Journal of Cellular Biochemistry 53:1–12 (1993)

ARTICLES

Multitude of Inverted Repeats Characterizes a Class of Anchorage Sites of Chromatin Loops to the Nuclear Matrix

Teni Boulikas and C.F. Kong

Institute of Molecular Medical Sciences, Palo Alto, California 94306

Abstract In order to understand the nature of DNA sequences that organize chromatin into domains or loops, we have cloned the nuclear matrix DNA (1.7% of the total DNA) from human myelogeneous leukemia cells in culture. Nuclear matrix is formed by interactions between specific stretches of DNA of about 0.1 to 5.0 kb with protein transcription factors, nuclear enzymes, and structural proteins. Nuclear matrix is believed to be the exclusive nuclear microenvironment in which initiation of DNA replication, transcription, and repair take place. The matrix attachment regions (MARs) of DNA have transcriptional enhancer activity, harbor the origins of replication of the human genome, and define the borders between neighboring chromatin loops. In this study we report the sequence of the human MAR fragment 19.2 of a size of 542 bp. Hum. MAR 19.2 is composed of TG-, CA-, CT-, and GA-rich blocks and shows 8 perfect and imperfect inverted repeats. Thus, we have identified a novel class of MARs with sequence characteristics divergent from the AT-rich class of MARs. The inverted repeats of the 19.2 sequence might be stabilized into their cruciform configuration by torsional strain and by specific transcription/replication protein factors. This MAR might function in the initiation of replication of the flanking chromatim domain and in the regulation of the transcriptional activity of the gene(s) that reside in this domain.

Key words: cloning, nuclear matrix, MARs, origins of replication, palindromes (inverted repeats), transcriptional enhancers, chromatin domains

MAR sequences (matrix attached regions) represent the attachment points of chromatin loops to the nuclear matrix [see Hancock and Boulikas, 1982; He et al., 1990; Berezney, 1991]. Thus MARs define the borders between chromatin domains, each harboring one gene or segment of a gene or groups of related genes [Mirkovitch et al., 1984; Cockerill and Garrard, 1986; Bode and Maass, 1988; Phi-Van and Strätling, 1988; Levy-Wilson and Fortier, 1989]. In addition to their structural involvement in the formation of the chromatin domain, MARs fill two of the most important functional roles: that of acting as sites of initiation of DNA replication [Berezney and Coffey, 1975; Pardoll et al., 1980; Carri et al., 1986; Vaughn et al., 1990] and that of acting as transcriptional enhancers [Blasquez et al., 1989; Xu et al., 1989]. Significant changes in the nuclear matrix protein pattern during osteoblast differentiation suggested that the nuclear matrix is involved in the developmental decision of cell type formation [Dworetzky et al., 1990]. The similarity of stretches of DNA in AT-rich MARs to the homeodomain protein recognition and binding site suggested that MARs are involved in differential gene expression, cell memory, and body plan formation in development [Boulikas, 1992a].

Inverted repeats (or palindromes or dyad symmetry motifs) can be detected by sequence analysis of the DNA, by electron microscopy, and from their rapid reassociation kinetics [e.g., Schmid et al., 1975]. The linear form of the inverted repeat may convert into the cruciform configuration, which could then act as initiation site for DNA replication [Hobom et al., 1979; Meijer et al., 1979]. Inverted repeats are found in replication origins (ORIs) of *Escherichia coli* [Meijer et al., 1979] and bacteriophages λ and ϕ80 [Hobom et al., 1979]. In addition, inverted repeats are found at the ORIs of mammalian viruses, including SV40 [Subramanian et al.,

Received December 21, 1992; revised April 6, 1993; accepted April 20, 1993.

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

1977; Deb et al., 1986], polyoma [Soeda et al., 1978; Prives et al., 1987], the human papovavirus BKV [Dhar et al., 1978], HSV1 [Stow and McMonagle, 1983; Weller et al., 1985], Epstein-Barr virus [Reisman et al., 1985], HSV2 [Lockshon and Galloway, 1986], and JC virus [Frisque, 1983]. Lin and Meyer [1987] have shown that DNA replication in the plasmid R1162 is initiated from two positions flanking a large (40-bp stem, 40-nt loop) inverted repeat, and have suggested that initiation is triggered from the cruciform structure. The cruciform configuration of the inverted repeat would be promoted and stabilized by the protein replication initiation complex [Lin and Meyer, 1987]. In addition, palindromes are found in the origins of replication of the monkey genome isolated and cloned by extrusion of newly synthesized DNA from replication forks (ors sequences) [Zannis-Hadjopoulos et al., 1985; Rao et al., 1990], the c-myc ORI [Iguchi-Ariga et al., 1988], and the myc and CAD genes origin of amplication [Ford and Fried, 1986]. The potential to form a cruciform also seems to be essential for the function of the ORI of the ColE1 [Masukata and Tomizawa, 1984] and the ORI of the Brome Mosaic Virus [Pogue and Hall, 1992].

