)
4 / 5 or 5 / 4	±1	2	(3)		
5 / 1 or 1 / 5	±1	1	(1)		
1 / 3 or 3 / 1	±2	10	(15)		
2 / 4 or 4 / 2	±2	15	(22)		
3 / 5 or 5 / 3	±2	15	(22)	59	(87)
1 / 4 or 4 / 1	±2	6	(9)		
2 / 5 or 5 / 2	±2	13	(19)		
Total		68	(100)	68	(100)

[a]Possible combination of subunits.

[b]Gain or loss of indicated numbers of subunit.

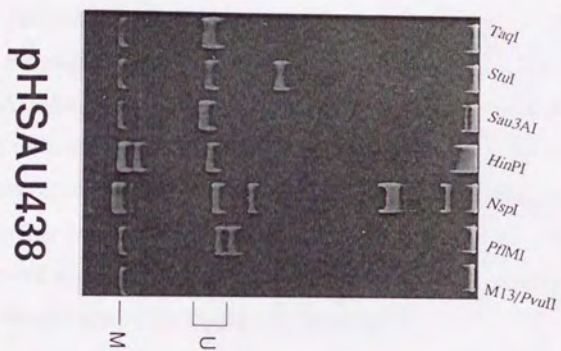3-2. Nucleosome phasing and DNA bending in the *Sau*3A family DNA

3-2-1. Mapping of DNA Bend Sites

The DNA bend sites in the human *Sau*3A family were determined by the circular permutation assay. Clones containing tandem duplicates of regions 79 to 517 (pHSAU438, 438 bp) and 517 to 79 (pHSAU411, 411 bp) were constructed from the parental clone pXR-4. These plasmids were digested with a series of restriction enzymes that cut the unit sequence once and thus produces fragments with identical lengths but have sequences of the unit in the permuted order. The restriction fragments were electrophoresed on polyacrylamide gels at 4 °C. The bend sites were defined as the regions between the two restriction sites located on both sides of the restriction site that produces the fragments of the highest mobility. Fig. 12 shows the results of the circular permutation assay with plasmid pXR-4, pHSAU411 and pHSAU438. The bend site within pHSAU411, shown in Fig. 12C, was likely to be located between *Mwo*I and *Tsp*509I sites, and close to *Ava*II site. The summary is shown in Fig. 13. Two bend sites were likely to be located between positions 282 to 499 (between *Hin*PI and *Pfl*MI sites) and 613 to 768 (between *Mwo*I and *Tsp*509I sites), from rough mapping. As shown in the figure, the retardation of the fragments carrying the bend sites in the middle ranged from 4.3 to 6.5 % upon electrophoresis. It was surprising that, even within the repetitive sequence, DNA bend sites existed and they were not located in every subunit (subunits 2 or 3 and 4 or 5). Since the sequence diversity in the *Sau*3A family subunits was mostly observed in the positions 1-22, these positions in the respective subunit were the candidates for DNA bending.
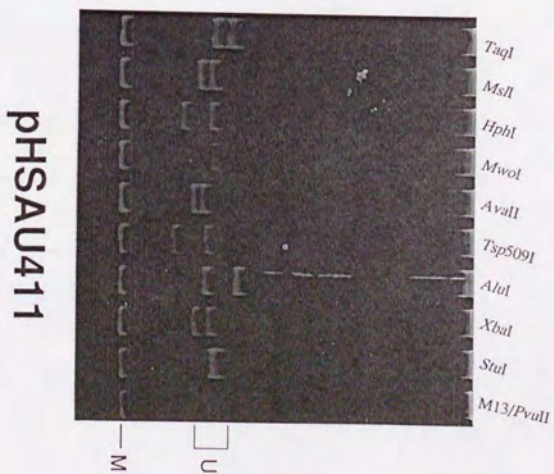
A   pXR-4

XbaI
MwoI
PflMI
NspI
HinPI
Sau3AI
M13/PvuII

M   u

B   pHSAU438

TaqI
StuI
Sau3AI
HinPI
NspI
PflMI
M13/PvuII

M   u

C   pHSAU411

TaqI
MslI
HphI
MwoI
AvaII
Tsp509I
AluI
XbaI
StuI
M13/PvuII

M   u

Fig. 12. **Mapping of the DNA bend sites in the *Sau*3A family.**

The circular permutation assay was performed at 4°C using the plasmid pXR-4 (A), pHSAU438 (B) and pHSAU411 (C) that contained either of four repeats of the whole unit, the region between 79 to 517 or the region between 517 and 79, respectively, according to the databases (entryname; HSRSSAU). By this assay, the bend center was likely to be located between *Mwo*I and *Tsp*509I sites and close to *Ava*II site. The positions of the fragments of the unit length are shown as U and the position of the marker is shown as M.
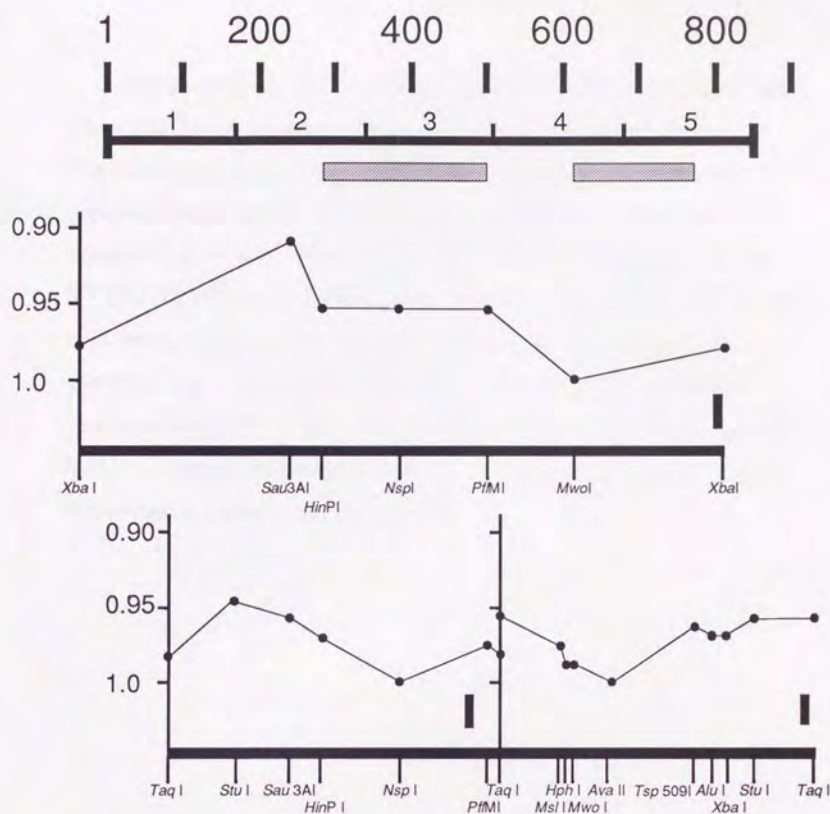
Fig. 13. **Summary of the DNA bend sites in the human *Sau*3A family DNA.**

The plasmids containing duplicates of the region shown as the thick horizontal line in each panel were digested with the enzymes shown below. The vertical bars in the panel show the average thickness of the bands relative to the migration distance, indicating a deviation of mobility. The bend sites, tentatively defined as the regions between the two restriction sites located on both sides of the restriction site that produces the fragments of the highest mobility, are shown as shadowed boxes.

Further mapping of DNA bend sites were performed by using uni-directionally-concatenated oligonucleotides (the positions of the oligonucleotides are shown in Fig. 14A). These regions were also correlated with the CENP-B binding site which was located within subunits 3 and 5 (the consensus sequence for CENP-B binding site is **TTTCGTTGGAAACGGG**A, where a 100% match is required for the core motif shown in bold) and with the recombination hotspot. As shown in Fig. 14B, SAUB-2 and SAUB-4 did not show retardation compared to $(dA)31+(dT)31$. On the other hand, SAUB-1, SAUB-3 and SAUB-5 showed retardation indicating the presence of the bend sites in these regions (summarized in Fig. 14C).

A

```
1    1 : CCTGTCCTCCAAAAGGAAAATCCTTCACAAAAAAACTAGATGGAAGCATT
2    1 : CCTACTGTAGTAAAGGAAATAACTTCATCTAAAAACCAAACGGAAGCATT
3    1 : ATTTCGTTGCAAACGGGATAAACTTCCC    AGAACTACACGGAAGCATG
4    1 : CCTTCCTTCGAAACGGGTATATCTTCACATCAAACCTAGACAGAAGCATT
5    1 : ATTTCGTTGGAAACGGGTTCATCTTCACAGAAAAACTAAACAGAAGCATT
```

1-22; region of a high sequence diversity

```
1   51 : CTCAGAAACTACTTTGTGATGATTGCATTCGACTCACAGAGTTGAACATT
2   51 : CACAGACAATTCTTAGTGATCATTGCATTGAACTAACAGAGCTGAACATT
3   51 : CTGAGAAACTTCTTTGTGATGTTTCCATTCAACTCACAGAGTTGAACCTT
4   51 : CTCAGAATGTTTCCTGTGATGAGTGCATTCAACTCACAGAGGTGAACAAT
5   51 : CTCAGAAACTACTTTGTGATGTTTGTGTTCCACTTCAAGAATTGAACTTT
```

```
1  101 : CCTATAGATAGAGCAGGTTGTAAACAATCTTTTTGTAGAATCTGCGATTG
2  101 : GCTTTAGATGGCGCAGTTTCCAAACCCACTTTCTGTAGAATCTGCAAGTG
3  101 : GCTTT CATAGTTCAGCTTTCAAACACTCTATTTGTAGAATCTGCAAGTG
4  101 : CCTGCTGATGGAGCAGTTTTGAAACTCTCTTTCTTTGGATTCTGCAAGTG
5  101 : CCTCTTGACAGAGCAGCTCTGAAACCCTCTTTTTCTAGAATCTGCAAGTG
```
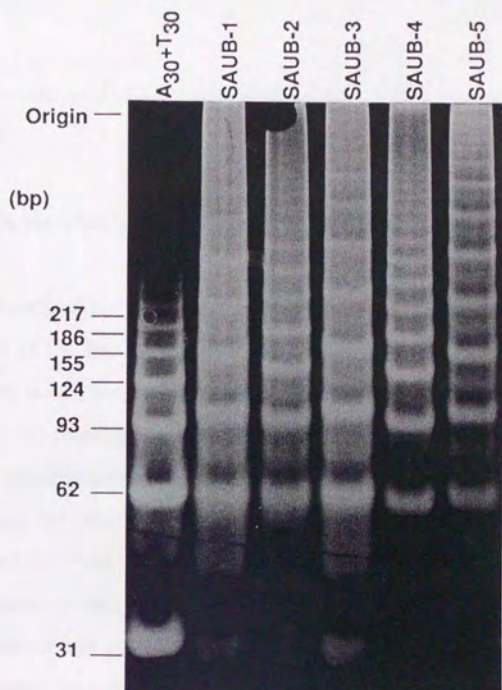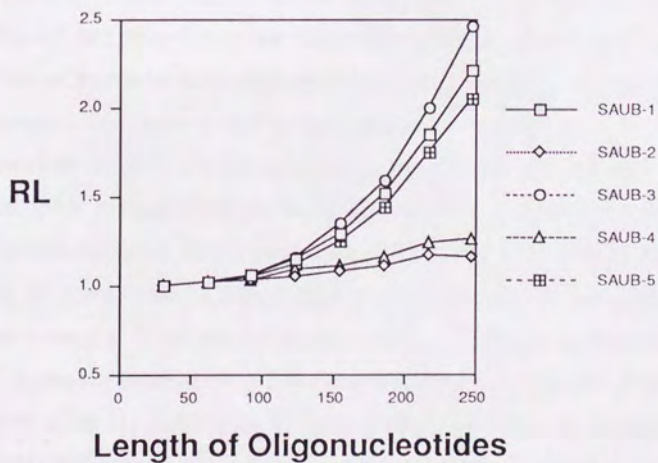
```
1  151 : GAGATTTGGACTGCTTTGAGG
2  151 : GATATTTGGACCTCTCTGAGG
3  151 : GATATTTGGACCACTTTGTGG
4  151 : G  ATATGGACCTCTGTGAAG
5  151 : GACATTTGGAGGGCTTTGAGG
```

Fig. 14. **Fine mapping of the bend sites in the *Sau*3A family.**
(A) The positions of the oligonucleotides used for the bending assay are shadowed. Each subunit was aligned according to the sequence homology and the oligonucleotides were located at the same position in the subunit structure. (B) Double-stranded oligonucleotides, 31 nucleotides long SAUB-1, SAUB-2, SAUB-3, SAUB-4 and SAUB-5, were electrophoresed after unidirectional ligation. Oligonucleotides $(dA)31+(dT)31$ were used as a control. (C) The relative mobilities of the multimers of $(dA)31+(dT)31$ to that of the multimers SAUB-1 to SAUB-5 are shown. R_L values are the ratio of apparent length to real length.

B

A₃₀+T₃₀ lanes: SAUB-1, SAUB-2, SAUB-3, SAUB-4, SAUB-5

Origin —

(bp)

217 —
186 —
155 —
124 —
93 —

62 —

31 —

C

RL vs Length of Oligonucleotides

- SAUB-1 (□)
- SAUB-2 (◇)
- SAUB-3 (○)
- SAUB-4 (△)
- SAUB-5 (⊞)

3-3. Nucleosomal phases around the translocation sites in the c-*myc* and Igμ loci

3-3-1. *In vitro* nucleosomal phases in the c-*myc* and immunoglobulin loci

I described the preference of the recombination sites in the subunit structure of the *Sau*3A family DNA (described in 2-1). Since the family DNA has subunits of 171 bp long, which correspond to the nucleosomal unit, this result suggests that the recombination preferentially takes place at specific locations in the nucleosomal phases. To examine this hypothesis for other types of recombination, for example, chromosomal translocation observed in lymphoma cell lines, I have determined the nucleosomal phases on both the c-*myc* and immunoglobulin loci.

