Journal of Cellular Biochemistry 84:590–600 (2002)

# Identification and Mapping of Nuclear Matrix-Attachment Regions in a One Megabase Locus of Human Chromosome 19q13.12: Long-Range Correlation of S/MARs and Gene Positions

**I.P. Chernov,* S.B. Akopov, L.G. Nikolaev, and E.D. Sverdlov**

Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia, 117871

**Abstract** The first draft human genome sequence now available allowed the identification of an enormous number of gene coding areas of the genomic DNA. However, a great number of regulatory elements such as enhancers, promoters, transcription terminators, or replication origins can not be identified unequivocally by their nucleotide sequences in complex eukaryotic genomes. One important subclass of these type of sequences is scaffold/matrix attachment regions (S/MARs) that were hypothesized to anchor chromatin loops or domains to the nuclear matrix and/or chromosome scaffold. We developed an experimental selection procedure to identify S/MARs within a completely sequenced one megabase (1 Mb) long gene-rich *D19S208-COX7A1* locus of human chromosome 19. A library of S/MAR elements from the locus was prepared and shown to contain ~20 independent S/MARs. Sixteen of them were isolated, sequenced, and assigned to certain positions within the locus. A majority of the S/MARs identified (11 out of 16) lie in intergenic regions, suggesting their structural role, i.e., delimitation of chromatin domains. These 11 S/MARs subdivide the locus into 10 domains ranging from 6 to 272 kb with an average domain size of 88 kb. The remaining five S/MARs were found within intronic sequences of *APLP1*, *HSPOX1*, *MAG*, and *NPHS1* genes, and can be tentatively characterized as regulatory S/MARs. The correspondence of the chromatin domains defined by the S/MARs to functional characteristics of the genes therein is discussed. The approach described can be a prototype of a similar search of long sequenced genomic stretches and/or whole chromosomes for various regulatory elements. J. Cell. Biochem. 84: 590–600, 2002. 

**Key words:** human genome; nuclear matrix; S/MARs; functional mapping; DNA sequence annotation

The first draft human genome sequence published recently [Lander et al., 2001] opened up a possibility of extensive analysis of genome functional elements, first of all coding sequences [Bentley, 2000]. The complete annotation of sequenced eukaryote genomes, however, should not be limited to genes, it should also include the positioning of all non-coding regulatory elements. This is a prerequisite to form a solid base for understanding complex functional interplay between genes and other functional elements. At present, the main approach to solving this problem consists in combining all the information available on the genome elements in a single database using modern bioinformatic technologies [Lewis et al., 2000].

Unfortunately, the data on genome positions of the multitude of regulatory sequences, like enhancers, promoters, transcription terminators, replication origins etc., are very limited, especially at the whole genome level. Therefore, the development of whole genome experimental approaches to the identification of the genome elements that can not be recognized solely by their nucleotide sequences seems highly desirable.

One important subclass of this type of sequences is matrix-attachment regions or MARs, i.e., the regions of chromatin loops or domains that are widely believed to anchor them to the nuclear matrix and/or chromosome scaffold (for recent reviews see [Berezney et al., 1995; Berezney and Wei, 1998; Hancock, 2000]).

Grant sponsor: Human Genome State Project of Russia.

*Correspondence to: I.P. Chernov, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow 117871, Russia. E-mail: igor@humgen.siobc.ras.ru

Received 19 July 2001; Accepted 25 September 2001



DOI 10.1002/jcb.10043

MARs are usually defined as short (100–1,000 bp) DNA sequences capable of binding specifically to the isolated nuclear matrix in vitro [Cockerill and Garrard, 1986; Boulikas, 1995]. Similarly defined, scaffold attachment regions (SARs) [Mirkovitch et al., 1984; Izaurralde et al., 1988] are most probably very much alike if not identical to MARs [Bode et al., 1996]. S/MARs were hypothesized to play a functional role in DNA packaging, replication, and gene expression [Jackson et al., 1992; Berezney et al., 1995; Bode et al., 1996, 2000a; Razin, 1999]. Experimental evidence indicates that S/MARs may physically separate ("structural" S/MARs) neighboring chromatin loops or domains that differ structurally and/or functionally, e.g., by their transcriptional activity and/or topological state [Geyer, 1997]. An attractive hypothesis ascribes to S/MARs an involvement into a large-scale regulation of genome activity by assuming that a specific fraction of them ("functional" S/MARs) can bind nuclear matrix under certain conditions and form chromatin domains depending on cell or tissue type [Jackson et al., 1992].

Other candidate elements participating in the formation of functional and structural chromatin domains were identified by their in vivo manifestation, namely insulators [Bell and Felsenfeld, 1999] and locus control regions [Li and Peterson, 1999]. Controversial data were obtained on interrelation of insulator elements and S/MARs. At least in some cases, insulators were capable of binding nuclear matrix [Nabirochkin et al., 1998; Namciu et al., 1998; Antes et al., 2001]. As an example, two different human S/MARs from the apolipoprotein B and alpha 1-antitrypsin loci were active in a *Drosophila* white gene position effect assay [Namciu et al., 1998]. The results of other groups, however, discriminated between insulator and matrix binding activities [Scott et al., 1999].

One possible reason for this controversy is that these elements were characterized for only a few functionally defined domains, like human beta-globin [Engel and Tanimoto, 2000] or interferon-beta [Bode et al., 1995]. Other domains may have different sets of functional modules contributing to the domain organization. To resolve this controversy, it would be helpful to find and correlate the positions of insulator, LCR, and S/MAR elements within long genome stretches spanning over several independently regulated genes, and therefore possibly including several functional domains.

Recently, we proposed a technique [Nikolaev et al., 1996] based on selection and cloning of genomic restriction fragments capable of binding specifically to isolated nuclear matrix in vitro. The technique was tested with human chromosome 19 and found to be applicable to the isolation and mapping of S/MARs from whole human chromosomes. In the present work, we used this procedure to select and map S/MAR elements within a one megabase (1 Mb) long sequenced *D19S208-COX7A1* locus of human chromosome 19q13.12. The locus contains 22 identified genes with characterized products, many of which are regulated in a tissue specific manner, and 6 genes coding for hypothetical proteins. Hopefully, the identification of S/MAR positions relative to genes within such a long stretch of the genome will help to get a deeper insight into the structural basis of S/MARs functioning. This approach can be further used for annotation of long sequenced genomic stretches and/or whole chromosomes by experimental mapping of functional genomic elements.

