Journal of Cellular Biochemistry 52:14–22 (1993)

Nature of DNA Sequences at the Attachment Regions of Genes to the Nuclear Matrix

Teni Boulikas

Institute of Molecular Medical Sciences, Palo Alto, California 94306

Abstract Matrix-attached regions (MARs) have been demonstrated to nest origins of replication and transcriptional enhancers. A set of 13 rules is proposed aimed at facilitating the classification of a DNA sequence as a matrix attachment region. These rules, which were deduced from a study of known MARs from other genes and some others identified in our laboratory, are (1) potential origin of replication are MARs; (2) the major class of MARs seclude clusters of AT-rich motifs and may harbor topoisomerase II binding and cleavage sites; (3) the AT-rich class of MARs may comprise the DNA sequence motifs ATTA and ATTTA representing core binding sites of homeotic proteins, implying that MARs may participate in the differential activation of origins of replication and in gene switch during development; (4) the habitat of MARs may include mass binding sites for protein transcription factors; even weak factor binding sites may lead to the formation of tight protein-DNA supramolecular structures; (5) MARs may contain intrinsically curved DNA, one type is oligo(dA) stretches of 3 to 7 nucleotides spaced every 10.5 nucleotides; (6) a class of MARs may contain kinked DNA, formed by CA, TG, and TA dinucleotides at distances of 5 or 10.5 nucleotides from their centers; the same dinucleotides, known to be abundant in protein recognition sites, may be overrepresented in a special class of MARs, (7) the AT-rich core of MARs may be flanked, at one or both sides, by sequences that can adopt the left-handed or triple-helical DNA structure; these include TG, TA, GC repeats and polypurine or polypyrimidine stretches; (8) palindromic (dyad symmetry) sequences, able to form cruciform structures when the DNA is under torsional strain may be found within MARs, and more so when the MAR is also an origin of replication; (9) transcriptional enhancers may be MARs; (10) a class of MARs may be composed of stretches of GA-rich DNA alternating with CT-rich stretches, 5–50 nucleotides long; (11) a class of MARs may be enriched in TG boxes, usually 6–12 nucleotides long, such as TGTTTTGGGG; this type of MAR occurs in the 3'-untranslated region of several genes, builds up the chromosome telomeres, and is highly recombinogenic; (12) a small fraction of Alu sequences might have MAR activity. This might depend on the number and distance from one another of DNA sequence motifs representing protein binding sites; and (13) MARs may coincide with the DNAse I hypersensitive sites of chromatin.

It is proposed here that MAR sequences can provide markers for mapping and sequencing the human, and other, genomes. Furthermore, it is proposed that large scale random cloning of MARs might advance our knowledge on the nature of DNA sequences that are used for the initiation of DNA replication, as transcriptional enhancers and as borders between chromatin domains.

Key words: nuclear matrix, homeotic protein site, palindromes, origins of replication, transcriptional enhancers, curved DNA, Z-DNA, human genome project, markers

It is well established that both interphase chromatin and mitotic chromosomes are organized into loops or domains. This organization is brought about by the anchorage of specific DNA sequence landmarks to a network of protein cross-ties termed chromosomal scaffolds at mitosis [Paulson and Laemmli, 1977; Mirkovitch et al., 1988] or nuclear matrix at interphase [see Hancock and Boulikas, 1982]. Such DNA segments, termed SARs or MARs (scaffold- or matrix-attached regions; see Cockerill and Garrard [1986a]) are supposed not only to facilitate the proper expression and replication of DNA during the cell cycle but also to govern the cell type-specific expression of genes and the activation of different origins of replication during development [see Bodnar, 1988; Fey et al., 1991; Berezney, 1991; Villarreal, 1991; Boulikas,

Abbreviations used HOMEO, homeotic protein binding site, MAR, matrix attachment region, ORI, origin of replication, DNase I HS site, DNase I hypersensitive site, kb, kilobase pairs, bp, base pairs, nt, nucleotides, 3' UTR, 3' untranslated region

Received October 28, 1992, accepted November 17, 1992

Address reprint requests to Dr Teni Boulikas, Institute of Molecular Medical Sciences, 460 Page Mill Road, Palo Alto, CA 94306

1992a]. MARs were inferred to govern the differential activation of origins of replication during development; the similarity of a major class of MARs to homeodomain protein binding sites suggested that MARs participate in the establishment of body segments and patterns formation [Boulikas, 1992b].

Two types of nuclear matrix structures have been described. First, an external part, lining the interior of the nuclear envelope, with lamins A, B, and C as the major components, apparently mediates the topological compartmentalization of the peripheral chromatin [see Hancock and Boulikas, 1982; Boulikas, 1987; Berezney, 1991]. Direct contacts between lamins and DNA, as determined by crosslinking with UV [Boulikas, 1986], seem to be the major determinants of this rather unspecific chromatin loop anchoring to the peripheral nuclear matrix. DNA sequence specificity could be conferred by the interaction of lamins and/or nearby DNA sequences with transcription factors. Second, an internal nuclear matrix component containing thick, highly branched polymorphic fibers might serve as the core structure for the internal matrix architecture [e.g., He et al., 1990]. Extraction of nuclease-digested nuclei with high salt in the presence of 2-mercaptoethanol yields empty nuclei shells depleted of their internal nuclear matrix and composed only of nuclear pore–lamina complexes [Kaufmann et al., 1983].