Nucleosomal phases were first determined with reconstituted nucleosomes (*in vitro* phases). After nucleosomes were reconstituted with plasmids pHmBAB40, pHmPS10, pHmBAB31, pHmHAN24 and pHmRRV8 for the c-*myc* locus, and pHIg5, pHIg6-1, pHIgXS, pHIgHT, pHIg-0.9 and pHIg-0.8 for the immunoglobulin locus, they were subjected to micrococcal nuclease digestion and the protected 146 bp fragments were recovered. The fragments were digested with restriction enzymes and the nucleosomal phases were determined (detailed in Materials and Methods). For example, 33 bp and 114 bp fragments appeared after digesting the 146 bp core DNA with *Sac*II (Fig. 15, lanes 1 and 2), indicating that the nucleosomal boundaries were located at 33 bp and 114 bp from the site. Nucleosomal phases in the region surrounding exon 1 of the c-*myc* gene (Fig. 15) and J5 to Sμ region of the Igμ locus (Fig. 17) were determined according to this strategy and are summarized in Figs. 16 and 18. Nucleosomes were aligned in a coordinated manner in both regions and their periodicity
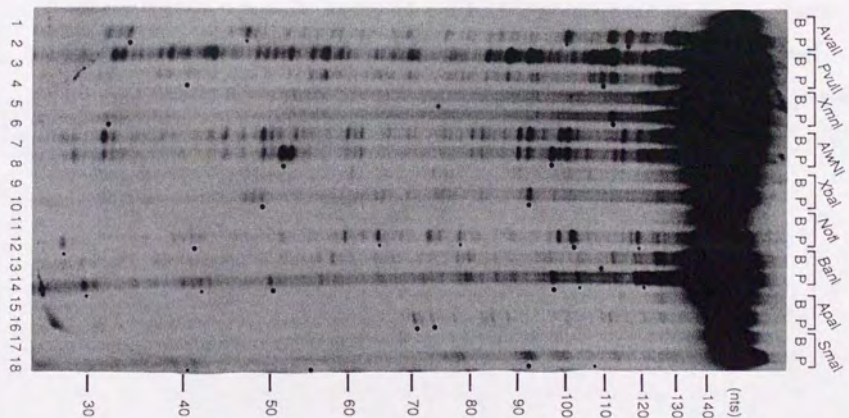
48

Fig. 15. **Mapping of the boundaries of reconstituted nucleosomes in the c-*myc* -1500 to 2000 region.**

Nucleosomes were reconstituted on the clones pHmBAB40, pHmPS10, pHmBAB31, pHmHAN24 and pHmRRV8 (shown in Fig. 22) and pBluescript (a control). The reconstituted nucleosomes were then subjected to micrococcal nuclease digestion and the protected 146 bp fragments were recovered. Each sample together with the control pBluescript was labeled with ^{32}P, digested with the restriction enzymes shown at the top of the figure and electrophoresed under denaturing conditions.

-1000   0   1000

c-myc

1

P1   P2

SacII
AvaII   HaeIII Apal   EcoNI   HaeIII HaeIII   SmaI NotI BsrBI AvaII PvuII   XmnI AlwNI   XbaI NotI BanI   ApaI SmaI

SacII 114+33
AvaII 74+30   35+115   50+*97   EcoNI 60+87   HaeIII 102+48   73+77   73+77   32+115   43+104   135+?   97+54   99+48   43+104   50+97   77+73   92+55
  HaeIII 72+85   SmaI 115+32   117+30

Fig. 16. **Summary of the nucleosomal phases in the c-*myc* locus.**
Nucleosomes are shown as open circles (minor ones are shown as shadowed
circles behind) and their positions were determined from the lengths of the
pairs of the fragments generated with each restriction enzyme (sizes are shown
below the nucleosomes). Shadowed circles on the lower line are the putative
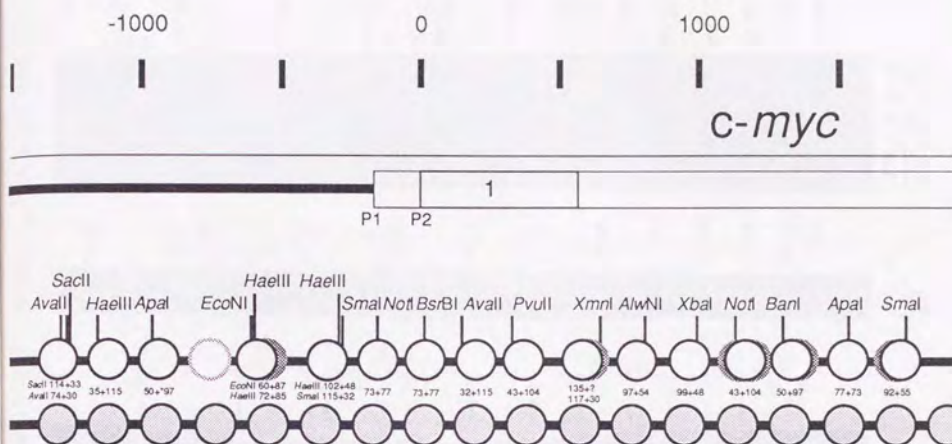nucleosomes with the 146 bp core and 24 bp linker regions.

Fig. 17. **Mapping of the boundaries of reconstituted nucleosomes in the immunoglbulin μ locus.**

Nucleosomes were reconstituted on the clones pHIg5 (lanes 3, 9, 11 and 13), pHIg6-1 (lanes 2, 5 and 7), pHIgXS (lanes 15, 17, 19 and 22), pHIgHT (lanes 20, 24 and 29), pHIg-0.9 (lane 26) and pHIg-0.8 (lane 28) (shown in Fig. 23) and pBluescript (a control: lanes 1, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 27). The reconstituted nucleosomes were analyzed according to the same strategy used in the c-*myc* locus.

J5

1000

J6

2000

(HindIII)
98+48

MscI BstYI
43+104 7+117
40+107

(HindIII)
(38+109)
110+37 Nspl 112-35
120-27
BpI BstUI

3000

enhancer

443bp repeat

4000

Bgl Eael
37+110
113+34

IgM heavy chain switch region

5000

117+ PvuII
SacI
+123

PvuII
55+95
(HindIII) Sau96I
47+100
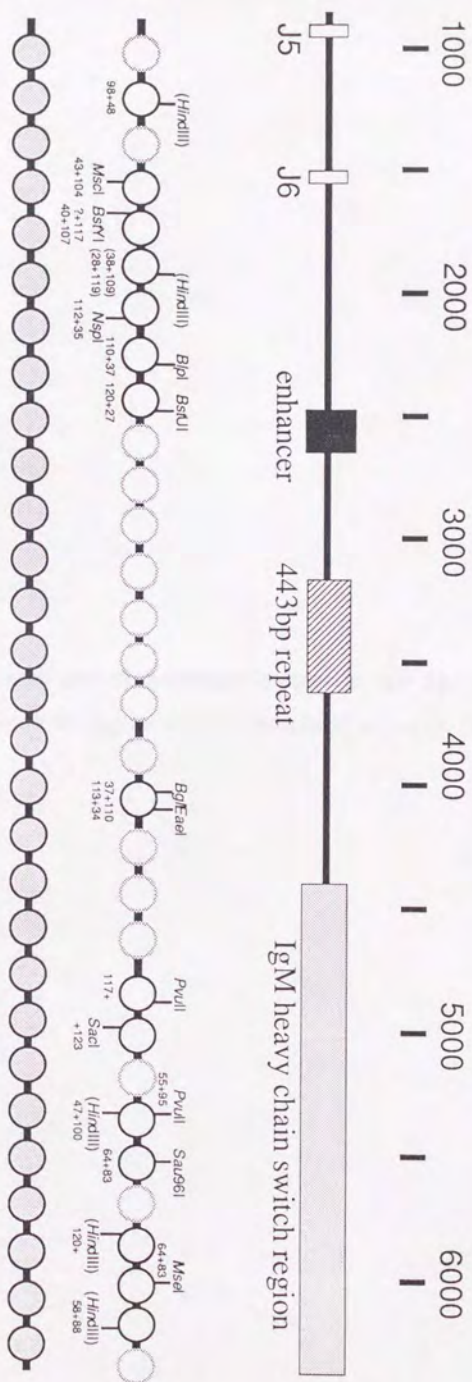64+83

6000

(HindIII) MseI
120+ 64+83
(HindIII)
58+88

Fig. 18. **Summary of the nucleosomal phases in the Igµ locus.**
Nucleosome phases in the Igµ locus are summarized as in Fig. 16.

closely resembled the putative phases shown below the figure. One bend center mapped at the P2 promoter the of c-*myc* gene was mapped to 10 to 40 bp from the nucleosomal boundary.

3-3-2. *In vivo* nucleosomal phases at the translocation junctions

In vivo nucleosomal phases were then determined by inverse PCR using HeLa mononucleosomal DNA as template. The scheme of determining the *in vivo* phases is shown in Fig. 19. First, the mononucleosomal DNA obtained by digesting the nuclei with micrococcal nuclease was blunt-ended by mung-bean nuclease and Klenow fragment. Then the samples were self-ligated and used as templates for inverse PCR. The primers were designed to be located inside of the nucleosomal core region according to the *in vitro* phases and to face each other only after self-circularization. Therefore, the PCR products obtained should contain both boundaries derived from the same nucleosomal cores.

By this method, determination of the *in vivo* nucleosomal phases is now in progress for two regions, each from the c-*myc* and immunoglobulin loci, containing the translocation junctions in a non-Hodgkin lymphoma cell line Manca (data not shown).
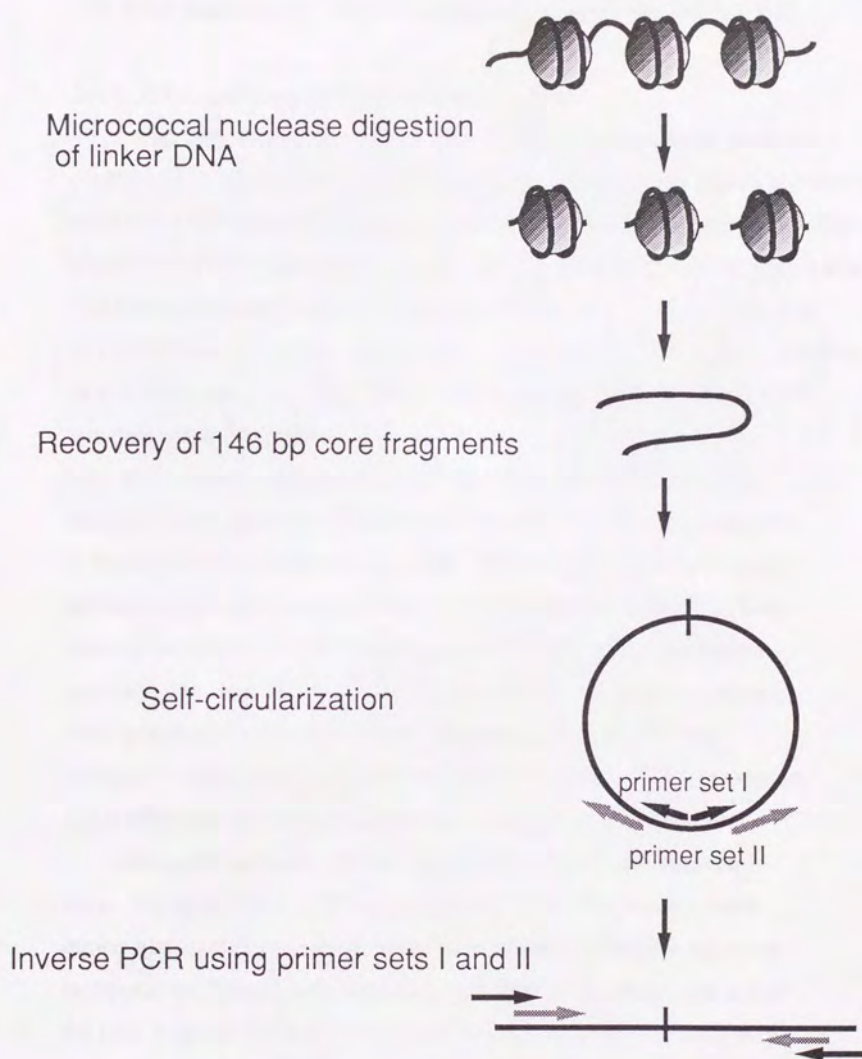
Micrococcal nuclease digestion of linker DNA

Recovery of 146 bp core fragments

Self-circularization

primer set I

primer set II

Inverse PCR using primer sets I and II

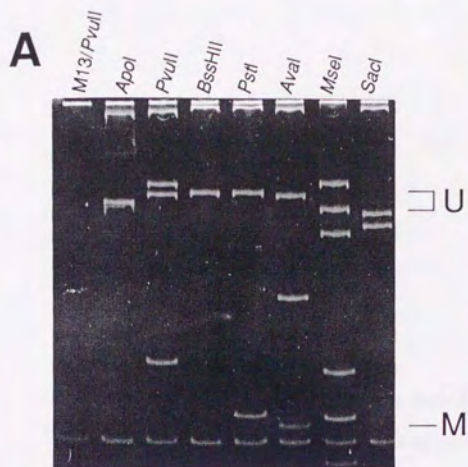Fig. 19. The scheme of the analysis of *in vivo* nucleosomal phases.

