Journal of Cellular Biochemistry 67:541–551 (1997)

Matrix Attachment Regions and Transcription Units in a Polygenic Mammalian Locus Overlapping Two Isochores

Maria-Aparecida Fernandez, Bruno Baron, Magali Prigent, Franck Toledo, Gérard Buttin, and Michelle Debatisse*

Unitéde Génétique Somatique (URA CNRS 1960), Institut Pasteur, Paris, France

Abstract Eukaryotic chromosomes are ponctuated by specialized DNA sequences (MARs) characterized by their ability to bind the network of nonhistone proteins that form the nuclear matrix or scaffold. We previously described an amplifiable cluster of genes with different tissue-specific expression patterns, located on Chinese hamster chromosome 1q. This model is especially appropriate to study the relationships between MARs and transcription units. We show here that four attachment regions, with sequences exhibiting motifs specific to MARs, are present within the 100 kb of screened DNA. Three of them are relatively short sequences localized in intergenic regions. The last one extends over one of the transcription units and contains a region previously identified as a recombination hot spot. Moreover, the analysis of a DNA sequence extending over some 50 Kb of this region and spanning at least four genes, disclosed a strikingly sharp change in G + C content. This strongly suggests that the studied region contains the boundary of two isochores. We propose that the frequency and the size of MARs are correlated to their localization in G + C rich or poor domains. J. Cell. Biochem. 67:541–551, 1997.

Key words: chromatin loops; chromosome organization; compositional mapping; gene cluster

The DNA of an eukaryotic cell can accomodate a variety of biological processes including transcription, replication, recombination, and distribution to daughter cells. The coordination of these complex functions is achieved by the operation of a network of regulatory mechanisms, including those operating at the chromatin level. The chromatin is organized in topologically constrained loops which are believed to be generated by the binding of specialized sequences to a network of nonhistone proteins, commonly refered to as the nuclear matrix [Bereznev and Coffey, 1974] or scaffold [Mirkovitch et al., 1984]. The attachment regions (MARs) generally extend over several hundreds of bp. Although no consensus sequence has emerged, MARs share several evolutionary conserved sequence characteristics and particular structural features including a narrow minor groove [Nelson et al., 1987] and a strong potential for unwinding [Bode et al., 1992; Laemmli et al., 1992; Probst and Herzog, 1985]. At present, however, the contribution of these sequences to the anchorage of chromatin loops is far from being clear since a recent work established that only a subset of MARs is localized at the base of the loops in vivo [Iarovaia et al., 1996].

MARs are considered to be involved in transcriptional control at different levels. In some cases, as examplified by the chicken lysozyme [Bonifer et al., 1990; McKnight et al., 1992; Phi-Van et al., 1990; Stief et al., 1989] and the human apolipoprotein B loci, regions in which the chromatin organization switches from an active to an inactive conformation have been mapped on both sides of the transcription units, and found to colocalize with MARs [Levy-Wilson and Fortier, 1989; Phi-Van and Stratling, 1988]. Constructs containing a gene flanked by these MARs exhibit a level of expression independent of the integration site, both in stable transfectants [Kalos and Fournier, 1995; Klehr et al., 1991; Phi-Van et al., 1990; Stief et al., 1989] and in transgenic animals [Bonifer et al., 1990; McKnight et al., 1992]. However, it has been shown more recently that transcriptional

Maria-Aparecida Fernandez' current address is Departamento de Biologia, Universidad Estadual de Maringa, Avenida Colombo, 3690, CEP 87.020, Maringa-Parana Brazil.

*Correspondence to: M. Debatisse, Unité de Génétique Somatique (URA CNRS 1960), Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cédex 15, France. E-mail: mdebat@pasteur.fr

Received 25 July 1997; Accepted 21 August 1997

insulation is achieved by specialized elements nested within, but separable from, the MAR regions of the chicken lysozyme gene [Phi-Van and Stratling, 1996]. In other model systems, MARs are colocalized with enhancer elements, a phenomenon described as cohabitation [Gasser and Laemmli, 1987], and some were found within transcription units. Recent work demonstrated that the intronic MAR of the immunoglobulin κ gene controls the level of gene expression but does not protect the transgenes from position effects [Blasquez et al., 1989; Xu et al., 1989]. Thus, the involvement of MARs in transcriptionnal insulation is becoming elusive.

The physiological functions of MARs are still unclear, in part because our current knowledge of the relationships between the localization of MARs and the general genome organization has been built from a relatively small number of examples: long-range mapping of MARs have been performed in three genetically defined chromosomal regions in Drosophila [Brun et al., 1990; Mirkovitch et al., 1986; Surdej et al., 1990], and in few gene clusters in chicken [Farache et al., 1990; Razin et al., 1991] and mammalian genomes [Cockerill, 1990; Hanson and Ley, 1992; Jarman and Higgs, 1988]. These works have shown that the distribution of MARs relative to the transcription units is highly variable from one cluster to the other. As an example, in human cells, the analysis of a 90 Kb region around the β-globin gene complex disclosed at least eight MARs whereas none has been detected within a 140 Kb region of the α-globin gene complex [Jarman and Higgs, 1988]. The latter result suggests that the functions of MARs can be either performed by other types of sequences or are dispensable in some chromosomal contexts.

We reported previously that mutants resistant to coformycin, which have amplified the gene encoding adenylate deaminase 2 (*AMPD2*), can be selected from the Chinese hamster fibroblastic line GMA 32. We showed that several genes are passively coamplified with the selected one in these mutants. A genetic map of the amplifiable domain, located on chromosome 1q, has been established and at least seven genes were identified [Debatisse et al., 1992]. Four of them, the genes encoding the αi3 (*GNAI3*) and αt2 (*GNAT2*) subunits of the GTP binding proteins, the *AMPD2* gene and one member of the gene family encoding the glutathione *S*-transferases (*GSTM*) are clustered in some 100 kb of cloned DNA, more than 50 Kb of which are sequenced. The *GNAI3* gene is ubiquitous while the others have a tissue-specific expression pattern: *GNAT2* is transcribed in retina cells only, *AMPD2* in a relatively broad range of tissues including the fibroblasts used here, and different members of the *GSTM* genes are expressed in different tissues. Our sequencing data allowed us to precisely map some of these transcription units [Baron et al., 1996, 1994]. Here we complete the mapping of the genes and their CpG islands. The analysis of the G + C content of the sequenced region disclosed a striking change in the G + C percentage along the studied domain. Because it is well known that the genomes of vertebrates are mosaics of isochores, originally identified as long DNA regions homogeneous in base composition [Bernardi et al., 1985; Macaya et al., 1976], our results strongly suggest that the cluster of genes under study spans the border of two isochores. We determined the distribution of MARs in this cluster of functionally unrelated genes with different expression patterns. We observed that MARs belonging to regions with different G + C contents exhibit striking differences both in their size and in their location relative to the genes. Our results strongly suggest that the functional importance of MARs largely depends on the overall base composition of the region in which the genes under study are embedded.