MARs may be permanent (constitutive) and be present in all cell types irrespective of gene activity or facultative (functional) appearing only in specific cell types in connection with transcriptional or repair activity [Razin et al., 1986; Levy-Wilson and Fortier, 1989; Boulikas, 1992b].

MARs were initially identified in the *Drosophila* histone gene repeat [Mirkovitch et al., 1984]. Since then, MARs were found in the 5′ and 3′ flanking regions of *fushi-tarazu*, *Sgs-4*, and alcohol dehydrogenase single-copy genes of *Drosophila* [Mirkovitch et al., 1986, 1988; Gasser and Laemmli, 1986b], the J-C intron of the mouse κ immunoglobulin gene [Cockerill and Garrard, 1986a,b; Cockerill et al., 1987; Blasquez et al., 1989b], the first intron of the human HPRT gene [Sykes et al., 1988], the chicken lysozyme gene domain [Phi-Van and Strätling, 1988], the human interferon-β gene [Bode and Maass, 1988; Mielke et al., 1990; Klehr et al., 1991], the human β-globin gene [Jarman and Higgs, 1988], the murine β-globin gene [Greenstein, 1988], the chicken α-globin gene [Farache et al., 1990], the apolipoprotein B gene [Levy-Wilson and Fortier, 1989], three root-specific tobacco genes [Hall et al., 1991], the human Ha-*ras* oncogene [Boulikas et al., in preparation] and the human poly(ADP-ribose) polymerase gene [Boulikas and Kong, submitted].

Overwhelming evidence shows that MARs may coincide with transcriptional enhancers or display enhancer activity in transfection and transgenic experiments [Xu et al., 1989; Blasquez et al., 1989a; Stief et al., 1989; Phi-Van et al., 1990; Klehr et al., 1991]. In addition, ends of transcription units are flanked, in some cases, by a MAR [Mirkovitch et al., 1984; Dijkwel and Hamlin, 1988; Phi-Van and Strätling, 1988]. Furthermore, initiation of DNA replication [e.g., Berezney and Coffey, 1975], as well as repair [McCready and Cook, 1984; Mullenders et al., 1988] seem to take place in association with the nuclear matrix. Nuclear matrix enriched in RNA polymerases is also the principal nuclear microenvironment where RNA synthesis occurs [Ciejek et al., 1982; Robinson et al., 1982; Leibovitch et al., 1983]. In agreement with these biochemical studies, electron microscopy autoradiography has shown that transcription occurs at the borders between hetero- and eu-chromatin [Fakan and Bernhard, 1971], presumably coinciding with the surface of some nuclear matrix structures.

In this paper, a set of rules are compiled to aid researchers in classifying already existing or newly sequenced stretches of DNA as MARs.

RULES DIAGNOSTIC OF MAR SEQUENCES

Rule 1: Potential origins of replication are MARs. That newly replicated DNA is specifically located on the nuclear matrix, has been demonstrated in several laboratories [Berezney and Coffey, 1975; Buongiorno-Nardelli et al., 1982; Aelen et al., 1983; van der Velden et al., 1984; Razin et al., 1986; Carri et al., 1986; Dijkwel and Hamlin, 1988; Vaughn et al., 1990]. According to the model of Pardoll and collaborators [1980], the DNA is reeled through its nuclear matrix attachment site during replication. Additional evidence for the close proximity or even coincidence of origins of replication and nuclear matrix anchorage sites derives from the work of Amati and Gasser [1988, 1990] showing that MARs from yeast coincide with putative origins of replication and that *Drosophila* MARs can drive the autonomous replication of plasmids in yeast. In addition, isolation and cloning

of putative origins of replication from monkey cells in culture [Frappier and Zannis-Hadjopoulos, 1987] has shown them to possess sequence homology with MARs [Rao et al., 1990]. These data rather conclusively demonstrate that origins of replication are nuclear matrix components. Furthermore, the work of Boulikas [1992b] pointing out that origins of replication, nuclear matrix anchorage sites, and homeotic protein recognition and binding sites share the ATTA, ATTTA, and ATTTTA motifs suggests that the differential activation of origins of replication, which has a highly important role in development [Callan, 1973; Spradling and Orr-Weaver, 1987] might be regulated on the nuclear matrix.

Rule 2: A major class of matrix attachment sites are known to be AT-rich sequences of about 100–1,000 bp in length [Mirkovitch et al., 1984, 1986; Gasser and Laemmli, 1986a,b] usually harboring topoisomerase II binding and cleavage sites [Cockerill and Garrard, 1986a; Gasser and Laemmli, 1986a; Käs and Chasin, 1987; Käs and Laemmli, 1992].

Rule 3: MARs may comprise recognition and binding sites of homeodomain proteins [Boulikas, 1992b]. Figure 1 shows a striking example of this kind; the constitutive MAR to the 3′ side of the human apolipoprotein B gene, which is present in cell types expressing the gene as well as in cell types harboring the gene into an inactive chromatin structure [Levy-Wilson and Fortier, 1989], appears to be a mosaic of homeodomain protein binding sites. The proper alignment of the nucleotide sequences (Fig. 1) shows that this MAR sequence is composed of a 30-nucleotide element, repeated 18.5 times. The 30-nucleotide element itself is a dimer containing the sequence TAATTA, a fusion of the TAAT and ATTA motifs. In addition, the 30 nt repeat comprises the TAAAAT, ATTTA, TAAAT, and ATTTTA elements (Fig. 1). These results support a role of the nuclear matrix in gene switch and differential activation of origins of replication in development [Dworetzky et al., 1990; Boulikas, 1992b].