## MATERIALS AND METHODS

### Basic Protocols

Growth and transformation of *E. coli* cells, preparation of plasmid DNA, agarose gel electrophoresis, blot-hybridization, and other standard manipulations were performed as described [Sambrook et al., 1989].

### Cosmid DNA Preparation

*E. coli* cells containing cosmids that represent the *D19S208-COX7A1* locus of human chromosome 19 were provided by Dr. Anne Olsen (Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory) and grown overnight at 37°C in 5 ml of LB medium supplemented with kanamycin (20 μg/ml). Cosmid DNA was isolated using a Wizard Plus Minipreps DNA Purification System (Promega) according to the manufacturer's recommendations.

### Nuclear Matrix Preparation and In Vitro Binding of S/MARs

Nuclear matrices (scaffolds) were prepared as described previously [Izaurralde et al., 1988] with 15 mM of lithium 3,5-diiodosalicylate from

HeLa S3 cells grown in suspension in S-MEM with 10% of fetal calf serum. DNA was fragmented with *Xho* I restriction endonuclease at a concentration of 1,500 U/mg of initial nuclear DNA.

Binding of selected fragments to isolated nuclear matrices in vitro was performed as described [Izaurralde et al., 1988]. Alpha-$^{32}$P-dATP was incorporated into S/MARs and positive/negative control DNA fragments in the course of PCR amplification [Nikolaev et al., 1996]. Binding coefficients were calculated as before [Nikolaev et al., 1996] and normalized to the coefficient of the positive control, the insert of a pUCMAR10 plasmid [Bode et al., 1992].

## S/MARs Library Construction

A general construction scheme is depicted in Figure 1. Each individual cosmid DNA (0.1 μg) was digested to completion with 20 U of either

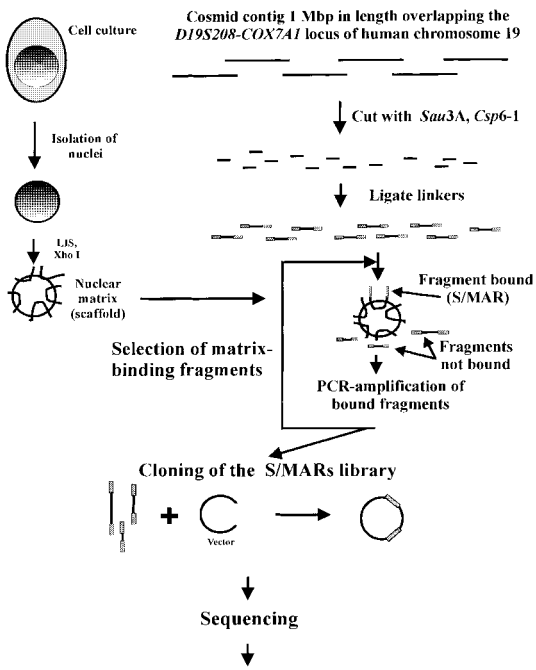



**Fig. 1.** Construction of the S/MARs library. A pool of short DNA fragments, covering the whole *D19S208-COX7A1* locus and containing the library primer at both ends, was mixed with isolated nuclear matrices (scaffolds), incubated, extensively washed to remove unbound DNA, and matrix-bound DNA was then PCR-amplified with the library primer and used for the next round of selection. The process was repeated five times. After the 5$^{th}$ round of binding with nuclear matrix the PCR-amplified mixture of the fragments was cloned, sequenced, and mapped within the *D19S208-COX7A1* locus.

*Sau*3A or *Csp*6I (Fermentas) in manufacturer's buffer for 3 h at 37°C. *Sau*3A and *Csp*6I digested cosmids were pooled separately, phenol-extracted and ethanol-precipitated. Ligation of the corresponding linkers and PCR-amplification were done as described earlier [Nikolaev et al., 1996]. The resulting *Sau*3A and *Csp*6I libraries were pooled together forming a short fragment library of the human chromosome 19 *D19S208-COX7A1* locus.

S/MARs were then selected from the short fragment library obtained using the in vitro selection procedure [Nikolaev et al., 1996], see Figure 1. Nuclear matrices ($1.5 \times 10^7$ initial cells) were washed three times with 500 μl of 40 mM Tris-HCl, pH 7.4, 0.1 mM spermine, 0.25 mM spermidine, 40 mM KCl, 140 mM NaCl, 4 mM EDTA, 0.2% digitonin, 0.2 mM PMSF, and resuspended in 50 μl of the same buffer. Fragmented *E. coli* DNA (100 μg) was then added to the matrices, the reaction volume was adjusted to 95–98 μl with water and the mixture was pre-incubated for 30 min at room temperature. A 5–25 ng of the short fragment library DNA in 2–5 μl of TE buffer was then added, and the incubation was continued for an additional 3 h. The nuclear matrix pellet was collected in a microcentrifuge (30 s; 10,000 rpm), washed with 500 μl of 20 mM Tris-HCl, pH 7.4, 0.05 mM spermine, 0.125 mM spermidine, 20 mM KCl, 70 mM NaCl, 2 mM EDTA, 0.1% digitonin, 0.1 mM PMSF, and resuspended in 100 μl of TE buffer with 0.5% SDS. Proteinase K was then added up to 50 μg/ml, and the suspension was incubated overnight at 56°C. The matrix-bound DNA was extracted with phenol/chloroform/isoamyl alcohol (25:24:1), precipitated with ethanol in the presence of 0.3 M NaOAc, washed with 70% ethanol, dried and dissolved in 20 μl of TE buffer. A 1–2 μl of the DNA solution was PCR-amplified (94°C/20 s; 60°C/30 s;72°C/75 s) with the library primer (ACTGAGCTCGAGTATCCATGAACA) for 17, 15, 13, 12, or 8 cycles (after the 1$^{st}$–5$^{th}$ selection round, respectively) and used for the next selection round. After five selection rounds, the resulting DNA was cloned into a pCR2.1 vector (Invitrogen) according to the manufacturer's protocol and clones were arrayed in 96-well microtiter plates.

