Dissection of the Ability of the Chicken Lysozyme Gene 5′ Matrix Attachment Region To Stimulate Transgene Expression and To Dampen Position Effects[†,‡]

Loc Phi-Van[§] and Wolf H. Strätling*[,||]

Institut für Kleintierforschung, Dörnbergstrasse 25, 29223 Celle, FR Germany, and Institut für Physiologische Chemie, Universitäts-Krankenhaus Eppendorf, Martinistrasse 52, 20246 Hamburg, FR Germany

Received February 16, 1996; Revised Manuscript Received May 29, 1996[⊗]

ABSTRACT: The chicken lysozyme gene domain is flanked by nuclear matrix attachment regions (MARs) on each side. We have previously shown that bilaterally flanking 5′ MARs in stably transfected artificial genetic units enhance expression of a reporter transgene and dampen position effects of the chromatin structure at the site of integration. The 5′ MAR was now dissected into smaller fragments that were monitored for effects on transgene expression in mouse 3T3 cells by a similar assay. Fragments, which contain 1.32 and 1.45 kb and represent the upstream and the downstream half, respectively, of the 5′ MAR, retained the ability to stimulate transgene expression as well as the ability to reduce the variation in the level of expression. However, a 452 bp subfragment (H1-*Hae*II), which still exhibits specific binding to nuclear matrices and contains two high-affinity binding sites for the abundant nuclear matrix protein ARBP, lost both of those abilities. A dimerized 177 bp sequence from fragment H1-*Hae*II, which also binds selectively to nuclear matrices and includes a duplicated ARBP binding site, was also unable to stimulate reporter gene expression. Furthermore, a 0.65 kb subfragment containing an intrinsically bent sequence did not affect an elevated reporter gene expression and its dampening. Our results show that the ability of MAR fragments to bind to nuclear matrices is not sufficient to enhance and insulate transgene expression in stably transfected cells.

In higher eukaryotes, the genome is organized into functional units, called domains, each containing one or several coding regions and flanking regulatory sequences (Benyajati & Worcel, 1976). Transcriptionally active domains are usually characterized by an "open" chromatin structure that exhibits an elevated nuclease sensitivity (Weintraub & Groudine, 1976). The important question of how it is controlled that regulatory sequences act only on the proper promoter but not on inappropriate ones in neighboring domains has attracted considerable attention. Domains are thought to be separated from one another by the attachment of flanking sequences to a proteinaceous framework structure, called the nuclear matrix in interphase nuclei and the scaffold in metaphase chromosomes, thus generating a loop organization (Paulson & Laemmli, 1977). Such sequences (called MARs and SARs) have been identified by in vitro studies for many genes in diverse species, including plants (Gasser & Laemmli, 1987; Phi-Van & Strätling, 1990). They are often localized at the boundaries of domains, supporting the idea that they act by insulating domains against effects of the surrounding chromatin.

A chromatin domain that has been extensively studied in the past is that of the chicken lysozyme gene. It includes ∼24 kb of elevated nuclease sensitivity and contains the coding sequence, a total of nine regulatory sites, and AT-rich MAR sequences (about 3 kb in length each) on each flanking side (Phi-Van & Strätling, 1988; Bonifer et al., 1991). The location of the chicken lysozyme MARs at the domain boundaries encouraged the notion that they have effects on gene expression. This notion was first tested by use of synthetic domains containing a reporter gene, which was controlled by the lysozyme gene promoter, regulated in some constructs by the lysozyme −6.1 kb enhancer, and flanked by the lysozyme 5′ MAR on each side. These constructs were used to stably transfect chicken myelomonocytes. It was found that the flanking boundary sequences (called A-elements in that study) elevate gene expression about 10-fold and render it proportional to the number of integrated copies (Stief et al., 1989). Similar experiments with constructs, which contain a reporter gene controlled by the herpes simplex virus thymidine kinase gene promoter, by use of rat fibroblasts as recipient cells showed that the lysozyme 5′ MAR can confer transcription-elevating effects in a position-less dependent manner also on a heterologous promoter and in heterologous cells (Phi-Van et al., 1990). Furthermore, the chicken lysozyme 5′ MAR reduces the variability of transgene expression also in tobacco plants (Mlynárová et al., 1994). In transgenic mice, the lysozyme 5′ MAR can confer position-independent and properly regulated expression of the whey acidic protein transgene (McKnight et al., 1992). A more complete analysis of the function that is exerted by the 5′ MAR and by the various 5′ regulatory sites on lysozyme gene expression in transgenic mice has been performed only very recently (Sippel et al.,

[†] This work was supported by grants to W.H.S. from the Deutsche Forschungsgemeinschaft.

[‡] The nucleotide sequence of the fragment B-1-H1 is deposited in the EMBL, GenBank, and DDBJ Nucleotide Sequence Databases under Accession Number X98408. The sequence of fragment H1-*Sac*I has been deposited along with previous reports (accession numbers X52989 and X84223) (von Kries et al., 1990; Buhrmester et al., 1995).

* Author to whom correspondence should be addressed. Telephone: 49-40-4717-2392. Fax: 49-40-4717-4862.

[§] Institut für Kleintierforschung.

[||] Institut für Physiologische Chemie.

[⊗] Abstract published in *Advance ACS Abstracts,* July 15, 1996.

S0006-2960(96)00378-9 CCC: $12.00

1993; Bonifer et al., 1994). A group of proximal regulatory sites (located between −1 and −3 kb) and a group of two distal ones (at −6.1 and −7.9 kb) are both required for high-level and position-independent expression. Deletion of the MARs had no influence on the copy number dependency of expression but significantly increased the incidence of ectopic expression. Since the effects of MARs are most probably mediated by specific proteins and the chromatin encompassing them, it is significant that proteins that bind to the 5′ MAR with high affinity have been identified, purified, and characterized (von Kries et al., 1991, 1994; Buhrmester et al., 1995).