We propose that an isolated AT-rich stretch of 4–20 nucleotides is not sufficient to specify a MAR. Instead, several specific sequence AT-rich stretches representing protein binding sites need to be clustered. Such sequences need to be spaced at multiples of 8–12 nucleotides (nt) to allow for the interaction of the proteins with each other

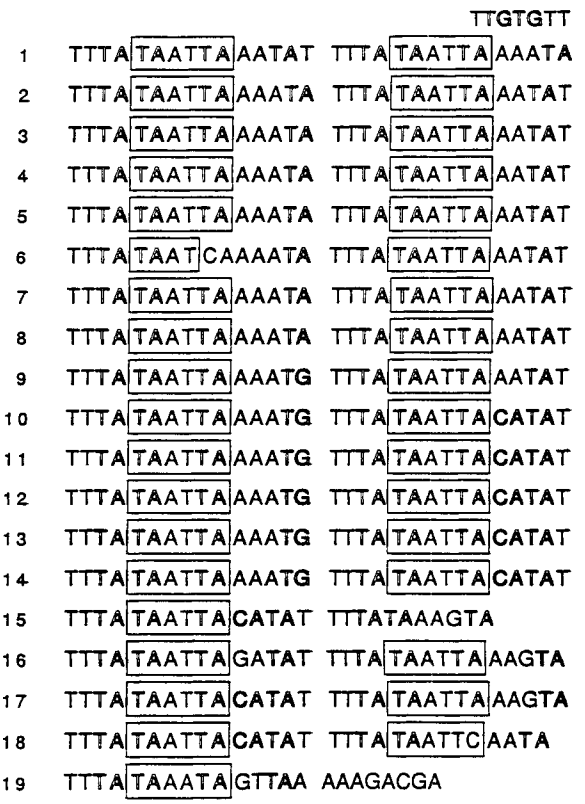

Fig. 1. The constitutive 3′ MAR of the human apolipoprotein B gene is a mosaic of TAAT, ATTA, TAAAT, ATTTTA, and TAAAAT motifs The sequence is from Levy-Wilson and Fortier [1989] but aligned The TAATTA motifs, core elements of homeotic protein recognition and binding sites, are boxed The TAATTA boxes are separated by the ATTTTA motif The TA, CA, and TG nucleotides which are characteristic of kinked DNA at a spacing of 0 5 or 1 0 helical turns between centers and which are overrepresented in protein binding sites on DNA (see text) are outlined This MAR is a 30-nucleotide sequence shown on a single line repeated 18 5 times Each 30mer is a dimer of a 15 nucleotide sequence TA + CA + TG account for 26 6% of all dinucleotides in this particular MAR compared with 18 75% expected for random DNA Thus, this particular MAR is a striking example of AT-rich MARs that contain putative homeodomain protein binding sites with a spacing of 1 5 helical turns and kinked DNA

on the same side of the double helix upon their binding to the AT-rich stretches. However, it should be stressed that this spacing of 8–12 nt is rather general and exceptions to this ought to be found. For example, single protein footprints on the DNA are usually 8–40 nucleotides long and their center to center spacing of footprints of neighboring proteins presumably interacting with each other about 22 nucleotides long (see Boulikas 1992b for other references). Examples are proteins interacting with DNA (and presumably with one another) in the 5′ locus control region (LCR) at the DNase I hypersensitive site

2 of the human β-globin gene complex [Talbot and Grosveld, 1991] and the six transcription factors which interact with the mouse albumin gene promoter and occupy a stretch of 133 bp of DNA [Lichtsteiner et al., 1987].

Rule 4: MARs may represent mass binding sites for protein transcription factors. This rule is merely a proposal. To date, there are no studies addressing this issue. A similar function, i.e., that of mass binding sites for proteins, has been proposed by Zuckerkandl and Villet [1988] for noncoding DNA sequences in general. The supramolecular DNA–protein structures remaining after treatment of nuclei with high salt or lithium diiodosalicylate to remove histones, and known to be enriched in topoisomerase II [Berrios et al., 1985; Earnshaw et al., 1985], poly(ADP-ribose) polymerase [Fakan et al., 1988], ARS consensus binding protein [Hofmann and Gasser, 1991], the ARBP protein [von Kries et al., 1991], matrin 3 [Belgrader et al., 1991], lamins [Boulikas, 1986], and a great number of other proteins, remind us of such mass binding sites and give some support to this proposal. To date, the distribution of protein factors among nuclear matrix and nonmatrix fractions has not been studied.

Rule 5: MARs harbor intrinsically curved DNA [Anderson, 1986; Sykes et al., 1988; Homberger, 1989; Boulikas and Kong, submitted]. Curved DNA motifs reflect a 10.4 nucleotide periodicity from the center to center of $(A)_n$ stretches; optimal curvature will be expected with four to five consecutive A residues, such as in the sequence $AAAAn_6AAAAn_7AAAA$; alternation between AAAA and TTTT stretches does not favor curving of DNA [Trifonov and Sussman, 1980; Trifonov, 1980, 1986, 1991; Wu and Crothers, 1984; Koo et al., 1986; Travers, 1987; Crothers et al., 1990]. In addition, motifs such as T_mA_n (but not A_nT_m) or GTTTAAAC (but not GAAATTTC) appear to have a pronounced bend into the major groove and a wide minor groove at the TA region; 5′ A_nT_m 3′ are essentially straight [Chuprina and Abagyan, 1988; Chuprina et al., 1991]. In addition to the AA and TT dinucleotides, the pairs AG, GA, GC, CG, CT and TC when phased correctly cause an appreciable curvature on the DNA, due to their large wedge values [Bolshoy et al., 1991]. Also, GGGGCCC repeats with a 10–11 nucleotide distance from their centers are curved [Brukner et al., 1991].