Monoclonal antibodies, directed against cruciforms occuring in ors sequences supposed to represent monkey origins of replication from unknown genes were shown to enhance DNA replication in permeabilized monkey cells [Zannis-Hadjopoulos et al., 1988]. This was interpreted to be a consequence of stabilization of the inverted repeats into their cruciform configuration by the antibody; such cruciforms were thought to be located at or near replication origins, resulting in multiple initiations of DNA replication from a single origin. Using monoclonal antibodies directed against cruciform and quantitative fluorescence flow cytometry, 3–5 $\times$ 10^5 cruciforms/nucleus were estimated for monkey CV-1 and human colon adenocarcinoma SW48 cells throughout the S phase; however, no cruciform–like structures could be detected during G0, G2/M or in metaphase chromosomes [Ward et al., 1990, 1991]. An increase in the number of the single-stranded DNA-specific S1 nuclease cutting sites was found during the S-phase by Collins [1979] and by Collins and coworkers [1982]. These data support a model of a transient conversion of inverted repeats into cruciform structures during initiation of DNA replication.

A fundamental question in understanding MAR function concerns the type of DNA sequences used by eukaryotic nuclei to set the borders between chromatin domains while using the same sequences or nearby motifs to initiate replication of the domain and to enhance the transcriptional activity of the genes nested in this particular domain. In the present study we analyze the nature of the DNA sequences of a MAR fragment cloned from human myelogeneous leukemia cells. This DNA sequence identifies a novel type of MAR enriched in inverted repeats.

MATERIALS AND METHODS

Cell Culture

K562 human erythroleukemia cells (American Type Culture Collection) were cultured in DMEM medium supplemented with 7% fetal bovine serum and antibiotics as described [Boulikas, 1988].

Nuclei Isolation and Nuclear Matrix Preparation

Nuclei were isolated using the ultrarapid procedure of Boulikas [1988]. According to this method, cultured cells were lysed in 1% Triton X-100 and nuclei were collected by centrifugation through a 60% glycerol cushion at 13,000*g* for 4 min in Eppendorf tubes [Boulikas, 1988]. Nuclei isolated from 2.5 $\times$ 10^7 cells were digested with 1,000 units micrococcal nuclease (MNase) at 37°C for 10 min in 0.5 ml 50 mM Tris · HCl, pH 7.5, 25 mM KCl, 1 mM $CaCl_2$, 4 mM $MgCl_2$. Nuclear matrices were isolated by an adaptation of the methods of Cockerill and Garrard [1986] and Boulikas [1986]. All procedures were at 0°C in the continuous presence of 0.1 mM N-p-tosyl-L-lysine chloromethyl ketone to minimize proteolysis and in Eppendorf tubes using short-time spins to avoid nuclear matrix disassembly. According to this method, digested nuclei were centrifuged in Eppendorf tubes for 2 min in a Microfuge (13,000*g*) to yield a supernatant fraction S1, enriched in active chromatin [Rose and Garrard, 1984; Garrard et al., 1988]. The digested nuclei pellet (P1) was then lysed in 0.5 ml 2 mM EDTA, 3 mM Tris, pH 7.5, followed by a centrifugation at 13,000*g* for 3 min. The supernatant fraction, S2, contained the bulk of nucleosomes mainly as mononucleosome particles due to the extensive treatment with MNase. The pellet fraction, P2, containing the residual of the chromatin loops and the nuclear matrix,

was resuspended in 0.5 ml 2M NaCl, 2 mM EDTA, 10 mM Tris, pH 7.5, and immediately centrifuged (microfuge 3 min) to separate the residual histones from the nuclear matrix (NM) fraction. The DNA content of the nuclear matrix fraction that can be monitored by the extent of digestion of nuclei by MNase [Boulikas, 1985] represented 1.7% of total DNA and had fragments ranging in size from 0.1 to 5.0 kb. The nuclear matrix fraction was resuspended in 0.2 ml 5M urea. Then 0.2 ml 2% SDS was added and the nuclear matrix proteins were removed with proteinase K (0.1 mg/ml), at 37°C for 16 h. DNA was extracted twice with phenol/chloroform and precipitated with 70% ethanol. NM DNA preparations were size-fractionated on low-melting agarose gels for cloning.

Agarose Gel Electrophoresis

The various nuclear fractions were adjusted to 1% SDS, 10% glycerol, 0.05% bromophenol blue in TBE buffer (89 mM Tris, 89 mM boric acid, pH 8.5, 1 mM EDTA) and loaded directly onto 1% low-melting agarose gels containing 10 μg/ml ethidium bromide, 0.1% SDS in TBE buffer. The electrophoresis buffer was TBE buffer containing 0.1% SDS.

Cloning of the Nuclear Matrix DNA

Pieces from the low-melting agarose gel containing 100–700 bp or 0.7–5.0 kbDNA fragments and about 3 μg of nuclear matrix DNA were excised under low-energy UV light. Exposure to UV was minimized to avoid formation of cyclobutane dimers and (6–4) photoproducts on the DNA that cause C → T transitions during replication [Vrieling et al., 1989; see Boulikas, 1992b].