3-4. DNA bend sites and nucleosomal phases in the c-*myc* and Igμ loci

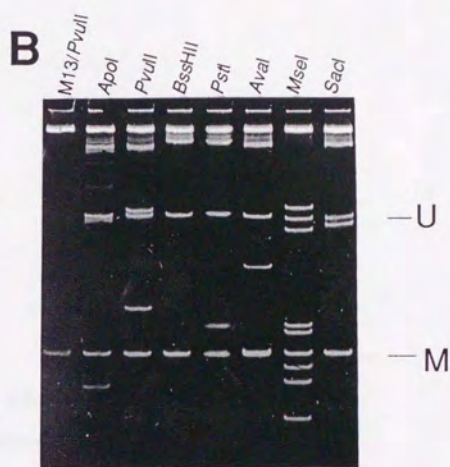#### 3-4-1. DNA bend sites in the c-*myc* and Igμ loci

Fig. 20A shows the result of a circular permutation assay using the plasmid, pHmA20, which contained tandem duplicates of the region between the positions +2184 and +3020 relative to the P2 cap site of the c-*myc* gene. The plasmid DNA was digested with restriction enzymes which cut the region at a single site. The restriction fragments of identical length were resolved by electrophoresis on polyacrylamide gels at 4°C (left) or 55°C (right). Anomaly of migration was observed at 4°C indicating the presence of a rigid DNA structure around the *Mse*I

site, which caused retardation of the fragments with *Mse*I site at the center (*Pvu*II fragments). The anomaly was lost at 55°C, indicating that it was due to DNA bending (Fig. 20B). The bend sites were tentatively defined as the regions between the two restriction sites located on both sides of the restriction site that produced the fragments of the highest mobility. The bend site in this region was therefore mapped between *Ava*I (position +2733) and *Sac*I (+2913) sites. According to this strategy, I mapped the bend sites that appeared in the c-*myc* gene region (total 8082 bp) and the immunoglobulin μ region (total 7427 bp).

Although most sites were mapped by the circular permutation assay, the bend sites in the Sμ region could not be mapped to a small region by this method because the region consisted of highly repetitive sequences and thus lacked restriction sites that cut at a single site within the unit. Instead, the bend sites of this region were analyzed after they were cloned into pBlVHD8 (Fig. 5), which contained duplicated 466 bp fragments from pBR322. Each clone was digested either with *Nhe*I, *Bam*HI or *Eco*NI, which cut at a single site within the 466 bp fragment but placed the insert in different orders. The *Nhe*I and *Eco*NI fragments

Fig. 20. **Mapping of the DNA bend site in pHmA20.**
The circular permutation assay was performed at 4°C (**A**) or 55°C (**B**)
using the plasmid pHmA20 containing the region between 2174 and
3010 relative to the P2 cap site of the c-*myc* oncogene. By this assay,
the bend center was likely to be located between *Mse*I and *Apo*I sites
and close to *Sac*I site. The results are summarized below the pattern of
the gel. The anomaly of migration was abolished at 55°C, confirming
that the anomaly was due to DNA bending.

have the insert close to the ends and the *Bam*HI fragment has the insert close to the center. Therefore, if the insert contains a bend site, migration of the *Bam*HI fragment will be retarded compared to the *Eco*NI and *Nhe*I fragments upon electrophoresis at 4°C (Fig. 21). The *Bam*HI fragments of the vector (a control) and pHIg0.8Nci showed relative migration rates of 0.97 to 0.98 compared to the *Nhe*I and *Eco*NI fragments indicating that there was no bend site in the cloned fragments. In contrast, the *Bam*HI fragments of pHIg0.9S and pHIg0.45S showed 0.95 to 0.96 migration rates, indicating the presence of bend sites in these regions. It should be noted that the retardation caused by the bend sites in the S region was smaller (approximately 5%) than those in other regions.

Figs. 22 and 23 show summaries of the bend sites in these regions (12 sites in the c-*myc* region and 11 sites in the Igμ region). They appeared continually as observed in the human β-globin locus. The average
interval of the bend sites were 693 ± 282.3 bp in the c-*myc* and 650 ± 206.8 bp in the Igμ region. However, the periodicity was disturbed around the exons and no bend sites appeared within the coding regions.

3-4-2. Sequence features of DNA bend sites
Previous studies regarding the sequence feature of DNA bend sites indicated the involvement of short $(dA)_n$ $(n \geq 2)$ tracts appeared at intervals of roughly ten, or multiples of ten nucleotides. Therefore, I focused upon the appearance of the short poly(dA) and poly(dT) tracts within the bend sites in the c-*myc* and immunoglobulin μ loci. As shown in Fig. 24, these tracts were frequently observed within the bend sites. In the S region, AA or TT dinucleotides appeared at intervals of 10 bp or its multiples. Since these tracts were not included in the conserved
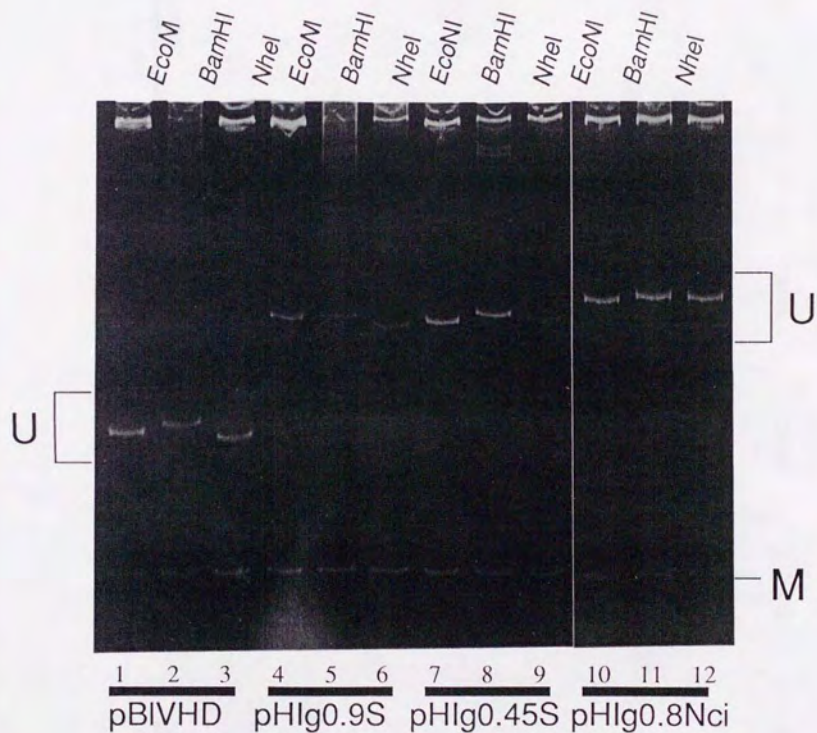
Fig. 21. **Mapping of the DNA bend sites within the Sμ region.**
Mapping of the bend sites were performed using the vector pBlVHD8. The
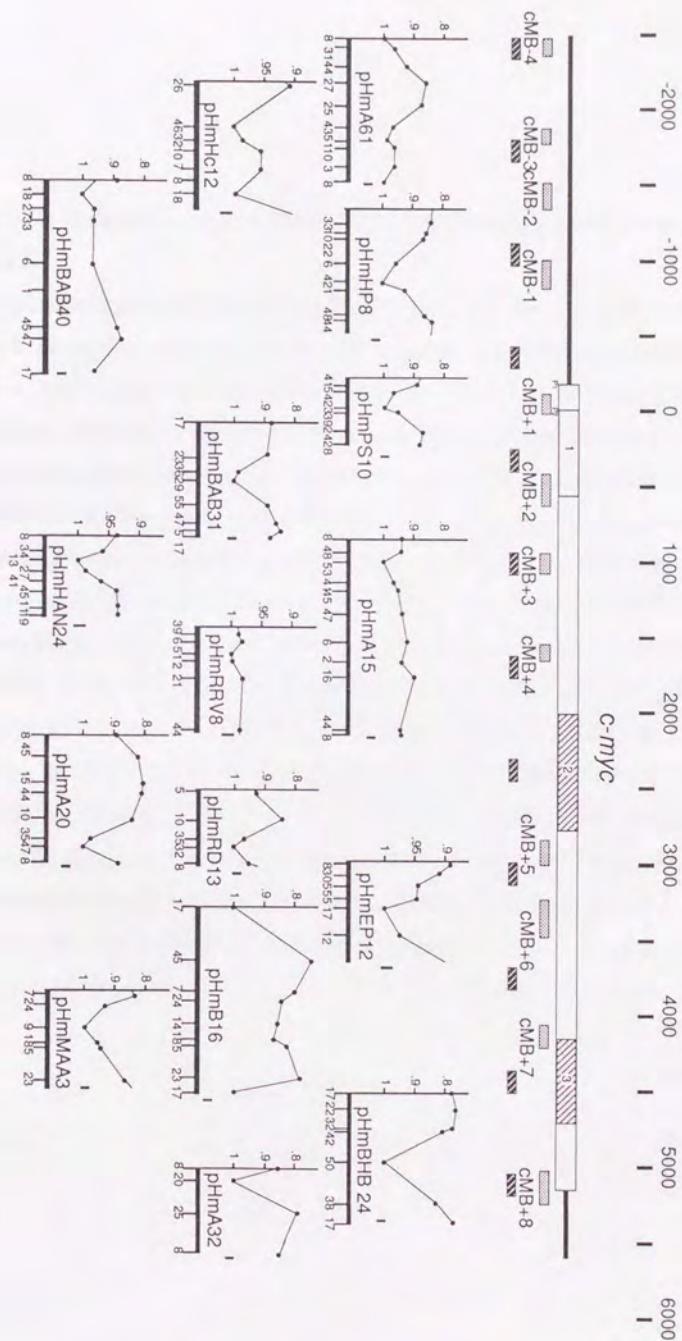pattern of the gel is shown.

Fig. 22. **Summary of the DNA bend sites in the human c-*myc* gene locus.**

The plasmid containing duplicates of the region shown as the thick horizontal line in each panel was digested with the enzymes shown below. The vertical bars in the panel show the average thickness of the bands relative to the migration distance, indicating a deviation of mobility. The bend sites, tentatively defined as the regions between the two restriction sites located on both sides of the restriction site that produces the fragments of the highest mobility, are shown as shadowed boxes. Boxes with stripe indicate putative 150 bp bend sites at 680 bp intervals. The restriction enzymes used are:1; *Acc*I, 2; *Afl*II, 3; *Afl*III, 4; *Alu*I, 5; *Alw*NI, 6; *Apa*I, 7; *Apa*LI, 8; *Apo*I, 9; *Ase*I, 10; *Ava*I, 11; *Ava*II, 12; Bam*HI, 13; *Ban*I, 14; *Ban*II, 15; *Bbv*I, 16; *Bcl*I, 17; *Bfa*I, 18; *Bgl*I, 19; *Bsr*GI, 20; *Bss*HII, 21; *Bst*71I, 22; *Bst*XI, 23; *Bst*YI, 24; *Cla*I, 25; *Dde*I, 26; *Dra*I, 27; *Dra*III, 28; *Eag*I, 29; *Ear*I, 30; *Eco*0109I, 31; *Eco*NI, 32; *Eco*RV, 33; *Fok*I, 34; *Hae*III, 35; *Hha*I, 36; *Hinc*II, 37; *Hinf*I, 38; *Hin*PI, 39; *Hpa*I, 40; *Hpa*II, 41; *Hph*I, 42; *Kpn*I, 43; *Mbo*I, 44; *Msc*I, 45; *Mse*I, 46; *Msl*I, 47; *Msp*I, 48; *Mun*I, 49; *Nae*I, 50; *Nar*I, 51; *Nhe*I, 52; *Nla*IV, 53; *Not*I, 54; *Nru*I, 55; *Nsi*I, 56; *Nsp*I, 57; *Pst*I, 58; *Pvu*II, 59; *Rsa*I, 60; *Sac*I, 61; *Sau*3AI, 62; *Sca*I, 63; *Ssp*I, 64; *Ssp*I, 65; *Sfc*I, 66; *Sty*I, 67; *Stu*I, 68; *Taq*I, 69; *Xba*I, 70; *Xho*I, 71; *Xmn*I.

Fig. 23. **Mapping of the DNA bend sites in the Igμ locus.**
DNA bend sites were determined by the circular permutation assay as described
for the c-*myc* locus. Additional restriction enzymes used are: 72; *Aat*II, 73;
*Aci*I 74;*Bgl*II, 75;*Blp*I, 76;*Bsa*I, 77;*Bsa*HI, 78;*Bsm*I, 79;*Bsp*HI, 80;*Bsp*MI,
81;*Bsr*BI, 82;*Cac*8I, 83;*Eco*RI, 84;*Hae*II, 85;*Hind*III, 86;*Mae*III, 87;*Nci*I,
88;*Nla*III, 89;*Pac*I, 90;*Pfl*MI, 91;*Sau*96I, 92;*Sma*I, 93;*Sph*I. The restriction sites
derived from the cloning vector are in parentheses.

sequences of the Sμ region (GAGCTGAGCTGGGGTGAGCT), the nucleotide sequence of the bend sites were not conserved. However, the sequence diversity of the region occurred so as to retain the DNA structure.