MAR elements from other genes and species, including plants, have insulating and transcription-enhancing activities comparable to those of the chicken lysozyme 5′ MAR (Blasquez et al., 1989; Xu et al., 1989; Mielke et al., 1990; Klehr et al., 1991; Breyne et al., 1992; Dietz et al., 1994; Thompson et al., 1994). As an example, the MAR elements that flank the enhancer of the immunoglobulin μ heavy chain locus are essential for transcription of the μ gene in B lymphocytes of transgenic mice (Forrester et al., 1994). Removal of the MARs decreased transcription 35−1000-fold and, furthermore, markedly increased the variation in the level of expression among the different mouse lines. However, it is presently not known how MARs act at the molecular level. Besides MARs, two other types of insulating elements have been described. First, the hypersensitive sites in the locus control region of the β-globin gene cluster confer high-level and copy number-dependent expression (Grosveld et al., 1987). Second, the scs and scs′ elements flanking the *Drosophila* hsp70 heat shock genes at locus 87A7 decrease position effect variation of reporter gene expression (Kellum & Schedl, 1991) and, when positioned between an upstream enhancer and the promoter, block the action of the enhancer (Kellum & Schedl, 1992). It is likely that all three types of elements have different underlying mechanisms of function.

In contrast to enhancers, MARs are extended sequences containing several hundred to a few thousand bp, e.g. 2.95 kb in the case of the lysozyme 5′ MAR. In this study, we therefore address the question of whether the effect of the lysozyme 5′ MAR to enhance transgene expression in a position-less dependent manner can be localized to shorter sequences. We found that 5′ MAR fragments containing 1.32 or 1.45 kb retained the abilities to stimulate transgene expression and to reduce the variation in the level of expression. However, two shorter fragments that still bind to the nuclear matrix and contain high-affinity binding sites for the abundant nuclear matrix protein ARBP proved to be unable to perform these activities. Thus, the ability of MAR fragments to bind the nuclear matrix is not sufficient to enhance and insulate transgene expression in stably transfected cells.

EXPERIMENTAL PROCEDURES

Recombinant DNA Plasmids. A plasmid (pLYSCAT2100) carrying the bacterial chloramphenicol acetyltransferase gene controlled by the chicken lysozyme gene promoter (+14 to −579) was obtained from A. E. Sippel (Theisen et al., 1986). An *Eco*RI-*Eco*RI 187 bp fragment containing the polyoma virus enhancer [cut out from pMC1Neo (Thomas & Capecchi, 1987)], ligated with synthetic *Eco*RI-*Pst*I adapters, was

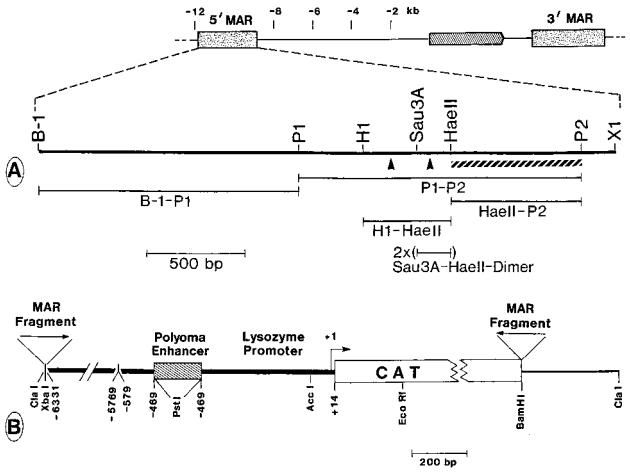

FIGURE 1: MAR fragments and vector map. (A) The upper part shows a map of the chicken lysozyme gene domain with the coding sequence (hatched arrow) and the flanking MARs (stippled blocks). Below, the 5′ MAR and the relative positions of the fragments analyzed are drawn in more detail. The dimer fragment contains a direct repeat of the sequence *Sau*3A-*Hae*II. The arrowheads indicate the location of two binding sites for the nuclear matrix protein ARBP. Fragment *Hae*II-P2 contains an intrinsically bent sequence (highlighted by a hatched bar). The restriction enzymes are as follows: B, *Bam*HI; H, *Hin*dIII; P, *Pvu*II; and X, *Xba*I. (B) The vector (a pML2 derivative) contains the chloramphenicol acetyltransferase (CAT) gene plus a SV40 polyadenylation sequence (open box) controlled by the lysozyme promoter (from +14 to −579) (heavy line). The 171 bp polyoma virus enhancer fragment was inserted into the *Pst*I site at bp −469 (hatched box). The vector also contains the chicken lysozyme −6.1 kb enhancer that is, however, not active in mouse 3T3 cells. MAR fragments were sequentially inserted into *Xba*I and *Bam*HI sites. Gel-purified *Cla*I-*Cla*I fragments, including 351 bp of vector sequence (thin line), were introduced into mouse 3T3 cells.

inserted into the *Pst*I site at −469 upstream of the transcription start site. Then, the 2955 bp 5′ MAR fragment B-1-X1 or various subfragments were inserted either by blunt-end ligation or by using synthetic linkers into the *Xba*I site at the 5′ end of the lysozyme promoter sequence and, after subcloning, into the *Bam*HI site at the 3′ end of the CAT gene (see Figure 1B). After identification by restriction analysis, constructs were selected for transfection that contained the MAR fragments at the *Xba*I site in the same orientation relative to the direction of transcription as the lysozyme gene domain but the MAR fragments at the *Bam*HI site in the opposite orientation. The MAR fragments were also used as hybridization probes in genomic Southern analyses. As a control sequence, we used an AT-rich 0.6 kb fragment (E9-B6) located ∼6 kb downstream of the 3′ MAR (Phi-Van & Strätling, 1988). The selectable marker pSV2-neo contained the *neo*^{r} gene controlled by the SV40 early promoter region (Southern & Berg, 1982).

Cell Culture, $CaPO_4$ Transfection, and Microinjection. NIH-3T3 cells, obtained from the American Type Culture Collection, were grown in Dulbecco's modified Eagle's medium (Boehringer, Mannheim) containing 8% fetal calf serum (Life Technologies Inc.), 0.1 g/L streptomycin, 0.1 unit/L penicillin, 0.1 g/L kanamycin, and 18.9 mM $NaCO_3$ at 37 °C and 5% CO_2. Cells were grown to a density of 5 × 10^5 cells per 8.5 cm plate. Gel-purified *Cla*I-*Cla*I fragments were $CaPO_4$ coprecipitated with *Eco*RI-linearized pSV2-neo in a molar ratio of 3:1. Transfection using $CaPO_4$ coprecipitation was performed as described previously (Phi-Van et al., 1990) using totally 2 μg of DNA per plate. Use

of carrier DNA was eliminated in order to keep the number of integrated copies low (Mielke et al., 1990). Cells were grown for 2 days without selection. G418 (GIBCO) was then added to a final concentration of 800 μg/mL. Clones were picked after 2 weeks and cultured at a lower concentration of G418 (400 μg/mL).

