2924–2930 *Nucleic Acids Research, 1999, Vol. 27, No. 14*

A bipartite sequence element associated with matrix/scaffold attachment regions

Cornelis M. van Drunen, Richard G. A. B. Sewalt, Rob W. Oosterling[1], Peter J. Weisbeek[1], Sjef C. M. Smeekens[1] and Roel van Driel*

E. C. Slater Instituut, BioCentrum Amsterdam, University of Amsterdam, Plantage Muidergracht 12, 1018 TV Amsterdam, The Netherlands and [1]Department of Molecular Cell Biology, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

Received March 8, 1999; Revised and Accepted May 21, 1999

ABSTRACT

We have identified a MAR/SAR recognition signature (MRS) which is common to a large group of matrix and scaffold attachment regions. The MRS is composed of two degenerate sequences (AATAAYAA and AWWRTAANNWWGNNNC) within close proximity. Analysis of >300 kb of genomic sequence from a variety of eukaryotic organisms shows that the MRS faithfully predicts 80% of MARs and SARs. In each case where we find a MRS, the corresponding DNA region binds specifically to the nuclear scaffold. Although all MRSs are associated with a SAR, not all known SARs and MARs contain a MRS, suggesting that at least two classes exist, one containing a MRS, the other not. Evidence is presented that the two sequence elements of the bipartite MRS occupy a position on the nucleosome near the dyad axis, together creating a putative protein binding site. The identification of a MAR- and SAR-associated DNA element is an important step forward towards understanding the molecular mechanisms of these elements. It will allow: (i) analysis of the genomic location of SARs, e.g. in relationship to genes, based on sequence information alone, rather than on the basis of an elaborate biochemical assay; (ii) identification and analysis of proteins that specifically bind to the MRS.

INTRODUCTION

Matrix/scaffold attachment regions (MARs/SARs) are genomic elements thought to delineate the structural and functional organisation of the eukaryotic genome. Originally, MARs and SARs were identified through their ability to bind to the nuclear matrix or scaffold (1). Binding cannot be assigned to a unique sequence element, but is dispersed over a region of several hundred base pairs. These elements are found flanking a gene or a small cluster of genes and are located often in the vicinity of *cis*-regulatory sequences (2). This has led to the suggestion that they contribute to higher order regulation of transcription by defining boundaries of independently controlled chromatin domains. There is indirect evidence to support this notion. In transgenic experiments MARs/SARs dampen position effects by shielding the transgene from the effects of the chromatin structure at the site of integration (3). Furthermore, they may act as boundary elements for enhancers, restricting their long range effect to only the promoters that are located in the same chromatin domain (4). Although a number of interacting proteins have been identified [lamins (5), ARBP (6), hnRNP-U/SafA (7), SafB (8), SatB1 (9,10) and Bright (11)], the molecular mechanism by which these proteins affect transcription regulation is still unclear.

From a structural point of view, SARs are thought to be involved in chromatin condensation and chromosome formation. A synthetic AT-hook protein, which specifically binds to MARs/SARs, interferes with proper chromatin condensation in *Xenopus laevis* egg extracts (12). Moreover, alignment of SARs around the central core of mitotic and meiotic chromosomes has been proposed to be required for correct condensation of DNA within these chromosomes (13). Preferential binding of histone H1 to these elements might be part of the same molecular mechanism, as recruitment of H1 to the nucleosome core particles results in a more compact form of chromatin (14,15). HMG1/Y can compete with histone H1 for binding (15), whereas HMG1 can compete with H1 for binding to four-way junctions (16). Competition between H1 and these HMG proteins may contribute to determining the global distribution of active and inactive chromatin. Also, histone acetylation has been linked to transcriptional regulation via MARs. Hyperacetylation is a hallmark of active regions in the genome, while hypo-acetylation is typical for regions that are transcriptionally inactive (17,18). Interestingly, the transcription potentiating effect of the histone deacetylase inhibitor sodium butyrate on transgenes is dependent on the presence of a MAR (19). The recent finding of topoisomerase II as part of the chromatin remodelling CHRAC complex is in line with the idea that these elements may influence transcription regulation via nucleosome remodelling (20). Binding sites for topoisomerase II could specifically target this complex to chromatin to promote local remodelling. Despite the fact that MARs/SARs are evolutionarily highly conserved (21), no MAR/SAR-associated sequence elements have been identified so far. Although DNA sequence repeats have been identified that are clustered in

*To whom correspondence should be addressed. Tel: +31 20 525 5150; Fax: +31 20 525 5124; Email: van.driel@chem.uva.nl

SARs and MARs (for a review see 22), none of these are truly specific. Additional clusters and individual elements can also be found elsewhere in the genome (23).

Here we present a bipartite sequence element that is unique for a large group of eukaryotic MARs/SARs. This MAR/SAR recognition signature (MRS) comprises two individual sequence elements that are <200 bp apart and may be aligned on positioned nucleosomes in MARs. The MRS can be used to correctly predict the position of MARs/SARs in plants and animals, based on genomic DNA sequence information only, thereby avoiding elaborate biochemical binding assays. Our results from the analysis of >300 kb of sequence data from several eukaryotic organisms show that wherever a MRS is observed in the DNA sequence, the corresponding genomic fragment is a biochemically identifiable SAR. The identification of the MRS is an important step forward, since it allows the localisation and distribution of these elements in genomes. Furthermore, it opens new avenues towards the unravelling of the mechanism by which MARs/SARs regulate gene activity through identifying proteins that associate with the MRS sequence.