Curved DNA is a characteristic element of replication origins of yeast [Snyder et al., 1986], SV40 DNA origin region I [Ryder et al., 1986], and bacteriophage λ [Zahn and Blattner, 1985]. Intrinsically curved DNA has been suggested to play an important role in nuclear processes involving specific protein–DNA interactions such as recombination [Ross et al., 1982], transcription [Bossi and Smith, 1984] and replication [Zahn and Blattner, 1987]. Even if the DNA motif, which is recognized by a protein, is not intrinsically bent, it becomes so upon its interaction with its protein [e.g., Hatfull et al., 1987].

Rule 6: A class of MARs may harbor kinked DNA. Kinks may be generated on DNA by the presence of TG, CA or TA dinucleotides separated from each other by two to four or 9–12 nucleotides. Kinked DNA may contain $TA_{n3}TG_{n3}CA$ motifs with TA, TG, and CA occurring in any order [McNamara et al., 1990; Trifonov, personal communication]. That a class of nuclear matrix anchorage sites may harbor kinked DNA [Boulikas and Kong, in press] is a novel observation.

Trifonov and Brendel [1993] have shown that the CA, TA, and TG dinucleotides are overrepresented in DNA sequences which are protein recognition sites. About 420 protein binding sites were used to obtain these results. Following our proposal in Rule 4, we have observed that some, not all MARs from published studies or identified in our laboratory may display an unusual richness in TA, TG, and CA (see Fig. 1).

Rule 7: We propose that a class of MAR motifs may be flanked at one or both sides by DNA sequences that can adopt left-handed or triple helical structures, when torsional strain is induced on the DNA. This proposal is based on a theoretical search of MARs over 77.3 kb of the human β-globin gene complex applying the rules which have been established as characteristic of MARs. Potential MAR sites in the β-globin gene complex were flanked by potential Z- or H-DNA stretches [Boulikas, in press]. Left-handed or Z-DNA motifs include alternating purines/pyrimidines such as TG, TA, CA, or CG repeats [Johnston and Rich, 1985]. Triple-helical or H-DNA structures are formed by DNA stretches containing polypurines on one strand and polypyrimidines on the other and may include poly(A), poly(G), poly(C), poly(T), poly(GA), or poly(CT) with a mirror symmetry [e.g., Johnston, 1988]. Several func-

tions have been suggested for the interspersed $(TG)_n$ microsatellite including those of a hotspot of recombination (see Boulikas, 1991, 1992a). Our present proposal points out that an additional function of $(TG)_n$, as well as of $(GA)_n$ and $(GC)_n$, might be to delineate the borders of matrix anchorage sites.

Rule 8: MARs may contain palindromic structures. Palindromic sequences are able to convert into cruciform structures upon introduction of torsional strain on the DNA [Panayotatos and Wells, 1981]. Many origins of replication contain palindromic sequences. These include viral DNA, such as SV40 [Hay and DePamphilis, 1982] and polyoma virus [Hendrickson et al., 1987], yeast ARS [Boulikas, in preparation], and mammalian ORI sequences [Frappier and Zannis-Hadjopoulos, 1987; Igushi-Ariga et al., 1988]. These observations together with numerous findings vigorously showing that origins of replication are attached to the nuclear matrix [e.g., Pardoll et al., 1980; Aelen et al., 1983; Razin et al., 1986] support the idea that palindromic sequences may indeed be characteristic structures of MARs.

Rule 9: Transcriptional enhancers may be MARs. A significant number of studies have indeed shown that MAR sequences are located near or at enhancer sites [e.g., Cockerill and Garrard, 1986a]. In addition, MAR sequences themselves were shown to act as transcriptional enhancers in experiments involving cells in culture and transgenic animals [Xu et al., 1989; Blasquez et al., 1989a; Stief et al., 1989; Phi-Van et al., 1990; Klehr et al., 1991].

Rule 10: A class of MAR sequences may contain short (5–50 nt) repeats of CT-rich stretches alternating with 5–50 nucleotide-long repeats of GA-rich stretches. This type of MAR has been found to the 3′ side of the Ha-ras oncogene [Boulikas et al., 1992] and in cloned and sequenced matrix anchorage region fragments from unknown genes [Boulikas and Kong, in press].

Rule 11: A class of MARs may be enriched in TG/CA motifs. This type of MAR has been identified at the 3′ untranslated region of the poly(ADP-ribose) polymerase gene and in two fragments isolated through random cloning of MARs [Boulikas and Kong, in press]. TG motifs are 7–10 nt long and are best represented by TGTTTTG, TGTTTTTTG, and TTTTGGGG. These motifs are very abundant in the 3′ UTR of a great number of other genes, seem to be signals at the recombination sites of immunoglobin genes [Hesse et al., 1989] and compose chromosome telomeres [Gray et al., 1991].