One volume of TE buffer (10 mM Tris, pH 8.0, 1 mM EDTA) was added to the gel piece, melted at 65°C, extracted once with phenol, once with phenol/chloroform, once with chloroform, and precipitated by the addition of 2 vol ethanol and 0.1 vol of 3M sodium acetate at −70°C overnight in Eppendorf tubes. DNA was collected by centrifugation at 0°C for 20 min (13,000*g*, microfuge), washed in 70% ethanol and resuspended in 20 μl TE. In some experiments the DNA fragments were phosphorylated at their 5′ ends using T4 polynucleotide kinase (BRL) and ATP as described [Wu et al., 1987]. The 3′ recessed ends generated by MNase [Sollner-Webb et al., 1978] were filled in with 1 μl (4 units) Klenow fragment of DNA polymerase (Amersham) in TE buffer, extracted with 1 vol phenol, once with phenol/chloroform and ethanol-precipitated. The DNA pellet was dried in a Speedvac for 5 min, resuspended in 7 μl H_2O, 1 μl 10× ligation buffer, 1 μl Bluescript vector cut with *Eco*RV, and 1 μl T4 ligase (USB) and incubated at 14°C for 16 h. DNA was directly transformed in JM109 competent *E. coli* cells and plated on LB/agar with ampicillin (100 μg/ml), X-gal, and IPTG. Plates were incubated at 37°C overnight and screened for white colonies. Individual colonies were picked and grown in 3 ml LB + ampicillin (100 μg/ml). Plasmid minipreps digested with *Eco*RI + *Hind*III were analyzed for the size of the insert on 1% agarose gels. About 150 colonies were screened in total. About 90 colonies had an identical size of insert (Hum. MAR 19), and upon partial sequencing (at least five of them) were found to have identical sequences. About 10 clones were determined to represent different sequences, and sequencing was pursued.

DNA Sequencing

DNA sequencing was performed using the dideoxy method and the Sequenase II kit (USB). For this purpose, double-stranded DNA plasmid was denatured, mixed with universal or reverse primers (USB), and DNA strand synthesis was performed according to provider's instructions and to the method of Mierendorf and Pfeffer [1987]. Synthesized DNA was analyzed on gels (4% acrylamide, 50% w/v urea, 60 cm long, 0.4 mm thick) and run at 65 Watts for 2.5–3 h. Gels were dried on Whatman 3 MM filter paper and exposed for autoradiography for about 24 h at room temperature using X-Omat film.

RESULTS

We have isolated the nuclear matrix DNA (MAR) from human K562 cells comprising 1–2% of nuclear DNA, after removal of bulk polynucleosomes with micrococcal nuclease. The MAR DNA was cloned in *E. coli*. In this article, using random cloning of the nuclear matrix DNA, we report a novel class of MAR sequences that diverge from the known AT-rich class of MARs [e.g., Mirkovitch et al., 1984; Cockerill and Garrard, 1986; Levy-Wilson and Fortier, 1989]. The new class of MARs, represented by the fragment Hum. MAR 19.2, is composed of TG-rich stretches and of polypurine stretches (shown in bold face) alternating on the same strand of DNA with polypyrimidine stretches (outlined)

(Fig. 1). This human MAR fragment is 542 bp in length derived from a large MAR locus of about 3.6 kb. The TG-rich (or their complementary CA-rich) motifs shown in larger characters in Figure 1 resemble sequence motifs found in recombination sites, in telomeres, in transcription and replication protein factor recognition and binding sequences, and in the 3′ untranslated region of a great number of genes (Table I).

The 19.2 sequence contains only a single ATTA motif (shown enlarged and underlined in Fig. 1), characteristic of homeodomain protein binding sites. The CCAATTAGCC stretch (position 92 in Fig. 1) that comprises the ATTA motif is a characteristic binding site of homeobox proteins with a glutamic acid at position 50 [Hanes and Brent, 1991] and represents a perfect recognition and binding site of the *Drosophila* engrailed protein occurring within intron 1 of the *engrailed* gene [Desplan et al., 1988]. Several ATTA motifs, properly spaced, signal the presence of a homeodomain protein-controlled regulatory region [e.g., Boulikas, 1992a]. Thus, if a homeodomain protein interacts in vivo with this stretch of DNA, it, no doubt, needs to interact with other transcription protein factors bound to vicinal sites on the 19.2 sequence.

The CAGAGATG motif occurs four times as a direct repeat. GA-rich blocks and TC-rich blocks on the same strand of DNA favor the occurrence of inverted repeats. We have manually screened the Hum. MAR 19.2 fragment for the presence of inverted repeats, shown in their cruciform structures in Figure 2. Some of these cruciforms, like those at positions 114, 158, 255, 354, and 440, might exist in vivo, provided that proteins able to stabilize these sequences in their hairpin structure are present. Several laboratories have postulated the presence of such proteins interacting with the stem, with the loop, or with the base of the cruciform that might function in the initiation of DNA replication (see Introduction and Discussion).

We have shown that the Hum. MAR 19.2 fragment and all other fragments of the 3.6-kb Hum. MAR 19 locus can be crosscomplexed with nuclear matrix proteins in vitro to form insoluble complexes easily removed by centrifugation (not shown). This in vitro assay has been widely used to show, for example, specific interaction of a fragment from the mouse κ immunoglobulin gene with nuclear matrix [Cockerill and Garrard, 1986]. Under the same conditions of matrix protein/DNA ratio, only specific fragments of the human Ha-ras oncogene disappear from the solution (Boulikas and Spandidos, in preparation). All fragments of our MAR clones are expected to bind to nuclear matrix proteins from K562 cells in vitro.

Contradictory data were reported concerning the copy number of MAR sequences in the genome. Some studies have determined that highly repetitive DNA is enriched in the nuclear matrix whereas others found that the DNA of the nuclear matrix is in single copies [see Boulikas, 1992a and references cited therein]. For this purpose we have screened the 19.2 MAR sequence for the presence of Alu repeats, L1, and the newly discovered medium reiteration frequency repeats such as MER [Kaplan et al., 1991; Jurka et al., 1992] using a program written by Jerzy Jurka. This search showed that the Hum. MAR 19.2 fragment does not contain any high or medium reiteration repeats.