MATERIALS AND METHODS

Plasmids and sequences

The appropriate *Bgl*II restriction fragments of the *Caenorhabditis elegans* cosmid M88 (24) were cloned in pBluescript (*Bam*HI). This resulted in the pCMx series [pCM04 (8374 bp), pCM05 (1759 bp), pCM06 (2208 bp), pCM07 (2716 bp), pCM08 (3729 bp) and pCM11 (8414 bp)]. The DNA sequences analysed for the MRS were from *Arabidopsis* [plastocyanin (z83321), *ATB*2 (z82043) and *ATH*1 (z83320)], *C.elegans* [cosmid M88 (z34802)], *Drosophila* [histone cluster (DMH1H3), *HSP*70 (DROHSP72A2), actin (DMACT5CA), *FTZ* (DMFTZUSE), *SGS*-4 (DROSGS4.01) and *ADH* (DMADHGC)], chicken [α-globin (m58749) and lysozyme (Dr C. Bonifer)], Chinese hamster [*DHFR* intron x06654], man [myc (AC004081), Ig(κ) (HSIGKA), interferon (Dr J. Bode) and β-globin locus (HUMHBB)], mouse [IgH (MMIG25), Ig(κ) (MMIG25), β-globin cluster (x14061) and *HPRT* (AF047825)], rabbit [β-globin cluster (RABBGLOB), Ig(κ1) (OCIG04), Ig(κ2) (OCIG05)], SV40 (SV40XX) and yeast [CENIII (SCCHRIII)].

Mapping of the MRS

Individual 16 (AWWRTAANNWWGNNNC) and 8 bp (AATAAYAA) sequence elements were mapped through a compiled list that included single mutations at all possible positions within the 16 bp consensus, using a standard DNA restriction analysis computer program. The MRS was defined as the region where 16 and 8 bp sequences are <200 bp apart.

Isolation of nuclei

Nuclei from rat liver cells were isolated as described by Izuarralde *et al.* (25) and were kept at –80°C in storage buffer (7.5 mM Tris–HCl, pH 7.4, 40 mM KCl, 1 mM EDTA, 0.25 mM spermidine, 0.1 mM spermine, 1% v/v thioglycol, 0.2 M sucrose, 50% v/v glycerol) at a density of 10^7 nuclei/ml.

Nuclear scaffold preparation

Procedures were essentially as described before (5). To obtain scaffolds for the binding assay rat liver nuclei were subjected to a lithium 3,5-diiodosalicylate (LIS) extraction protocol described by Mirkovitch *et al.* (1). Nuclei of 10^7 cells were washed once in 10 ml washing buffer (3.75 mM Tris–HCl, pH 7.4, 20 mM KCl, 0.5 mM EDTA, 0.125 mM spermidine, 0.05 mM spermine, 1% v/v thioglycol, 0.1% w/v digitonin and 20 μg/ml aprotinin). After pelleting (300 *g* for 10 min at 4°C) nuclei were gently resuspended in 0.5 ml washing buffer and stabilised by incubation for 20 min at 42°C. Non-scaffold proteins were extracted by adding 10 ml of 10 mM LIS in extraction buffer (20 mM HEPES–KOH, pH 7.4, 100 mM lithium acetate, 1 mM EDTA, 0.1 mM PMSF, 0.1% w/v digitonin and 20 μg/ml aprotinin) followed by incubation for 15 min at 25°C. The resulting nuclear halos were collected by centrifugation (15 000 *g* for 5 min at 4°C) and washed four times with 10 ml of digestion buffer (20 mM Tris–HCl, pH 7.4, 70 mM NaCl, 20 mM KCl, 10 mM $MgCl_2$, 0.125 mM spermidine, 0.05 mM spermine and 10 μg/ml aprotinin). For the *in vitro* SAR binding assay rat nuclear scaffolds were obtained by restriction of the genomic DNA of the halos in 1 ml digestion buffer containing 1000 U each of *Eco*RI, *Hin*dIII and *Xho*I for 2 h at 37°C.

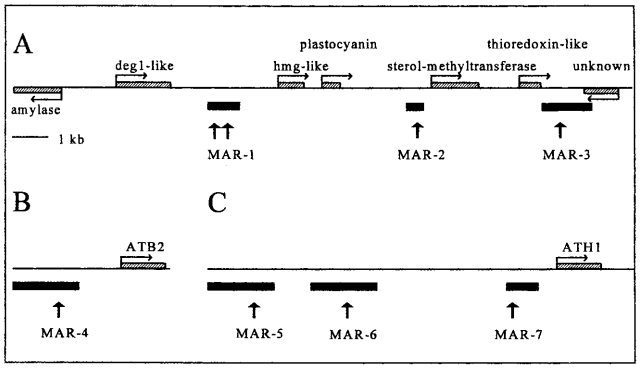
SAR binding experiments

The rat liver scaffold preparation was adjusted to a final concentration of 15 mM EDTA and 120 μg/ml *Escherichia coli* competitor DNA. To identify SARs in *C.elegans* cosmid M88, nuclear scaffolds from 10^6 cells were incubated overnight at 37°C with 15 ng of the appropriate [α-^{32}P]dATP end-labelled restriction fragments where *Xho*II was used to release the whole *Bgl*II insert from the pCMx plasmid series. After separation into pellet and supernatant fractions by centrifugation (15 000 *g* for 30 min at 4°C) DNA was purified by incubation at 37°C for 60 min with 0.1% SDS and 50 μg/ml proteinase K, followed by phenol/chloroform extraction. DNA was precipitated, dissolved in 50 μl TE and subsequently half of the pellet, supernatant or input fractions were loaded on a 1.2% agarose gel. After electrophoresis the gel was dried on Whatman 3MM paper, followed by overnight autoradiography on Kodak X-Omat S film. The quality of our matrix preparations was checked with the H1–H3 SAR of *Drosophila melanogaster*.

RESULTS

Arabidopsis thaliana SARs contain a conserved sequence element

The first indication that an evolutionarily conserved MRS exists came from our previous work on the genomic organisation in *A.thaliana* (23). In three independent genomic regions of >30 kb we have mapped SARs in relation to potential transcription units. A detailed sequence analysis of the seven SARs that we identified earlier revealed that they share a 21 bp degenerate sequence element (TAWAWWWNNAWWRTAANNWWG). We showed that the 21 bp sequence is made up of two individual sequences (TAWAWWW and AWWRTAANNWWG), which are found in a number of configurations that differ in the distance between the two elements (23).