Microinjections were carried out via glass micropipettes prepared from glass capillaries on a micropipette puller. The DNA solution (0.006 μg/μL *Cla*I-*Cla*I fragment and 0.0013 μg/μL linearized pSV2-neo, equivalent to a molar ratio of approximately 3:1) was injected under visual control of an inverted phase contrast microscope (Leitz Orthoplan, 400×). The solution was forced into the cells with the aid of a Leitz micromanipulator by gentle air pressure from a syringe connected to the micropipette (Graessmann & Graessmann, 1983).

Isolation of DNA and Genomic Southern Analysis. To isolate genomic DNA from G418-resistant clones, cells were washed twice with 1 × PBS, harvested, and lysed in a solution containing 20 mM Tris-HCl (pH 7.5), 20 mM EDTA, 300 mM NaCl, and 0.4% SDS followed by incubation with 500 μg/mL proteinase K at 37 °C overnight. The lysate was extracted successively with phenol/chloroform/isoamyl alcohol (12:12:1) and with chloroform/isoamyl alcohol (24:1). Resulting probes were precipitated with ethanol, and DNA pellets were washed with 80% ethanol, dried under vacuum, and dissolved overnight in 450 μL of TE [10 mM Tris-HCl (pH 7.5) and 1 mM EDTA]. DNA samples were then digested with 50 μg/mL RNase A for 2.5 h at 37 °C followed by digestion with 100 μg/mL proteinase K for 2 h at 37 °C. Then samples were extracted, ethanol precipitated, and dissolved as before.

Seven-microgram samples were digested with *Pst*I, and fragments were resolved by electrophoresis on 1% agarose gels run in TPE [36 mM Tris, 30 mM NaH_2PO_4 (pH 8.0), and 1 mM EDTA] and transferred to nitrocellulose filters by standard procedures (Sambrook et al., 1989). Filters were hybridized to ^{32}P-random oligonucleotide-labeled probes (Sambrook et al., 1989) [specific activity = approximately $1-2 \times 10^8$ cpm/μg DNA (Cerenkov counting)] in buffer containing 4 × SSC, 5 × Denhardt's solution (1 × Denhardt's = 0.02% (w/v) Ficoll, 0.02% (w/v) poly(vinylpyrrolidone), and 0.02% (w/v) bovine serum albumin], 50 mM sodium phosphate buffer (pH 6.5), and 100 μg/mL salmon sperm carrier DNA at 65 °C for 20−28 h. Hybridized filters were washed at 65 °C with two changes of 2 × SSC, 25 mM sodium phosphate buffer (pH 6.5), and 0.1% SDS for 60 min and then with two changes of 1 × SSC, 25 mM sodium phosphate buffer (pH 6.5), and 0.1% SDS for 60 min. After washing, air-dried filters were exposed to Kodak XAR-5 films at −80 °C. As a reference, all filters contained 7 μg of *Pst*I-cut genomic chicken DNA (ch lanes in Figure 2), which harbors the lysozyme 5′ MAR in a 4.8 kb fragment. To quantitate copy numbers, autoradiograms were scanned with a Bio-Rad model 620 video densitometer and the signals were corrected for DNA loading by a factor obtained by scanning the negatives of the photographs taken of the ethidium bromide-stained agarose gels and dividing the resulting value for chicken DNA by those for each sample DNA lane. Then the autoradiographic signal for the chicken DNA was set as two copies of the 5′ MAR fragment per diploid genome and was used as a reference to calculate copy numbers of the 3T3 clones after a final correction for the 2.17 times larger genome of mouse cells.

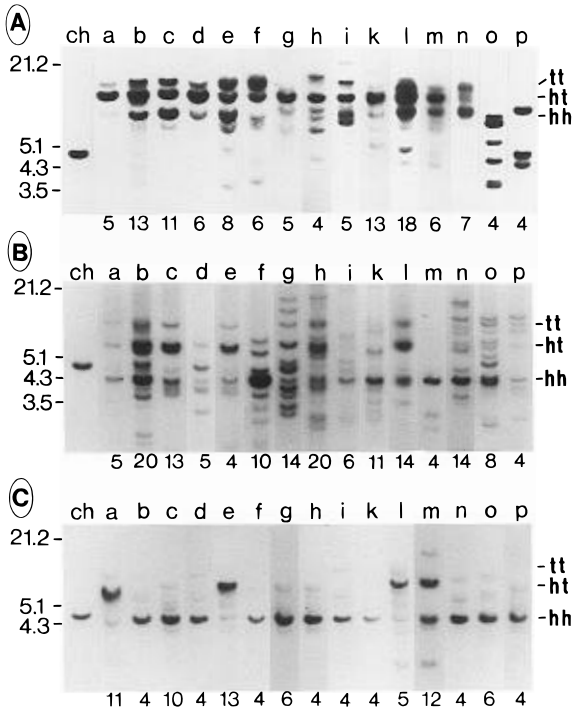

FIGURE 2: Genomic Southern analysis of G418-resistant 3T3 clones. Genomic DNA was isolated from 3T3 cell clones cotransfected with the B-1-X1- (A), B-1-P1- (B), and P1-P2-flanked (C) construct, digested with *Pst*I, resolved on 1% agarose gels, and Southern blotted. The resulting filters were hybridized with a B-1-X1 (A), B-1-P1 (B), and P1-P2 (C) probe, respectively. Lanes marked ch contained *Pst*I-cut genomic chicken DNA. Shown on the left are the migration positions of size standards corresponding to a *Hin*dIII plus *Eco*RI digest of bacteriophage λ DNA. The band positions characteristic for hh, ht, and tt tandem integrations are indicated on the right. Copy numbers are shown at the bottom.