DISCUSSION

Cruciforms and MARs

Two independent lines of evidence suggest that cruciform structures might be highly enriched in MARs. First, HMG1, which specifically binds to the single-stranded loop region of the cruciform [Bianchi et al., 1989, 1992], is a component of the nuclear matrix [Ivanchenko and Avramova, 1992]. Second, origins of replication, characterized by an enrichment in cruciforms (see Introduction) are nested in matrix anchored regions [Aelen et al., 1983; Buongiorno-Nardelli et al., 1982]. However, there is one experiment arguing against an enrichment of cruciforms in MARs: in an attempt to detect cruciform structures in the nuclear matrix, Price and collaborators [Ward et al., 1991] have used DNase I digestion of nuclei treated with anticruciform monoclonal antibodies; the fraction of the DNA associated with the nuclear matrix was found to be resistant to DNase I, and it was thus concluded that cruciforms either are not present in the nuclear matrix or are buried in the scaffold remaining inaccessible to DNase I [Ward et al., 1991]. The present work, using random cloning of the nuclear matrix attached regions from human cells has identified a novel class of non-AT-rich MARs; these sequences are highly enriched in inverted repeats with the potential to form cruciform structures under torsional strain [Dayn et al., 1992]. Our data support the model according to which inverted repeats are highly enriched in the attachment regions of chroma-

Hum MAR 19.2

GGGAAAGCTT GCATGCCTGC AGACCCGAGA 30

CTGCCTAGGT CACAACGCCT CC AGCAGGCC 60

CAGAGTCTCC T GCCCTGGCC ACAGGCAGCT 90

CCC AATTAGC CCAGATGCAT CCT GGAGCAC 120
Dros. eng homeo site

AGGGCCTTAA AGG CTCC ATT CACCCACGCA 150

TGCATTCCTT TCCCTCC ACC AACGGAGTGA 180

GGGAATTCAC CCTACT CCAC ACC AGAGATG 210

TGGCCGGAAT GCAGAGATGA AGTCACTGAG 240

CACAGTAGGT ATGGTCTTAC CCTT GAGAAG 270

CTAGTGGGAA AGAGAGAGTT TGATTGGCAA 300

AGCACTTATC T ATTGGGTTC C AGGCAGGAA 330

CAAAGGCAGC ACTCT GCAGC AACCCCTGGT 360

GGATTTGGAT TGCCCAGGGT GGTCAAGTCT 390

GGCTTAGCAG TTGACAACAG ACTCATGAGA 420

TGCAGACCTA CTCTT AGGAA CCA CTCTCTC 450

CATATGGGCC AGAGAGACAT GTGTCAGCAC 480

GTACAGTGAG TAGAGACTGG TCACGTGCAT 510

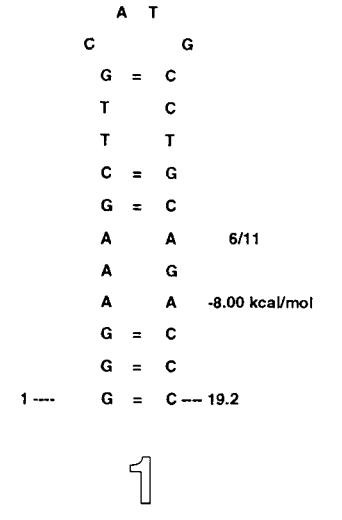
GAGACCTATA CTCTGTGTAG CTT GCTCTAT 540

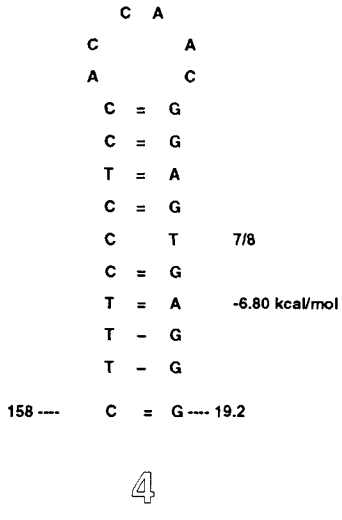
GG 542

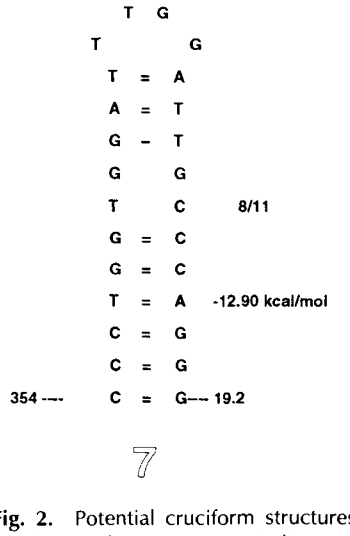
Fig. 1. Sequence of the human MAR fragment 19.2. Polypurine stretches of more than four nucleotides are in bold face; polypyrimidine stretches of more than four nucleotides are outlined. Three purines separated by a single pyrimidine from a block of three or more purines and vice versa are also considered. The TG boxes, i.e., stretches of TG (or its complementary CA) of more than six nucleotides, are enlarged. The CCAATTAGCC motif, recognition and binding site of the *Drosophila* engrailed homeotic protein, is enlarged and underlined (position 92).