3-4-3. Molecular basis of DNA bending

As I mentioned in Introduction, the molecular basis for DNA bending is not fully understood. To examine the sequences that cause DNA bending in the region between -452 and +49 of the c-*myc* locus, I constructed four clones that contained serial deletions encompassing this site and determined the bend center within this region. Fig. 25A shows the map of the region in pHSR0.5-1 containing the bend site cMB+1. The clones pHmNa, pHmSb, pHmMc and pHmLd contained the regions -452 to -102, -452 to -52, -452 to -2 and -452 to +49, respectively. The whole bend site was retained in the plasmid pHmLd and a portion in pHmMc and pHmSb, while the site was completely deleted in pHmNa. When each of the plasmid DNA digested with *Bam*HI or *Sma*I was resolved by electrophoresis at 4°C, plasmids pHmLd and pHmMc exhibited lower migration rates for the *Sma*I fragments (S) than the *Bam*HI fragments (B), indicating the presence of the bend site in the cloned fragments (Fig. 4B). On the other hand, both of the *Sma*I and *Bam*HI fragments derived from pHmSb and pHmNa migrated equally, demonstrating that the bend site was absent in these clones. As summarized in Fig. 25C, the *Sma*I fragments derived from the plasmids containing the positions -53 to -2 (plasmids pHmMc and pHmLd) were retarded by ~ 7% compared to the *Bam*HI fragments, indicating the presence of the bend center in this region.

To further determine the DNA bend center in this region, DNA bending was examined using 30-mer, unidirectionally concatenated

oligonucleotides containing sequences -51 to -22 (LA+LT) and -31 to -2 (RA+RT). The pattern of the polyacrylamide gel is shown in Fig. 25C and the summary of the results is shown below, where the R_L values (ratio of apparent length to real length) of each band were plotted against their real length. As shown in the figure, only RA+RT marked showed retardation compared to $(dA)30+(dT)30$, suggesting that the DNA bend center of cMB+1 was located within the positions -31 to -2. The A4N5T4N6T3 sequence was likely to be the bend center, because it contained poly(dA) and poly(dT) tracts at intervals of roughly 10 bp.

cMB-4	cMB-3	cMB-2	cMB-1	cMB+1	cMB+2
-2461	-1859	-1501	-953	-97	434
GAATTTTTTT	ACATGCTATA	AAATTTTAAT	GGCCCCACGG	CTCGAGAAGG	CACTTTAATG
TTCGTCTATG	CACACACCCC	TAGCTCAAGA	AAGCCTGAGC	GCAGGGCTTC	CTGAGATGAG
TACTTGTGAA	TTTCCCCCGA	CTCCCCCCCC	AGGCGGGGCA	TCAGAGGCTT	TCGAATGCCT
TTATTTCACG	ATTGTTTTCT	CCCCCAAAAA	GGAGGGGCGG	GGCGGGAAAA	AAATAGGGTG
TTTGCCATTA	CTTTTGGAGG	AAGGCACGGA	TATCTGCTGC	AGAACGGAGG	TCTTTTCTCC
CCGG -2408	TGGTGGAGGG	AGTAATACTC	TTTGGCAGCA	GAGGGATCGC	CATTCCTGCG
	AGAGAAAAGT	CTCTCCTCTT	AATTGGGGGA	GCTGAGTATA	CTATTGACAC
	TTACTTAAAA	CTTTGATCAG	CTCAGTCTGG	AAAGCCGGTT	TTTTCTCAGA
	TGCCTTTGGG	AATCGATGCA	GTGGAAGGTA	TTCGGGGCTT	GTAGTTATGG
	TGAGGGACCA	TTTTTTGTGC	TCCAATCCAG	TATCTAACTC	TAACTGGGGC
	AGGATGAGAA	ATGACCGCAT	ATAGCTGTGC	GCTGTAGTAA	TGGGGTGGGG
	GAATGTTTTT	TTCCAATAAT	ATACATAATG	TTCCAGCGAG	GGTAATCCAG
	TGTTTTTCAT	AAAAGGGGAA	CATAATACAT	AGGCAGAGGG	AACTGGATCG
	GCCGTGGAAT	AGAGGACCTG	GACTCCCCCC	AGCGAGCGGG	GGGTAAAGTG
	AACACAAAAT	GAAAGGAATT	AACAAATGCA	CGGCCGGGC 51	ACTTGTCAAG
	AAAAAATCCC	AAACGTCCGG	ATGGGAGTTT		ATGGGAGAGG
	GAG -1697	TTTGTCCGGG	ATTCATAACG		AGAAGGCAGA
		GAGGAAAGAG	CGCTCTCCAA		GGGAAAACGG
		TTAAC -1317	GTATAC -767		GAATGGTTTT
					TAAGACTACC
					CTTTCGAGAT
					TTCTGCCTTA
					TGAATATATT
					C 664

cMB+3	cMB+4	cMB+5	cMB+6	cMB+7	cMB+8
953	1555	2860	3252	4087	5047
GATCCTTTTA	GGGCCCCGGC	TTAACGGGCC	CAGGGCCTGC	TGTACAGCAT	GAATTTCAAT
AGAAGTTGGC	GCGGAGCGGG	ACTCTTATTA	CTGAGTGCGG	TAATCTGGTA	CCTAGTATAT
ATTTGGCTTT	GTTCACGCAG	GGAAGGAGAG	GAGCCAGTGA	ATTGATTATT	AGTACCTAGT
TTAAAAAGCA	CCGCTAGCGC	ATAGCAGATC	ACTGCCTCAA	TTAATGTAAC	ATTATAGGTA
ATAATACAAT	CCAGGCGCCT	TGGAGAGATT	GAGTGGGTGG	CTTGCTAAAG	CTATAAACCC
TTAAAACCTG	CTCGCCTTCT	TGGGAGCTCA	GCTGAGGAGC	GAGTGATTTC	TAATTTTTTTT
GGTCTCTAGA	CCTTCAGGTG	TCACCTCTGA	TGGGATCTTC	TATTTCCTTT	TATTTAAGTA
GGTGTTAGGA	GCGCAAAACT	AACCTTGGGC	TCAGCCTATT	CTTAAAGAGG	CATTTTGCTT
CGTGGTGTTG	TTTGTGCCTTG	TTTAGCGTTT	TTGAACACTG	AGGAACAAGA	TTTAAAGTTG
GGTAGGCGCA	GATTTTGGCA	CCTCCCATCC	AAAAGCAAAT	AGATGAGGAA	ATTTTTTTCT
GGCAGGGGAA	AATTGTTTTC	CTTCCCCTTA	CCTTGCCAAA	GAAATCGATG	ATTGTTTTTA
AAGGGAGGCG	CTCACCGCCA	GACTGCCCAT	GTTGGACTTT	TTGTTTCTGT	GAAAAAATAA
AGGATGTGTC	CCTCCCGCGG	GTTTGCAGCC	TTTTTTTCTT	GGAAAAGAGG	AATAACTGGC
CGATTCTCCT	CTTCTTAAG	CCCCTCCCCG	TTATTCCTTC	CAGGCTCCTG	AAATATATCA
GGAATCGTTG	1693	TTTGTCTCCC	CCCCGCCCTC	GCAAAAGGTC	TTGAGCCAAA
AC 1104		ACCCCTCAGG	TTGGACTTTT	AGAG 4245	TCTTAAGTTG
		AATTT 3024	GGCAAAACTG		TGAATGTTTT
			CAATTTTTTTT		GTTTCGTTTC
			TTTTTTATTT		5335
			TTCATTTCCA		
			GTAAAATAGG		
			GAGTTGCTAA		
			AGTCATACCA		
			AGCAATTTGC		
			AGC 3494		

IgB+1

381
CCGGCCCGGG
ACAGTCGGAG
AGTCAGGTTT
TTGTGCACCC
CTTAATGGGG
CCTCCCACAA
TGTGGCTACT
TTGACTACTG
GGGCCAAGGG
ACCCTGGTCA
CCGTCTCCTC
AGGTGAGTCC
TCACAACCTC
TCTCCTCCTT
TAACTCTGAA
GGGTTTTGTT
GACTTTTGGG
GGAATAAGGG
TGCTGGGGGC
CTGCCAAGAG
AGCCCCGGAG
CAGCCCTGGG
GGCTGCAGGA
GGCCTGAGGC
AACAGCGGCA
CACACAGACG
AGGGGCAAGG
GTCTCCAGAT
GCTCCTTCCT
CCTGAGCCAG
CAGCACGGGT
TCGTCTCGGC
GCCAGGGCCA
CCC 714

IgB+2

1167
CAGGGGCTGT
CGTGATGATT
GCGTGGTGAC
TCTGTCCCGC
TCCAAGACCC
GCTCTCTGGG
CGGGTGCCCC
CCGGGGTTTT
TGGACTCCTG
GGGGTAACTT
GCGGCCGTCT
GCTTGCGGTT
GGACTTCCCA
GCCGACAGTG
GTGGTCTGGC
TTCTGAGGGG
TCAGGCCAGA
ATG 1339

IgB+3

1890
AATATTTTCT
TTAGAATTAT
GAGGTGCCTT
TGCACTGCAT
CTAATCTAT
1938

IgB+4

2109
GCATGCTTAC
TGTTAAAAGA
CAGGATATGT
TTGAAGTGGC
TTCTGAGAAA
AATGGTTAAG
AAAATTATGA
CTTAAAAATG
TGAGAGATTT
TCAAGTATAT
TAATTTTTTT
AACTGTCCAA
GTATTTGAAA
TTCTTATCAT
TTGATTAACA
CCCATGAGTG
ATATGTGTCT
GGAATTGAGG
CCAAAGCAAG
CTCAGCTAAG
AAATACTACA
GTGCTGTCGG
CCCCGATGCG
GGACTGCGTT
TTGACCATCA
TAAATCAAGT
TTATTTTTT
2377

IgB+6

3666
CTAAGTCATT
GACTGTAGGT
CATCATCGCA
CCCTTGAAAG
TAGCCCATGC
CTTCCAAAGC
GATTTATGGT
AAATGGCAGA
ATTTTAAGTG
GCAAATTCAG
ATAAAATGCA
TTTCTTGGTT
GTTTCCAATG
ATGACTGTTA
TCTAGA 3806

IgB+7

4027
CCAGCCTGGC
TGTGCAGGAA
CCCGGCAATG
AGATGGCTTT
AGCTGAGACC
AAGCAGGGTC
TGGTGGGCTG
ACATTTCTGG
CCATGACACT
CATCCAGCTT
TCAGAAATGG
ACTCAGATGG
CTAAACTGAG
CCTAAGCTGA
GCCTAGACTA
ACAGGCTGAA
CTGGGCTGAG
CTGAGCTGAA
CTGGGCTGAG
TTGAACTGGG
TTGAGCTGAG
CTGAGCTGAG
CTGGGCTAAG
TTGCACCAGG
4266

IgB+8

4699
CGGGTTGAGC
GTGCTGTGCT
GGGCTGAGCC
AAGCTAGGCT
GAGCTGAGCC
AAGTTGAGCT
TAGCTGGGCT
GAGCTAACCT
GGGCAGGGCT
GAGCTGGGCT
GAGCTAACCT
GGACTGGGCT
GAGCTAACCT
GGGCAGAGCT
GAGCTGGGCT
GAGCTAACCT
GGGCTGGGCT
GAGCTAACCT
GGGCTGGCGT
CAGCTGAGCT
GACGTACGCT
GGGCTGGGCT
GGGCTGAGCC
GAGCTGAACT
GGGCTGAGCA
GGCTGTGTCG
GGCTGAGCCA
AGCTGG 4974

IgB+9

5703
ACTGAGCTGA
CCTAGGCTGA
GCTGAGCTGA
GC 5724

IgB+10

6014
TTAAGCTAAG
CTGAGCTGAA
CTGGGCCTGT
GCTGAGCTAG
GCTGGGCTGA
GCTGAGCTGG
GCTGGGCTGG
GCTGAGCCAG
ATTGTGCCTG
GCTGAACTGA
GCTGGGCTAA
GCTGAGCTGG
GCTGAGCTGA
GCTGGGCTGA
GTGGGGCGGG
GCTGAGCTGA
GCCGGACTGG
GCTGGGCTGG
GCTC 6207

IgB+11

6973
GCTCTGCTGT
GCTGTGCTGA
GCAGGGCTGA
GCTGAACTGG
GCTGAGCTGG
GCTGAGCCGG
GCTGAGTTGA
GCAGAGCTGG
GTTGAGCAGA
GCTGGGCTGA
GTTGAGCCAG
GCTGACCTGG
GCTGAGCCAA
GCTGGGTTGA
GCCAAACTGG
GCTGGGCTGG
GCTGAGCCAG
GCTGGATTGA
GCTGGGCTGG
GCTGGGCTAA
GCTGGGCTGT
GCT 7195

Fig. 24. Sequence features of DNA bend sites in the c-*myc* and Igμ loci.