MATERIALS AND METHODS

Cell Lines and Growth Conditions

Line 42 is a clonal cell population derived from the lung fibroblastic Chinese hamster GMA32 line through two steps of selection with increasing coformycin concentrations. Compared to the wild-type cells, a 15-fold amplification of the *AMPD2* gene and surrounding sequences was disclosed in these cells [Debatisse et al., 1992]. Cells of line 42 grow as monolayer in Eagle's medium supplemented with 10% horse serum, 2×10^{-5} M azaserine, 5×10^{-5} M adenine, 5×10^{-5} M uridine, and 5 μg/ml coformycin, at 37°C in humidified incubator with 8% CO_2 atmosphere.

CpG Island Characterization

Ten μg of DNA were digested with a chosen restriction enzyme and a 5 μg aliquot of restricted DNA was further digested with 20 units

of *Bst* UI for 1 h at 37°. This DNA and the DNA untreated with *Bst* UI were run on 0.6% agarose gels and transfered onto Hybond-N membrane (Amersham, Buckinghamshire, UK). The DNA was cross linked to the membrane by UV irradiation and probed with cloned fragments labeled with $^{32}\alpha$P-dCTP (3000 Ci/mmol) using the Ready-to-go kit (Pharmacia, Gaithersburg, MD). Probes containing repetitive sequences were annealed with sonicated DNA from GMA32 cells as previously described [Ardeshir et al., 1983].

MAR Assay

In a typical experiment, Petri dishes (10 cm diameter) were each seeded with 5×10^5 cells. Exponentially growing cells ($2–3 \times 10^6$ cells/plate) were collected after 40 h of growth in regular medium. Nuclear halos were then prepared as previously described [Dijkwell and Hamlin, 1988] with few modifications. The plates were washed with 10 ml of cold CWB buffer (5 mM Tris-HCl pH 7.4; 50 mM KCL; 0.05 mM spermine; 0.125 mM spermidine; 0.5% thiodiglycol; 0.25 mM PMSF; 0.5 mM EDTA), the cells were then scrapped off, resuspended in 10 ml of cold CWBD buffer (CWB complemented with 0.05% digitonin; soluble form, Sigma, St. Louis, MO), and the suspension was layered over 4 ml of 20% glycerol in cold CWBD buffer. The nuclei were recovered by centrifugation at 4°C for 10 min at 2,000 rpm and stabilized 20 min at 4°C in 10 ml of CWB without EDTA, complemented with 0.003 mM $CuSO_4$. The suspension was then poured into 90 ml of LIS buffer (10 mM lithium diiodosalicylate; 100 mM lithium acetate; 0.05% digitonin; 0.05 mM spermine; 0.125 mM spermidine; 0.25 mM PMSF; 20 mM HEPES-KOH pH 7.4), incubated for 10 min at 25°C and the nuclear halos were collected by centrifugation at 2.500 rpm for 5 min at 4°C. The pellet was washed at least three times with cold appropriate restriction buffer. Two classical assays were used: in assay A, nuclear halos corresponding to 10^8 cells were digested overnight with 1,000 units of *Sau*3AI in 10 ml of restriction buffer (10 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 100 mM NaCl, 1 mM 2-mercaptoethanol) at 37°C, then 25 µg of RNAse A were added 30 min before the end of the incubation. The naked matrices were recovered by centrifugation at 3,000 rpm for 5 min, and washed at least twice with cold BB buffer (20 mM KCl; 50 mM NaCl; 2 mM EDTA; 10 mM Tris-HCl pH 7.4; 0.25 mM PMSF; 100 mM sucrose; 0.1 mg/ml BSA). Matrices derived from 10^7 cells were incubated with 20 ng of each preparation of ^{32}P-end-labeled DNA fragments [Sambrook et al., 1989] in BB buffer supplemented with various concentrations of sonicated *E. coli* competitor DNA, in a final volume of 200 µl. Binding reactions were carried out for 6–8 h at 37°C, the matrix and bound DNA were pelleted by centrifugation at 4°C, 13.000 rpm for 5 min in an Eppendorf centrifuge. The pellets were washed once with BB buffer and incubated overnight at 30°C with proteinase K (1 mg/ml) in PK buffer. Aliquots were directly loaded on 0.8% agarose gel. The radioactivity of the dried gels was quantified with the ImageQuant program (Phosphorimager; Molecular Dynamics, Sunnyvale, CA) to determine the percentage of bound fragments. In assay B, nuclear halos corresponding to 10^7 cells were incubated with 1,000 units of restriction enzyme(s) in 1 ml of restriction buffer at 37°C for 45 min. Ten µg of RNAse A were added and incubation was continued for another 45 min. The loop and the matrix fractions were separated by centrifugation at 3.000 rpm for 5 min. The pellet containing the matrix and bound DNA was redissolved and incubated at 37°C with 1 mg/ml of proteinase K in PK buffer (10 mM Tris-HCl pH 8; 1 mM EDTA; 300 mM NaCl). The DNA of the two fractions was extracted, ethanol precipitated and redissolved in TE. Two µg of DNA from the loop and the pellet fractions were run on 0.6% agarose gels.

Sequencing

Sequence analyses were performed as previously described [Baron et al., 1994]. The ATATTT motif, oligo A or T tracts at least 11 bp long, Drosophila "in vitro" topoisomerase II cleavage consensus sites (GTNWAYATTNATNNR) and the G + C percentages of the studied sequences were searched either with GCG (Program Manual for the Wisconsin Package, Version 8, September 1994) or DNASIS programs.

RESULTS

Screening for CpG Islands Along 100 Kb of Cloned Sequence

The previously determined physical map of the cloned region is represented on Figure 1a. To complete the analysis of the transcription units belonging to this gene cluster, we local-