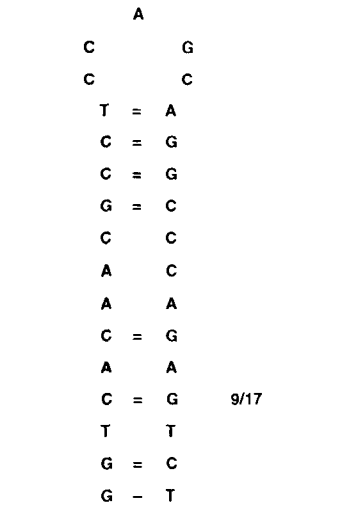
TABLE I. General Occurence and Function of TG Boxes

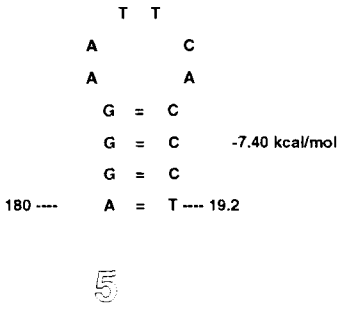
SEQUENCE	FUNCTION or OCCURENCE	REFERENCE
CACAAC CACCCAC CCACCAAC CCACACCA GTTTGaTTGG TTGGGTT AACAAA CAACCCCTGGTGG TTTGG GGGTGGT ACAACA	Hum. MAR 19.2 sequence	Boulikas and Kong, 1993 This study
GTGTCAC GTGACAC	Heptamer consensus sequence in immunoglobulin gene rearrangement	Hesse et al., 1989
TGTTTTTGG CCAAAAACA	Nonamer consensus sequence in immunoglobulin gene rearrangement	Hesse et al., 1989
$(TTTTGGGG)_n$	Telomeres of *Oxytricha nova* macronuclear chromosomes	Gray et al., 1991
GTTTT AAAAC	Core pentanucleotide recognized by the general transcription/ replication factor GFI in yeast	Dorsman et al., 1990
GTTTT or AAAAC	Core recognition sequence of FP4, zeste 3, GAGA2 and protein B within the *Drosophila Ultrabithorax* promoter.	Biggin and Tjian, 1988
ACACCCAAATATGGCGAC GTCGCCATATTTGGGTGT	chicken CBF	Boxer et al., 1989 Walsh, 1989
TGGTATGATTTTGTAATGGGG CCCCATTACAAAATCATACCA	rat DBP	Mueller et al., 1990
AACAAACACAAA TTTGTGTTTGTT	human H1TF1	Gallinari et al., 1989 van Wijnen et al., 1988
TRTTTGY RCAAAYA	rat HNF-5	Grange et al., 1991
GGGTGGG CCCACCC	human PuF	Postel et al., 1989
TCTGTGGTTAA TTAACCACAGA	mouse SEF-1	Thornell et al., 1988 Boral et al., 1989
GGGTGTGG CCACACCC	human TEF-2	Davidson et al., 1988
AAGTGTTTGC GCAAACACTT	human TGT3	Ben-Levy et al., 1989
AAGATAAAACC GGTTTTATCTT	human, mouse, rat, RVF	Yan and Hung, 1991
TTGTTTTG GTGTTGT TGTGGGG GGGGTTG	Frequently occurring in the 3' untranslated region of genes	e.g., Kurosaki et al., 1987
$(TGTGTG)_n$ or Z-DNA	Recombination hot spot	see Boulikas, 1991 for references
GCTGGTGG	chi recombination sequence in *E. coli*	Smith et al., 1981
5' TTATCCTAAACCA) 3' AATAGGATTTGGT	Palindromic sequence, hixC consensus recombination site of all known invertase systems related to Hin	Moskowitz et al., 1991

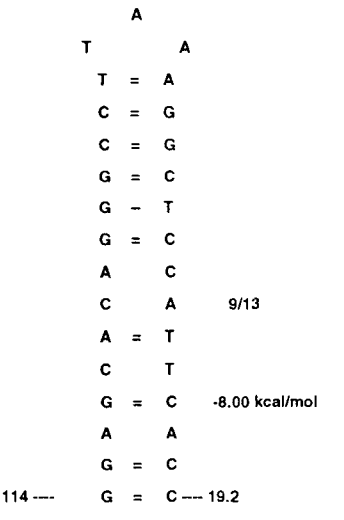


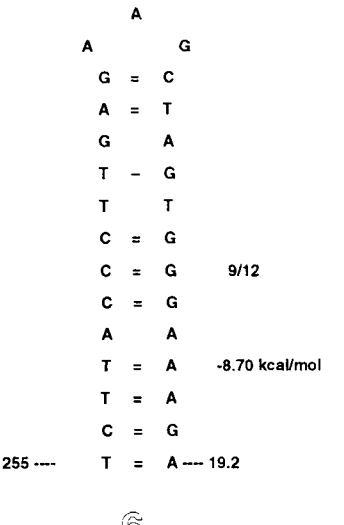












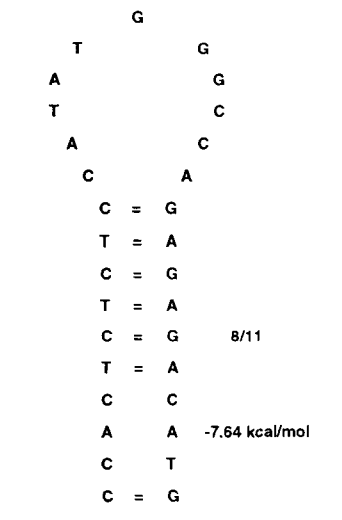


Fig. 2. Potential cruciform structures in Hum. MAR 19.2 fragment. The number to the 5′ start site indicates its position in the 19.2 human MAR fragment and the two numbers to the right hand side of the stem, the number of matches over the total number of nucleotides in the stem of the cruciform. The G-T pairs are shown with one dash line since they can form one hydrogen bond. Free energy for cruciform formation in kcal/mol in 1M Na^+, pH 7.0, 25°C were estimated using the DNASIS program (Hitachi, San Bruno, CA), which does not consider energy from TG pairs.