## Sequencing and Computer Analysis

Clone inserts were sequenced using an ALFexpress II automated DNA sequencer

(Amersham-Pharmacia Biotech). The sequences were mapped to the GenBank deposits (nucleotide, EST, and HTGS divisions) using the BLAST [Altschul et al., 1997] server at NCBI (www.ncbi.nlm.nih.gov/BLAST). The data were further analyzed with the help of the Draft Human Genome Browser (http://genome.ucsc.edu/goldenPath/hgTracks.html).

## RESULTS

### S/MARs Library Construction

A general scheme of the S/MARs library construction is shown in Figure 1. A pool of short (200–1,000 bp) fragments with the library primer at both ends, obtained as described in Materials and Methods, was used for the selection of S/MARs by the in vitro matrix binding assay. Nuclear matrices (scaffolds) were mixed with this pool in the presence of excess amounts of blocking *E. coli* DNA, incubated, extensively washed to remove unbound DNA, and matrix-bound DNA fragments were isolated. The matrix-bound DNA was then PCR-amplified with the library primer and used for the second round of matrix binding. The process was repeated five times. The result of each successive binding step is shown in Figure 2. The initial DNA gave a smear of about 4,000 expected fragments (lane 1). After successive matrix binding-purification rounds, the smear was gradually replaced by a ladder of a limited number of fragments, indicating preferential selection of a putative matrix binding fraction of the initial DNA.

After the fifth round of binding with the nuclear matrix, the PCR-amplified mixture of the fragments was cloned into a plasmid vector. Transformed *E. coli* cells were plated on X-gal/IPTG agar plates and 184 white colonies were arrayed on 96-well clusters.

### Hybridization and Sequencing

The presence and the size of the inserts in library clones were checked by PCR with the library primer. 168 PCR-amplified inserts were resolved in an agarose gel and blotted to nylon filters. The filters were hybridized to a random primer labeled pool of the fragments selected after the fifth selection round (Fig. 3). Twenty-nine clones having given strong hybridization signals and therefore abundant in the library were selected for sequencing. The clones with weak hybridization signals (like 2 and 10 in Fig. 3) were considered non-specific and excluded from further analysis.

### Map of the S/MARs

Twenty-nine clones selected from the library of putative S/MARs were sequenced yielding 16 independent sequences. To estimate the number of independent clones in the library, we

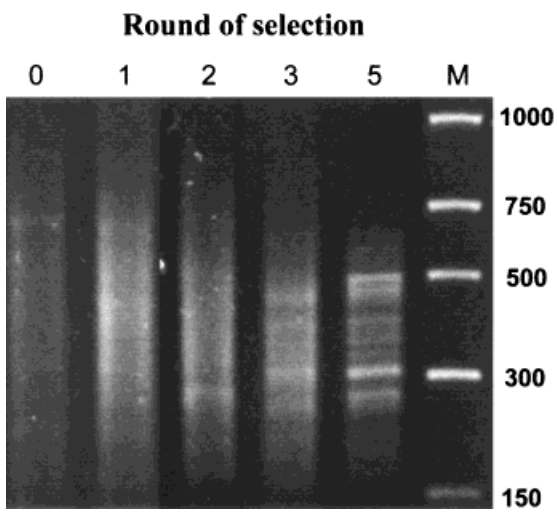



**Fig. 2.** Selection of DNA fragments binding preferentially to nuclear matrix. Agarose gel electrophoresis (1%) of the initial short fragment library of the *D19S208-COX7A1* locus (**lane 0**) and fragments after 1st, 2nd, 3rd, and 5th rounds of the selection (**lanes 1, 2, 3,** and **5**, respectively). M-DNA length marker (Novagen). Note that the initial smear gradually transforms into a ladder of distinct bands.

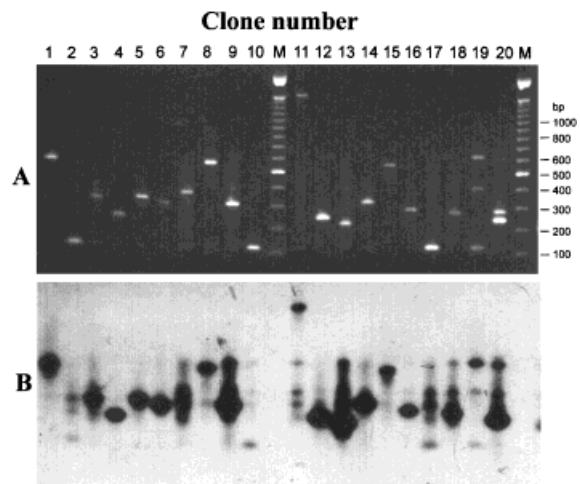



**Fig. 3.** Agarose gel electrophoresis (1%) of the PCR-amplified inserts from the library of putative S/MARs after staining with ethidium bromide (**panel A**) and blot-hybridization of the same fragments to the random primer labeled pool of the putative S/MARs selected after the 5th selection round (**panel B**). M-DNA length marker.

**TABLE I. Properties of S/MAR Sequences**

| S/MAR no. | GenBank accession no. of the S/MAR containing sequence | Position of the S/MAR in GenBank sequence | G + C (%) | A + G (%) | Density of AT-rich regions (%)[a] | Density of inverted repeats (%)[b] | Matrix binding efficiency (%)[c] |
|---|---|---|---|---|---|---|---|
| 1 | AC002128 | 28543–28841 | 40 | 59 | 29 | 9 | 40 |
| 2 | AC002132 | 9575–10020 | 50 | 46 | 0 | 0 | 50 |
| 3 | AC002132 | 14934–15333 | 51 | 41 | 0 | 17 | 75 |
| 4 | AC002997 | 23799–22822 | 41 | 61 | 30 | 44 | 20 |
| 5 | AC002997 | 30367–31030 | 39 | 39 | 44 | 68 | 20 |
| 6 | AC002115 | 6361–7037 | 45 | 43 | 11 | 20 | 30 |
| 7 | U95090 | 3681–3352 | 44 | 59 | 26 | 15 | 10 |
| 8 | AC002133 | 18499–18938 | 46 | 34 | 10 | 13 | 120 |
| 9 | AC002133 | 31379–31559 | 54 | 48 | 0 | 24 | 70 |
| 10 | AD000864 | 11263–10625 | 50 | 35 | 3 | 17 | 90 |
| 11 | AF038458 | 24698–24511 | 57 | 30 | 0 | 15 | 80 |
| 12 | AF038458 | 29101–29468 | 41 | 68 | 32 | 8 | 40 |
| 13 | AF038458 | 71228–71670 | 50 | 36 | 5 | 6 | 45 |
| 14 | AC002116 | 2683–2239 | 46 | 39 | 20 | 27 | 60 |
| 15 | AC004144 | 27978–28345 | 49 | 41 | 13 | 19 | 40 |
| 16 | AD001527 | 22629–22125 | 51 | 51 | 17 | 22 | 50 |