Nucleosome sequences of the DNA bend sites cMB-4 to cMB+8 from the c-*myc* locus and IgB+1 to IgB+11 form the Igμ locus are shown to highlight the presence of short poly(dA) or poly(dT) tracts. Poly(dA) or poly(dT) tracts of more than or equal to three nucleotides long are lightly shadowed for cMB-4 to cMB+8 in the c-*myc* locus and IgB+1 to IgB+11 in the Igμ locus. For the bend sites in the S region, poly(dA) or poly(dT) tracts of more than or equal to two nucleotides long are heavily shadowed.
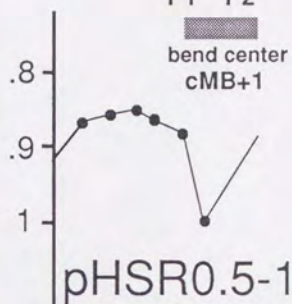
A

c-myc

P1 P2

bend center
cMB+1

.8
.9
1

pHSR0.5-1

BssHII BsmAI SmaI BstYI NotI XhoI

SmaI BamHI

pHmNa
pHmSb
pHmMc
pHmLd

-70   -60   -50   -40   -30   -20   -10   -1
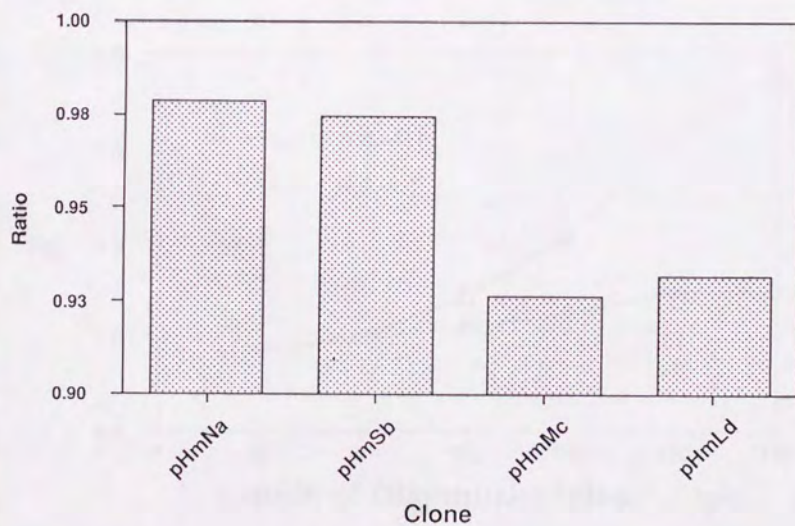GGCTTGGCGGGAAAAAGAACGGAGGGAGGGATCGCGCTGAGTATAAAAGCCGGTTTTCGGGGCTTTATCTAAC
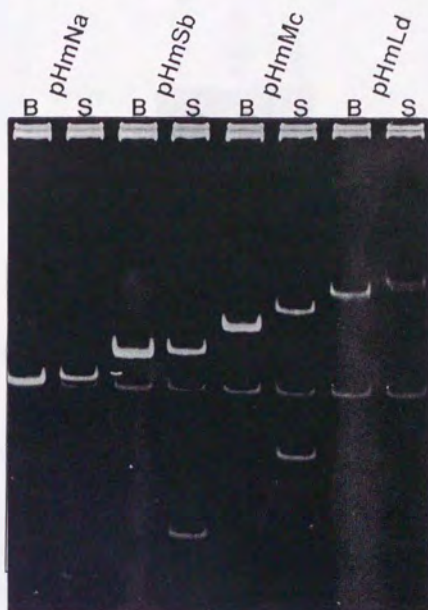
← pHmSb →

← pHmMc

LA+LT   RA+RT

1   10   20   30   40   50   60
TCGCTGTAGTAATTCCAGCGAGAGGCAGAGGGAGCGAGCGGGCGGCCGGCTAGGGTGGAAGAGCC

← pHmLd

B

C

Origin

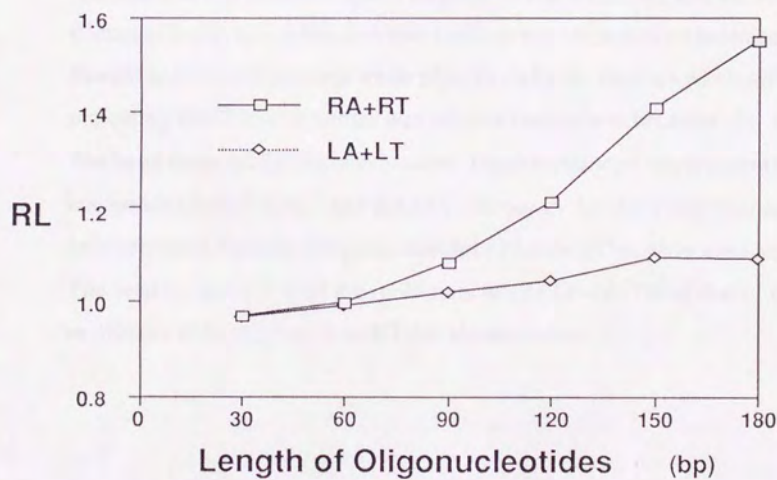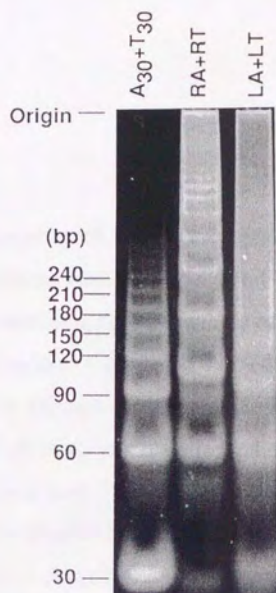A₃₀⁺T₃₀   RA+RT   LA+LT

(bp)
240
210
180
150
120
90
60
30

Fig. 25. **Fine mapping of the bend center near the P2 promoter.**
(**A**) Deletion constructs near the DNA bend center cMB+1 (positions -97 to
51). Each clone containing tandem dimers of the regions between -458 and -102
(pHmNa), -52 (pHmSb), -2 (pHmMc), or +49 (pHmLd) were created by PCR
using 28-nucleotide primers containing *Bam*HI site at the 5' end and cloned into
the *Bam*HI site of pBluescript. Therefore, the *Bam*HI digestion produces
fragments of one unit long. (**B**) The circular permutation assay. Each clone was
digested either with *Bam*HI or *Sma*I and electrophoresed on a polyacrylamide
gel at 4°C. Since each clone has *Bam*HI site at both ends and *Sma*I site in
between, the *Bam*HI fragment of the clone which has the bend center near the
ends show a higher mobility than that of the *Sma*I fragment. The relative
mobilities of the *Sma*I to *Bam*HI fragment were calculated and are shown.
Clones pHmMc and pHmLd showed differences in mobilities between the
*Bam*HI and *Sma*I fragments while pHmSb and pHmNa showed no difference,
indicating that the bend center was located somewhere between -51 and -2. (**C**)
The bend assay using oligonucleotides. Double-stranded oligonucleotides, 30
nucleotides long LA+LT and RA+RT (shown in **A**), were electrophoresed after
unidirectional ligation. Oligonucleotides (dA)30+(dT)30 were used as a control.
The relative mobilities of the multimers of (dA)30+(dT)30 to that of the
multimers of LA+LT or RA+RT are shown below.

4. Discussion

4-1. Preference of the recombination sites involved in excision of the human alphoid *Sau*3A repeat family

I described here the nature of the recombination involved in the rearrangement of a human repetitive sequence family, the alphoid *Sau*3A family. To efficiently clone the fragments containing the recombination junctions, I used PCR amplification (Figs. 6 to 8). The PCR products were then cloned and subjected to the sequence analysis. To avoid the amplification of the fragments with complete units, EC DNA was first digested with *Sau*3AI, *Bfa*I or *Xba*I which cut the family sequence, and the fragments without these sites were amplified by PCR with the primers located on both sides of the restriction sites. Therefore, only the fragments which lacked the sites by rearrangement or point mutation could be amplified. As shown in Fig. 8, all the cases of PCR amplification of EC DNA showed at least one additional band which was not present in the PCR products from chromosomal DNA. Since the amounts used for PCR were same between EC DNA and chromosomal DNA in Fig. 8, the additional bands observed in the products from EC DNA were not the artifacts generated by PCR with specific primers. These bands were also observed in a separate experiment using the unamplified EC DNA as a source of PCR (Fig. 10).

Since only fragments including recombination junctions derived from the region without the restriction site (*Sau*3AI, *Bfa*I or *Xba*I) or fragments that lost the sites by point mutation were enriched, there might be a case of representing the recombination junctions from too small numbers of DNA fragments. To exclude this possibility, I used three restriction enzymes (*Sau*3AI, *Bfa*I or *Xba*I) which cut different

sites, and three pairs of specific primers (PCR1 to 3). When the recombination junctions were analyzed, clones obtained by PCR1 contained many types of recombination junctions (Fig. 10), indicating that they represented a wide variety of recombination junctions. On the other hand, clones from PCR2-A and especially PCR3 were likely to be derived from a few major species in the population, probably because both of the PCR schemes amplify a smaller region, and therefore, reduce the number of recombination events that could be detected.

Sequence analysis of the recombination junctions revealed that the subunit structure was retained through the homologous recombination and, based on comparison of scattered mismatches between the subunits, I mapped the recombination junctions (Figs. 9 and 10). The distribution of the 68 junctions showed that most (55/68) of the sites were centered at positions 10-25 (A), 40-50 (B), 85-90 (C) and 135-160 (D). As expected, regions with high sequence homology between subunits exhibited higher recombination frequencies (regions B, C and D). Interestingly, region A did not correspond to those with high sequence homology between the subunits and region D showed higher recombination frequency compared with regions B and C, suggesting that another factor(s) affects the recombination frequency. The higher recombination frequency compared with the regional homology (shown in Fig. 11A) in both regions A and D is statistically significant as judged by $p<0.05$. This tendency was absent in the regions B and C.

It has been shown that the alphoid (or α for primates) satellite DNA has several sequence features characteristic of protein binding motifs. Strauss and Varshavsky (78) reported that α protein binds preferentially to three regions (I to III), which are located close to each other spatially on the nucleosomes, and regions II and III are located next to the linker region. This protein is presumably essential for the

function of positioning nucleosomes along the subunits. On the other hand, there are sequence motifs for CENP-B (79, 80), a protein which is a part of kinetochore structure and binds specifically to alphoid satellite DNA, at positions 2 to 18 of subunits 3 and 5. This protein might play an important role in aligning the subunits. The preference of ± 2 over ± 1 subunit shift (Table 2) could be explained by this specificity. Meanwhile, the recombination hotspots described here were located next to the linker region and overlapped with these protein binding sites (Fig. 26). Furthermore, analysis of the recombination sites in the chromosomal DNA also indicated one of these regions (positions 135-160) to be recombinogenic (81). Therefore, these regions are not only susceptible to binding proteins such as α protein and CENP-B, but also to the proteins which mediate DNA rearrangements. Such preference of the recombination events at the linker region has been suggested from nonrandom size distribution of EC DNA in several cell types (32). In contrast, Kawasaki *et al.* (82) reported that no such preference of the recombination sites was observed for artificially constructed tandem repeats of the primate a satellite DNA. This could be explained by the involvement of a protein specific to the human alphoid DNA which enhances the regional recombination frequency, or by differences between the artificial and the intact repeat constructs.

4-2. DNA bend sites and the chromatin structure in the *Sau*3A family

DNA bend sites have been studied with respect to recombination, transcription, replication and the chromatin structure including nucleosome phasing (58-67). I have demonstrated here the presence of DNA bend sites within the *Sau*3A family repeats. Although nucleosome phasing in alphoid DNA was well documented (83, 84), it has not been discussed in association with DNA bending. Several satellite DNAs,
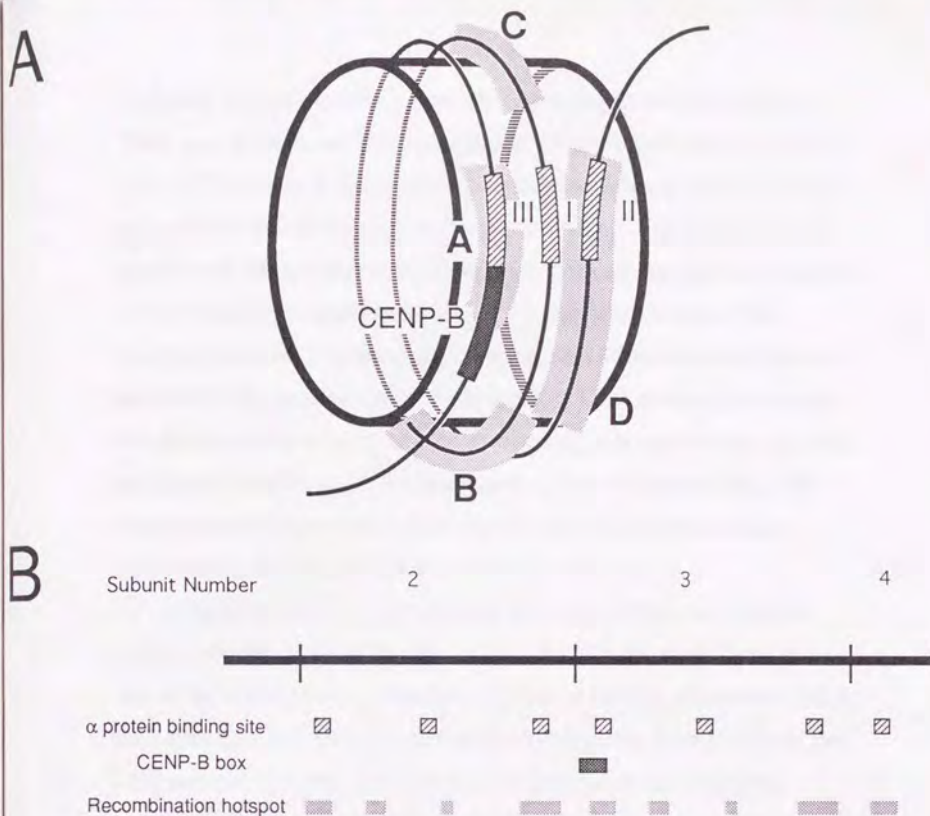
Fig. 26. **Schematic illustration of the recombination hotspots and the potential binding sites of α protein and CENP-B (A) and their map in the subunit structure (B). (A)** Recombination hotspots (shadowed) are illustrated along with the α protein binding sites (striped) and the CENP-P box (solid) on the nucleosome structure. **(B)** Position of the α protein binding sites (74-80 for I, 151-157 for II and 17-23 for III), the CENP-B box (2-18) and the recombination hotspots are mapped on the *Sau*3A family subunits. The linker DNA of the nucleosome (positions 157-17) is located between the sites II and III.

including α-satellite, have repeat units of multiples of the length of DNA in a nucleosome. For example, the 119 bp repeat from *Drosophila*, the 372 bp repeat from rat and the α-satellite from primates (172 bp repeat) have nucleosome repeat lengths of 240 bp, 174 bp and 172 bp, respectively (85), which are roughly the same size (primates), or half of (rat) or twice (*Drosophila*) the size of the nucleosome units. This suggests that a nucleosome phasing signal, possibly a sequence-directed specific DNA structure, exists in the satellite DNAs. By performing the circular permutation assay and the assay using oligonucleotides (detailed in Materials and Methods, and in Results), I have mapped three DNA bend sites, which could be a candidate for the nucleosome phasing signal, within the *Sau*3A family unit (discussed below).