Various Assays. CAT activity was measured by incubating 5−80 μL of cell extract for 1−5 h (with repeated additions of acetylcoenzyme A) at 37 °C as described previously (Phi-Van et al., 1990). The steady state level of reporter gene transcripts was determined by an RNase protection assay (Melton et al., 1984). Poly(A)$^+$ RNA was isolated using an oligo(dT)-cellulose adsorption method (Rahmsdorf et al., 1987). Four micrograms of RNA was hybridized to a ^{32}P-labeled RNA probe covering 364 nucleotides from an *Eco*RI site in the CAT gene to an *Acc*I site in the lysozyme gene promoter (see map in Figure 1B). Prior to the RNase protection assay, the integrity and concentration of the poly(A)$^+$ RNA preparations were controlled by determining the level of β-actin mRNA in a Northern blot analysis by use of a ^{32}P-labeled genomic rat β-actin DNA probe containing the first exon (Figure 4B). Hybridization was performed according to the method described by Church and Gilbert (1984). Binding of labeled fragments to nuclear matrices from chicken HD11 cells was assayed as described previously (Cockerill & Garrard, 1986; Phi-Van & Strätling, 1988).

RESULTS

Experimental Design. The influence of the chicken lysozyme gene 5′ MAR on gene expression has been previously explored by stable transfection of artificial domains and monitoring transgene expression in individual

Table 1: The 5′ MAR and Its Halves Elevate CAT Transgene Expression in 3T3 Cells

flanking MAR fragment	route of gene transfer	number of clones	CAT activity [pmol of CA acetylated (μg of protein)$^{-1}$ h^{-1}] (mean ± SD)	mean copy number (variation)	normalized CAT activity (CAT activity/copy) (mean ± SD)
control fragment	calcium phosphate	40	0.54 ± 0.61	7.3 (1−8)	0.17 ± 0.16
	injection	8	0.29 ± 0.42	5.8 (2−13)	0.12 ± 0.13
B-1-X1	calcium phosphate	66	3.23 ± 1.70	5.2 (1−25)	0.91 ± 0.39
	injection	22	12.57 ± 1.89	7.4 (2−14)	1.53 ± 0.44
B-1-P1	calcium phosphate	60	4.52 ± 1.85	6.9 (1−20)	1.05 ± 0.33
P1-P2	calcium phosphate	79	5.85 ± 3.09	4.9 (2−13)	1.46 ± 0.66
	injection	12	3.46 ± 2.14	4.3 (2−12)	1.23 ± 0.61
H1-*Hae*II	calcium phosphate	55	1.01 ± 1.68	10.7 (6−19)	0.09 ± 0.17
*Hae*II-P2	calcium phosphate	108	0.41 ± 0.57	8.9 (1−25)	0.06 ± 0.10
*Sau*3A-*Hae*II dimer	calcium phosphate	43	0.75 ± 0.85	7.3 (1−14)	0.13 ± 0.13

cell clones (Stief et al., 1989; Phi-Van et al., 1990). The domains used included a reporter gene and flanking 5′ MAR elements on each side. It was found that the 5′ MAR elevated gene expression in a copy number-dependent manner. Compared to enhancer elements, the lysozyme 5′ MAR is fairly long (2.95 kb) and, furthermore, it exhibits several intriguing characteristics. Two high-affinity binding sites of the abundant nuclear matrix protein ARBP, marked by arrowheads in Figure 1A, are localized in subfragment H1-*Hae*II, which efficiently binds to the nuclear matrix (Buhrmester et al., 1995). The 3' abutting sequence, which binds to the nuclear matrix only weakly, contains an intrinsically bent sequence, marked by a hatched bar (von Kries et al., 1990). We therefore attempted to dissect the 5′ MAR and to identify the constraints that govern the position-independent stimulation of transgene expression.

In analogy to our previous experiments, we used artificial domains which included the CAT gene with the chicken lysozyme gene promoter (+14 to −579), the polyoma virus enhancer (inserted at −469), and bilaterally flanking 5′ MAR elements or various subfragments therefrom (Figure 1B). Figure 1A illustrates the subfragments utilized: (i and ii) the halves of the 5′ MAR (fragments B-1-P1 and P1-P2), (iii) fragment *Hae*II-P2 containing the intrinsically bent sequence element, (iv) fragment H1-*Hae*II containing the two high-affinity ARBP binding sites, and (v) the dimerized fragment *Sau*3A-*Hae*II containing the downstream ARBP binding site. The flanking fragments were inserted in their genomic orientations in the upstream *Xba*I site as well as in their reverse orientations in the downstream *Bam*HI site. As a control (non-MAR) fragment, we utilized an AT-rich (60%) 0.6 kb fragment located ∼6 kb downstream of the 3′ MAR; this fragment does not exhibit any affinity to the nuclear matrix from hen oviduct (Phi-Van & Strätling, 1988). *Cla*I-*Cla*I fragments, free of vector sequences except 351 bp, were introduced into mouse 3T3 cells by unlinked cotransfection with plasmid pSV2-neo containing the selectable marker gene for neomycin phosphotransferase using the $CaPO_4$ coprecipitation technique (Southern & Berg, 1982). This avoids several potential influences: (i) that of prokaryotic sequences, (ii) influences of the MAR fragments on the expression of the *neo*^{r} gene and the selection process, (iii) the silencing effect of the *neo*^{r} gene on nearby promoters (Artelt et al., 1991), and (iv) influences by preselection for insertion into "active" chromatin expected in linked transfections (Blasquez et al., 1989). Furthermore, we omitted carrier DNA to obtain low-copy number clones, which are desirable to avoid artifacts by high-copy number integrations (Mielke et al., 1990). We additionally chose microinjection as a means to introduce three constructs in order to verify the results obtained by $CaPO_4$ transfection.