[a]Percentage of continuous AT-rich (> 75% A + T) regions longer than 20 bp.
[b]Percentage of perfect inverted repeats (including palindromes) longer than 6 bp.
[c]As compared to the positive control—the insert of a pUCMAR10 plasmid (see Materials and Methods).

assumed that the frequency of cloned sequences occurrence in library samples fits the Poisson distribution. Using the least-squares method, we adjusted a Poisson curve to fit the data obtained. From this Poisson curve, the library size can be calculated as (the number of selected clones)/$q$, where $q$ is a parameter of the Poisson distribution. The estimated number was found to be about 20. It means that 16 sequences found may represent most, if not all, potential S/MARs in our library.

All independent sequences were compared to the GenBank database (Table I). As expected, all 16 sequences were unambiguously mapped into sequences of the human chromosome 19 *D19S208-COX7A1* locus. Accession numbers of the corresponding genomic sequences and positions of the isolated sequences within them are presented in Table I. A resulting map indicating the positions of all S/MAR elements relative to genes identified in the locus is presented in Figure 4.

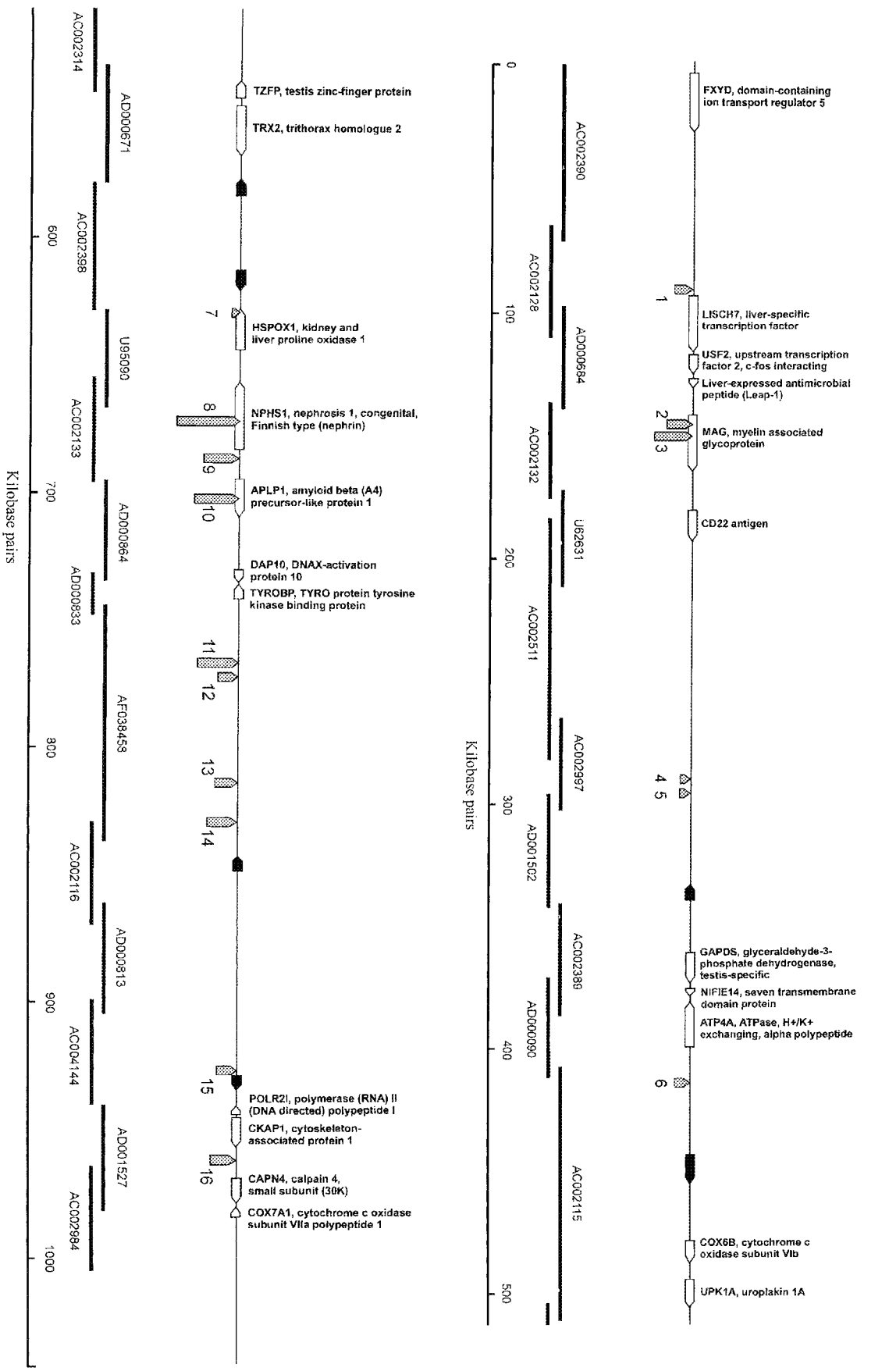
### Analysis of the Sequences and Binding Coefficients

Nuclear matrix binding coefficients for 16 selected S/MARs were determined as described earlier [Nikolaev et al., 1996] and normalized to an assumed 100% binding efficiency of the positive control–the insert of a pUCMAR10 plasmid (a pUC19 vector with a ten times repeated synthetic 25 bp sequence of an S/MAR located 3′ to the IgH enhancer [Bode et al., 1992]). The results are shown in Figure 5, and mean binding coefficients obtained using at least two independent matrix preparations are summarized in Table I.

It should be noted that the values of binding coefficients, and to a lesser extent, their normalized values, can vary for different nuclear matrix preparations, and therefore, we consider these data as only semi-quantitative. On the other hand, all S/MARs do bind the nuclear matrix stronger than the negative control in all experiments, and their relative binding efficiency is generally conserved from one experiment to another.