D			Distance
MAR-1	ATAGTAAAAATGACAA (C)	AATAACAA (C)	+ 5
	AATGTAACCAAAGAGC (C)	AATAACAA (C)	+153
	ATTATAAGGAAGAAAC (W)		+143
	ATTAAAAATATGTCAC (C)		+125
MAR-2	ATTATAACCAAGTGCT (W)	AATAACAA (C)	+ 17
MAR-3	AAAGTAATGTAGTGGC (W)	AATAACAA (C)	- 64
MAR-4	AAAGTAAAAACGTTTC (W)	AATAATAA (W)	+ 46
		AATAATAA (W)	+165
MAR-5	GTAGTAATATTGATCC (W)	AATAATAA (C)	-108
MAR-6	AAAATAATAAAATAAC (C)	AATAACAA (C)	Overlap
		AATAATAA (W)	+ 37
MAR-7	AAAATAATTTTGTCCT (W)	AATAATAA (W)	- 32
	AWWRTAAnnWWGnnnC	AATAAYAA	

Figure 1. Scaffold attachment regions from three genomic regions of *A.thaliana* contain a conserved sequence element. The distribution of potential open reading frames (hatched bars), SARs (solid bars) and the position of the MRSs (vertical arrows) in the genomic regions around (**A**) the plastocyanin gene, (**B**) the *ATB2* gene and (**C**) the *ATH1* gene of *Arabidopsis*. (**D**) The sequence make-up of the MRS with its alignment [R = A or G, Y = C or T, W = A or T and N = G, A, T or C], relative orientation [(C), Crick strand; (W), Watson strand] and number of bases between the two parts of the MRS. A negative distance indicates that the 8 bp sequence precedes the 16 bp sequence.

The novel idea of a combination of two sequences that might constitute a MAR/SAR-associated sequence motif prompted us to look for similar and other combinations of two sequence elements. This approach revealed an additional conserved element within the *Arabidopsis* SARs (Fig. 1A–C) that resembles, but is distinct from, the 21 bp sequence. This new motif is an extension of the latter half of the 21 bp sequence (AWWRTAANNWWGNNNC) in combination with a second sequence (AATAAYAA). Like the two parts of the *Arabidopsis*-specific 21 bp MRS, these new sequences are also found within a relatively short distance (~200 bp) of each other, although the maximal distance between these elements is larger than that for the 21 bp MRS (Fig. 1D). Each of the two sequence elements alone can be found in numerous other positions throughout the genome. The combination of the two sequences, however, is found only in the SARs and not elsewhere in the genomic regions of *A.thaliana* that were investigated.

The MRS correctly predicts the location of SARs in a *C.elegans* contig

To test whether the observed correlation between the MRS and the position of a SAR holds true in organisms other than *Arabidopsis*, we screened a 33 kb fragment of genomic DNA from *C.elegans*. The rationale is to locate the MRSs in this sequenced contig (Table 1) and test corresponding restriction fragments for binding to the nuclear scaffold. Our operational definition of the MRS is the presence of the AWWRTAANNWWGNNNC sequence (where we allow one mismatch within the 16 bp consensus) within a distance of 200 bp of the AATAAYAA sequence. Figure 2 depicts the organisation of ORFs and the location of all MRSs in cosmid M88 (24). On the basis of the distribution of the MRSs we predict six SARs. As a control for the quality of our scaffold preparations we used the H1–H3 intergenic histone cluster SAR of *D.melanogaster* (1,26) located on a 1017 bp *Pst*I–*Eco*RI fragment (Fig. 2, panel control).

Table 1. Potential MARs in cosmid M88 of *C.elegans*

Potential MAR	Position MRS	AWWRTAAnnWWGnnnC	AATAAYAA	Distance
MRS-A	3880	AAAATAACAATGAAAA (W)	AATAACAA (W)	overlap
	4097	ATAGTAAAAAAAGTTC (W)	AATAACAA (W)	+170
MRS-C	11482	ATAATAAATTTGATAC (W)	AATAATAA (W)	overlap
	11650	AAAATAAATTTTGATC (C)	AATAATAA (C)	- 20
MRS-D	22543	ATAATAATAAAATTTC (W)	AATAATAA (W)	overlap
MRS-E1	26346	AATGTAAATTTGTTGC (C)	AATAATAA (W)	+184
MRS-E2	27275	AAAATAATAATGCTTT (C)	AATAATAA (C)	overlap
			AATAATAA (C)	+176
MRS-F	29274	ATAATACTATTGATTC (W)	AATAACAA (W)	-128

Positions of the MRSs with their sequence composition, relative orientation (C, Crick strand; W, Watson strand) and distance between the two sequence elements of the *C.elegans* cosmid M88 that was used in the *in vitro* SAR binding assay. A negative distance indicates that the 8 bp sequence precedes the 16 bp sequence.

In these binding assays all MRSs map to biochemically identifiable SARs. Figure 2A–F correspond to the assays leading to the identification of SARs A–F depicted in the upper part of Figure 2. The first SAR (A) is located in a 2208 bp *Bgl*II fragment that spans the intergenic region between the F25F2.2 and M88.1 genes. Fine mapping narrowed this SAR down to a 1343 bp *Bgl*II–*Eco*RI fragment (Fig. 2A) with both MRSs on the 3'-edge of this fragment. The second set of MRSs that point to a SAR are located in a large 8414 bp *Bgl*II fragment that spans several genes. The SAR could be assigned to the region around the MRSs in the relatively small 828 bp *Eco*RI–*Pst*I fragment (Fig. 2C). MRS D is found in a large, 8374 bp *Bgl*II fragment. This restriction fragment also contains a SAR as shown by the strong retention of this fragment by the nuclear scaffold (Fig. 2D). The SAR is located in the 3178 *Xba*I–*Cla*I fragment. This region overlaps with the large third intron of the M88.5 gene that contains the MRS. In the case of MRS E1 and MRS E2 found within the 3729 bp *Bgl*II fragment we could not determine whether these sequences map in one or two SARs. The relatively small distance of 950 bp between these two MRSs and the average size of SARs of several hundred base pairs contribute to this uncertainty. The last MRS of the *C.elegans* M88 cosmid is located in a 1759 bp *Bgl*II fragment which overlaps with the promoter and the 5'-region of the M88.6 gene. This MRS F, within the small fourth intron, correctly predicts a SAR (Fig. 2F).