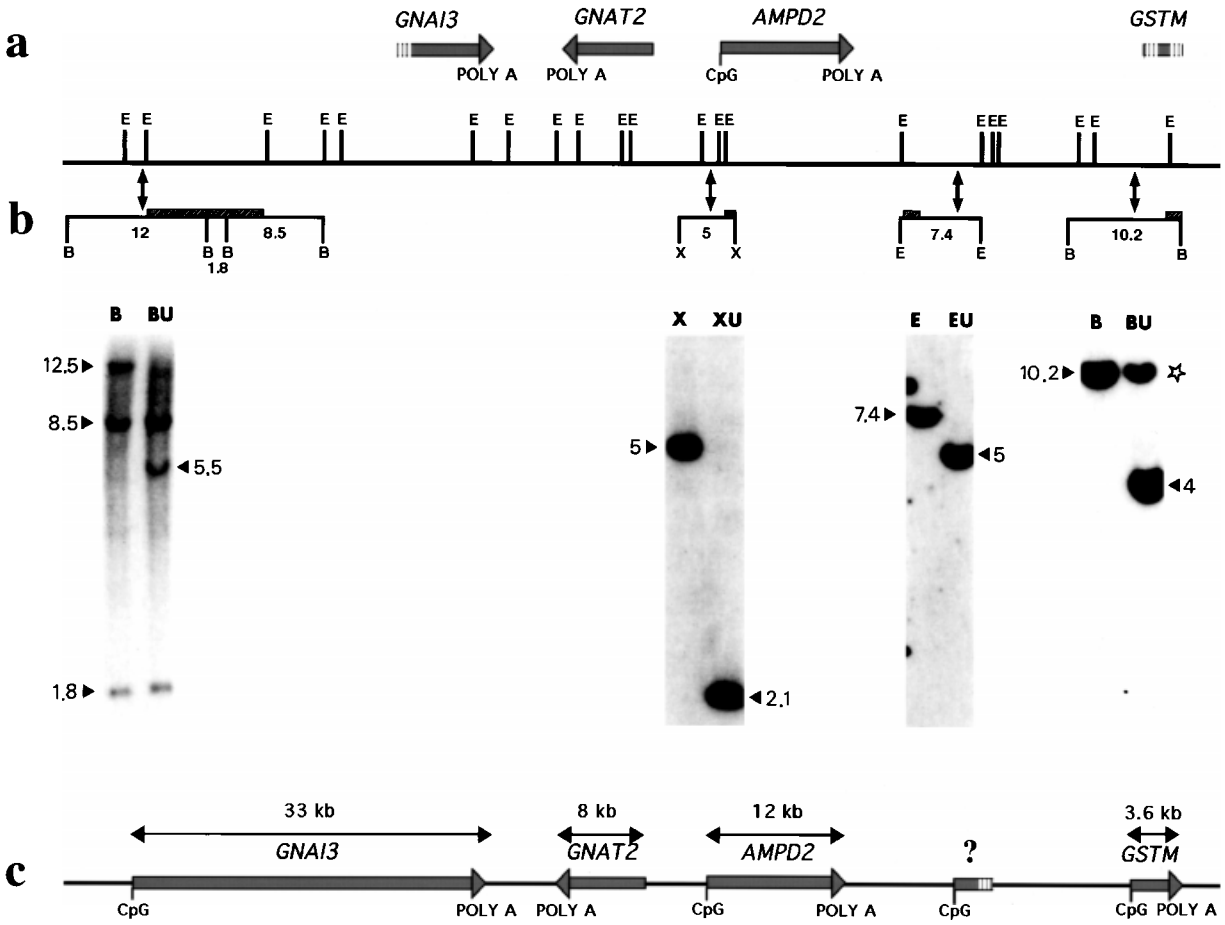

Fig. 1. Gene mapping in the studied locus. E: *Eco*RI, B: *Bam*HI, X: *Xba*I, U: *Bst*UI restriction sites. **a:** Previously established genetic map along with the *Eco*RI physical map of the region. **b:** CpG island (U sites) localization. The fragments of interest and the corresponding probes (black boxes) are represented below the *Eco*RI map, the double arrows localize the CpG islands. *Point to a band resulting from a partial cleavage. **c:** Limits and sizes of the genes as determined here.

ized *Bst*UI sensitive sites in the genomic DNA: this enzyme recognizes and cuts CGCG sequences only when the C residues are unmethylated, a situation mostly encountered in CpG islands. Four sites were disclosed (Fig. 1b), one of them corresponding to the previously identified 5′ end of the *AMPD2* gene [Baron et al., 1996]. The leftmost site was identified in a 2 Kb *Eco*RI/*Eco*RI fragment which was partially sequenced, a CpG island surrounding a sequence showing 94% identity with the human *GNAI3* exon I was identified. Since we previously mapped the 3′ end of this gene [Baron et al., 1994], we now establish that the Chinese hamster *GNAI3* gene spreads over 33 Kb, on the left side of the physical map (Fig. 1c). The last two *Bst*UI sensitive sites were localized on the right side of the studied region. We analyzed the sequences surrounding these sites and we found that the rightmost one maps within a CpG island associated with the 5′ end of a member of the *GSTM* gene family. Sequence analysis also permitted to identify the 3′ end of this gene (Fig. 1c). The sequence surrounding the last *Bst*UI site was shown to contain also a CpG island, but significant homology with a functionally identified gene was not found. However, the existence of a *DNAse*1 hypersensitive sites (not shown) associated with this unmethylated CpG island suggests that a gene lies in this region. Among the transcription units mapped within the 100 Kb region under study, only the *GNAT2* gene is devoid of a CpG island.

Screening for Nuclear Matrix Attachment Sites

Two "in vitro" assays have been routinely used to identify sequences which are able to bind the nuclear matrix. In both cases the nu-

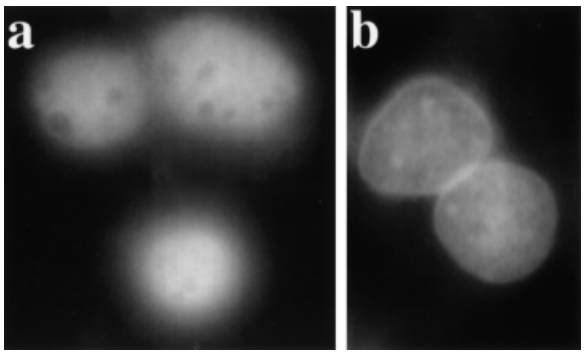

Fig. 2. Exemples of nuclear halos (**a**) and naked matrices (**b**).

clei were isolated, and partially deproteinized to form nuclear halos (Fig. 2a). In the first assay, designated below assay A, the DNA associated with the halos was extensively digested (Fig. 2b), and the ability of cloned fragments to bind the naked matrices was determined. In the second assay, designated B, the DNA bound to the matrices (M) was separated from the so called loop DNA (L) after digestion with restriction enzymes. The two fractions were probed with cloned fragments to determine in which fraction segregate the sequences detected by the probe. The 100 Kb domain was scanned using these two methods. Figure 3 shows, as an example, the results obtained with assay A for the 3′ region of the *AMPD2* gene. A rough comparison of lane 0—without competitor DNA— to the control line containing 1% of the input mixture of fragments indicates that 1 to 5% of each fragment were bound to the naked matrices under our experimental conditions. To estimate the relative affinity of the various fragments for the nuclear matrix, the binding reactions were performed in the presence of increasing concentrations of double-stranded or single-stranded *E. coli* competitor DNA, sonicated to an average size of 2 kb. The 1.4 kb *Xba*I/*Ava*I and the 1.7 kb *Xba*I/*Eco*RI fragments were almost completely displaced with respectively 100 and 50 µg/ml of double-stranded and single-stranded competitor DNA, while the 2.5 kb *Xba*I/*Xba*I and especially the 1.1 kb *Ava*I/*Xba*I fragments were still tightly attached at much higher concentrations of either competitor. The quantification of the signals with the ImageQuant program confirms that the 2.5 kb *Xba*I/*Xba*I and the 1.1 kb *Ava*I/*Xba*I fragments have a significant affinity for the nuclear matrix, the latter fragment beeing particularly strongly attached. These experiments also show that single-stranded DNA is a more powerful competitor than double-stranded DNA.