Although S/MARs of the now classical type have been reported to be generally AT rich, recent studies assign a more important role to a regular distribution of "90% AT boxes" [Michalowski et al., 1999]. We analyzed S/MAR sequences using the PC/Gene package. Out of 16, 10 characterized S/MARs were found to

**Fig. 4.** A map of the *D19S208-COX7A1* locus with positions of S/MARs indicated by shaded vertical arrows. The lengths of the arrows roughly correspond to the nuclear matrix binding strength of the indicated S/MARs. Empty horizontal arrows indicate identified genes and their direction of transcription, solid horizontal arrows indicate positions of hypothetical genes. Clone coverage of the locus is presented under the map together with the GenBank accession numbers of the respective cosmid sequences.


FXYD, domain-containing ion transport regulator 5
LISCH7, liver-specific transcription factor
USF2, upstream transcription factor 2, c-fos interacting
Liver-expressed antimicrobial peptide (Leap-1)
MAG, myelin associated glycoprotein
CD22 antigen
GAPDS, glyceraldehyde-3-phosphate dehydrogenase, testis-specific
NIFIE14, seven transmembrane domain protein
ATP4A, ATPase, H+/K+ exchanging, alpha polypeptide
COX6B, cytochrome c oxidase subunit VIb
UPK1A, uroplakin 1A
AC002390
AC002128
AD000684
AC002132
U62631
AC002511
AC002997
AD001502
AC002389
AD000090
AC002115
0
100
200
300
400
500
Kilobase pairs
TZFP, testis zinc-finger protein
TRX2, trithorax homologue 2
HSPOX1, kidney and liver proline oxidase 1
NPHS1, nephrosis 1, congenital, Finnish type (nephrin)
APLP1, amyloid beta (A4) precursor-like protein 1
DAP10, DNAX-activation protein 10
TYROBP, TYRO protein tyrosine kinase binding protein
POLR2I, polymerase (RNA) II (DNA directed) polypeptide I
CKAP1, cytoskeleton-associated protein 1
CAPN4, calpain 4, small subunit (30K)
COX7A1, cytochrome c oxidase subunit VIIa polypeptide 1
AC002314
AD000671
AC002398
U95090
AC002133
AD000864
AD000833
AF038458
AC002116
AD000813
AC004144
AD001527
AC002984
600
700
800
900
1000
Kilobase pairs

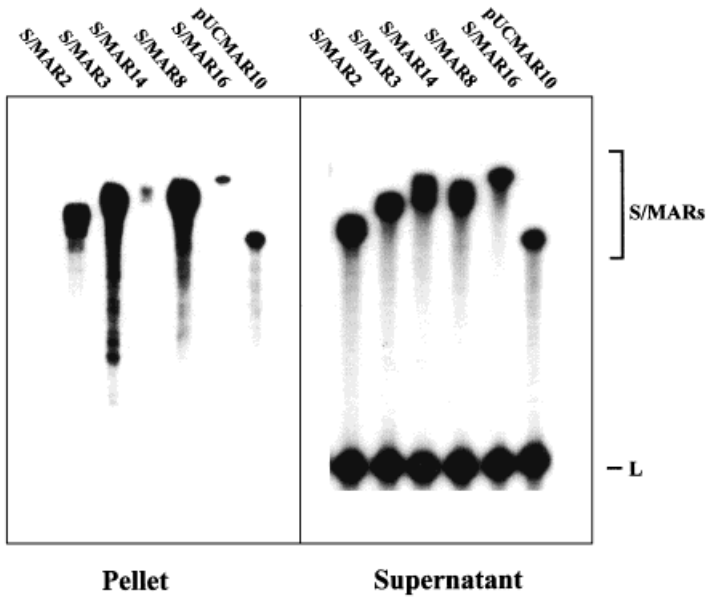



**Fig. 5.** Specific binding of the S/MARs obtained to the nuclear matrix. Radioactively labeled S/MARs were incubated with the isolated nuclear matrix. The fractions bound (pellet) and not bound (supernatant) to the nuclear matrix were separated in a 10% denaturing polyacrylamide gel. pUCMAR10–positive control, a synthetic sequence of an S/MAR located 3′ to the IgH enhancer; L–negative control, a fragment of lambda phage genome.

contain more than 10% of $\geq$ 20 bp long regions with an AT content of more than 75% (Table I; see [Boulikas, 1995]). However, four S/MARs (2, 3, 9, 11) with high affinity for the nuclear matrix did not contain any AT-rich sequences longer than 20 bp. Similarly, the number of inverted repeats (potential hairpins or cruciform structures, [Boulikas and Kong, 1993]) within the S/MARs also did not correlate with the matrix affinity: e.g., S/MARs 8 and 10 with high affinity contained a much lesser number of inverted repeats than S/MAR 5, a clone with a comparatively low affinity (Table I).

## DISCUSSION

Now, when sequences of several metazoan genomes, including the human genome, are almost finished, an old problem of structural basis of the functional organization of the genomes can get a principally new impetus. One of the first steps toward this goal could be the construction of a comprehensive physical map of various functional elements within the genome context.

One of the challenges for such a mapping is definition and positioning of autonomous functional units also called functional gene expression domains [Boulikas, 1995; Dillon and Sabbattini, 2000]. Many authors are inclined to consider the functional domains structurally and functionally relevant to series of individual loops in the bases of which lie S/MARs (see Introduction). One of the roles of S/MARs is the insulation of neighboring domains in such a way that active transcription and regulatory elements (e.g., enhancers) of different domains would not interfere with each other [Bode et al., 1992]. Another concept is based on the idea that the propagation and maintenance of open (active) chromatin could well be governed via transcription factor binding sites distributed across the region [Dillon and Grosveld, 1994; Dillon and Sabbattini, 2000] or cellular compartmentalization of DNA and regulatory proteins [Stein et al., 2000].

Most approaches used so far to define chromatin domains had a serious limitation: they dealt with domains containing either a single gene or a cluster of related genes. To avoid this limitation, it is necessary to map all the S/MARs within long gene-rich DNA stretches and to correlate their positions with the genomic location of variously regulated genes. Such an attempt has been made recently for a ~150 kb long region containing the human serpin gene cluster at 14q32.1 [Rollini et al., 1999].

We mapped S/MARs within a 1 Mb long stretch of human chromosome 19 DNA located between markers *D19S208* and *COX7A1*. The *D19S208-COX7A1* locus was chosen as one of sufficiently long completely sequenced loci, that contains 22 identified genes with characterized products and 6 genes coding for hypothetical proteins. Gene names, positions, and transcription directions are presented in Figure 4 and Table II.