Genomic Southern Analysis Detects Nonrandom Tandem Integrations. To determine the number of integrated copies and their organization, we performed genomic Southern analysis on DNA isolated from 493 G418-resistant, chloramphenicol acetyltransferase-expressing clones. *Pst*I, which asymmetrically cuts the constructs at the site of the polyoma enhancer insertion into two portions of different sizes (see Figure 1B), was used for DNA digestion, and the appropriate MAR fragment was utilized as the hybridization probe along with *Pst*I-cut chicken DNA as a reference for genomic DNA with two integrated MAR copies per genome. Figure 2 illustrates such an analysis for some clones containing the construct with the complete 5′ MAR (panel A) as well as clones containing the constructs with the MAR halves B-1-P1 and P1-P2 (panels B and C). Comparison of the hybridization signals with that of chicken DNA (ch lanes) allowed quantitation of the number of integrated copies. Typically, this number was found to vary between 1 and 25 with a mean of 4.3−10.7 (Table 1). Head-to-head (hh), head-to-tail (ht), and tail-to-tail (tt) tandem integrations would give diagnostic 7.4, 9.0, and 10.6 kb fragments for the B-1-X1 construct, 4.1, 5.7, and 7.3 kb fragments for the B-1-P1 construct, and 4.4, 6.0, and 7.6 kb fragments for the P1-P2 construct. From the data shown, we conclude that most of the copies were integrated in tandem but that the orientation is highly nonrandom; hh, ht, and hh plus ht tandem integrations were significantly preferred in each group of transfected clones (Table 2). By contrast, clones with solely tt tandem integrations were lacking, and mixed hh plus tt and ht plus tt tandems were strongly disfavored. The mixture of all three orientations occurred at an intermediate frequency. Surprisingly, head-to-head tandems were also preferred in those clones, which contained multiple copies of the constructs (see e.g. Figure 2C). Consequently, these clones contained independently integrated head-to-head doublets of the constructs. The number of these doublets could often be simply estimated by counting the number of junction bands (see e.g. clones n and o in Figure 2B). Clones with single integrations appeared rarely, with the exception of the group containing B-1-X1 constructs, where 35% of the clones contained single integrations but often at multiple sites (see e.g. clones o and p in Figure 2A). For most of the clones (excluding those containing single integrations), the diagnostic bands for hh, ht, and tt orientations were the most prominent ones, indicating that intervening integration of the pSV2-neo plasmid occurred rarely, as expected from the 3:1 ratio of experimental transgene/SV2-neo. With three con-

Table 2: Nonrandom Tandem Integration of Transfected Constructs

flanking MAR fragment	route of gene transfer	orientation						
		hh	ht	hh + ht	hh + tt	ht + tt	hh + ht + tt	single copy
control fragment	calcium phosphate	2	8	10	5	−	15	−
	injection	2	5	1	−	−	−	−
B-1-X1	calcium phosphate	5	20	13	1	−	4	23
	injection	2	7	−	−	−	13	−
B-1-P1	calcium phosphate	40	3	3	−	−	14	−
P1-P2	calcium phosphate	50	8	11	−	−	10	−
	injection	4	8	−	−	−	−	−
H1-*Hae*II	calcium phosphate	17	14	10	−	−	14	−
*Hae*II-P2	calcium phosphate	53	5	20	6	7	15	2

structs, we also obtained clones by microinjection. The distribution of orientations among these generally reflected that among the clones obtained by $CaPO_4$ transfection; tt alone, hh plus ht, hh plus tt, and ht plus tt integrations were disfavored. The basis for the nonrandom distribution of orientations, in particular the disfavored appearance of tt integrations, is not clear. It may be speculated that the unequal ends of the constructs, which contained on the 5′ side a copy of the repeated MAR sequence but on the 3′ side a unique vector sequence, are responsible for the effect (see Figure 1B).

Only the Halves of the 5′ MAR Duplicated the Enhancing Effect on Gene Expression. To determine the effects that the selected MAR subfragments exert on gene expression, CAT activities were divided on a per clone basis by the number of integrated copies to obtain normalized CAT activities. The means of normalized CAT activities for each MAR fragment are listed in Table 1. The complete MAR (B-1-X1) resulted in a 5.4-fold enhancement of the mean level of CAT expression. Statistical analysis by a parametric analysis (Student's *t* test) and a nonparametric test (Mann−Whitney U test) revealed that this increase is significant ($p < 0.05$). These results confirm our previous findings obtained with different constructs, different cell types, and a lower number of clones (Stief et al., 1989; Phi-Van et al., 1990). When the construct was microinjected, the mean level of normalized CAT expression was elevated 12.7-fold relative to those of clones injected with the control construct. Clones containing constructs in which the MAR was replaced by the upstream (B-1-P1) and downstream (P1-P2) half exhibited a 6.2- and 8.6-fold increase, respectively, in the mean level of normalized CAT expression. Thus, the halves of the 5′ MAR duplicated the enhancing effect of the complete MAR on gene expression.

The 452 bp subfragment H1-*Hae*II retains the ability to bind to the nuclear matrix (Phi-Van & Strätling, 1988), and furthermore, it contains two high-affinity binding sites for the abundant nuclear matrix protein ARBP (Buhrmester et al., 1995). Surprisingly, this fragment on the flanks of integrated constructs did not elevate the level of normalized CAT expression (Table 1). Furthermore, we generated a fragment, *Sau*3A-*Hae*II dimer, that contains two copies of a downstream 177 bp sequence in fragment H1-*Hae*II and thus includes two copies of the downstream high-affinity ARBP binding site (see Figure 1A). Fragment *Sau*3A-*Hae*II dimer bound to the nuclear matrix with an activity comparable to that of the larger fragments B-1-X1 and P1-P2 (Figure 3). However, on the flanks of stable constructs, this fragment had no stimulating activity on normalized CAT expression. In fact, the mean normalized CAT activities for the H1-*Hae*II and the *Sau*3A-*Hae*II dimer constructs were 53 and 76%, respectively, relative to the control construct, yet these decreases were not statistically significant. We conclude that the retained ability of shorter MAR fragments to bind to the nuclear matrix is not sufficient to enhance transgene expression of stably transfected constructs.