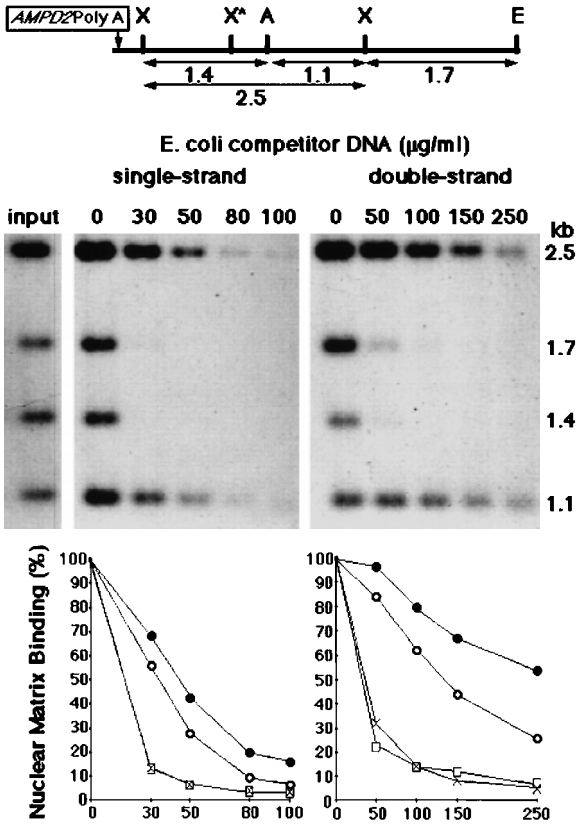

Fig. 3. Identification of an attached region close to the 3′ end of the *AMPD2* gene with assay A. **A:** *Ava*I, E: *Eco*RI, X: *Xba*I, X*: site cut in DNA prepared from hamster cells but not cut, albeit identified at the sequence level, in plasmid DNA prepared from the DH5α *E. coli* strain. This probably reflects differential methylation patterns in the two organisms and explains the different *Xba*I maps presented in Figures 3 and 4. The size and the localization of the four analyzed fragments (double arrows) are shown below the physical map of the region. Input: line loaded with 1% of the initial mixture of fragments added in each reaction. Left: Competition with increasing concentrations of single-stranded *E. coli* DNA. Right: Competition with increasing concentrations of double-stranded *E. coli* DNA. The crude autoradiograms (up) and the graphs derived from the quantification of each band, expressed in % of the DNA bound without competitor DNA (down), are shown. ●: 1.1, ○: 2.5, x: 1.7, □: 1.4 kb fragments.

This assay was used to study 81 fragments covering the whole 100 kb region under study. The 1.4 kb *Xba*I/*Ava*I and the 1.7 kb *Xba*I/*Eco*RI fragments described above were considered as negative controls. Thus, residual bindings up to 15% in the presence of 100 µg/ml of double-stranded competitor DNA or up to 5% in the presence of either 250 µg/ml of double-stranded or 50 µg/ml of single-stranded competi-