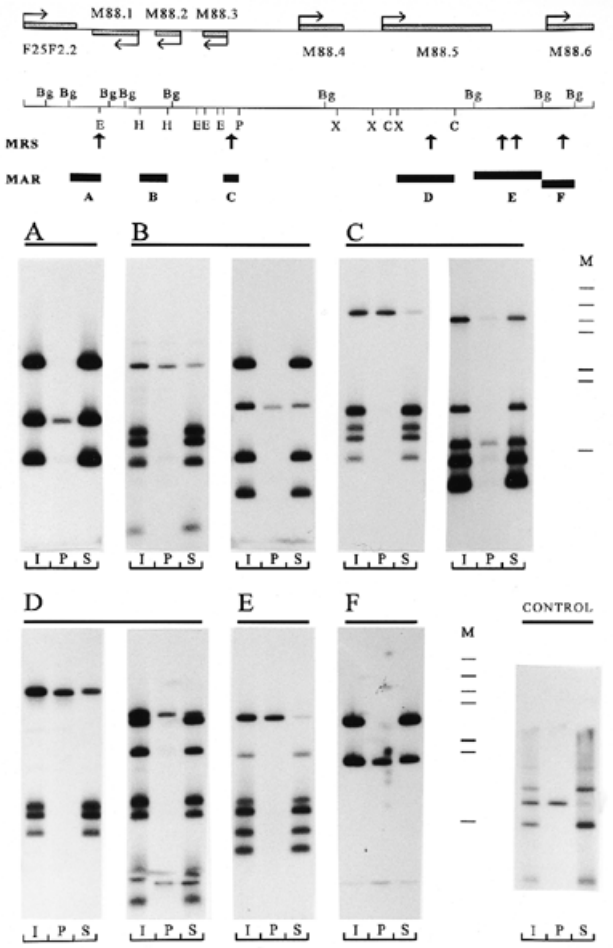


Figure 2. Identification of SARs within the M88 cosmid of *C.elegans*. The top panel indicates the position of potential open reading frames (hatched bars) with their direction of transcription (horizontal arrows) within cosmid M88. The position of the MRS (vertical arrows) in relation to the restriction map (Bg, *Bgl*II; C, *Cla*I; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; X, *Xba*I) and the identified SARs A–F (solid bars). SAR assay of cloned *Bgl*II fragments from the cosmid M88 (I, input; P, pellet; S, supernatant) where SARs A–F are mapped in the corresponding panels. In these experiments the *Xho*II or an additional site from the polylinker (indicated as *Bgl*II) were used to release the *Bgl*II inserts from the vector (a *Bgl*II–*Bam*HI fusion). This results in three additional fragments from the vector: (**A**) *Bgl*II/*Eco*RI digest of pCM06 (2208 bp); (**B**) *Xho*II and *Bgl*II/*Hin*dIII digest of pCM07 (2716 bp); (**C**) *Xho*II and *Eco*RI/*Pst*I digest of pCM11 (8414 bp); (**D**) *Xho*II and *Xba*I/*Cla*I digest of pCM04 (8374 bp); (**E**) *Xho*II digest of pCM08 (3729 bp); (**F**) *Xho*II digest of pCM05 (1759 bp). The positive control is an *Ava*II/*Eco*RI/*Hin*dIII/*Pst*I digest of the *D.melanogaster* intergenic histone SAR. The size marker (M) is a lambda *Hin*dIII digest with fragments of 23.1, 9.4, 6.6, 4.4, 2.3, 2.0 and 0.6 kb.

Our search for SARs in the *C.elegans* cosmid M88 revealed that, in addition to the six predicted SARs, the 2700 bp *Bgl*II fragment, which contains no MRS, also binds to the nuclear matrix (Fig. 2B). We therefore conclude that the MRS is diagnostic for a large class of SARs but evidently not for all SARs.

The MRS is present in many eukaryotic MARs/SARs

Table 2 shows that seven out of eight previously identified SARs of *Drosophila* contain a MRS (1,2,27,28). Only in the case of the development-specific adult SAR of the alcohol dehydrogenase gene could we not detect a MRS. We could extend these observations by showing that the MARs/SARs from DNA virus SV40 (29), yeast (30), chicken (31,32), Chinese hamster (33,34), mouse (21,26,35,36–39), rabbit (39) and man (40–44) also map at the position of a MRS (Table 2). Not all MARs/SARs from these organisms contain a MRS, suggesting that the MRS-containing elements constitute a specific class of SARs and that this class can be found in many eukaryotic species.

If MARs/SARs mediate some essential function in the regulation of gene expression, one would expect that a MAR or SAR associated with a given gene, including its MRS, is conserved between species. Observations related to the immunoglobin κ gene support this idea. The Ig(κ) gene, which has been cloned from a number of different organisms, contains a MAR in the J–C intronic region (21). In both man and mouse this MAR contains a MRS. The rabbit genome contains two Ig(κ) genes of which the Ig(κ1) gene does and the Ig(κ2) gene does not contain a MAR (43). Significantly, the Ig(κ1) intron does and the Ig(κ2) intron does not contain a MRS. So, in the Ig(κ) gene family the presence of a MRS sequence and a MAR are tightly linked and evolutionarily conserved. Conservation of the MRS is not a trivial consequence of sequence similarities that are expected between homologous genes. The DNA sequence, the relative orientation and the distance between the two sequence elements that make up the MRSs are all different for each of the Ig(κ) MARs.