Rule 12: Very few *Alu* repeats are proposed to be attached to the nuclear matrix. Such sequences may only comprise those members of *Alu,* which contain a sufficiently high number of ATTA, ATTTA, and other protein binding sequences properly distanced between their centers. This idea suggests that the discrepancy on whether *Alu* sequences may or may not function as origins of replication in primate cells [Johnson and Jelinek, 1986; Rao et al., 1990] and may or may not have MAR activity [Small et al., 1982] might depend on the presence of ATTA, ATTTA, and protein factor motifs in their sequences.

Preliminary data indicate that only a fraction of the 900,000 total *Alu* sequence repeats, which are dispersed in the human genome, are expected to function as matrix attachment sites (Boulikas and Jurka, unpublished observations). Of the 1825 *Alu* sequences in the Genbank (as of September 1991), there were: 1633 containing ATTA; 2403, TAAT; 99, ATTTA; 532, TAAAT; 37, ATTTTA; 306, TAAAAT. Only three out of the 1825 *Alu* sequences had five ATTA motifs, six had four ATTA and nineteen had three ATTA motifs. We are searching for complete homeodomain protein and other factor binding sites in *Alu* sequences in order to identify the best *Alu* candidates for ORI, MAR, and enhancer activity in those *Alu* sequences whose primary structure and genomic location is known.

Rule 13: DNase I-hypersensitive sites might be diagnostic of MARs. Indeed, the data of Käs and Laemmli [1992] suggest that the DNase I-hypersensitive sites in the *Drosophila* histone gene repeat coincide with topoisomerase II cleavage sites and the MAR sector in the H1–H3 intergenic region. This proposal can be tested by correlating the great number of the mapped DNase I–HS sites in the literature with data on identification of new MAR sequences.

TRENDS AND CONCLUSIONS

The rules that are systematically set in this study might help to define noncoding sequences of other genes as potential MARs. This knowledge is important for a better understanding of the function of enhancers and their varying

location with respect to the coding region of genes. Recent studies have shown that at least 10 kb of flanking regions of genes are essential for their correct developmental expression in transgenic experiments; such sequences might include MARs. Indeed, MAR sequences may fall in proximal and distal sites flanking the 5′ and 3′ ends of genes (Phi-Van and Strätling, 1988; Bode and Maass, 1988; Levy-Wilson and Fortier, 1989), but also in introns (Cockerill and Garrard, 1986a; Käs and Chasin, 1987; Sykes et al., 1988), in the intergenic region between related genes [Mirkovitch et al., 1984], in a GC-rich promoter region [Marilley and Gassend-Bonnet, 1989], and in the 3′ untranslated region of genes [Boulikas and Kong, in press]. Such sequences may represent mass binding sites for regulatory nonhistone proteins. MARs harbor potential origins of replication and might play a dynamic role in the developmental expression of genes and in the timing and order of activation of different ORIs during development and cell cycle [see Boulikas, 1992b]. Thus, the prediction of MAR sequences along a stretch of a cloned gene might help our understanding of control mechanisms governing transcription and replication during development.

The project of mapping the human genome makes use of sequence markers that help define the location of the various genes along the DNA molecule. Such markers include Alu sequences and other repetitive elements [Brooks-Wilson et al., 1990], restriction fragment length polymorphisms [Botstein et al., 1980], and some polymorphic microsatellites [e.g., Weber and May, 1989; Baron et al., 1992]. We propose that MAR sequences can provide excellent markers for mapping the human genome. Markers used in genome mapping need to be polymorphic, very common (10^5–10^6 copies per mammalian genome) and evenly dispersed, permitting the establishment of high-resolution maps of the entire genome [e.g., Dietrich et al., 1992]. One major class of MAR sequences were inferred to be polymorphic because of the unusual stretches they contain such as Z-DNA, palindromes, AT-rich DNA, repetitive motifs, etc. [see Boulikas, 1992c,d]. Such stretches of DNA might be damaged at higher rates by mutagens, might be refractory to repair enzymes (like Z-DNA) and susceptible to a higher rate of errors during DNA replication than average sequences, e.g., by formation of slippage structures [see Boulikas, 1992c,d].

Large-scale sequencing of MARs from individual chromosomes is an easily realizable project proposed to aid the Human Genome Project. Indeed, only about 5,000 MARs are estimated on the average per human chromosome each MAR having an average length of 500 bp. On the basis of a large enough number of MAR sequences we expect to gain insights into the structuring of genomes into domains. Such results are expected to have notable implications with respect to our knowledge of the control of initiation of DNA replication and the nature of transcriptional enhancers.

ACKNOWLEDGMENTS

The valuable input of Emile Zuckerkandl, Ed Trifonov, Jerzy Jurka, and Bernhard Hirt is greatly appreciated. Special thanks to Dawn Brooks for help in preparing the manuscript.

REFERENCES

Aelen JMA, Opstelten RJG, Wanka F (1983): Organization of DNA replication in *Physarum polycephalum*. Attachment of origins of replicons and replication forks to the nuclear matrix. Nucl Acids Res 11:1181–1195.

Amati BB, Gasser SM (1988): Chromosomal ARS and CEN elements bind specifically to the yeast nuclear scaffold. Cell 54:967–978.

Amati B, Gasser SM (1990): *Drosophila* scaffold-attached regions bind nuclear scaffolds and can function as ARS elements in both budding and fission yeasts. Mol Cell Biol 10:5442–5454.