Sixteen S/MAR elements were identified and mapped within this locus. Their in vitro nuclear matrix (scaffold) binding affinity ranged from 10 to 120% of that for a well characterized S/MAR-element derived from the mouse IgH locus [Cockerill et al., 1987]. Since, the estimated total number of individual S/MARs in the library is about 20, the majority of the locus S/MARs were mapped, even though some S/MAR-containing fragments could be lost during the preparation procedure.

Out of 16, 11 identified S/MARs were found to lie in the intergenic regions, consistent with their role in delimitation of chromatin domains (Fig. 4, Table II). These 11 S/MARs subdivide the locus into 10 putative domains of 6–272 kb long with an average domain size of 88 kb, in good accord with the previous estimations of the chromatin loop size [Jackson et al., 1990; Jack and Eggert, 1992; Boulikas, 1995; Iarovaia and Razin, 1996]. At the same time, 4 of the 10 domains lacked any characterized or hypothetical genes. Three of these "pseudodomains" are short (6–16 kb) and may just be something similar to a long genomic regulatory element containing several proximal S/MAR elements,

**TABLE II. Characteristics of Putative Domains Delimited by S/MARs Located in Intergenic Regions of the *D19S208-COX7A1* Locus of Human Chromosome 19**

| Region between S/MARs denoted by their nos. | Length of the region (kb) | Genes assigned to the region | Tissue specificity of the gene expression |
|---|---|---|---|
| 1–4 | 199 | *LISH7*, liver-specific transcription factor | Liver |
| | | *USF2*, upstream transcription factor 2, c-fos interacting | Many tissues |
| | | *LEAP-1*, liver-expressed antimicrobal peptide | Liver |
| | | *MAG*, myelin associated glycoprotein | Nervous tissues |
| | | *CD22* antigen | B-lymphocytes |
| 4–5 | 6 | None | |
| 5–6 | 120 | *GAPDS*, glyceraldehyde-3-phosphate dehydrogenase, testis-specific | Testis |
| | | *NIFIE14*, seven-transmembrane domain protein | ND |
| | | *ATP4A*, ATPase, $H^+/K^+$ exchanging, alpha-polypeptide | Gastric parietal cells |
| 6–9 | 272 | *COX6B*, cytochrome C oxidase, subunit VIb | Many tissues, mitochondrial enzyme |
| | | *UPKA1*, uroplakin 1A | Urothelium |
| | | *TZFP*, testis zinc-finger protein | Testis |
| | | *TRX2*, trithorax homologue 2 | ND |
| | | *HSPOX1*, kidney and liver proline oxidase 1 | Kidney, liver (less in brain, heart, no in other tissues) |
| | | *NPHS1*, nephrosis 1, congenital, Finnish type (nephrin) | Kidney glomerular epithelial cells |
| 9–11 | 81 | *APLP1*, amyloid beta (A4) precursor-like protein 1 | Nervous tissues |
| | | *DAP10*, DNAX activation protein 10 | NK and T-cells |
| | | *TYROBP*, TYRO protein tyrosine kinase binding protein | NK cells |
| 11–12 | 6 | None | |
| 12–13 | 41 | None | |
| 13–14 | 16 | None | |
| 14–15 | 97 | Hypothetical gene | ND |
| 15–16 | 35 | *POLR2I*. polymerase (RNA) II (DNA-directed) polypeptide 1 | Housekeeping |
| | | *CKAP1*, cytosceleton associated protein 1 | Many tissues |

ND, not determined.

like a locus control region [Li and Peterson, 1999]. In turn, the longer "pseudodomain" may contain not yet identified genes.

The map demonstrates that genes with significantly different tissue specificity can be found within one and the same domain. An example is *COX6B* coding for mitochondrial cytochrome C oxidase 6B subunit expressed ubiquitously [Carrero-Valenzuela et al., 1991], and uroplakin 1A gene which is expressed predominantly in urothelium [Sun et al., 1996]. These results can be explained in different, but not necessarily mutually exclusive ways:

I. There are data suggesting that S/MARs can be subdivided into two groups [Phi-Van and Stratling, 1990; Jackson et al., 1992]. The first one is usually termed "stable," "constitutive," or "structural" S/MARs; this group is thought to be responsible for the formation of the chromatin domains common for all cells and tissues. In contrast, the second group named "dynamic," "tissue-specific," or "function-dependent" S/MARs is supposed to form temporary domains in particular tissues or at definite stages of the cell cycle, thus being involved in the large-scale genetic regulation [Razin and Vassetzky, 1992; Bode et al., 1995]. There is strong evidence for the existence of tissue-specific S/MARs: e.g., it was shown that a human apolipoprotein gene is bound to the nuclear matrix only in HepG2 cells actively expressing this gene [Levy-Wilson and Fortier, 1989]. Another example are S/MARs of the malic enzyme gene which are anchored to the nuclear matrix in thymus cells, where the enzyme is active, but not in reticulocytes [Brotherton et al., 1991]. Tissue specific proteins possessing S/MAR binding activity have also been described (review: [Boulikas, 1995]). In this connection, it would be very interesting to compare distributions of S/MARs in the same genomic locus, but in different cell types.
II. Apart from S/MARs, various regulatory factor binding sites distributed across the active regions are needed to determine functional chromatin domains. In this case, S/MARs can play a role of delimiters of large open chromatin stretches, but whether a particular gene is expressed or silent in this region depends on gene or tissue specific factors interacting with regulatory elements.
III. Some of the S/MARs located between genes with different tissue specificity were missed.
IV. S/MARs are not the only type of boundary elements, as was previously suggested [Dillon and Grosveld, 1994; Dillon and Sabbattini, 2000].

Additional experiments are needed to test these possibilities, such as finding a correlation of S/MARs and insulators positions in the region [Bode et al., 2000b], and distribution of DNase hypersensitive sites which is now in progress.