Unexpectedly, they did not exist in every subunit but appeared only in subunits 1, 3 and 5 (Figs. 13 and 14). The distances between any two of the bend sites are, therefore, 170 bp or 340 bp. In contrast, DNA bend sites appear at intervals of 680 bp in the human β-globin locus and other loci (58-60). This difference in the interval of the bend sites probably reflects on the difference in the chromatin structure between the centromere region (for the *Sau*3A family) and other regions. Since the bend sites generally exhibit a high affinity to histones (86), one could assume that a higher frequency of the bend sites could facilitate a specific condensation of the nucleosomes at the centromere by tight association of DNA to the core histones.

Fitzgerald *et al.* demonstrated by a computer analysis that conserved patterns of DNA bending exist in the satellite DNAs (85). They studied 57 types of satellites from a number of species and found a conserved structure in the satellite units. This structure, in which two bending elements of 50-60 bp long are separated by a 20-30 bp non-bending element, resembled the structure exhibited by the DNA within

the nucleosomes. One of the two bending elements determined by the computer analysis included the region of the bend sites shown in Figs. 12 to 14.

4-3. DNA bend sites and recombination in the *Sau*3A family

There were reports that suggest involvement of bend sites in recombination. For example, human immunodeficiency virus (HIV) and murine leukemia virus prefer to integrate at DNA bend sites, and the IHF protein, which functions at the integration of λ phage, introduces DNA bending at the integration site (65, 67). One of the recombination hotspots (A; positions 10 to 25) of the *Sau*3A family DNA described above, overlapped with the bend sites (Figs. 11A and 14), suggesting that the bend sites are also favored for the recombination of the *Sau*3A family. Furthermore, there were also reports that a specific position within the nucleosome structure is exposed to recombination machinery. HIV prefers to integrate into nucleosomal DNA rather than naked DNA, with a significant preference for the integration site (67). HIV integrates into the nucleosomal DNA at the distorted major groove which is located at the positions approximately 15, 35 and 55 bp away from the dyad axis. Hence, these studies support the idea that recombination is affected by the chromatin or nucleosome structure. The recombination observed in the *Sau*3A family occurs frequently at the nucleosomal linker-core region so as to retain the subunit structure, whereby retaining the nucleosomal phases. This indicates a close relationship among recombination, nucleosomal phases and DNA bend sites.

Recombination of the *Sau*3A family discussed above was also shown to have a preference in the subunit usage (shown in Table 2 and Fig. 11B). The intervals of the bend sites in the *Sau*3A family are 340 bp (between the subunits 1 and 3 and between the subunits 3 and 5) or

81

170 bp (between the subunits 5 and 1). Among the combinations of the subunits used in recombination, most of them positioned the bend sites after recombination at the intervals of 340 bp or 170 bp. However, in the case of recombination using subunit combinations 2/1, 2/3, 4/3 and 4/1, one of the intervals between the bend sites after the recombination is 510 bp. Although 13 or 14 cases were expected to occur in 68 recombination events for the combinations of 2/1, 2/3 4/1 and 4/3 if they occur randomly, only two such cases were shown to occur (X-120 and X-122; although the recombination in clone X-125 involved subunits 4 and 1, it did not create a 510 bp interval) (Fig. 10). This indicates that creating a 510 bp distance was not favored in these recombination events. The interval of the bend sites observed in the human β–globin locus was 680 bp on average (58-60). The computer analysis also revealed that the intervals of the potential bend sites are about 700 bp with a minor peak at 350 bp (87). It seems, therefore, that the interval of 510 bp may cause a structural instability and would not be favored in the cells.

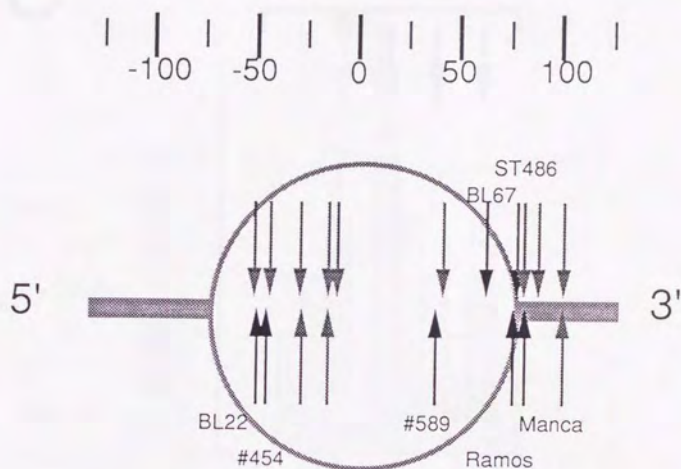4-4. Nucleosomal phases around the translocation sites in the c-*myc* and Igμ loci

In Burkitt's lymphoma, the c-*myc* oncogene located on chromosome 8 is frequently translocated into the immunoglobulin heavy chain locus located on chromosome 14 (20). This tranlocation causes deregulation of the c-*myc* gene expression, driven by the adjacent elements on the immunoglobulin locus, including the locus control region located 3' of Cα gene, or by disruption of the transcription-regulatory elements present in the 5' region and in the first intron of the c-*myc* gene (20). I have determined the nucleosomal phases in the c-*myc* and Igμ regions using reconstituted nucleosomes (Figs. 15 to 18). The

nucleosomes were phased continually, nearly as the hypothetical phases, in both regions. Nucleosomal phases at the translocation sties in these regions are summarized in Fig. 27. Surprisingly, the translocation junctions within the nucleosomes were likely to be centered at three distinct positions. Positions I and II were located near the nucleosomal core-linker junction and overlapped with the recombination hotspots observed in the recombination of the *Sau*3A family to form EC DNA, which are shown by shadowed arrows in Fig. 27. (A and D, shown in Fig. 11A). On the other hand, positions II and III were located at those preferred by HIV integrase within the nucleosome structure, which are shown by solid arrows in Fig. 27 (67). The observation that the specific positions within the nucleosome were preferred in translocation further strengthens the idea that the recombination is influenced by the chromatin structure.
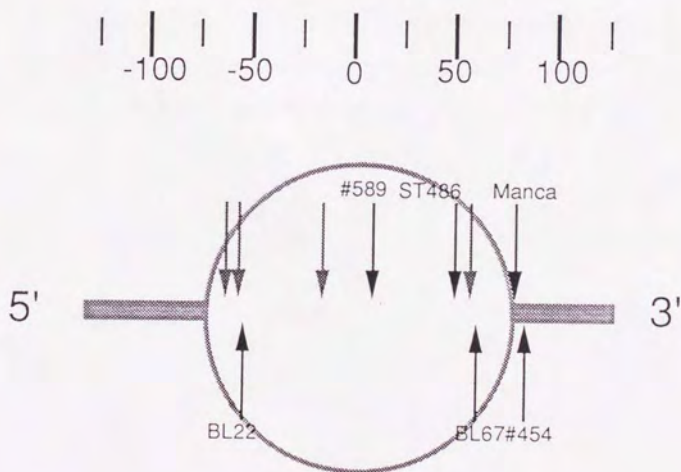
4-5. DNA bend sites in the c-*myc* and Igμ loci

Here, I mapped the DNA bend sites in the c-*myc* and the Igμ loci (Figs. 22 and 23). A total of 12 bend sites in the 8082 bp c-*myc* gene region and 11 sites in the 7427 bp Igμ region were mapped by the circular permutation assay. The average intervals of the bend sites were 693 ± 282.3 bp for the c-*myc* locus and 650 ± 206.8 bp for the immunoglobulin μ locus. These values were close to those obtained in the ε-globin (682.5 ± 132.0 bp), β–globin (685.5 ± 267.7 bp) and ψβ– globin gene (642 ± 178.8 bp) regions (60), indicating that the periodicity of DNA bend sites at intervals of approximately 700 bp on average also exists in these regions. However the periodicity was disturbed around the exons and no bend sites were mapped within the coding regions. Analysis of the sequence features of the bend sites revealed that short poly(dA)·poly(dT) tracts appeared invariably.

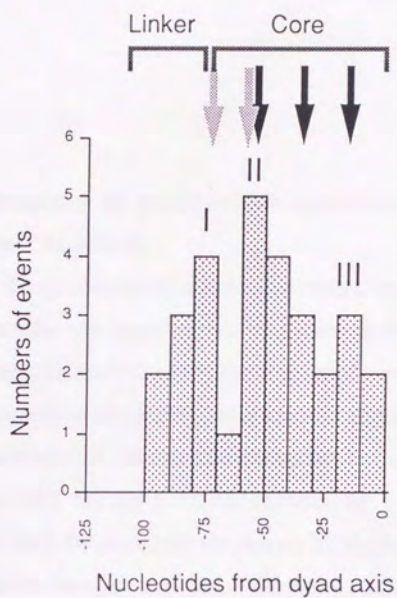A c-myc locus

B Immunoglobulin μ locus

Fig. 27. **Summary of translocation junctions within the nucleosome structure.**

(A and B) The positions of the translocation sites within the nucleosome structure were mapped according to the nucleosomal phases shown in Figs. 16 and 18 and to the reports of translocations junctions (3-u). The positions (shown by solid arrows) and the names of the cell lines are indicated for the translocation junctions observed in both the c-*myc* and Igμ loci. For the junctions between the c-*myc* or the Igμ and other locus, only the positions are shown by shadowed arrows. **(C)** The summary of the translocation sites within the nucleosomal structure. Three peaks (I to III) indicate the hotspots for translocation. Shadowed arrows indicate the positions of the recombination hotspot (A and D) observed in the recombination of the *Sau*3A family, while the solid arrows indicate the sites where HIV integration preferentially occurs. The positions shown below the graph indicate the distance from the dyad axis of the nucleosome structure.

Detailed analysis of a bend site, cMB+1, by the oligonucleotide bending assay revealed the bend center at the P2 promoter of the c-*myc* gene (Fig. 25). The A4N5T4N6T3 sequence observed in this region was most likely to cause bending. This bend center was mapped inside the nucleosomal core region, 10 to 40 bp away
from the core-linker junction. Interestingly, these positions in the nucleosome structure were very similar to those observed in the β-globin locus (12 to 33 bp inside from the core-linker junction) (60). This further supports the hypothesis that the bend sites function as a signal for phasing nucleosomes, which has also been suggested by the observation that both of the DNA within the phased nucleosome and bent DNA share a common sequence feature of short poly(dA)·poly(dT) tracts (42, 47, 68, 69).

In Fig. 28, I schematically illustrated the nucleosomal array along with the bend sites based on the hypothesis that four nucleosomes can be placed between the periodically appearing bend sites at 680 bp intervals. As shown in Fig. 28, 5 out of 6 translocation events (83 %) involved the nucleosomes containing the bend site in at least one of the loci. However, when these events occur randomly between these loci, the probability of such cases would be 44% (7/16, provided that every four nucleosomes in each locus contains the bend site). The higher frequency of the events suggests that although the bend sites are not required on both loci, they are needed in the translocation process. On the other hand, there were no cases that resulted in the bend sites located next to each other. Such constructs may not be favored in the process.