Anderson JN (1986): Detection, sequence patterns and function of unusual DNA structures. Nucl Acids Res 14:8513–8533.

Belgrader P, Dey R, Berezney R (1991): Molecular cloning of matrin 3. A 125-kiloDalton protein of the nuclear matrix contains an extensive acidic domain. J Biol Chem 266: 9893–9899.

Berezney R (1991): The nuclear matrix: A heuristic model for investigating genomic organization and function in the cell nucleus. J Cell Biochem 47:109–123.

Berezney R, Coffey DS (1975): Nuclear protein matrix: Association with newly synthesized DNA. Science 189:291–293.

Berrios M, Osheroff N, Fisher PA (1985): In situ localization of DNA topoisomerase II, a major polypeptide component of the *Drosophila* nuclear matrix fraction. Proc Natl Acad Sci USA 82:4142–4146.

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

Blasquez VC, Sperry AO, Cockerill PN, Garrard WT (1989b): Protein:DNA interactions at chromosomal loop attachment sites. Genome 31:503–509.

Bode J, Maass K (1988): Chromatin domain surrounding the human interferon-β gene as defined by scaffold-attached regions. Biochemistry 27:4706–4711.

Bodnar JW (1988) A domain model for eukaryotic DNA organization A molecular basis for cell differentiation and chromosome evolution J Theor Biol 132 479–507

Bolshoy A, McNamara P, Harrington RE, Trifonov EN (1991) Curved DNA without A-A Experimental estimation of all 16 DNA wedge angles Proc Natl Acad Sci USA 88 2312–2316

Baron B, Poirier C, Simon-Chazottes D, Barnier C, Guenet J-L (1992) A new strategy useful for rapid identification of microsatellites from DNA libraries with large size inserts Nucl Acids Res 20 3665–3669

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphism Am J Hum Genet 32 314–331

Bossi L, Smith DM (1984) Conformational change in the DNA associated with an unusual promoter mutation in a tRNA operon of *Salmonella* Cell 39 643–652

Boulikas T (1986) Protein–protein and protein–DNA interactions in calf thymus nuclear matrix using cross-linking by ultraviolet irradiation Biochem Cell Biol 64 474–484

Boulikas T (1987) Nuclear envelope and chromatin structure Int Rev Cytol Suppl 17 599–684

Boulikas T (1991) Relation between carcinogenesis, chromatin structure and poly(ADP-ribosylation) Anticancer Res 11 489–528

Boulikas T (1992a) Poly(ADP-ribosyl)ation, repair, chromatin and cancer Curr Persp Mol Cell Oncol 1 1–109

Boulikas T (1992b) Homeotic protein binding sites, origins of replication and nuclear matrix anchorage sites share the ATTA and ATTTA motifs J Cell Biochem 50 111–123

Boulikas T (1992c) Evolutionary consequences of preferential damage and repair of chromatin domains J Mol Evol 35 156–180

Boulikas T (1992d) Chromatin and nuclear matrix in development and in carcinogenesis A theory Int J Oncol 1 357–372

Boulikas and Kong (in press) Multitude of inverted repeats characterize a class of anchorage sites of chromatin loops to the nuclear matrix

Brooks-Wilson AR, Goodfellow PN, Povey S, Nevanlinna HA, de Jong PJ, Goodfellow PJ (1990) Rapid cloning and characterization of new chromosome 10 DNA markers by *Alu* element-mediated PCR Genomics 7 614–620

Brukner I, Jurukovski V, Konstantinovic M, Savic A (1991) Curved DNA without AA/TT dinucleotide step Nucl Acids Res 19 3549–3551

Buongiorno-Nardelli M, Micheli G, Carrì MT, Marilley M (1982) A relationship between replicon size and supercoiled loop domains in the eukaryotic genome Nature 298 100–102

Callan HG (1973) DNA replication in the chromosomes of eukaryotes Cold Spring Harbor Symp Quant Biol 38 195–203

Carri MT, Micheli G, Graziano E, Pace T, Buongiorno-Nardelli M (1986) The relationship between chromosomal origins of replication and the nuclear matrix during the cell cycle Exp Cell Res 164 426–436

Chuprina VP, Abagyan RA (1988) Structural basis of stable bending in DNA containing A_n tracts Different types of bending J Biomol Struct Dyn 6 121–138

Chuprina VP, Fedoroff OY, Reid BR (1991) New insights into the structure of A_n tracts and B -B bends in DNA Biochemistry 30 561–568

Ciejek EM, Tsai M-J, O'Malley BW (1983) Actively transcribed genes are associated with the nuclear matrix Nature 306 607–609

Cockerill PN, Garrard WT (1986a) 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, Garrard WT (1986b) Chromosomal loop anchorage sites appear to be evolutionarily conserved FEBS Lett 204 5–7

Cockerill PN, Yuen M-H, 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

Crothers DM, Haran TE, Nadeau JG (1990) Intrinsically bent DNA J Biol Chem 265 7093–7096

Dietrich W, Katz H, Lincoln SE, Shin H-S, Friedman J, Dracopoli NC, Lander ES (1992) A genetic map of the mouse suitable for typing intraspecific crosses Genetics 131 423–447

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

Dworetzky SI, Fey EG, Penman S, Lian JB, Stein JL, Stein GS (1990) Progressive changes in the protein composition of the nuclear matrix during rat osteoblast differentiation Proc Natl Acad Sci USA 87 4605–4609