Out of 16, 5 S/MARs were found within intronic sequences of four known genes: amyloid beta precursor-like protein 1 (*APLP1*) [Bayer et al., 1999], kidney and liver proline oxidase 1 (*HSPOX1*), myelin associated glycoprotein (*MAG*) belonging to the immunoglobulin superfamily [Konat, 1996; Schachner and Bartsch, 2000], and nephrin (*NPHS1*) [Lenkkeri et al., 1999]. Interestingly, all four genes are suggested to be involved in the pathogenesis of various heritable diseases, like multiple sclerosis (*MAG*), Alzheimer disease (*APLP1*), type I hyperprolinemia (*HSPOX1*), and congenital nephrotic syndrome of the Finnish type (*NPHS1*). Exact locations of the S/MARs within these genes are depicted in Figure 6.

S/MARs were found in introns of several genes, including those coding for the mouse kappa immunoglobulin light chain, mouse heavy chain μ, human beta-globin, hamster dihydrofolate reductase, and human topo I (reviewed in [Boulikas, 1995; Phi-Van and Stratling, 1996; Bode et al., 1998]). Intronic MARs are often associated with important cis-regulatory elements. For example, two intronic S/MARs of the mouse kappa immunoglobulin light chain gene and T-cell receptors genes are

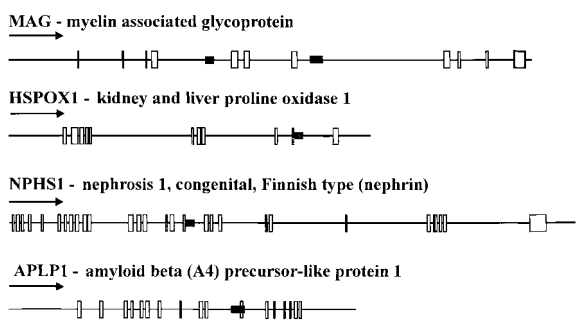



**Fig. 6.** Location of S/MAR-elements (black rectangles) in the intronic sequences of four human genes. Exons are denoted by open vertical rectangles. Arrows indicate the direction of transcription. See text for detail.

flanking a tissue-specific enhancer, and both elements are required for proper regulation of the gene in development [Lichtenstein et al., 1994; Yi et al., 1999; Zhong et al., 1999].

The occurrence of S/MARs in introns imposes certain restrictions on these elements. In particular, since these S/MARs are transcribed, they should not interfere with read-through by RNA polymerase II [Phi-Van and Stratling, 1990; Bode et al., 1998]. Their interaction with nuclear matrix may be tissue specific and/or realized transiently during certain periods of the cell life. Therefore, these S/MARs are probably the best candidates for "functional" or "dynamic" elements [Jackson et al., 1992]. It should be also noted that the expression of known genes harboring our S/MARs within their introns is highly tissue-specific: two of them (*APLP1* and *MAG*) are expressed in neuronal tissues [Pedraza et al., 1991; Lenkkeri et al., 1998], and two other in kidney. All of these genes are not expected to be expressed in HeLa cells used in this work for nuclear matrix preparation.

The mapping of matrix-attachment regions in a megabase long locus of human chromosome 19 provides the first insight into the functioning of the locus as a whole. The information obtained presents a useful conceptual framework for planning future experiments aimed at deciphering functions of already identified S/MARs, where the structural elements will be tested both in vitro and in transgenic animals containing individual elements or their combinations.

## ACKNOWLEDGMENTS

The authors thank V. Potapov and N. Skaptsova for oligonucleotide synthesis, and J. Bode and B. Glotov for discussions and helpful suggestions. Our special thanks are due to A. Carrano, E. Branscomb, A. Olsen, L. Ashworth, and other colleagues from the Lawrence Livermore National Laboratory for cosmid clones and indispensable help in the mapping.

## REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

Antes TJ, Namciu SJ, Fournier RE, Levy-Wilson B. 2001. The 5′ boundary of the human apolipoprotein b chromatin domain in intestinal cells. Biochemistry 40: 6731–6742.

Bayer TA, Cappai R, Masters CL, Beyreuther K, Multhaup G. 1999. It all sticks together–the APP-related family of proteins and Alzheimer's disease. Mol Psychiatry 4:524–528.

Bell AC, Felsenfeld G. 1999. Stopped at the border: Boundaries and insulators. Curr Opin Genet Dev 9:191–198.

Bentley DR. 2000. The Human Genome Project–An overview. Med Res Rev 20:189–196.

Berezney R, Wei X. 1998. The new paradigm: Integrating genomic function and nuclear architecture. J Cell Biochem Suppl 31:238–242.

Berezney R, Mortillaro MJ, Ma H, Wei X, Samarabandu J. 1995. The nuclear matrix: A structural milieu for genomic function. Int Rev Cytol 162A:1–65.

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.

Bode J, Schlake T, Rios-Ramirez M, Mielke C, Stengert M, Kay V, Klehr-Wirth D. 1995. Scaffold/matrix-attached regions: Structural properties creating transcriptionally active loci. Int Rev Cytol 162A:389–454.

Bode J, Stengert-Iber M, Kay V, Schlake T, Dietz-Pfeilstetter A. 1996. Scaffold/matrix-attached regions: Topological switches with multiple regulatory functions. Crit Rev Eukaryot Gene Expr 6:115–138.

Bode J, Bartsch J, Boulikas T, Iber M, Mielke C, Schubeler D, Seibler J, Benham C. 1998. Transcription-promoting genomic sites in mammalia: Their elucidation and architectural principles. Gene Therapy Mol Biol 1:551–580.

Bode J, Benham C, Knopp A, Mielke C. 2000a. Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements). Crit Rev Eukaryot Gene Expr 10:73–90.

Bode J, Schlake T, Iber M, Schubeler D, Seibler J, Snezhkov E, Nikolaev L. 2000b. The transgeneticist's toolbox: novel methods for the targeted modification of eukaryotic genomes. Biol Chem 381:801–213.

Boulikas T. 1995. Chromatin domains and prediction of MAR sequences. Int Rev Cytol 162A:279–388.

Boulikas T, Kong CF. 1993. Multitude of inverted repeats characterizes a class of anchorage sites of chromatin loops to the nuclear matrix. J Cell Biochem 53:1–12.

Brotherton T, Zenk D, Kahanic S, Reneker J. 1991. Avian nuclear matrix proteins bind very tightly to cellular DNA of the beta-globin gene enhancer in a tissue-specific fashion. Biochemistry 30:5845–5850.

Carrero-Valenzuela RD, Quan F, Lightowlers R, Kennaway NG, Litt M, Forte M. 1991. Human cytochrome c oxidase subunit VIb: Characterization and mapping of a multigene family. Gene 102:229–236.