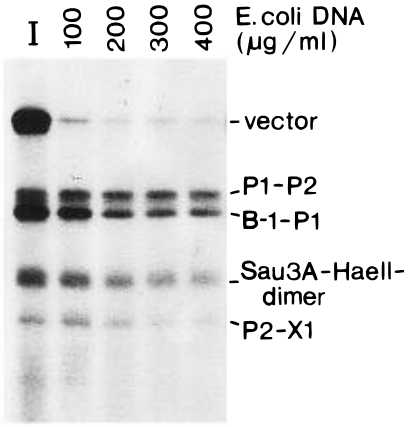

FIGURE 3: *Sau*3A-*Hae*II dimer fragment binds to the nuclear matrix. A plasmid containing the complete 5′ MAR sequence B-1-X1 was cleaved with *Bam*HI and *Pvu*II to generate the MAR fragments B-1-P1, P1-P2, and P2-X1 and a vector fragment. Second, a plasmid containing the *Sau*3A-*Hae*II dimer was cleaved with *Pvu*II to generate an 858 bp fragment containing the dimer and a vector fragment. Labeled fragments were incubated with nuclear matrices from HD11 cells in the presence of increasing concentrations of *E. coli* competitor DNA as indicated. The bound fragments were separated on agarose gels and autoradiographed. Lane I (input) contains 5% of the labeled fragments.

The downstream flank (sequence *Hae*II-P2) of the 5′ MAR exhibits weak binding to the nuclear matrix but contains an intrinsically bent sequence (von Kries et al., 1990). We included this sequence in our analysis of the effects of 5′ MAR subsequences on gene expression and found that fragment *Hae*II-P2 on the flanks of integrated constructs did not affect an increase in normalized CAT activity (Table 1). Thus, sequence bending per se, like nuclear matrix binding activity, is insufficient to enhance transgene expression in stably transfected cells.

The relative level of correctly initiated reporter gene transcripts from four randomly selected clones of each group of clones was determined in an RNase protection analysis. After normalization of the CAT transcript signals in Figure 4 per integrated copy, we found that the complete MAR elevated transcript levels, on the average, 16.9-fold, the upstream half-fragment B-1-P1 9.3-fold, and the downstream half-fragment P1-P2 23.0-fold. In contrast, subfragment H1-*Hae*II did not increase transcript levels significantly (1.3-fold). Similar results were obtained with the abutting

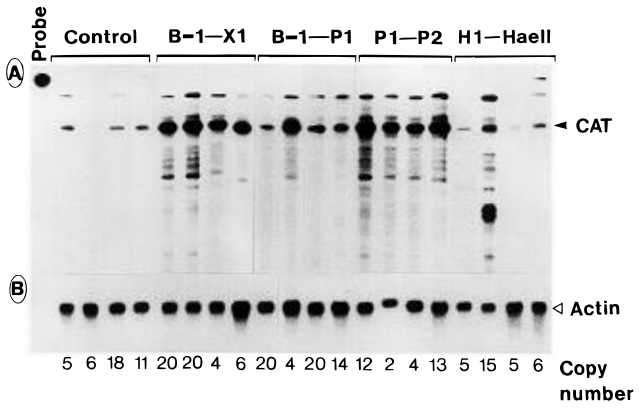

FIGURE 4: Quantitation of CAT transcripts. In an RNase protection analysis (A), 4 μg of poly(A)$^+$ RNA from four randomly selected clones of the indicated groups was hybridized to a ^{32}P-labeled RNA probe. The filled arrowhead indicates the position of correctly initiated CAT transcripts. The integrity and loading of the poly-(A)$^+$ RNA preparations was controlled by determining the level of β-actin mRNA in a Northern blot analysis (B). Copy numbers for each clone are indicated below each lane.

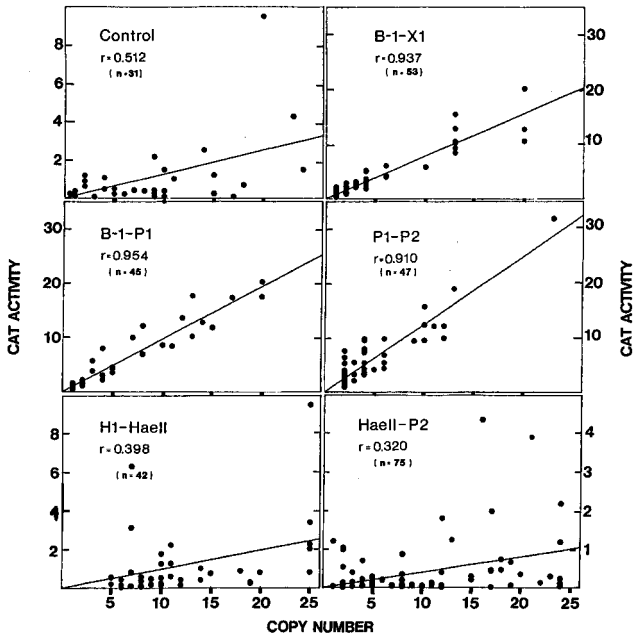

FIGURE 5: 5′ MAR and its halves generate copy number-proportional transgene expression. CAT activities for each group of clones were plotted versus copy numbers. The individual groups of clones are identified by the flanking MAR sequence in the construct used for transfection. The correlation coefficients are shown. CAT activity units are the same as in Table 1.

subfragment *Hae*II-P2 (data not shown). These results reinforce our previous conclusion that the halves of the 5′ MAR duplicate the expression enhancing effect of the complete MAR but that the ability of subfragments to bind to the nuclear matrix is not sufficient to enhance transgene transcription. Figure 4 also shows that the relative transcript levels on a per copy basis varied among the clones within a group. A similar variation was observed for the level of CAT activity (see below and Figure 5). However, the number of clones (four) was too low to determine whether MARs have an effect on the extent of variation at the transcript level.