tin loops to the nuclear matrix and these sequences, in their cruciform configuration, play an important role in the initiation of DNA replication.

MARs as Markers in Genome Mapping

Mapping of genomes makes use of sequence markers that help define the location of the various genes along the DNA molecule, such as Alu sequences, restriction fragments with length polymorphisms, and some polymorphic microsatellites [e.g., Baron et al., 1992]. Such markers need to be polymorphic, very common (10^5–10^6 copies per mammalian genome) and evenly dispersed. We propose that MAR sequences can provide markers for mapping genomes. One major class of MAR sequences was inferred to be polymorphic because of the unusual stretches they contain, such as Z-DNA, palindromes, AT-rich DNA, and repetitive motifs [Boulikas, 1992c]. These unusual sequences are damaged at higher rates by mutagens, seem to be refractory to repair enzymes and prone to a higher rate of errors during DNA replication than average sequences [reviewed by Boulikas, 1992b,c]. Our estimate of MAR sequences is about 60,000 permanent MARs per nucleus, which together with the facultative class of MARs may amount to a total of 100,000 or more MARs per mammalian nucleus. This gives an average spacing of one MAR every 30,000 bp of DNA. Thus, MARs may fulfill the three criteria for their potential use as markers.

Random Cloning of MARs

We suggest that MARs are underrepresented in genomic libraries because of the relative resistance of their complexes with chromosomal proteins to solubilization by proteinase K. Our method for random cloning of the nuclear matrix DNA, undertaken on a large scale, is expected to identify the sequences that are used as borders between chromatin domains and are expected to harbor transcriptional enhancers and the origins of replication of the human genome.

Methods for MAR Identification

Two different methods have been devised to study the presence of MARs within a gene locus. According to the first method, isolated nuclei were treated with lithium diiodosalicylate to extract all histones and several nonhistone proteins, giving "nuclear halos" where extended chromatin loops puffing out of nuclei were still attached to the insoluble matrix. DNA was then cleaved with a restriction enzyme giving soluble loop fragments ($\sim$75% of DNA) and insoluble fragments ($\sim$25% of DNA) attached to the nuclear scaffold. DNA fragments from the two fractions were electrophoresed in equal amounts on agarose gels and blotted with specific gene probes [Mirkovitch et al., 1984; Cockerill and Garrard, 1986; Jarman and Higgs, 1988]. However, prolonged incubation of "nuclear halos" at 37°C for restriction enzyme digestion may cause partial disassembly of nuclear matrix. Also, the restriction enzyme digestion may often be incomplete; furthermore, some restriction sites might be "masked" by matrix proteins bound to them. Furthermore, the fragments remaining bound to the nuclear matrix or scaffold are rather long and therefore do not immediately reveal the actual attachment sites of the DNA to the matrix.

According to the second method, a mixture of DNA restriction fragments from a cloned gene were incubated with isolated nuclear matrix or scaffold proteins and the bound DNA fragments were separated from the unbound DNA by a simple centrifugation step [Cockerill and Garrard, 1986]. The second method also suffers from the fact that entire restriction fragments of DNA, rather than the actual attachment sites, are identified as MARs. The actual MAR sites were proposed to include binding sites of protein transcription factors [Boulikas, 1992a,c].

Our method does circumvent the problems associated with the other methods for MAR identification. First, micrococcal nuclease (MNase), an enzyme able to digest DNA in nuclei to a very high extent, has been used. This nuclease attacks linker DNA between nucleosomes, initially giving a nucleosome ladder. The precise structure of the nuclear matrix, especially at its borders with nucleosomes, is unknown. However, our experiments on nuclear matrix [Boulikas, 1985; this study] suggest that a protection similar to that conferred by nucleosomal histones to DNA, is conferred by nuclear matrix protein to DNA toward digestion by MNase. Thus, using MNase and monitoring the extent of digestion, we can end up with a small percentage of the total DNA remaining in the insoluble NM, representing about 0.5–2% of the total DNA. This approach may reveal the actual attachment points of the chromatin loops to the

matrix. Second, the time of nuclease digestion used here is short (37°C, 10 min), compared to that of other methods (3 h, 37°C, in Mirkovitch et al. [1984]) and avoids nuclear matrix disassembly into protein and DNA components. Third, the nonspecificity of MNase toward DNA sequences, which is in contrast with that of restriction endonucleases, can reveal the true nature of the bases of chromatin loops. MNase does display an initial preference for AT-rich sequences [Hörz and Altenburger, 1981] and active chromatin fractions [Bloom and Anderson, 1978], but we believe that this property of the enzyme does not substantially interfere with its ability to reveal the true matrix anchoring sites in nuclei.