The distribution of MRSs alludes to eukaryotic genomic organisation

In three genomic regions of in total 32 kb of *Arabidopsis* we found on average one SAR every 4.5 kb (23), while in this paper we show that for the 33 kb *C.elegans* M88 cosmid this is once every 5.5 kb. The distribution of MRSs in the two large sequenced contigs from these organisms revealed a similar organisation. In a 250 kb region of chromosome IV of *Arabidopsis* (ESSA project of Drs Murphy and Bancroft, John Innes Centre, Norwich) on average we find one MRS per 6.5 kb. Likewise, in a 2.1 Mb contig of *C.elegans* chromosome III (24), we find one MRS per 5.1 kb. Interestingly, the average distance between the MRSs is similar to the average gene density (one ORF per 5.6 kb) for both these organisms.

The distribution of the MRS can be used to investigate the basic genomic organisation, although we have shown that some MARs/SARs may be missed. The evolutionarily conserved β-globin cluster is an interesting region in this respect. This region with its developmentally regulated genes has been cloned and sequenced from several organisms. Furthermore, in the case of the human locus the distribution of SARs has been determined (43). The human β-globin cluster contains seven MRSs (Fig. 3) that all map to biochemically identified SARs. One in the locus control region (LCR) between HS5 and HS4, one just upstream of the ε gene, three around the pseudo-β gene and one downstream of the β-globin gene. A similar picture emerges for the distribution of MRSs in the homologous β-globin clusters of galago (45), rabbit (45) and mouse (46) (Fig. 3). A small number of MRSs are found dispersed throughout these loci, marking potential chromatin domains. Each of these domains contains a single gene. Notably, in mouse the active genes and their pseudogenic counterparts map to individual SAR-bounded domains. Also, MRSs are present in the LCR. Interestingly, even more prominent than in

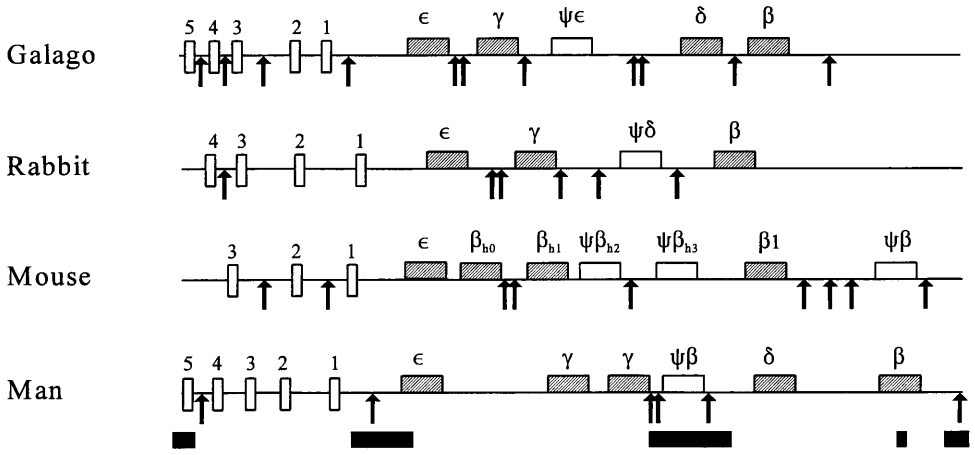


Figure 3. The genomic organisation of the mammalian β-globin gene clusters. Schematic representation of the β-globin clusters of galago, rabbit, mouse and man with the DNase I hypersensitive sites of the LCR (open rectangles), functional genes (hatched bars) and pseudogenes (open bars). The vertical arrows indicate the positions of the MRSs with the solid bars below the human cluster representing the previously identified SARs in this cluster.

man, several DNase I hypersensitive sites in the LCR are separated by MRSs. The data suggest that not only the functional genes, but also the genomic organisation of these loci has been conserved in evolution.

DISCUSSION

Identification of a MAR/SAR-associated sequence element

In the present study we describe the discovery of a sequence element that is associated with a large group of eukaryotic MARs/SARs. The MRS is a bipartite sequence element that consists of two individual sequences of 8 (AATAAYAA) and 16 bp (AWWRTAANNWWGNNNC) within a 200 bp distance from each other. Our analysis of a 33 kb *C.elegans* cosmid showed that the MRS correctly predicts the position of six SARs on the basis of genomic sequence information. All fragments that contain a MRS bind specifically and with high affinity to the nuclear matrix if tested in an *in vitro* SAR assay. The same is true for three non-related genomic loci in *Arabidopsis*. Here, seven MRSs in >30 kb genomic sequences correctly identified seven SARs. Significantly, all MRSs that have been identified so far map to SARs. We never found a MRS that did not relate directly to a biochemically defined SAR.

Not all SARs contain a MRS. Biochemical analysis of the *Arabidopsis* and *C.elegans* genomic loci revealed that one of the *C.elegans* SARs does not contain a MRS. A database survey of MARs and SARs that have been identified by others led to a similar conclusion. Table 2 shows that although many (20 out of 27) of these elements did contain the MRS, some (7 out of 20) did not. This suggests that at least one other type of MAR/SAR may exist which does not contain a MRS. Alternatively, the MRS consensus sequence may still have to be adapted to cover all of the known MAR and SAR sequences. Since we could not find a consensus sequence that covers a larger fraction of known elements, despite considerable efforts, we favour the former possibility. It is important to note that the MRS is found in MAR/SAR elements that have been characterised using a variety of experimental systems, employing different types of matrix and scaffold preparations and cell types. These also include *in vivo* assays that confirm the presence of MAR/SAR elements initially identified using an *in vitro* binding assay. Evidently, the observed link between the MRS and MARs/SARs is not dependent on the type of biochemical assay used to identify them.

The MRS may have a special spatial arrangement on a nucleosome

We have shown that the MRS is a combination of two sequence elements within a short distance. The observed variation in the distance is suggestive of a relation of the MRS with nucleosomes. In the case of the *Drosophila* histone cluster the positions of the nucleosomes have been mapped (2,47,48). The SAR located between the histone H1 and H3 genes contains a number of nucleosomes with a defined position and two MRSs. Within the resolution of the nucleosome mapping data (~15 bp), the two MRSs seem to occupy a similar position on their respective nucleosomes. The first MRS, where the bipartite sequence elements are in close proximity, is found in the vicinity of a dyad centre of a nucleosome. In the case where they are 145 bp apart both elements are located near the entry and exit site of a nucleosome. So, due to the wrapping of the DNA around the histone protein complex, the two bipartite sequences appear to be physically close together.