Earnshaw WC, Halligan B, Cooke CA, Heck MMS, Liu LF (1985) Topoisomerase II is a structural component of mitotic chromosome scaffolds J Cell Biol 100 1706–1715

Fakan S, Leduc Y, Lamarre D, Brunet G, Poirier GG (1988) Immunoelectron microscopical distribution of poly(ADP-ribose) polymerase in the mammalian cell nucleus Exp Cell Res 179 517–526

Fakan S, Bernhard W (1971) Localisation of rapidly and slowly labelled nuclear RNA as visualized by high resolution autoradiography Exp Cell Res 67 129-141

Farache G, Razin SV, Targa FR, Scherrer K (1990) Organization of the 3 -boundary of the chicken α globin gene domain and characterization of a CR 1-specific protein binding site Nucl Acids Res 18 401–409

Fey EG, Bangs P, Sparks C, Odgren P (1991) The nuclear matrix Defining structural and functional roles CRC Crit Rev Euk Gene Express 1 127–143

Frappier L, Zannis-Hadjopoulos M (1987) Autonomous replication of plasmids bearing monkey DNA origin-enriched sequences Proc Natl Acad Sci USA 84 6668–6672

Gasser SM, Laemmli UK (1986a) The organization of chromatin loops Characterization of a scaffold attachment site EMBO J 5 511–518

Gasser SM, Laemmli UK (1986b) Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of *D melanogaster* Cell 46 521–530

Gray JT, Celander DW, Price CM, Cech TR (1991) Cloning and expression of genes for the *Oxytricha* telomere-binding protein Specific subunit interactions in the telomeric complex Cell 67 807–814

Greenstein RJ (1988) Constitutive attachment of murine erythroleukemia cell histone-depleted DNA loops to nuclear scaffolding is found in the β major but not the α1-globin gene DNA 7 601–607

Hall G Jr, Allen GC, Loer DS, Thompson WF, Spiker S (1991) Nuclear scaffolds and scaffold-attachment regions in higher plants Proc Natl Acad Sci USA 88 9320–9324

Hancock R, Boulikas T (1982) Functional organization in the nucleus Int Rev Cytol 79 165–214

Hatfull GF, Noble SM, Grindley NDF (1987) The γδ resolvase induces an unusual DNA structure at the recombinational crossover point Cell 49 103–110

Hay RT, DePamphilis ML (1982) Initiation of SV40 DNA replication in vivo Location and structure of 5 ends of DNA synthesized in the *ori* region Cell 28 767–779

He D, Nickerson JA, Penman S (1990) Core filaments of the nuclear matrix J Cell Biol 110 569–580

Hendrickson EA, Fritze CE, Folk WR, DePamphilis ML (1987) The origin of bidirectional DNA replication in polyoma virus EMBO J 6 2011–2018

Hesse JE, Lieber MR, Mizuuchi K, Gellert M (1989) V(D)J recombination A functional definition of the joining signals Genes Dev 3 1053–1061

Hofmann JF-X, Gasser SM (1991) Identification and purification of a protein that binds the yeast ARS consensus sequence Cell 64 951–960

Homberger HP (1989) Bent DNA is a structural feature of scaffold-attached regions in *Drosophila melanogaster* interphase nuclei Chromosoma 98 99–104

Igushi-Ariga SMM, Okazaki T, Itani T, Ogata M, Sato Y, Ariga H (1988) An initiation site of DNA replication with transcriptional enhancer activity present upstream of the c-*myc* gene EMBO J 7 3135–3142

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

Johnson EM, Jelinek WR (1986) Replication of a plasmid bearing a human Alu-family repeat in monkey COS-7 cells Proc Natl Acad Sci USA 83 4660–4664

Johnston BH (1988) The S1-sensitive form of $d(C\text{-}T)_n$ $d(A\text{-}G)_n$ Chemical evidence for a three-stranded structure in plasmids Science 241 1800–1804

Johnston BH, Rich A (1985) Chemical probes of DNA conformation Detection of Z-DNA at nucleotide resolution Cell 42 713–724

Kas E, Chasin LA (1987) Anchorage of the Chinese hamster dihydrofolate reductase gene to the nuclear scaffold occurs in an intragenic region J Mol Biol 198 677–692

Kas E, Laemmli UK (1992) In vivo topoisomerase II cleavage of the *Drosophila* histone and satellite III repeats DNA sequence and structural characteristics EMBO J 11 705–716

Kaufmann SH, Gibson W, Shaper JH (1983) Characterization of the major polypeptides of the rat liver nuclear envelope J Biol Chem 258 2710–2719

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

Koo H-S, Wu H-M, Crothers DM (1986) DNA bending at adenine thymidine tracts Nature 320 501–506

Leibovitch SA, Leibovitch MP, Hillion J, Kruh J, Harel J (1983) A destabilized DNA conformation associated with tightly bound nuclear proteins in active genes of rat myoblast Nucl Acids Res 11 4035–4047

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

Lichtsteiner S, Wuarin J, Schibler U (1987) The interplay of DNA-binding proteins on the promoter of the mouse albumin gene Cell 51 963–973

Marilley M, Gassend-Bonnet G (1989) Supercoiled loop organization of genomic DNA A close relationship between loop domains, expression units and replicon organization in rDNA in *Xenopus laevis* Exp Cell Res 180 475–489

McCready SJ, Cook PR (1984) Lesions induced in DNA by ultraviolet light are repaired at the nuclear cage J Cell Sci 70 189–196