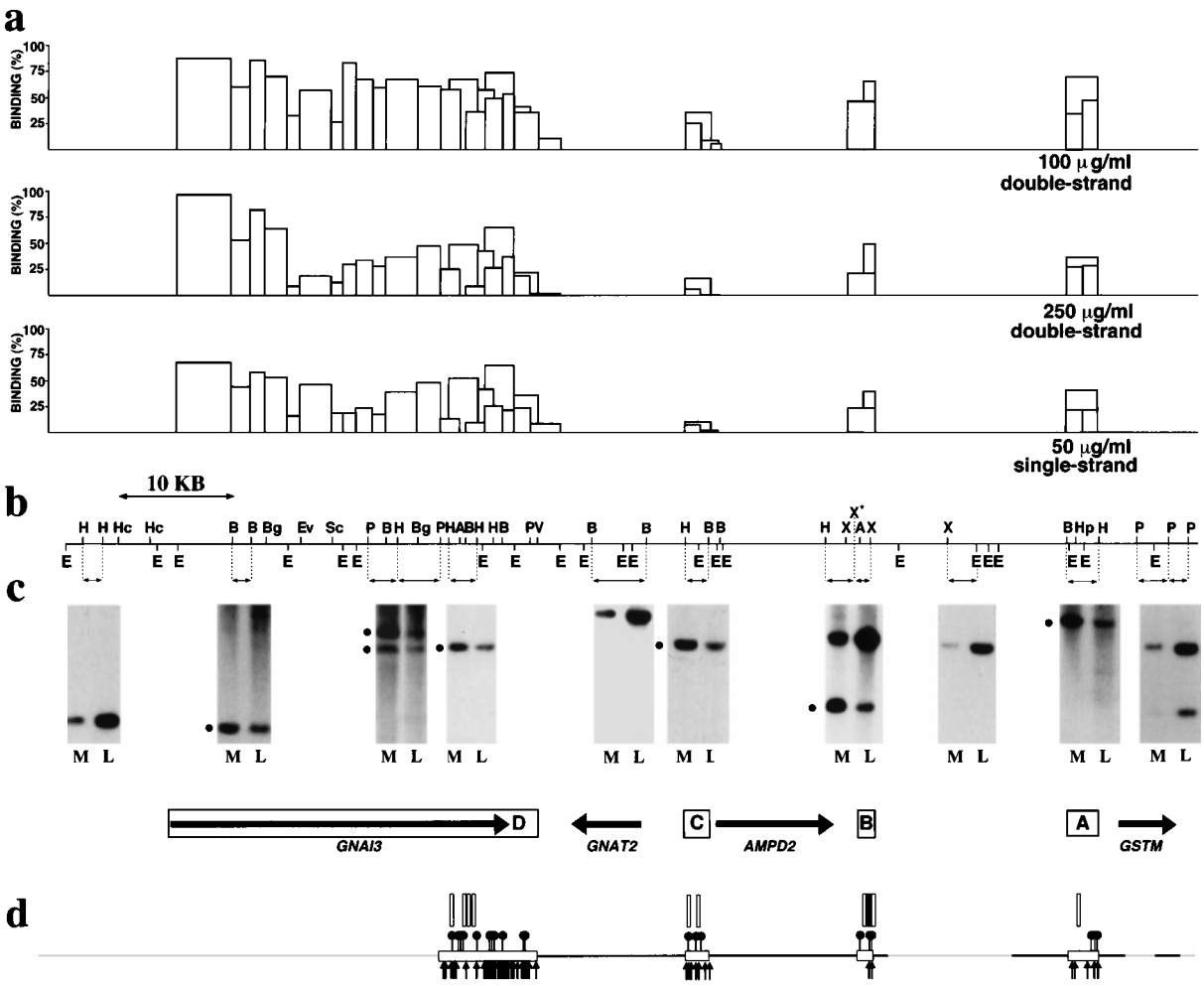

Fig. 4. Analysis of the whole region with assays A and B. **a:** Results of assay A histograms show the binding ability of the different fragments under three competition conditions. Background values were deduced from the crude results as indicated in the text. **b:** *Eco*RI map and localization of other restriction sites used in this study; A: *Ava*I, B: *Bam*HI, Bg: *Bgl*II, E: *Eco*RI, Ev: *Eco*RV, Hc: *Hinc*II, H: *Hind*III, Hp: *Hpa*I, P: *Pst*I, V: *Pvu*II, Sc: *Sca*I, X: *Xba*I, X*: as in Figure 3. The scale is given above the map. **c:** Results of assay B. M: Matrix fraction, L: supernatant (loop) fraction, (●): fragments preferentially recovered in fraction M. Below the autoradiograms is presented the summary of the results. The boxes labelled A, B, C, and D indicate the localization and extent of the MARs, along with the map of the transcription units. **d:** sequence analysis. Black line: sequenced region, grey line: unsequenced region. The horizontal open boxes represent the attached fragments or regions. The intervening sequences were also studied but the results are not presented (see text). □: Homopolymeric oligo dA/dT tracts ≥11 bp, ▮: 44 bp alterned dA/dT tract, ⚲: ATATTT boxes, ↑: "in vitro" Drosophila topoisomerase II consensus cleavage sites.

tor DNA were considered as background values. These values were substracted from each crude result to construct the histograms presented in Figure 4a. The binding ability of each studied fragment in the presence of 100 and 250 µg/ml of double-stranded, or 50 µg/ml of single-stranded competitor DNA, was plotted along the physical map of the locus (Fig. 4b). Four regions, called below A, B, C, and D were identified. In regions A, B, and C, the analysis of the binding ability of various overlapping restriction fragments points to relatively narrow MARs, extending over few kb, and lying in intergenic regions. On the contrary, in region D, a variety of overlapping fragments, extending over about 36 kb of contiguous sequence spanning the *GNAI3* transcription unit, all exhibited some ability to bind the matrix so that the attachment region cannot be more precisely defined.

The whole domain was next screened using assay B. A few results, corresponding to the analysis of regions previously identified as MARs and of some control regions, are shown on Figure 4c. With this assay, not all the regions are amenable to quantification, essentially be-

cause some fragments are rich in repetitive sequences and give rise to a high level of background when used as probes (eventhough prehybridization of the probes with sheared wild type DNA was routinely performed to remove the repeated sequences). However, the results are in good agreement with the conclusions of assay A: the very same regions exhibit a significant affinity for the nuclear matrix, and the A, B, and C MARs appear as narrow regions located outside the genes while region D extends largely within the *GNAI3* gene.

Sequence Analysis

Sequences of regions A, B, C, and of the rightmost 8 Kb of region D were determined and analyzed. The A + T richness was calculated and motifs frequently correlated with MARs in previous works were searched, in particular ATATTT boxes [Cockerill and Garrard, 1986; Probst and Herzog, 1985], homopolymeric oligo A/T tracts [Adachi et al., 1989; Käs et al., 1989], and sequences related to the "in vitro" Drosophila topoisomerase II consensus cleavage site [Adachi et al., 1989; Cockerill and Garrard, 1986; Gasser and Laemmli, 1987]. In the latter case, only the sequences presenting 100% identity with the underlined core sequence (GTNWAYATTNATNNR) and at least 12/15 bp matches with the consensus site were considered. Figure 4d shows a summary of our results. MAR A maps within a 3.2 kb *Bam*HI/*Hind*III fragment which encompasses eight subregions of 28 to 45 bp each, with an A + T content comprised between 79 and 90%. Within these subregions, three TATATTT motifs, four topoisomerase II consensus sites and a 12 bp long homopolymeric oligo A/T tract were found. In region B, two 150 bp subregions 73% and 81% A + T rich were found in the 1.1 kb *Ava*I/*Xba*I fragment which has a strong binding potential (Fig. 3). Each of them contains a 11 bp homopolymeric A/T tract and one encompasses a 44 bp long alterned A + T motif, two ATATTT boxes and a topoisomerase II consensus site. In region C, two subregions with sequence characteristics of MARs are present within the attachment prone 2.5 kb *Hind*III/*Bam*HI fragment. The first one contains a 62% A + T rich 350 bp long region in which a 27 bp homopolymeric A/T tract, three topoisomerase II sites and an ATATTT box were disclosed. The second one is a 67% A + T rich 300 bp long sequence containing a 11 bp homopolymeric A/T tract, two ATATTT motifs and two topoisomerase II consensus sites. Region D is not completely sequenced at that time, but numerous 100 to 300 bp sequences with an A + T content of 65 to 73% were identified all along the 8 kb of available sequence (which have an overall 60% A + T richness). On the left part of this sequenced region, six topoisomerase II consensus sites, five ATATTT boxes, and four homopolymeric A/T tracts respectively 12, 15, 16, and 17 bp long are present. The right part of the sequenced region, which overlaps the 3′ end of the *GNAI3* gene, is devoid of purine tracts but contains 21 topoisomerase II consensus sites and 7 ATATTT boxes. The same analysis was performed for the available sequence in control regions; some of these motifs were found unfrequently and dispersed in subregions which are not highly enriched in A + T. For example, only two homopolymeric A/T tracts were found along some 43 Kb of sequence lying between the MARs while nine such motifs were identified within 14 Kb of analysed sequence in the MAR regions (not shown).