This configuration of the two elements of the MRS could be a common feature. In SV40 (49) a nucleosome with its dyad centre at position 4020 aligns the 16 bp sequence (4012) at the dyad centre with the 8 bp sequence (3957) at the entry/exit site of the nucleosome, located some 80 bp away (Table 2). For the γ-globin SAR in the human β-globin cluster (43) and the chicken α-globin SAR (35) the positions of nucleosomes have been predicted on the basis of sequence characteristics (22,50). In the γ-globin SAR the 8 bp sequence is located at the dyad centre and the 16 bp sequence 60 bp downstream at the entry/exit site. The chicken α-globin SAR contains two overlapping sequence elements of the MRS at the dyad centre. The same

Table 2. The distribution of MRSs in matrix attachment regions

Organism	locus	sequence and orientation				distance
		AWWRTAANNWWGNNNC		AATAAYAA		
chicken	α-globin	AATATGATGTTGAACC	(W)	AATAATGA	(W)	overlap
	lysozyme [3']	AATATAAAAATGGTCA	(C)	AATAATAA	(C)	+134
	lysozyme [5']		no MRS			
chin. hamster	dhfr [intron]	ATTATAATAAAACAAC	(W)	AATAATAA	(W)	+3
	dhfr [5'+3']		no MRS			
drosophila	α-actin	AAAATAAATAAGTCTA	(W)	AATAACAA	(W)	+200
	adh [larval]	ATAATAAAATTAAATC	(C)	AATAATAA	(C)	overlap
	adh [adult]		no MRS			
	ftz	ATAGTAACTAAGAATT	(W)	AATAATAA	(C)	+68
	histone	CAAGTAATAAAGTATC	(C)	AATAACAA	(W)	+145
		AATATAAAATAGCTAG	(W)	AATAATAA	(W)	+6
	hsp70	ATAATAATCGAGTTCC	(C)	AATAATAA	(C)	overlap
	sgs4	ATTGTAATTAAAGCCC	(C)	AATAATAA	(C)	-39
human	Ig(κ)	AAAATAAGATATATTC	(W)	AATAACAA	(W)	-133
	myc		no MRS			
	β-globin [3']	ATAATAAAAAAGAATA	(W)	AATAATAA	(W)	overlap
		AAAATATATAAGAAAC	(W)	AATAACAA	(W)	+11
	β-intron		no MRS			
	ϵ-globin	CAAGTAAGAATGTCTC	(C)	AATAATAA	(W)	-59
	ψβ [5']	ATAATAAGCCTGCGCC	(W)	AATAATAA	(W)	overlap
		AAAATAAATGAGGAGC	(C)	AATAATAA	(W)	+144
	ψβ [3']	ATTATAAATATGTTTC	(C)	AATAATAA	(W)	+109
	ifnβ [E]		no MRS			
mouse	hprt	ATAATAAAAATCAGAC	(C)	AATAATAA	(C)	overlap
	Ig(κ)	AATATAAATTTGTGAC	(W)	AATAATAA	(W)	+67
	IgH		no MRS			
rabbit	Ig(κ1)	AAATTAATTTAGAGCC	(W)	AATAATAA	(C)	+132
SV40	large T	ATTAAAATCATGCTCC	(W)	AATAACAA	(W)	-81
yeast	cen-III	ATAACAATAATGCAAC	(W)	AATAACAA	(W)	overlap
		ATTATCACGTTGATTC	(W)	AATAACAA	(W)	+33

Positions of the MRSs in previously identified SARs and MARs from a variety of eukaryotic organisms with their sequence composition, relative orientation (C, Crick strand; W, Watson strand) and distance between the two sequence elements. A negative distance indicates that the 8 bp sequence precedes the 16 bp sequence.

configuration of the two sequence elements of the MRS might also apply for the other MARs/SARs discussed in this paper.

Possible functions of the MRS

The association of the MRS with MARs and SARs may be a result of two different mechanisms. Either the MRS is required for binding or it represents an interaction site for proteins that mediate a specific function other than binding to the nuclear matrix/scaffold. These possible functions of the MRS are currently under investigation. It is intriguing that the MRS seems to occupy a specific position on nucleosomes, whereas the nuclear scaffolds used for detection of the SARs are most likely fully stripped of nucleosomes. When the MRS is involved in binding to the nuclear matrix, the nucleosomal context is evidently not required. Alternatively, the MRS could be involved in some MAR/SAR-specific function other than matrix/scaffold binding.

An indication concerning its function might come from its position on nucleosomes. As described above, the two sequence elements of the MRS seem to come physically close together near the dyad axis of the nucleosome. Interesting in this respect are the elegant experiments of Laemmli and co-workers, who attempted to identify proteins that interact with the H1–H3 intergenic SAR of *D.melanogaster* (51). Using *Exo*III to map the position of SAR-binding proteins they identified four strong stops for *Exo*III. The positions of the individual 16 and 8 bp components of the two MRSs present in this SAR precisely map to the position of these *Exo*III stops. These observations suggest that the 8 and 16 bp elements of the MRS constitute a target for nuclear proteins.

Wrapping DNA around a nucleosome to create a protein binding site employing two sequence elements non-adjacent on the linear DNA may be a more general function of nucleosomes in transcription regulation. A similar situation can be found in the *Drosophila* hsp26 promoter. Here, binding sites for the GAGA and the heat shock factors, required for activation of the promoter, are brought close together through a positioned nucleosome located between these sites (52).