McNamara PT, Bolshoy A, Trifonov EN, Harrington RE (1990) Sequence-dependent kinks induced in curved DNA J Biomol Struct Dyn 8 529–538

Mielke C, Kohwi Y, Kohwi-Shigematsu T, Bode J (1990) Hierarchical binding of DNA fragments derived from scaffold-attached regions Correlation of properties in vitro to function in vivo Biochemistry 29 7475–7485

Mirkovitch J, Gasser SM, Laemmli UK (1988) Scaffold attachment of DNA loops in metaphase chromosomes J Mol Biol 200 101–109

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

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

Mullenders LHF, van Kasteren van Leeuwen AC, van Zeeland AA, Natarajan AT (1988) Nuclear matrix associated DNA is preferentially repaired in normal human fibroblasts, exposed to a low dose of ultraviolet light but not in Cockayne's syndrome fibroblasts Nucl Acids Res 16 10607–10623

Panayotatos N, Wells RD (1981) Cruciform structures in supercoiled DNA Nature 289 466–470

Paulson JR, Laemmli UK (1977) The structure of histone-depleted metaphase chromosomes Cell 12 817–828

Pardoll DM, Vogelstein B, Coffey DS (1980) A fixed site of DNA replication in eucaryotic cells Cell 19 527–536

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

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

Rao BS, Zannis-Hadjopoulos M, Price GB, Reitman M, Martin RG (1990) Sequence similarities among monkey *ori*-enriched (*ors*) fragments Gene 87 233–242

Razin SV, Kekelidze MG, Lukanidin EM, Scherrer K, Georgiev GP (1986) Replication origins are attached to the nuclear skeleton Nucl Acids Res 14 8189–8207

Robinson SI, Nelkin BD, Vogelstein B (1982) The ovalbumin gene is associated with the nuclear matrix of chicken oviduct cells Cell 28 99–106

Ross W, Shulman M, Landy A (1982) Biochemical analysis of *att*-defective mutants of the phage lambda site-specific recombination system J Mol Biol 156 505–529

Ryder K, Silver S, DeLucia AL, Fanning E, Tegtmeyer P (1986) An altered DNA conformation in origin region I is

a determinant for the binding of SV40 large T antigen Cell 44 719–725
Small D, Nelkin B, Vogenstein B (1982) Nonrandom distribution of repeated DNA sequences with respect to supercoiled loops and the nuclear matrix Proc Natl Acad Sci USA 70 5911–5915
Snyder M, Buchman AR, Davis RW (1986) Bent DNA at a yeast autonomously replicating sequence Nature 324 87–89
Spradling A, Orr-Weaver T (1987) Regulation of DNA replication during *Drosophila* development Annu Rev Genet 21 373–403
Stief A, Winter DM, Stratling WH, Sippel AE (1989) A nuclear DNA attachment element mediates elevated and position-independent gene activity Nature 341 343–345
Sykes RC, Lin D, Hwang SJ, Framson PE, Chinault AC (1988) Yeast ARS function and nuclear matrix association coincide in a short sequence from the human HPRT locus Mol Gen Genet 212 301–309
Talbot D, Grosveld F (1991) The 5′ HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites EMBO J 10 1391–1398
Travers AA (1987) DNA bending and nucleosome positioning Trends Biochem Sci 12 108–112
Trifonov EN (1986) Curved DNA CRC Crit Rev Biochem 19 89–106
Trifonov EN (1980) Sequence-dependent deformational anistropy of chromatin DNA Nucl Acids Res 8 4041–4053
Trifonov EN (1991) DNA in profile Trends Biochem Sci 16 467–470
Trifonov EN, Brendel V (1993) Gnomic A Dictionary of Genetic Codes 2nd Ed New York VCH
Trifonov EN, Sussman JL (1980) The pitch of chromatin DNA is reflected in its nucleotide sequence Proc Natl Acad Sci USA 77 3816–3820
van der Velden HMW, Van Willigen G, Wetzels RHW, Wanka F (1984) Attachment of origins of replication to the nuclear matrix and the chromosomal scaffold FEBS Lett 171 13–16
Vaughn JP, Dijkwel PA, Mullenders LHF, Hamlin JL (1990) Replication forks are associated with the nuclear matrix Nucl Acids Res 18 1965–1969
Villarreal LP (1991) Relationship of eukaryotic DNA replication to committed gene expression General theory for gene control Microbiol Rev 55 512–542
von Kries JP, Buhrmester H, Stratling WH (1991) A matrix/scaffold attachment region binding protein Identification, purification and mode of binding Cell 64 123–135
Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction Am J Hum Genet 44 388–396
Wu H-M, Crothers DM (1984) The locus of sequence-directed and protein-induced DNA bending Nature 308 509–513
Xu M, Hammer RE, Blasquez VC, Jones SL, Garrard WT (1989) Immunoglobulin κ gene expression after stable integration II Role of the intronic MAR and enhancer in transgenic mice J Biol Chem 264 21190–21195
Zahn K, Blattner FR (1985) Sequence-induced DNA curvature at the bacteriophage λ origin of replication Nature 317 451–453
Zahn K, Blattner FR (1987) Direct evidence for DNA bending at the lambda replication origin Science 236 416–422
Zuckerkandl E, Villet R (1988) Generation of high specificity of effect through low-specificity binding of proteins to DNA FEBS Lett 231 291–298