4-6. DNA bend sites and DNase I hypersensitive sites

In the recent studies on the chromatin structure in association with transcription and replication, the assays were mainly based on the accessibility
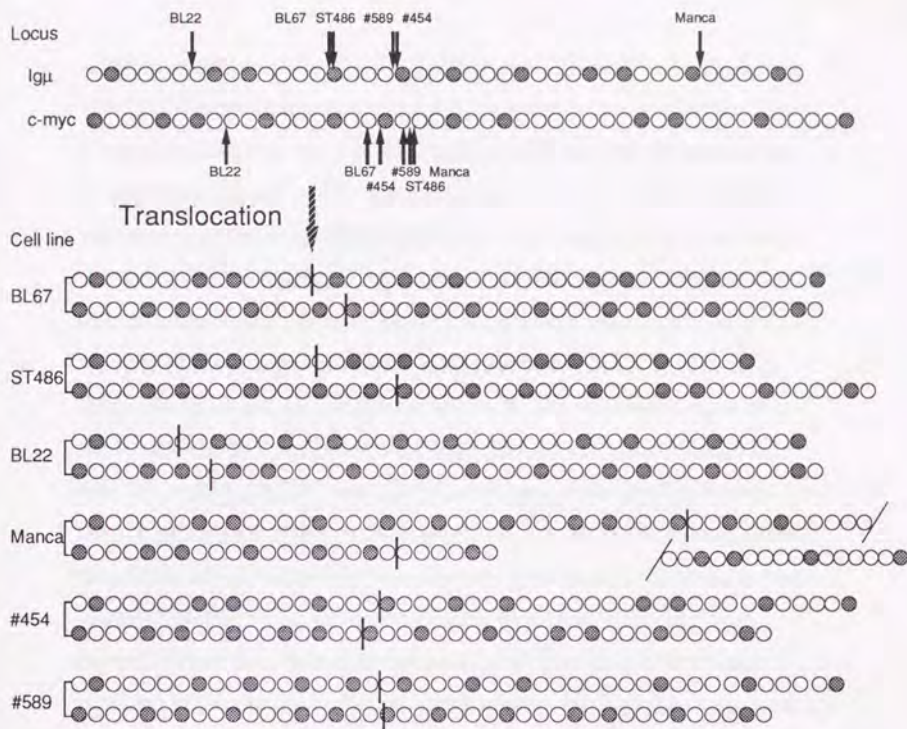
Fig. 28. **A scheme of the translocation between the c-*myc* and immunoglobulin heavy-chain μ loci.**

The derivative 14 chromosome (above) and derivative 8 (below) are shown. The translocation junctions are shown by solid arrows in the intact Igμ and c-*myc* loci and by vertical bars in the derivative chromosomes. The positions of the nucleosomes (shown by circles) were deduced by the hypothesis that bend sites are the signal for nucleosome phasing and phases four nucleosomes at a 680 bp interval. Although most of them are hypothetical, some were experimentally demonstrated (Figs. 16 and 18). The nucleosomes including the bend sites are shown as shadowed circles.

of the enzymes, such as micrococcal nuclease, DNase I and S1 nuclease (88-93). The developmental and cell-type specific changes in the DNase I hypersensitive sites have been extensively studied and the implication as to their relationship with the transcription and replication activities has been elucidated (89-93). The DNase I hypersensitive sites are often located at critical regions for gene expression such as the binding site of transcription factors. Seven developmentally and transcriptionally regulated DNase I hypersensitive sites (positions -2012, -1541, -912, -285, -70 +639 and +1639 relative to the P2 cap site) exist in the 5' region and the first intron of the c-*myc* gene (3-r). Among them, five sites (II1, , III1, III2, IV and V) were mapped within the DNA bend sites (Fig. 22), while the remaining two sites (I and II2) were mapped in the middle of the two bend sites. The same results were obtained in the human β-globin locus, where four minor hypersensitive sites were mapped within the DNA bend sites and other four major sites were localized in the middle of the two bend sites. In the β-globin locus, the sites that were located at the bend sites were the major hypersensitive sites where protein binding motif is located, and those located in the middle were located in the promoter region and exhibited an open chromatin structure. There were several reports that hypersensitive sites appear either at the binding sites for specific *trans*-acting factors or at the positions where tightly-phased nucleosomes exist (93). Thus, one could speculate that the hypersensitive sites located in between the two bend sites and those located at the bend sites are formed by the different molecular mechanisms.

There are several reports that demonstrated the stability of the chromatin structure after translocation. For example, DNase I hypersensitive sites of the c-*myc* locus were conserved in the lymphoma cell lines containing the translocation points in the upstream or

downstream of the c-*myc* gene. No changes in the pattern were observed in all hypersensitive sites which are associated with the c-*myc* gene after translocation, even though the translocation occurred in the promoter region of the c-*myc* gene (91). This further supports the idea that a mechanism of stabilizing the chromatin structure, by keeping the bend sites at appropriate distances for example, exists in the translocation process.

4-7. Conclusions and future prospects

In this thesis, I described two types of recombination with respect to the chromatin structure. One is the homologous recombination that occurs at the excision of the *Sau*3A family DNA to form EC DNA. The other is the recombination observed in the translocation of the c-*myc* locus into the immunoglobulin μ locus, which occurs in the process of neoplastic transformation of lymphocytes. The results described here revealed that the chromatin structure plays a crucial role in both types of recombination. The stability of the chromatin structure should be precisely controlled to maintain the hierarchy of the genome structure in the nuclei. Further studies on the chromatin structure in other regions also critical for recombination would provide more evidence for this hypothesis.

References

1. Gelmann, E. P., Psallidopoulos, M. C., Papas, T. S. and Favera, R. D. (1983) Identification of reciprocal translocatiom sites within the c-*myc* oncogene and immunoglobulin m locus in a Burkitt lymphoma. *Nature*, **306**, 799-803.

2. Tsujimoto, Y., Gorham, J., Cossman, J., Jaffe, E. and Croce, C. M. (1985) The t(14;18) chromosome trnalocations involved in B-cell neoplasms result from mistakes in VDJ joining. *Science*, **229**, 1390-1393.

3. Haluska, F. G., Finver, S., Tsujimoto, Y. and Croce, C. M. (1986) The t(8;14) chromosomal translocation occurring in B-cell malignancies results from mistakes in V-D-J joining. *Nature*, **324**, 158-161.

4. Ikeda, H. (1986) Illegitimate recombination mediated by T4 DNA t opoisimerase *in vitro*. Recombinants between phage and plasmid molecules. *Mol. Gen. Gent.*, **202**, 518-520.

5. Low, K. B., and Porter, R. D. (1978) Modes of gene transfers and recombination in bacteria. *Annu. Rev. Genet.*, **12**, 259-287.

6. Pryciak, P. M., Sil, A. and Varmus, H. E. (1992) Retroviral integration into minichromosomes *in vitro*. *The EMBO Journal*, **11**, 291-303.

7. Wu, T.-C. and Lichten, M. (1994) Meiosis-induced double strand break sites determined by yeast chromatin structure. *Science*, **263**, 515-518.

8. Tsukiyama, T., Becker, P. B. and Wu, C. (1994) ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature*, **367**, 525-532.

9 Wilson, C. J., Chao, D. M., Imbalzano, A. N., Schnitzler, G. R., Kingston, R. E. and Young, R. A. (1996) RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell*, **84**, 235-244.

10 Honjo, T., and Kataoka, T. (1978) Organization of immunoglobulin heavy chain genes and allelic deletion model. *Proc. natl. Acad. Sci. USA*, **75**, 2140-2144.

11. Sakano, H., Huppi, K., Heinrich, G. and Tonegawa, S. (1979) Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature*, **280**, 288-294.

12. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W., and Tonegawa, S. (1980) The two types of somatic recombination necessary for generation of complete immunoglobulin heavy chain genes. *Nature*, **286**, 676-683.

13. Marcu, K. G., Lang, R. B., Stanton, L. W. and Harris, L. J. (1982) A model for the molecular requirements of immunoglobulin heavy chain class switching. *Nature*, **298**, 87-89.

14. Oettinger, M. A., Schatz, D. G., Gorka, C. and Baltimore, D. (1990) RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science*, **248**, 1517-1523.

15. Stanhope-Baker, P., Hudson, K., M., Shaffer, A., L., Constantinescu, A. and Schlissel, M., S. (1996) Cell type-specific chromatin structure determines the targeting of V(D)J recombinase activity in vitro. *Cell*, **85**, 887-897.

16. Shinohara, A., Ogawa, H. and Ogawa, T. (1992) Rad51 protein involved in repair and recombination in S. cevisiae is a RecA-like protein. *Cell*, **69**, 457-470.

17. Shinohara, A., Ogawa, H., Matsuda, Y., Ushio, N. and Ogawa, T. (1993) Cloning of human, mouse and fission yeast recombination genes homologous to Rad51 and recA. *Nature genetics*, **4**, 239-243.

18. Hollis, G. F., Mitchell, K. F., Battey, J., Potter, H., Taub, R., Lenoir, G. M. and Leder, P. (1984) A variant translocation places the λ immunoglobulin genes 3' to the c-*myc* oncogene in Burkitt's lymphoma. *Nature*, **307**, 752-755.

19. Kato, S., Tachibana, K., Takayama, N., Kataoka, H., Yoshida, M. C. and Takano, T. (1991) Genetic recombination in a chromosomal translocation t(2;8)(p11;q24) of a Burkitt's lymphoma cell line KOBK101. *Gene*, **97**, 239-244.

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

21. Brys, A. and Maizels N. (1994) LR1 regulates c-*myc* transcription in B-cell lymphomas. *Proc. Natl. Acad. Sci. USA*, **91**, 4915-4919.

22. Madisen, L. and Groudine, M. (1994) Identification of a locus control region in the immunoglobulin heavy-chain locus that deregulates c-*myc* expression in plasmacytoma and Burkitt's lymphoma cells. *Genes and Development*, **8**, 2212-2226.

23. Ohki, R., Oishi, M. and Kiyama, R. A Whole Genome Analysis of Allelic Changes in Renal Cell Carcinoma Using In-gel Competitive Reassociation. submitted

24. Greenwald, B. D., Harpaz, N., Yin, J., Huang, Y., Tong, Y., Brown, V. L., McDaniel, T., Newkirk, C., Resau, J. H., and Meltzer, S. J. Loss of heterozygosity affecting the *p53* , *Rb*, and *mcc/apc* tumor suppressor gene loci in dysplastic and cancerous ulcerative colitis. Cancer Res. *52:* 741-745, 1992.

25. Deng, G., Chen L.-C., Schott, D. R., Thor, A., Bhargava, V., Ljung, B.-M., Chew, K., and Smith, H. S. Loss of heterozygosity and *p53* gene mutations in breast cancer. Cancer Res. *54:* 499-505, 1994.

26. Mor, O., Messinger, Y., Rotman, G., Bar-Am, I., Ravia, Y., Eddy, R.L., Shows, T.B., Park, J.-G., Gazdar, A. F., and Shiloh,Y. Novel DNA sequences at chromosome $10q26$ are amplified in human gastric carcinoma cell lines: molecular

cloning by competitive DNA reassociation. Nucl. Acids. Res. *19:* 117-123, 1991.

27. Yunis, J. J. and Soreng, A. L. (1984) Constitutive fragile sites and cancer. *Science*, **226**, 1199-1204.

28. Ohta, M., Inoue, H., Cotticelli, M. G., Kastury, K., Baffa, R., Palazzo, J., Siprashvili, Z., Mori, M., McCue, P., Druck, T., Croce, C. M. and Huebner, K. (1996) The FHIT gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. *Cell*, **84**, 587-597.

29. Flavell, A. J. and Ish-Horowicz, D. (1983) The origin of extrachromosomal circular copia elements.*Cell,* **34**, 415-419.

30. Kiyama, R., Matsui, H., Okumura, K. and Oishi, M. (1987) A group of repetitive families that is characterized by extrachromosomal oligomers and restriction-fragment length polymorphism. *J. Mol. Biol.,* **193**, 591-597.

31. Fujimoto, S. and Yamagishi, H. (1987) Isolation of an excision product of T-cell receptor alpha-chain gene rearrangements. *Nature,* **327**, 242-243.

32. Gaubatz, J. W. (1990) Extrachromosomal circular DNAs and genomic sequence plasticity in eukaryotic cells. *Mut. Res.,* **237**, 29-36.

33. Iwasaki, T., Ohki, R., Kiyama R. and Oishi, M. (1995) Analysis of recombination junctions in extrachromosomal circular DNA obtained by in-gel competitive reassociation. *FEBS Letters,* **363**, 239-245.

34. Jones, R. S. and Potter, S. S. (1985) L1 sequences in HeLa extrachromosomal circular DNA: evidence for circularization by homologous recombination. *Proc. Natl. Acad. Sci. U.S.A.,* **82**,1989-1993.

35. Iwasato, T., Shimizu, A., Honjo T. and Yamagishi, H. (1990) Circular DNA is excised by immunoglobulin class switch recombination. *Cell,* **62**,143-149.

36. Kornberg, R. (1974) Chromatin structure: a repeating unit of histones and DNA. *Science,* **184**, 868-871.

37. Kornberg, R. and Thomas, J. O. (1974) Chromatin structure: oligomers of histones. *Science,* **184**, 865-868.

38. Noll, M. (1974) Subunit structureof chromatin. *Nucl. Acids Res.,* **1**, 1573-1578.

39. Lutter, L. (1978) Kinetic analysis of deoxyribonuclease I cleavage sites in the nucleosome core: evidence for a DNA superhelix. *J. Mol. Biol.,* **124**, 391-420.

40. Gale, J. M. and Smerdon, M. J. (1988). Photofootprint of nuclosome core DNA in intact chromatin having different structural states. *J. Mol. Biol.* **204**, 949-958.

41. Hayes, J. J., Tullius, T. D. and Wolffe A. P. (1990) The structure of DNAin a nucleosome. *Proc. Natl. Acad. Sci. USA,* **87**, 7405-7409.

42. Hayes, J. J., Clark, D. J. and Wolffe A. P. (1990) Histone contribution to the structure of DNA in a nucleosome. *Proc. Natl. Acad. Sci. USA,* **88**, 6829-6833.

43. Klug, A. and Lutter, L. C. (1981) The helical periodicity of DNA on the nucleosome. *Nuc. Acids Res.,* **9**, 4267-4283.

44. Satchwell, S. C., Drew, H. R. and Travers, A. A. (1986) Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.,* **191**, 659-675.