The Ability of 5′ MAR Subfragments To Enhance Transgene Expression Correlated with Their Insulating Activity. Expression of stably transfected genes is, in general, highly variable and dependent on the chromosomal site of integration (position effect) (Palmiter & Brinster, 1986). We have previously shown that clones containing transgenes, which are flanked by 5′ MARs, show reduced variation in expression of the transgene among the clones (Stief et al., 1989; Phi-Van et al., 1990). In Figure 5 (upper panels), plots of non-normalized CAT activities versus copy numbers for the clones containing the control construct and the B-1-X1-flanked construct confirm that flanking 5′ MARs affect a tight correlation of the level of expression with copy numbers ($r = 0.937$), while expression among the clones with the control construct is highly variable ($r = 0.512$) (upper panels). In the two groups of clones, which contain transgenes flanked by the halves of the 5′ MAR (B-1-P1 and P1-P2), CAT activities also tightly correlated with copy numbers ($r = 0.954$ and $r = 0.910$, respectively) (middle panels). This shows also that the 5′ MAR halves were highly efficient in dampening position effects. By contrast, no correlation was observed for the clones containing transgenes flanked by subsequences H1-*Hae*II and *Hae*II-P2 ($r = 0.398$ and $r = 0.320$, respectively) (lower panels). Furthermore, no correlation was found for the transgene flanked by the duplicated subsequence *Sau*3A-*Hae*II (data not shown). Thus, the activity of the 5′ MAR halves to enhance selectively transgene expression paralleled their specific ability to insulate gene expression from effects of the chromatin structure at the site of integration. In addition, the ability to bind to the nuclear matrix is not sufficient to insulate transgene activity.

DISCUSSION

We previously found that the chicken lysozyme gene 5′ MAR that includes 2.95 kb enhances transgene expression in cultured cells in a copy number-dependent manner (Stief et al., 1989; Phi-Van et al., 1990). Here we show that the halves of the 5′ MAR containing 1.32 and 1.45 kb, respectively, quantitatively duplicated these effects, while shorter subfragments proved to be inactive. Since the halves of the 5′ MAR share no sequence homology, the observed effects on gene expression were not determined by a specific sequence. This feature is reminiscent of a general property of MARs. Although MARs are AT-rich, they do not share sufficient sequence similarity to allow cross-hybridization (Gasser & Laemmli, 1987; Phi-Van & Strätling, 1990). Nevertheless, MARs have significant biological effects in common, e.g. stimulating and insulating effects on transgene expression (Blasquez et al., 1989; Xu et al., 1989; Klehr et al., 1991; Breyne et al., 1992; Dietz et al., 1994; Forrester et al., 1994). Surprisingly, the inactive subfragments H1-*Hae*II and *Sau*3A-*Hae*II dimer exhibit strong nuclear matrix binding activities in an in vitro binding assay (Cockerill & Garrard, 1986). Thus, the activity to enhance transgene expression in a copy number-dependent manner requires more sequence information than the ability to bind to the nuclear matrix. Subfragments H1-*Hae*II and *Sau*3A-*Hae*II dimer each contain two high-affinity binding sites for the nuclear matrix protein ARBP. Most likely, the capacity of the MARs to bind to the nuclear matrix requires the presence of only one or a few DNA binding sites for nuclear matrix constituents. The group of proteins that mediate MAR binding includes (among others) SATB1, p120 (also named SAF-A, SP120, and hnRNP U), and matrin F/G (Dickinson et al., 1992; von Kries et al., 1994; Hakes & Berezney, 1991).

The determinants that govern a stimulation of transgene expression in a copy number-dependent manner are more complex than those that affect binding of MARs to the nuclear matrix or purified nuclear matrix proteins. Most likely, integrated MARs, when they are active as transgene stimulators, are packaged into nucleosomes and higher-order structures (Saitoh & Laemmli, 1994). This possibly causes the requirement for longer sequences (three and more nucleosomal lengths). As a second explanation, it may be postulated that several different features of MAR DNA are *collectively* required for the transgene-stimulating effect. If these features reside in different subfragments, it would be conceivable that longer DNA fragments are required. Besides binding sites for nuclear matrix proteins, DNA bends and the property to unwind DNA may be critical (Bode et al., 1992). Furthermore, several different MAR binding proteins may act together in a larger complex, some components having preferentially architectural functions by the stabilization of protein−protein interactions, while others possibly contact enhancer binding proteins or other transcription constituents.

It should be noted that shorter fragment lengths are required for other types of insulating elements. For example, the gypsy retrotransposon requires 340 bp of DNA, and a fragment of the scs element flanking the 87A7 hsp locus requires 600 bp of DNA for enhancer-blocking activity (Cai & Levine, 1995; Vazquez & Schedl, 1994). Kalos and Fournier (1995) reported that tandem arrangements of constructs bilaterally flanked by the apolipoprotein B MARs failed to express the transgene. We did not observe a comparable phenomenon among our clones. Instead, we could show that flanking MARs in multicopy clones affected a copy number dependence of transgene expression (see Figure 5). Yet we note that only CAT-positive clones were included in the present analysis, allowing the possibility that in some (missed) clones a tandem arrangement caused silencing of expression. We have been unable to separate the expression-stimulating activity of the chicken lysozyme MAR from its ability to confer copy number-independent expression. This might not be a universal feature of MAR/SARs, since e.g. *Drosophila* SARs have been reported to only stimulate gene expression (Poljak et al., 1994). Bonifer and co-workers observed that, in transgenic mice, deletion of the MARs from the otherwise intact chicken lysozyme gene domain increased the incidence of ectopic expression (Sippel et al., 1993; Bonifer et al., 1994). Since integration probably occurred into active chromatin, this increase is likely caused by the lack of the insulating activity of the MARs. An expression-stimulating activity of the lysozyme MARs was not observed in the transgenic mice studies. Perhaps such an activity was minor compared to that of the orchestra of regulatory sites upstream of the lysozyme gene (Bonifer et al., 1991). Since the constructs we used to stably transfect cultured cells lacked this orchestra, we specifically focused on the effects of the 5′ MAR in a heterologous context.

Figure 2 and Table 2 document tandem integrations of DNA repeats in a nonrandom distribution of orientations. hh repeats are favored, while tt repeats are strongly underrepresented. The nonrandom distribution could either result from a preference of distinct preintegration ligations of two molecules or indicate some dynamic changes within integrated repeats. The close neighborhood of MARs in arrays containing tandem integrations of constructs with 5′ MARs on each side may be a likely source of inversions. We also point out that, since each repeat in an array of tandem integrations is bilaterally flanked by MARs and thus represents an intact minidomain, different orientations should not affect transcription rates.

Beyond the basic biological question of how complex genomes are organized to allow efficient use of the encoded information, an understanding of the features that determine chromatin organization may be of high potential practical significance. In the emerging fields of human gene therapy and improvement of agricultural crops, the inclusion of elements (like MARs) that regulate chromatin structure in transgene constructs may be helpful.