Predicting MARs and SARs in genomic loci

The conserved function of MARs and SARs has led to an intensive search for methods to predict these elements based on genomic sequence information alone, rather than on a laborious biochemical assay. The field has been pioneered by Boulikas and co-workers, describing a large number of sequence elements that may be enriched in MARs and SARs (for a review see 22). Common elements, such as A and T boxes and topoII sites, were discovered through direct comparison of known MAR and SAR sequences. Although these sequences were found clustered in MARs/SARs, similar clusters could also be found at other sites in the genome that do not have any affinity for the nuclear skeleton. This has been demonstrated in the region around the plastocyanin gene in *Arabidopsis* (23). Methods using a combination of sequence elements were the first to have some success in predicting the position of MARs and SARs (53,54). Such statistical methods are likely to focus on general structural aspects, rather than some unique feature. Although, the MAR-finder program developed by Kramer and co-workers (52) allows the identification of MARs in the human TNP and β-globin loci, it uses two different sets of rules for the two genomic regions. Evidently, this approach has not resulted in a unique set of rules for MAR/SAR identification.

Recently, a new approach was developed on the basis of the DNA-unwinding potential of MARs and SARs (4,38). A computer program was developed that calculates the stress-induced DNA-unwinding potential of a given sequence (40). Applying this criterion to sequences that are known to contain a SAR revealed regions of strong unwinding potential to coincide with promoters, poly(A) signals and SARs. In the case of the IgH gene (37) the program correctly identified the local unwinding regions within this SAR (37,38). An advantage of this approach is that it directly focuses on a well-defined functional aspect of MARs/SARs, i.e. their local DNA-unwinding potential. A complication is that one needs a detailed understanding of the genomic structure to discriminate between the different types of DNA-unwinding regions. In cases where MARs or SARs are located in the vicinity of promoters and enhancers one might not be able to tell these different regions apart. Here we show that the MRS can be used to faithfully predict the genomic MARs and SARs. In every instance where we have mapped a MRS, the corresponding genomic region binds to the nuclear scaffold. There is no correlation of the MRS with other MAR- or SAR-associated sequence elements, such as ATC sequences, i.e. the sequence composition of the MRS does not show the typical asymmetrical distribution of GC nucleotides (38,55).

In summary, our observations show that the MRS is a new and potential powerful tool in the unravelling of the molecular mechanism of action of MARs/SARs. Data from ongoing sequencing projects can now be used to assess the distribution of MARs/SARs in a genome and allow one to assess the higher order organisation of a particular genomic region in terms of the chromatin loop domain model. The identification of proteins that bind to the MRS will be a next important step towards the elucidation of the molecular mechanisms by which SARs and MARs control gene activity.

ACKNOWLEDGEMENTS

The authors would like to thank Drs T. Kowhi-Shigematsu (Lawrance Berkeley National Laboratory, Berkeley, CA) and J. Bode (GBF, Braunschweig, Germany) for their stimulating discussions and ideas on this work prior to publication. Drs G. Murphy and I. Bancroft (John Innes Centre, Norwich, UK) kindly made the DNA sequence of a 250 kb region of chromosome IV of *A.thaliana* from the ESSA project (European Scientists Sequencing *Arabidopsis*) available prior to publication. The *C.elegans* cosmid M88 was obtained through Dr R. Plasterk (NKI, Amsterdam, The Netherlands). The plasmid containing the intergenic histone MAR of *D.melanogaster* was a gift of Dr U. K. Laemmli (University of Geneva, Geneva, Switzerland). The DNA sequences of the interferon and chicken lysozyme MARs were kindly provided by Drs J. Bode (GBF, Braunschweig, Germany) and C. Bonifer (University of Leeds, Leeds, UK). This work was supported by a NWO/ALW (17.261) and a HFSP grant (RG0360/1995-M).