This sequence analysis suggested that some discontinuity in the global base composition may exist along the studied domain. To confirm these preliminary informations, we performed a compositional mapping of the genes which are completely or partially sequenced. We found an average G + C content of 39%, 44%, 54%, and 53%, respectively in the GNAI3, GNAT2, AMPD2, and the GSTM genes. Such base compositions suggest that the left part of the studied region lies in a G + C poor isochore while the subregion located on the right of the map belongs to a G + C rich isochore [Bernardi, 1995]. These observations were extended to the analysis of the G + C richness within coding sequences, at the third codon position: because the third position is not under a stringent coding pressure, it tends to reflect the base composition of the region in which the gene is nested. This analysis allowed us to take into account the gene encoding the ζ subunit of the proteasome for which the coding, but not the genomic sequence is available. The gene coding the ζ subunit of the proteasome was identified by comparing the partial sequence of a 750 bp hamster cDNA clone to the sequences deposited in the Genome Data Base (GDB) sequence library. A high level of homology was found with the human PSMA5 gene and the rat homologue (unpublished results). In Chinese hamster ge-

nome, we previously mapped this gene on the left side of the *GNAI3* gene, outside the 100 Kb region described here [Debatisse et al., 1986]. Recently, we showed that the clustering and the order of this group of genes is conserved in the human genome [Mayau et al., submitted]. The analysis of the sequences determined here and in previous work, or deposited in the data bases, allowed us to compare the G + C content of these genes in man and Chinese hamster (Fig. 5). Very similar compositional characteristics were found in the two species. Moreover, the analysis of the G + C content of 10 genes belonging to this cluster confirms the existence of a strong fluctuation in the base composition of the studied locus, and suggests an assignment of the genes located on the left and on the right side of the *GNAT2-AMPD2* region to a L2 or H1 isochore and to an H3 isochore, respectively.

DISCUSSION

The clustering of functionally unrelated genes in the region under study offers an interesting model to investigate the functional aspects of chromatin organization. Moreover, the availability of amplified mutants gives a unique tool to analyze those processes which are difficult to study in single copy regions. For example, amplification has proved useful to study the replication pattern in the DHFR amplicon [Dijkwell and Hamlin, 1988], and in the present gene cluster [Toledo et al., submitted]. Thus, our model system should help to elucidate the relationships between transcription and replication units in mammalian cells, and to correlate these functional units to the MARs and to larger chromosomal domains such as isochores.

Four MARs were identified in the analyzed chromosomal locus. Among them, the A, B, and C regions have rather similar properties, including intergenic localization and relatively small sizes. Suprisingly, sequence D spans the whole *GNAI3* gene and overlaps part of the *GNAI3-GNAT2* intergenic region (Fig. 4). We compared the efficiency of *E. coli* single-stranded or double-stranded DNA to compete with different fragments for matrix attachment. We found that an efficient competition can be achieved with either competitor (Figs. 3, 4) as previously shown in other systems [Kay and Bode, 1994; Luderus et al., 1994]. For all studied fragments, we observed that single-stranded DNA is a better competitor than double-stranded DNA, a property previously reported for the binding of MARs to lamins [Luderus et al.,

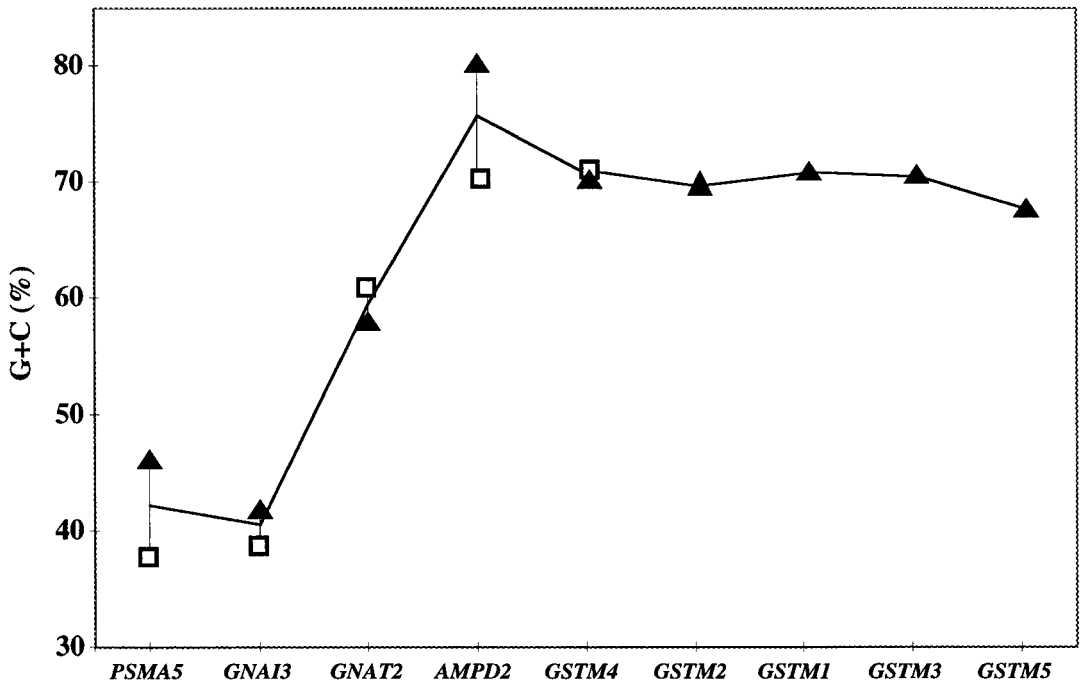

Fig. 5. Base composition along the region. A compilation of the G + C content (in %) of the third codon position of human (▲) and/or Chinese hamster (□) cDNAs is presented along with a map of nine studied genes. The distances between the genes below the diagram are arbitrary.

1994]. Thus, with these two types of competitors, the four MARs analyzed here behave like the MARs described in other systems. Moreover in each region, A + T rich sequences and clustering of characteristic motifs have been found (Fig. 4d). However, because of the unusual properties of region D, it may be asked whether "in vitro" binding of such sequences has a physiological significance or wether it reflects a binding ability common to any A + T rich sequence under these experimental conditions. Four lines of evidence support the first hypothesis: 1°) the general A + T content of MARs A, B, and C is comparable to the one of the *GNAT2* gene, the tested fragments of which were never found associated with the matrix. This clearly indicates that the general A + T richness per se is not responsible for the affinity of a DNA sequence for the nuclear matrix 2°) the sequenced part of region D contains multiple motifs characteristic of MARs. Even in this 60% A + T rich context, the probability to find at random several 12 to 17 bp long homopolymeric oligo A/T tracts clustered in some 4 Kb of sequence is extremely low 3°) the comparison of the sequenced part of region D to the equivalent region of the human genome shows that homopolymeric A/T tracts, several ATATTT boxes and Drosophila topoisomerase II consensus sites are present, like in hamster gene, in human intron 7 and noncoding exon 9. Such a conservation in noncoding sequences strongly suggests that these motifs are selected for because they have a physiological role 4°) Sperry et al. [1989] have shown that some MARs are recombination prone "in vivo," and suggested that the recombination events could be initiated by double strand-breaks resulting from a dysfunction of topoisomerase II. It is striking that a recombination hot spot has been previously identified in region D [Hyrien et al., 1986]. This property further supports the idea that region D does behave as a MAR "in vivo."