45. Shrader, T. E. and Crothers, D. M. (1989) Artificial nucleosome positioning sequences. *Proc. Natl. Acad. Sci. USA,* **86**, 7418-7422.

46. Simpson, R. T. (1991) Nucleosome positioning occurrence, mechanisms and functional consequences. *Progr. Nucl. Acids Res. Mol. Biol.*, **40**, 143-184.

47. FitzGerald, P. C. and Simpson, R. T. (1985). Effects of sequence alterations in a DNA segment containing the 5S rRNA gene from *lythechinus variegatus* on positioning of a nucleosome core particle *in vitro. J. Biol. Chem.*, **260**, 15318-24.

48. Meerssemann, G., Pennings, S. and Bradbury, E. M. (1991) Chromatosome positioning on assembled long chromatin. Linker histones affect nucleosome placement on 5S DNA. *J. Mol. Biol.*, **220**, 89-100.

49. Prunell, A. (1982) Nucleosome reconstitution on plasmid-inserted poly(dA).poly(dT). *EMBO J.*, **1**, 173-179.

50. Garner, m. M. and Felsnfeld, G. (1987). Effect of Z-DNA on ncleosome placement. *J. Mol. Biol.*, **196**, 581-590.

51. Dong, F. and van Holde, K. E. (1991). Nucleosome positioning is determined by the (H3-H4)2 tetramer. *Proc. Natl. Acad. Sci. USA*, **88**, 10596-10600.

52. Allan, J., Cowling, G. J., Harborne, N., Cattani, P., Cragie, R. and Gould, H. (1981). Regulation of the higher-order structure of chromatin by histones H1 and H5. *J. Cell Biol.,* **90**, 279-288.

53. Thoma, F., Koller, T. and Klug, A. (1979) Involvement of histone H1 in the organiziation of the nucleosome and the salt-dependent superstructures of chromatin. *J. Cell Biol.*, **83**, 402-427.

54. Earnshow, W. C. (1991) Large scale chromosome structure and organization. *Curr. Opin. Struct. Biol.*, **1**, 237-244.

55. Dickinson, L. A., Joh, T., Kohwi, Y. and Kohwi Shigematsu, T. (1992) A tissue specific MAR/SAR DNA-binding protein with unusual binding site recognition. *Cell*, **70**, 631-645.

56. von Kries, J. P., Buhrmester, H. and Stratling, W. H. (1991) A matrix/scaffold attachment region binding protein: Identification, purification, and mode of binding. *Cell*, **64**, 123-135.

57. Gasser, S. M. and Laemmli, U. K. (1986) The organization of chromatin loops; characterization of a scaffold attachment site. *EMBO J.* **5**, 511-518.

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

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

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

61. Schultz, S. C., Sheilds, G. C. and Steitz, T. A. (1991) Crystal structure of a CAP-DNA complex: the DNA is bent by 90°. *Science*, **253**, 1001-1007.

62. Kim, J. L., Nikolov, D. B. and Burley, S. K. (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, **365**, 520-527.

63. Goodman, S. D. and Nash, H. (1989) Functional replacement of a protein-induced bend in a DNA recombination site. *Nature*, **341**, 251-254.

64. Churchill, M. E., Jones, D. N., Glaser, T., Hefner, H., Searles, M. A. and Travers, A. A. HMG-D is an architecture-specific protein that preferentially binds to DNA containing the dinucleotide TG. (1995) *EMBO J.*, **14**, 1264-75.

65. Muller, H. P., Varmus, H. E. (1994) DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. *EMBO J., 13*, 4704-14.

66. Mazin, A., Milot, E., Devoret, R. and Chartrand, P. (1994) KIN17, a mouse nuclear protein, binds to bent DNA fragments that are found at illegitimate recombination junctions in mammalian cells. *Mol. & Gen. Genet., 244*, 435-8.

67. Pruss, D., Bushman, F. D. and Wolffe, A. P. (1994) Human immunodeficiency virus integrase directs integration to sites os severe DNA distortion within the nucleosome core. *Proc. Natl. Acad. Sci. USA*, **91**, 5913-5917.

68. Koo, H.-S., Wu, H.-M. and Crothers, D. M. (1986) DNA bending at adenine. timine tracts. *Nature*, **320**, 501-506.

69. Hagerman, P. J. (1986) Sequence-directed curvature of DNA. *Nature*, **321**, 449-450.

70. Kiyama, R., Okumura, K., Matsui, H., Bruns, G. A., Kanda, N. and Oishi, M. (1987) Nature of recombination involved in excision and rearrangement of human repetitive DNA. *J. Mol. Biol., 198*, 589-598.

71. Kiyama, R., Matsui, H. and Oishi, M. (1986) *Proc. Natl. Acad. Sci. U.S.A., 83*, 4665-4669.

72. Gelmann, E. P., Psallidopoulos, M. C., Papas, T. S. and Dalla-Favera, R. (1983) Identification of reciprocal translocation sites within the c-*myc* oncogene and immunoglobulin mu locus in a Burkitt lymphoma. *Nature* **306**, 799-803.

73. Otsu, M., Katamine, S., Uno, M., Yamaki, M., Ono, Y., Klein, G., Sasaki, M. S., Yaoita, Y. and Honjo, T. (1987) Molecular characterization of novel reciprocal translocation t(6;14) in an Epstein-Barr virus-transformed B cell precursor. *Mol. Cell. Biol.* 7, 708-717.
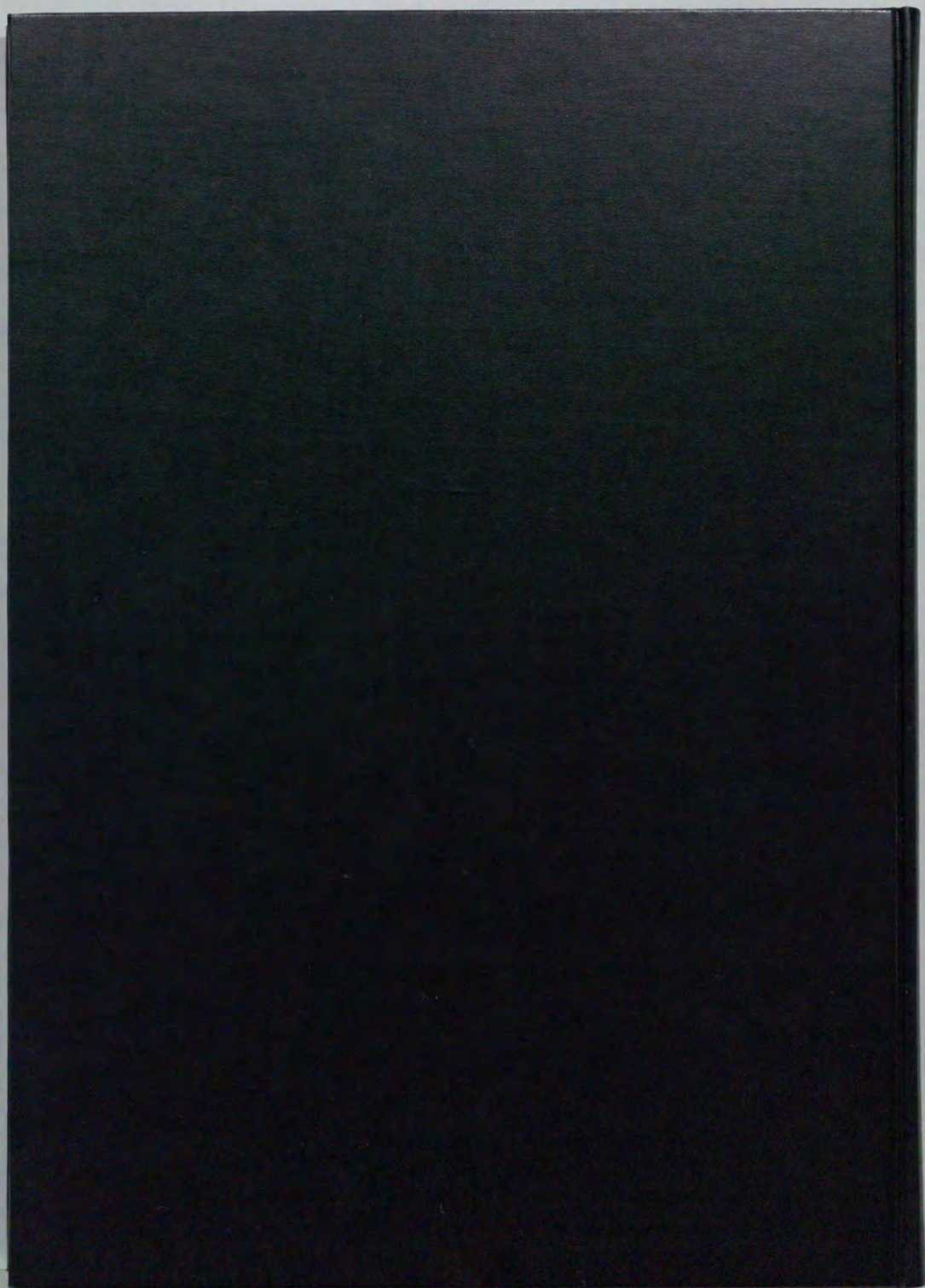
74. Showe, L. C., Ballantine, M., Nishikura, K., Erikson, J., Kaji, H. and Croce, C. M.(1985) Cloning and sequencing of a c-*myc* oncogene in a Burkitt's lymphoma cell line that is translocated to a germline alpha switch region. *Mol. Cell. Biol.* **5**, 501-509.

75. Battey, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G. and Leder, P. (1983) The human c-*myc* oncogene: structural consequences of translocation into the IgH locus in Burkitt lymphoma. *Cell* **34** (3), 779-787 (1983)

76. Ye, B. H., Chaganti, S., Chang, C. C., Niu, H., Corradini, P., Chaganti, R. S. and Dalla-Favera, R. Chromosomal translocations cause deregulated BCL6 expression by promoter substitution in B cell lymphoma. (1995) *EMBO J.* 14 (24), 6209-6217.

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

78. Strauss, F. and Varshavsky, A. A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. (1984) *Cell*, **37**, 889-901.

79. Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. and Okazaki, T. (1989) Centromere protein B assembles human centromeric α-satellite at the 17-bp sequence, CENP-B box. *J. Cell Biol.,* **109**, 1963-1973.

80. Cooke, C. A., Bernat, R. L. and Earnshaw, W. C. (1990) CENP-B: a major human centromere protein localized beneath the kinetochore. *J. Cell Biol.,* **110**, 1475-1488.

81. Warburton, P. E., Waye, J. S. and Willard, H. F. Nonrandom localization of recombination events in human alpha satellite repeat unit variants: implications for higher-order structural characteristics within centromeric heterochromatin. (1993) *Mol. Cell. Biol.,* **13**, 6520-6529.
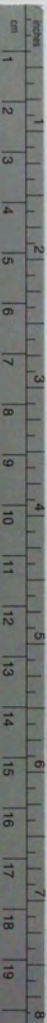
82. Kawasaki, I., Bae, Y. S., Eki, T., Kim, Y. and Ikeda, H. (1994) Homologous recombination of monkey a-satellite repeats in an in vitro siman virus 40 replication system: possible association of recombination with DNA replication. *Mol. Cell. Biol.,* **14**, 4173-4182.

83. Wu, K., Strauss, F. and Varshavsky, A. (1983) Nucleosome arrangement in green monkey α-satellite chromatin.*J. Mol. Biol.,* **170**, 93-117.

84. Zhang, X. Y., Fittler, F., and Horz, W. (1983) Eight different highly specific nucleosome phases on α-satellite DNA in african green monkey. *Nucl. Acids Res.* **11**, 4287-4305.

85. Fitzgerald, D. J., Dryden, G. L., Bronson, E. C., Williams, J. S. and Anderson, J. N. (1994) Conserved patterns of bending in satellite and nucleosome positioning DNA. *J. Biol. Chem.,* **269**, 21303-21314.

86. Shrader, T. E. and Crothers, D. M. (1989) Artificial nucleosome positioning sequences. *Proc. Natl. Acad. Sci. U.S.A.,* **86**, 7418-422.

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

88. Larsen, A., and Weintraub, H. (1987) An altered DNA conformation detected by S1 nuclease occurs at specific regions in active chick globin chromatin. *Cell,* **29**, 609-622.

89. Mills, F. C., Fisher, L. M., Kuroda, R., Ford, A. M. and Gould, H. J. (1983) DNase I hypersensitive sites in the chromatin of human μ immunoglobulin heavy-chain genes. *Nature,* **306**, 809-812.

90. Siebenlist, U., Henninghausen, L., Battey, J., Leder, P. (1984) Chromatin structure and protein binding in the putative regulatory region of the c-*myc* gene in Burkitt lymphoma. *Cell,* **37**, 381-391.

91. Dyson, P. J., Rabbitts, T. H. (1985) Chromatin structure around the c-*myc* gene in Burkitt lymphomas with upstream and downstream translocation points. *Proc. Natl. Acad. Sci. U.S.A.,* **82**, 1984-1988.

92. Roque, M. C., Smith, P. A. and Blasquez, V. C. (1996) A developmentally modulated chromatin structure at the mouse immunoglobulin κ 3' enhancer. *Mol. Cell. Biol.,* **16**, 3138-3155.

93. Almer, A. and Horz, W. (1986) Nuclease hypersensitive regions with adjacent positioned nucleosomes mark the gene boundaries of the PHO5/PHO3 locus in yeast. *EMBO J.,* **5**, 2681-2697.

Acknowledgments