REFERENCES

1. Mirkovitch,J., Mirault,M.-E. and Laemmli,U.K. (1984) *Cell*, **39**, 223–232.
2. Gasser,S.M. and Laemmli,U.K. (1986) *Cell*, **46**, 521–530.
3. Allen,G.C., Hall,G.E., Childs,L.C., Weissinger,A.K., Spiker,S. and Thomson,W.F. (1993) *Plant Cell*, **5**, 603–615.
4. Bode,J., Schlake,T., Rios-Ramirez,M., Mielke,C., Stengert,M., Kay,V. and Klehr-Wirth,D. (1995) In Berezney,R. and Jeon,K.W. (eds), *Structural and Functional Organization of the Nuclear Matrix*, Vol. 162A. Academic Press, San Diego, CA, pp. 389–454.
5. Ludérus,M.E.E., Degraaf,A., Mattia,E., Den Blaauwen,J.L., Grande,M.A., De Jong,L. and Van Driel,R. (1992) *Cell*, **70**, 949–959.
6. Von Kries,J.P., Rosorius,O., Buhrmester,H. and Stratling,W.H. (1994) *FEBS Lett.*, **342**, 185–188.
7. Fackelmayer,F.O., Dahm,K., Renz,A., Ramsperger,U. and Richter,A. (1994) *Eur. J. Biochem.*, **221**, 749–757.
8. Renz,A. and Fackelmayer,F.O. (1996) *Nucleic Acids Res.*, **24**, 843–849.
9. Cunningham,J.M., Purucker,M.E., Jane,S.M., Safer,B., Vanin,E.F., Ney,P.A., Lowrey,C.H. and Nienhuis,A.W. (1994) *Blood*, **84**, 1298–1308.
10. Nakagomi,K., Kohwi,Y., Dickinson,L.A. and Kohwi-Shigematsu,T. (1994) *Mol. Cell. Biol.*, **14**, 1852–1860.
11. Herrscher,R.F., Kaplan,M.H., Lelsz,D.L., Das,C., Scheuermann,R. and Tucker,P.W. (1995) *Genes Dev.*, **9**, 3067–3082.
12. Strick,R. and Laemmli,U.K. (1995) *Cell*, **83**, 1137–1148.
13. Saitoh,Y. and Laemmli,U.K. (1993) *Cold Spring Harbor Symp. Quant. Biol.*, **58**, 755–765.
14. Izaurralde,E., Kas,E. and Laemmli,U.K. (1989) *J. Mol. Biol.*, **210**, 573–585.
15. Zhao,K., Kas,E., Gonzalez,E. and Laemmli,U.K. (1993) *EMBO J.*, **12**, 3237–3247.
16. Varga-Weisz,P., Vanholde,K. and Zlatanova,J. (1994) *Biochem. Biophys. Res. Commun.*, **203**, 1904–1911.
17. Belyaev,N.D., Keohane,A.M. and Turner,B.M. (1996) *Hum. Genet.*, **97**, 573–578.
18. Turner,B.M. (1991) *J. Cell Sci.*, **99**, 13–20.
19. Klehr,D., Schlake,T., Maass,K. and Bode,J. (1992) *Biochemistry*, **31**, 3222–3229.
20. Varga-Weisz,P., Wilm,M., Bonte,E., Dumas,K., Mann,M. and Becker,P.B. (1997) *Nature*, **388**, 598–602.
21. Cockerill,P.N. and Garrard,W.T. (1986) *FEBS Lett.*, **204**, 5–7.
22. Boulikas,T. (1995) In Berezney,R. and Jeon,K.W. (eds), *Structural and Functional Organization of the Nuclear Matrix*, Vol. 162A. Academic Press, San Diego, CA, pp. 279–388.
23. Van Drunen,C.M., Oosterling,R.W., Keultjes,G.M., Weisbeek,P.J., Van Driel,R. and Smeekens,S.C.M. (1997) *Nucleic Acids Res.*, **25**, 3904–3911.
24. Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J., Coulson,A., Dear,S., Du,Z., Durbin,R., Favello,A., Fraser,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laisster,N., Latereille,P., Lightning,J., Lloyd,C., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A., Saunders,D., Shownkeen,R., Sims,M., Smaldon,N., Smith,A., Smith,M., Sonnhammer,E., Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Sproat,J. and Wohldman,P. (1994) *Nature*, **168**, 32–38.
25. Izaurralde,E., Mirkovitch,J. and Laemmli,U.K. (1988) *J. Mol. Biol.*, **200**, 111–125.
26. Cockerill,P.N. and Garrard,W.T. (1986) *Cell*, **44**, 273–282.
27. Eggert,H. and Jack,R.S. (1991) *EMBO J.*, **10**, 1237–1243.
28. Mirkovitch,J., Gasser,S.M. and Laemmli,U.K. (1988) *J. Mol. Biol.*, **200**, 101–109.
29. Pommier,Y., Cockerill,P.N., Kohn,K.W. and Garrard,W.T. (1990) *J. Virol.*, **64**, 419–423.
30. Amati,B.B. and Gasser,S.M. (1988) *Cell*, **54**, 967–978.
31. Phi-Van,L. and Stratling,W.H. (1988) *EMBO J.*, **7**, 655–664.
32. Stief,A., Winter,D.M., Stratling,W.H. and Sippel,A.E. (1989) *Nature*, **341**, 343–345.
33. Dijkwel,P.A. and Hamlin,J.L. (1988) *Mol. Cell. Biol.*, **8**, 5398–5409.
34. Kas,E. and Chasin,L.A. (1987) *J. Mol. Biol.*, **198**, 677–692.
35. Avramova,Z. and Paneva,E. (1992) *Biochem. Biophys. Res. Commun.*, **182**, 78–85.
36. Cockerill,P.N. (1990) *Nucleic Acids Res.*, **18**, 2643–2648.
37. Cockerill,P.N. and Yuen,M.H. (1987) *J. Biol. Chem.*, **262**, 5394–5397.
38. Kohwi-Shigematsu,T. and Kohwi,Y. (1990) *Biochemistry*, **29**, 9551–9560.
39. Sperry,A.O., Blasquez,V.H. and Garrard,W.T. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 5497–5501.
40. Benham,C., Kohwi-Shigematsu,T. and Bode,J. (1997) *J. Mol. Biol.*, **274**, 181–196.
41. Bode,J. and Maass,K. (1988) *Biochemistry*, **27**, 4706–4711.
42. Gromova,I., Thomsen,B. and Razin,R.V. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 102–106.
43. Jarman,A.P. and Higgs,D.R. (1988) *EMBO J.*, **7**, 3337–3344.
44. Whitehurst,C., Henney,H.R., Max,E.E., Schroeder,W.H., Stuber,F., Siminovitch,K.A. and Garrard,W.T. (1992) *Nucleic Acids Res.*, **20**, 4929–4930.
45. Slightom,J.L., Bock,J.H., Tagle,D.A., Gumucio,D.L., Goodman,M., Stojanovic,N., Jackson,J., Miller,W. and Hardison,R. (1997) *Genomics*, **39**, 90–94.
46. Reitman,M., Grasso,J.A., Blumenthal,R. and Lewit,P. (1993) *Genomics*, **18**, 616–626.
47. Samal,B. and Worcel,A. (1981) *Cell*, **23**, 401–409.
48. Worcel,A., Gargiulo,G., Jessee,B., Udvardy,A., Louis,C. and Schedl,P. (1983) *Nucleic Acids Res.*, **11**, 421–439.
49. Ambrose,C., Lowman,H., Rajadhyaksha,A., Blasquez,V. and Bina,M. (1990) *J. Mol. Biol.*, **214**, 875–884.
50. Wadakiyama,Y. and Kiyama,R. (1994) *J. Biol. Chem.*, **269**, 22238–22244.
51. Gasser,S.M. and Laemmli,U.K. (1986) *EMBO J.*, **5**, 511–518.
52. Lu,Q., Wallrath,L.L. and Elgin,S.C.R. (1995) *EMBO J.*, **14**, 4738–4746.
53. Kramer,J.A., Singh,G.B. and Krawetz,K.A. (1996) *Genomics*, **33**, 305–308.
54. Singh,G.B., Kramer,J.A. and Krawetz,K.A. (1997) *Nucleic Acids Res.*, **25**, 1419–1425.
55. Bode,J., Kohwi,Y., Dickinson,L., Joh,T., Klehr,D., Mielke,C. and Kohwi-Shigematsu,T. (1992) *Science*, **255**, 195–197.