Our analysis strongly suggests that the gene cluster under study overlaps an isochore boundary. This conclusion is based on the G + C content of both the available genomic and cDNA sequences. The comparison of the human and Chinese hamster sequences at this locus disclosed a remarquable similarity in the global base composition of the genes in the two species. This is rather unexpected since the genome of rodents are generally more G + C rich than the genome of primates [Bernardi, 1995]. Our results indicate that Chinese hamster, as previously shown for the guinea pig, represents an exception among the rodents. The *GNAI3* gene belongs to a G + C poor isochore and its own G + C content is particularly low (Fig. 5). This situation is unusual since genes are not distributed at random among the different isochore families: gene concentration is low in G + C poor isochores, increases with increasing G + C richness, and reaches a maximum within the very G + C rich H3 isochores [Bernardi, 1995; Gardiner, 1996]. This seems specially true in the case of house-keeping genes, which are mostly confined to the H3 isochores. The special base composition of the *GNAI3* gene could require specific features to establish and/or to maintain a transcription competent chromatin state since histone H1, which has been shown to be involved in transcription silencing [Garrard, 1991], binds preferentially to A + T rich DNA. MARs are A + T rich regions, containing homopolymeric dA/dT tracts which have a particularly high affinity for histone H1. However, it has been proposed that histone H1 is specifically titrated by proteins like HMG-I/Y when bound to homopolymeric dA/dT tracts. A shift toward the open chromatin state would be initiated from these sequences and may spread along the chromatin fiber [Käs et al., 1993; Zhao et al., 1993]. We propose that the existence of a large CpG island at the 5′ end of the *GNAI3* gene, a type of sequence for which histone H1 has weak affinity, is not sufficient to open the chromatin along this 33 kb long gene in such an A + T rich context. The presence of a MAR extending all along the gene could be required to reach a transcription competent state in this case. On the contrary, a situation relatively favorable for chromatin opening is encountered in the *AMPD2* and the *GSTM* genes which exhibit some 70% G + C at the third codon position and have their 5′ end embedded in a CpG island. Our experiments disclosed short MARs between but not within the transcription units in these cases.

We analyzed sequences available in the data bases to determine the base composition of genes localized within the human β globin gene cluster, in which at least eight MARs were found between and within the genes [Jarman and Higgs, 1988]. We found that these genes are 66% G + C rich at the third codon position, a base composition which represents, like the MARs distribution, an intermediate situation

between the *GNAI3* domain and the *AMPD2* or *GSTM* domains. No MARs have been found in the α globin gene cluster [Jarman and Higgs, 1988], the α, θ, and ζ genes of which have each a CpG island and a G + C content at the third codon position comprised between 89 and 95%. Together, these two features could be sufficient to leave the domain essentially free of histone H1 and the chromatin in an open conformation even in cells in which the genes are not expressed, making MARs dispensable in this cluster. Thus, the α globin genes could illustrate an extreme situation just opposite to the one described here for the *GNAI3* gene. These analyses suggest that the size and the distribution of MARs in a chromosomal domain essentially depend on the base composition of the chromosomal context—the isochore—and that these sequences are mainly involved in the establishment of a transcription competent state of the chromatin.

ACKNOWLEDGMENTS

This work was supported in part by the Université Pierre et Marie Curie, the Ligue Nationale Française contre le Cancer (Comité de Paris), and the Association pour la Recherche sur le Cancer. M.-A.F. was supported by a post-doctoral fellowship from Brazilian Conselho Nacional de Desenvolvimento Cientifico e Tecnologico (CNPq).

REFERENCES

Adachi Y, Käs E, Laemmli UK (1989): Preferential, cooperative binding of DNA topoisomerase II to scaffold-associated regions. EMBO J 8:3997–4006.

Ardeshir F, Giulotto E, Zieg J, Brison O, Liao WSL, Stark G (1983): Structure of amplified DNA in different Syrian hamster cell lines resistant to *N*-(phosphonacetyl)-L-aspartate. Mol Cell Biol 3:2076–2088.

Baron B, Fernandez MA, Carignon S, Toledo F, Buttin G, Debatisse M (1996): GNAI3, GNAT2, AMPD2, GSTM are clustered in 120 kb of Chinese hamster chromosome 1q. Mammalian Genome 7:429–432.

Baron B, Fernandez MA, Toledo F, Le Roscouët D, Mayau V, Martin N, Buttin G, Debatisse M (1994): The highly conserved Chinese hamster GNAi3 gene maps less than 60 kb from the AMPD2 gene and lacks the intronic U6 snRNA present in its human counterpart. Genomics 24: 288–294.

Bereznev R, Coffey D (1974): Identification of a nuclear protein matrix. Biochem Biophys Res Commun 60:1410–1419.

Bernardi G (1995): The human genome: Organization and evolutionary history. Annu Rev Genetics 29:445–476.

Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J (1985): The mosaic genome of warm-blooded vertebrates. Science 228:953–958.

Blasquez VC, Xu M, Moses SC, Garrard WT (1989): Immunoglobulin kappa gene expression after stable integration. I. Role of the intronic MAR and enhancer in plasmacytoma cells. J Biol Chem 264:21183–21189.

Bode J, Kohwi Y, Dickinson L, Joh T, Klehr D, Mielke C, Kohwi-Shigematsu T (1992): Biological significance of unwinding capability of nuclear matrix-associating DNAs. Science 255:195–197.

Bonifer C, Vidal M, Grosveld F, Sippel AE (1990): Tissue specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice. EMBO J 9:2843–2848.

Brun C, Dang Q, Miassod R (1990): Studies of a 800 kilobase DNA stretch of the Drosophila X chromosome: Comapping of a subclass of scaffold-attached regions with sequences able to replicate autonomously in Saccharomyces Cerevisiae. Mol Cell Biol 10:5455–5463.

Cockerill PN (1990): Nuclear matrix attachment occurs in several regions of the IgH locus. Nucl Acids Res 18:2643–2648.

Cockerill PN, Garrard WT (1986): Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. Cell 44:273–282.

Debatisse M, Hyrien O, Petit-Koskas E, Robert de Saint Vincent B, Buttin G (1986): Segregation and rearrangement of coamplified genes in different lineages of mutant cells that overproduce adenylate deaminase. Mol Cell Biol 6:1776–1781.

Debatisse M, Toledo F, Robert De Saint Vincent B, Buttin G (1992): Amplification of the adenylate deaminase 2 and linked genes in Chinese hamster lung fibroblasts. In Kellems RE (ed): "Gene Amplification in Mammalian Cells—A Comprehensive Guide." New York: Marcel Dekker Inc., pp 173–183.