Types of Nuclear Matrix Structures

Two types of nuclear matrix structures have been described. The first is the peripheral matrix, lining the interior of the nuclear envelope, with lamins as principal components, apparently mediating 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. The second type is an internal nuclear matrix component containing thick, highly branched polymorphic fibers that 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 shells of nuclei depleted of their internal nuclear matrix and composed only of nuclear pore–lamina complexes [Kaufmann et al., 1983]. This finding suggests that disulfide bonds stabilize the internal matrix. We do not know whether the non-AT-rich type of MAR described here arises from the peripheral or from the internal nuclear matrix.

Types of MAR Activity

Two types of MAR activities have been described with respect to their functional role: The permanent or constitutive MAR activity, present in all cell types, and the facultative or transient MAR activity that relates to the transcriptional or repair activities of a cell and that appears, for example, in active or damaged genes but is absent from the same gene in a cell type not expressing or repairing this gene [Boulikas, 1992a,b]. The MARs described here were derived from human chronic myelogenous leukemia K562 cells that have been dedifferentiated [Lozzio et al., 1981] and are thus expected to have a higher number of active origins of replication with a smaller replicon size than normal human lymphocytes [Hatton et al., 1988]. Consequently, since MARs have ORI activity [e.g., Vaughn et al., 1990; Boulikas, 1992a], a greater number of MARs per nucleus are expected to be found in K562 cells than in normal lymphocytes. It remains to be determined whether or not the Hum.MAR19.2 fragment described here, which is recognized by matrix proteins from K562 cells, is also recognized by matrix proteins from normal human lymphocytes. If it is, then this sequence ought to be classified as a constitutive MAR. If not, it should be classified as a facultative MAR specific to the dedifferentiated, tumor state.

Sequence Characteristics of MARs

MARs may also be classified into at least three different types with respect to their DNA sequence characteristics: (1) most MAR sequences identified in other laboratories fall into the AT-rich class of MARs [e.g., Levy-Wilson and Fortier, 1989; Blasquez et al., 1989b; reviewed by Boulikas, 1992a]; (2) one example of a GC-rich type of MAR has been identified in the promoter region of the ribosomal RNA gene of *Xenopus* [Marilley and Gassend-Bonnet, 1989]; and (3) our present study identifies a third type of MAR sequence containing homopurine and homopyrimidine blocks as well as TG (or its complementary CA) -rich blocks. This third type of MARs is also enriched in inverted repeats. The potential role of this MAR fragment as a human origin of replication, its chromosomal location, copy number and the particular human gene it is linked to are under investigation.

ACKNOWLEDGMENTS

Special thanks to Jerzy Jurka, Emile Zuckerkandl, Maria Zannis-Hadjopoulos, and Andrea Todd for their valuable input and to Dawn Brooks and Michael Ramsey-Perez for typing 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.

Baron B, Poirier C, Simon-Chazottes D, Barnier C, Guénet 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.

Ben-Levy R, Faktor O, Berger I, Shaul Y (1989): Cellular factors that interact with the Hepatitis B virus enhancer. Mol Cell Biol 9:1804–1809.

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

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.

Bianchi ME, Beltrame M, Paonessa G (1989): Specific recognition of cruciform DNA by nuclear protein HMG1. Science 243:1056–1059.

Bianchi ME, Falciola L, Ferrari S, Lilley DMJ (1992): The DNA binding site of HMG1 protein is composed of two similar segments (HMG boxes), both of which have counterparts in other eukaryotic regulatory proteins. EMBO J 11:1055–1063.

Biggin MD, Tjian R (1988): Transcription factors that activate the *Ultrabithorax* promoter in developmentally staged extracts. Cell 53:699–711.

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.

Bloom KS, Anderson JN (1978): Fractionation of hen oviduct chromatin into transcriptionally active and inactive regions after selective micrococcal nuclease digestion. Cell 15:141–150.

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

Boral AL, Okenquist SA, Lenz J (1989): Identification of the SL3-3 virus enhancer core as a T-lymphoma cell-specific element. J Virol 63:76–84.

Boulikas T (1985): Interaction of histone H1 with oligonucleosomes in vitro provides a convenient method for isolating mononucleosomes. Can J Biochem Cell Biol 63:1022–1032.

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 (1988): At least 60 ADP-ribosylated variant histones are present in nuclei from dimethylsulfate-treated and untreated cells. EMBO J 7:57–67.

Boulikas T (1992a): Homeotic protein binding sites, origins of replication and nuclear matrix anchorage sites share the ATTA and ATTTA motifs. J Cell Biochem 50:1–13.

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

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

Boxer LM, Prywes R, Roeder RG, Kedes L (1989): The sarcomere actin CArG-binding factor is indistinguishable from the *c-fos* serum response factor. Mol Cell Biol 9:515–522.

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

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.

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

Collins JM (1979): Transient structure of replicative DNA in normal and transformed human fibroblasts. J Biol Chem 254:10167–10172.

Collins JM, Glock MS, Chu AK (1982): Nuclease S_1 sensitive sites in parental deoxyribonucleic acid of cold- and temperature-sensitive mammalian cells. Biochemistry 21:3414–3419.

Davidson I, Xiao JH, Rosales R, Staub A, Chambon P (1988): The HeLa cell protein TEF-1 binds specifically and cooperatively to two SV40 enhancer motifs of unrelated sequence. Cell 54:931–942.