ACKNOWLEDGMENT

We express our special gratitude to K. Müller, K. Weber, and B. Bettin for excellent and patient technical assistance. We are furthermore indebted to G. Buschhausen-Denker for his help in plasmid injection experiments and U. Nudel for a generous gift of plasmid DNA.

REFERENCES

Artelt, P., Grannemann, R., Stocking, C., Friel, J., Bartsch, J., & Hauser, H. (1991) *Gene 99*, 249−254.
Benyajati, C., & Worcel, A. (1976) *Cell 9*, 393−407.
Blasquez, V. C., Xu, M., Moses, S. C., & Garrard, W. T. (1989) *J. Biol. Chem. 264*, 21183−21189.
Bode, J., Kohwi, Y., Joh, T., Klehr, D., Mielke, C., & Kohwi-Shigematsu, T. (1992) *Science 255*, 195−197.
Bonifer, C., Hecht, A., Saueressig, H., Winter, D. M., & Sippel, A. E. (1991) *J. Cell. Biochem. 47*, 99−108.
Bonifer, C., Yannoutsos, N., Krüger, G., Grosveld, F., & Sippel, A. E. (1994) *Nucleic Acids Res. 22*, 4202−4210.
Breyne, P., Van Montagu, M., Depicker, A., & Gheysen, G. (1992) *Plant Cell 4*, 463−471.
Buhrmester, H., von Kries, J. P., & Strätling, W. H. (1995) *Biochemistry 34*, 4108−4117.
Cai, H., & Levine, M. (1995) *Nature 376*, 533−536.
Church, G. M., & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1991−1995.
Cockerill, P. N., & Garrard, W. T. (1986) *Cell 44*, 273−282.
Dickinson, L. A., Joh, T., Kohwi, Y., & Kohwi-Shigematsu, T. (1992) *Cell 70*, 631−645.
Dietz, A., Kay, V., Schlake, T., Landsmann, J., & Bode, J. (1994) *Nucleic Acids Res. 2*, 2744−2751.
Forrester, W. C., van Gendersen, C., Jenuwein, T., & Grosschedl, R. (1994) *Science 265*, 1221−1225.
Gasser, S. M., & Laemmli, U. K. (1987) *Trends Genet. 3*, 16−22.
Graessmann, M., & Graessmann, A. (1983) *Methods Enzymol. 101*, 482−492.
Grosveld, F., Blom van Assendelft, G., Greaves, D. R., & Kollias, G. (1987) *Cell 51*, 975−985.
Hakes, D. J., & Berezney, R. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 6186−6190.
Kalos, M., & Fournier, R. E. K. (1995) *Mol. Cell. Biol. 15*, 198−207.
Kellum, R., & Schedl, P. (1991) *Cell 64*, 941−950.
Kellum, R., & Schedl, P. (1992) *Mol. Cell. Biol. 12*, 2424−2431.
Klehr, D., Maass, K., & Bode, J. (1991) *Biochemistry 30*, 1264−1270.
McKnight, R. A., Shamay, A., Sankaran, L., Wall, R. J., & Henninghausen, L. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6943−6947.
Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K., & Green, M. R. (1984) *Nucleic Acids Res. 12*, 7035−7056.
Mielke, C., Kohwi, Y., Kohwi-Shigematsu, T., & Bode, J. (1990) *Biochemistry 29*, 7475−7485.
Mlynárová, Ľ., Loonen, A., Heldens, J., Jansen, R. C., Keizer, P., Stiekema, W. J., & Nap, J.-P. (1994) *Plant Cell 6*, 417−426.

Palmiter, K. D., & Brinster, R. L. (1986) *Annu. Rev. Genet. 20*, 465−499.
Paulson, J. R., & Laemmli, U. K. (1977) *Cell 12*, 817−828.
Phi-Van, L., & Strätling, W. H. (1988) *EMBO J. 7*, 655−664.
Phi-Van, L., & Strätling, W. H. (1990) *Prog. Mol. Subcell. Biol. 11*, 1−11.
Phi-Van, L., von Kries, J. P., Ostertag, W., & Strätling, W. H. (1990) *Mol. Cell. Biol. 10*, 2302−2307.
Poljak, L., Seum, C., Mattioni, T., & Laemmli, U. K. (1994) *Nucleic Acids Res. 22*, 4386−4394.
Rahmsdorf, H. J., Schönthal, A., Angel, P., Litfin, M., Rüther, U., & Herrlich, P. (1987) *Nucleic Acids Res. 15*, 1643−1659.
Saitoh, Y., & Laemmli, U. K. (1994) *Cell 76*, 609−622.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular cloning: a laboratory manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Plainview, NY.
Sippel, A. E., Schäfer, G., Faust, N., Saueressig, H., Hecht, A., & Bonifer, C. (1993) *Cold Spring Harbor Symp. Quant. Biol. 58*, 37−44.
Southern, P. J., & Berg, P. (1982) *J. Mol. Appl. Genet. 1*, 327−341.
Stief, A., Winter, D. M., Strätling, W. H., & Sippel, A. E. (1989) *Nature 341*, 343−345.
Theisen, M., Stief, A., & Sippel, A. E. (1986) *EMBO J. 5*, 719−724.
Thomas, K. R., & Capecchi, M. R. (1987) *Cell 51*, 503−512.
Thompson, E. M., Christians, E., Stinnakre, M.-G., & Renard, J.-P. (1994) *Mol. Cell. Biol. 14*, 4694−4703.
Vazquez, J., & Schedl, P. (1994) *EMBO J. 13*, 5984−5993.
von Kries, J. P., Phi-Van, L., Diekmann, S., & Strätling, W. H. (1990) *Nucleic Acids Res. 18*, 3881−3885.
von Kries, J. P., Buhrmester, H., & Strätling, W. H. (1991) *Cell 64*, 123−125.
von Kries, J. P., Buck, F., & Strätling, W. H. (1994) *Nucleic Acids Res. 22*, 1215−1220.
Weintraub, H., & Groudine, M. (1976) *Science 193*, 848−856.
Xu, M., Hammer, R. E., Blasquez, V. C., Jones, S. L., & Garrard, W. T. (1989) *J. Biol. Chem. 264*, 21190−21195.

BI9603783