Dijkwell PA, Hamlin JL (1988): Matrix attachment regions are positioned near replication initiation sites, genes, and an interamplicon junction in the amplified dihydrofolate reductase domain of Chinese hamster ovary cells. Mol Cell Biol 8:5398–5409.

Farache G, Razin SV, Rzeszowska-Wolny J, Moreau J, Recillas Targa F, Scherrer K (1990): Mapping of structural and transcription-related matrix attachment sites in the alpha-globin gene domain of avian erythroblasts and erythrocytes. Mol Cell Biol 10:5349–5358.

Gardiner K (1996): Base composition and gene distribution: Critical patterns in mammalian genome organization. TIG 12:519–524.

Garrard W (1991): Histone H1 and the conformation of transcriptionally active chromatin. BioEssays 13:87–88.

Gasser SM, Laemmli UK (1987): A glimpse at chromosomal order. TIG 3:16–22.

Hanson RD, Ley TJ (1992): A-T-rich scaffold attachment regions flank the hematopoietic serine protease genes clustered on chromosome 14q11.2. Blood 79:610–618.

Hyrien O, Debatisse M, Buttin G, Robert De Saint Vincent B (1986): A hotspot for novel amplification joints in a mosaic of *Alu*-like repeats and palindromic A + T-rich DNA. EMBO J 6:2401–2408.

Iarovaia O, Hancock R, Lagarkova M, Miassod R, Razin SV (1996): Mapping of genomic DNA loop organization in a 500-Kilobase region of the Drosophila X chromosome by the topoisomerase II-mediated DNA loop excision protocol. Mol Cell Biol 16:302–308.

Jarman AP, Higgs DR (1988): Nuclear scaffold attachment sites in the human globin gene complexes. EMBO J 7:3337–3344.

Kalos M, Fournier RE (1995): Position-independent transgene expression mediated by boundary elements from the apolipoprotein B chromatin domain. Mol Cell Biol 15:198–207.

Käs E, Izaurralde E, Laemmli UK (1989): Specific inhibition of DNA binding to nuclear scaffolds and histone H1 by distamycin: The role of oligo(dA)-oligo(dT) tracts. J Mol Biol 210:587–599.

Käs E, Poljak L, Adachi Y, Laemmli UK (1993): A model for chromatin opening: Stimulation of topoisomerase II and restriction enzyme cleavage of chromatin by distamycin. EMBO J 12:115–126.

Kay V, Bode J (1994): Binding specificity of a nuclear scaffold: Supercoiled, single-stranded, and scaffold-attached-region DNA. Biochemistry 33:367–374.

Klehr D, Maass K, Bode J (1991): Scaffold-attached regions from three human interferon b domains can be used to enhance the stable expression of genes under the control of various promoters. Biochemistry 30:1264–1270.

Laemmli UK, Käs E, Poljak L, Adachi Y (1992): Scaffold-associated regions: cis-acting determinants of chromatin structural loops and functional domains. Curr Opin Cell Biol 2:275–285.

Levy-Wilson B, Fortier C (1989): The limits of the DNase I-sensitive domain of the human apolipoprotein gene coincide with the locations of chromosomal anchorage loops and define the 5′ and the 3′ boundaries of the gene. J Biol Chem 264:21196–21204.

Luderus MEE, Den Blaauwen JL, De Smit OJB, Compton DA, Van Driel R (1994): Binding of matrix attachment regions to lamin polymers involves single-stranded regions and the minor groove. Mol Cell Biol 14:6297–6305.

Macaya G, Thiery J, Bernardi G (1976): An approach to the organization of eukaryotic genomes at a macromolecular level. J Mol Biol 108:237–254.

McKnight RA, Shamay A, Sankaran L, Wall RJ, Hennighausen L (1992): Matrix-attachment regions can impart position-independent regulation of a tissue-specific gene in transgenic mice. Proc Natl Acad Sci USA 89:6943–6947.

Mirkovitch J, Mirault ME, Laemmli UK (1984): Organization of the higher-order chromatin loop: Specific DNA attachment sites on nuclear scaffold. Cell 39:223–232.

Mirkovitch J, Spierer P, Laemmli UK (1986): Genes and loops in 320,000 base pairs of the Drosophila melanogaster chromosome. J Mol Biol 190:255–258.

Nelson HCM, Finch JT, Luisi BF, Klug A (1987): The structure of an Oligo(dA)-Oligo(dT) tract and its biological implications. Nature 330:221–226.

Phi-Van L, Stratling WH (1988): The matrix attachment regions of the chicken lysozyme gene co-map with the boundaries of the chromatin domain. EMBO J 7:655–664.

Phi-Van L, Stratling WH (1996): Dissection of the ability of the chicken lysozyme gene 5′ matrix attachment region to stimulate transgene expression and to dampen position effects. Biochemistry 35:10735–10742.

Phi-Van L, Von Kries JP, Ostertag W, Stratling WH (1990): The chicken lysozyme 5′ matrix attachment region increases transcription from a heterologous promoter in heterologous cells and dampens position effects on the expression of transfected genes. Mol Cell Biol 10:2302–2307.

Probst H, Herzog R (1985): DNA regions associated with the nuclear matrix of Ehrlich ascites cells expose single-stranded sites after deproteinization. Eur J Biochem 146:167–171.

Razin SV, Petrov P, Hancock R (1991): Precise location of the alpha-globin gene cluster within one of the 20 ts 300-kilobase DNA fragments released by cleavage of the chicken chromosomal DNA at topoisomerase II sites in vivo: Evidence that the fragments are DNA loops or domains. Proc Natl Acad Sci USA 88:8515–8519.

Sambrook J, Fritsch EF, Maniatis T (1989): New York:

Sperry AO, Blasquez VC, Garrard WT (1989): Dysfunction of chromosomal loop attachment sites: Illegitimate recombination linked to matrix association regions and topoisomerase II. Proc Natl Acad Sci USA 86:5497–5501.

Stief A, Winter AM, Stratling WH, Sippel AE (1989): A nuclear DNA attachment element mediates elevated and position-independent gene activity. Nature 341:343–345.

Surdej P, Got C, Rosset R, Miassod R (1990): Supragenic loop organization: Mapping in Drosophila embryos of scaffold-associated regions on a 800 kilobase DNA continuum cloned from the 14B-15B first chromosome region. Nucl Acids Res 18:3713–3722.

Xu M, Hammer RE, Blasquez VC, Jones SL, Garrard WT (1989): Immunoglobulin kappa gene expression after stable integration. II. Role of intronic MAR and enhancer in transgenic mice. J Biol Chem 264:21190–21195.

Zhao K, Käs E, Gonzalez E, Laemmli UK (1993): SAR-dependent mobilization of histone H1 by HMG-I/Y in vitro: HMG-I/Y is enriched in H1-depleted chromatin. EMBO J 12:3237–3247.