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.

Cockerill PN, Yuen MH, Garrard WT. 1987. The enhancer of the immunoglobulin heavy chain locus is flanked by presumptive chromosomal loop anchorage elements. J Biol Chem 262:5394–5397.

Dillon N, Grosveld F. 1994. Chromatin domains as potential units of eukaryotic gene function. Curr Opin Genet Dev 4:260–264.

Dillon N, Sabbattini P. 2000. Functional gene expression domains: Defining the functional unit of eukaryotic gene regulation. Bioessays 22:657–665.

Engel JD, Tanimoto K. 2000. Looping, linking, and chromatin activity: new insights into beta-globin locus regulation. Cell 100:499–502.

Geyer PK. 1997. The role of insulator elements in defining domains of gene expression. Curr Opin Genet Dev 7:242–248.

Hancock R. 2000. A new look at the nuclear matrix. Chromosoma 109:219–225.

Iarovaia OV, Razin SV. 1996. Comparison of positions of attachment sites of DNA loops to nuclear matrix, MAR-elements, and autonomously replicating sequences in the elongated region of Drosophila melanogaster X-chromosome. Mol Biol (Mosk) 30:1184–1192.

Izaurralde E, Mirkovitch J, Laemmli UK. 1988. Interaction of DNA with nuclear scaffolds in vitro. J Mol Biol 200:111–125.

Jack RS, Eggert H. 1992. The elusive nuclear matrix. Eur J Biochem 209:503–509.

Jackson DA, Dickinson P, Cook PR. 1990. Attachment of DNA to the nucleoskeleton of HeLa cells examined using physiological conditions. Nucleic Acids Res 18:4385–4393.

Jackson DA, Dolle A, Robertson G, Cook PR. 1992. The attachments of chromatin loops to the nucleoskeleton. Cell Biol Int Rep 16:687–696.

Konat GW. 1996. Chromatin structure and transcriptional activity of MAG gene. Acta Neurobiol Exp 56:281–285.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. International Human Genome Sequencing Consortium. Nature 409:860–921.

Lenkkeri U, Kestila M, Lamerdin J, McCready P, Adamson A, Olsen A, Tryggvason K. 1998. Structure of the human amyloid-precursor-like protein gene APLP1 at 19q13.1. Hum Genet 102:192–196.

Lenkkeri U, Mannikko M, McCready P, Lamerdin J, Gribouval O, Niaudet PM, Antignac CK, Kashtan CE, Homberg C, Olsen A, Kestila M, Tryggvason K. 1999. Structure of the gene for congenital nephrotic syndrome of the finnish type (NPHS1) and characterization of mutations. Am J Hum Genet 64:51–61.

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

Lewis S, Ashburner M, Reese MG. 2000. Annotating eukaryote genomes. Curr Opin Struct Biol 10:349–354.

Li Q, Peterson KR. 1999. Locus control regions: Coming of age at a decade plus. Trends Genet 15:403–408.

Lichtenstein M, Keini G, Cedar H, Bergman Y. 1994. B cell-specific demethylation: A novel role for the intronic kappa chain enhancer sequence. Cell 76:913–923.

Michalowski SM, Allen GC, Hall GE Jr, Thompson WF, Spiker S. 1999. Characterization of randomly-obtained matrix attachment regions (MARs) from higher plants. Biochemistry 38:12795–12804.

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.

Nabirochkin S, Ossokina M, Heidmann T. 1998. A nuclear matrix/scaffold attachment region co-localizes with the gypsy retrotransposon insulator sequence. J Biol Chem 273:2473–2479.

Namciu SJ, Blochlinger KB, Fournier RE. 1998. Human matrix attachment regions insulate transgene expression from chromosomal position effects in Drosophila melanogaster. Mol Cell Biol 18:2382–2391.

Nikolaev LG, Tsevegiyn T, Akopov SB, Ashworth LK, Sverdlov ED. 1996. Construction of a chromosome specific library of human MARs and mapping of matrix attachment regions on human chromosome 19. Nucleic Acids Res 24:1330–1336.

Pedraza L, Frey AB, Hempstead BL, Colman DR, Salzer JL. 1991. Differential expression of MAG isoforms during development. J Neurosci Res 29:141–148.

Phi-Van L, Stratling WH. 1990. Association of DNA with nuclear matrix. Prog Mol Subcell Biol 11:1–11.

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.

Razin SV. 1999. Chromosomal DNA loops may constitute basic units of the eukaryotic genome organization and evolution. Crit Rev Eukaryot Gene Expr 9:279–283.

Razin SV, Vassetzky YS. 1992. Domain organization of eukaryotic genome. Cell Biol Int Rep 16:697–708.

Rollini P, Namciu SJ, Marsden MD, Fournier RE. 1999. Identification and characterization of nuclear matrix-attachment regions in the human serpin gene cluster at 14q32.1. Nucleic Acids Res 27:3779–3791.

Sambrook J, Fritsch EF, Maniatis T. 1989. "Molecular cloning. A laboratory manual." Cold Spring Harbour: CSHL Press.

Schachner M, Bartsch U. 2000. Multiple functions of the myelin-associated glycoprotein MAG (siglec-4a) in formation and maintenance of myelin. Glia 29:154–165.

Scott KC, Taubman AD, Geyer PK. 1999. Enhancer blocking by the *Drosophila gypsy* insulator depends upon insulator anatomy and enhancer strength. Genetics 153:787–798.

Stein GS, Montecino M, van Wijnen AJ, Stein JL, Lian JB. 2000. Nuclear structure-gene expression interrelationships: Implications for aberrant gene expression in cancer. Cancer Res 60:2067–2076.

Sun TT, Zhao H, Provet J, Aebi U, Wu XR. 1996. Formation of asymmetric unit membrane during urothelial differentiation. Mol Biol Rep 23:3–11.

Yi M, Wu P, Trevorrow KW, Claflin L, Garrard WT. 1999. Evidence that the *Igkappa* gene MAR regulates the probability of premature V-J joining and somatic hypermutation. J Immunol 162:6029–6039.

Zhong XP, Carabana J, Krangel MS. 1999. Flanking nuclear matrix attachment regions synergize with the T cell receptor delta enhancer to promote V(D)J recombination. Proc Natl Acad Sci USA 96:11970–11975.