Dayn A, Malkhosyan S, Mirkin SM (1992): Transcriptionally driven cruciform formation in vivo. Nucl Acids Res 20:5991–5997.

Deb S, DeLucia AL, Baur C-P, Koff A, Tegtmeyer P (1986): Domain structure of the Simian Virus 40 core origin of replication. Mol Cell Biol 6:1663–1670.

Desplan C, Theis J, O'Farrell PH (1988): The sequence specificity of homeodomain–DNA interaction. Cell 54: 1081–1090.

Dhar R, Lai C-J, Khoury G (1978): Nucleotide sequence of the DNA replication origin for human papovavirus BKV: sequence and structural homology with SV40. Cell 13:345–358.

Dorsman JD, van Heeswijk WC, Grivell LA (1990): Yeast general transcription factor GFI: Sequence requirements for binding to DNA and evolutionary conservation. Nucl Acids Res 18:2769–2776.

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.

Ford M, Fried M (1986): Large inverted duplications are associated with gene amplification. Cell 45:425–430.

Frisque RJ (1983): Nucleotide sequence of the region encompassing the JC virus origin of DNA replication. Virology 46:170–176.

Gallinari P, La Bella F, Heintz N (1989): Characterization and purification of H1TF2, a novel CCAAT-binding protein that interacts with a histone H1 subtype-specific consensus element. Mol Cell Biol 9:1566–1575.

Garrard WT, Cockerill PN, Hunting DW, McDaniel-Gerwig D, Szent-Gyorgyi C, Xu M, Gross DS (1988): In Adolph KW (ed): "Chromosome and Chromatin Structure." Boca Raton, FL: CRC Press, Vol 1, pp 133–178.

Grange T, Roux J, Rigaud G, Pictet R (1991): Cell-type specific activity of two glucocorticoid responsive units of rat tyrosine aminotransferase gene is associated with multiple binding sites for C/EBP and a novel liver-specific nuclear factor. Nucl Acids Res 19:131–139.

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.

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

Hanes S, Brent R (1991): A genetic model for interaction of the homeodomain recognition helix with DNA. Science 251:426–430.

Hatton KS, Dhar V, Brown EH, Iqbal MA, Stuart S, Didamo VT, Schildkraut CL (1988): Replication program of active and inactive multigene families in mammalian cells. Mol Cell Biol 8:2149–2158.

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

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.

Hobom G, Grosschedl R, Lusky M, Scherer G, Schwarz E, Kössel H (1979): Functional analysis of the replicator structure of lambdoid bacteriophage DNAs. Cold Spring Harb Symp Quant Biol 43:165–178.

Hörz W, Altenburger W (1981): Sequence specific cleavage of DNA by micrococcal nuclease. Nucl Acids Res 9:2643–2658.

Iguchi-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.

Ivanchenko M, Avramova Z (1992): Interaction of MAR-sequences with nuclear matrix proteins. J Cell Biochem 50:190–200.

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

Jurka J, Walichiewicz J, Milosavljevic A (1992): Prototypic sequences for human repetitive DNA. J Mol Evol 35:286–291.

Kaplan DJ, Jurka J, Solus JF, Duncan CH (1991): Medium reiteration frequency repetitive sequences in the human genome. Nucl Acids Res 19:4731–4738.

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

Kurosaki T, Ushiro H, Mitsuuchi Y, Suzuki S, Matsuda M, Matsuda Y, Katunuma N, Kangawa K, Matsuo H, Hirose T, Inayama S, Shizuta Y (1987): Primary structure of human poly(ADP-ribose) synthetase as deduced from cDNA sequence. J Biol Chem 262:15990–15997.

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.

Lin L-S, Meyer RJ (1987): DNA synthesis is initiated at two positions within the origin of replication of plasmid R1162. Nucl Acids Res 15:8319–8331.

Lockshon D, Galloway DA (1986): Cloning and characterization of ori_{L2}, a large palindromic DNA replication origin of Herpes Simplex Virus Type 2. Virology 58:513–521.

Lozzio BB, Lozzio CB, Bamberger EG, Feliu AS (1981): A multipotential leukemia cell line (K-562) of human origin (41106). Proc Soc Exp Biol Med 166:546–550.

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.

Masukata H, Tomizawa J-I (1984): Effects of point mutations on formation and structure of the RNA primer for ColE1 DNA replication. Cell 36:513–522.

Meijer M, Beck E, Hansen FG, Bergmans HEN, Messer W, von Meyenburg K, Schaller H (1979): Nucleotide sequence of the origin of replication of the *Escherichia coli* K-12 chromosome. Proc Natl Acad Sci USA 76:580–584.

Mierendorf RC, Pfeffer D (1987): Direct sequencing of denatured plasmid DNA. Methods Enzymol 152:556–562.

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.

Moskowitz IPG, Heichman KA, Johnson RC (1991): Alignment of recombination sites in Hin-mediated site-specific DNA recombination. Genes Dev 5:1635–1645.

Mueller CR, Maire P, Schibler U (1990): DBP, a liver-enriched transcriptional activator, is expressed late in ontogeny and its tissue specificity is determined posttranscriptionally. Cell 61:279–291.

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

Phi-Van L, Strätling 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.

Pogue GP, Hall TC (1992): The requirement for a 5′ stem-loop structure in Brome Mosaic Virus replication supports a new model for viral positive-strand RNA initiation. Virology 66:674–684.

Postel EH, Mango SE, Flint SJ (1989): A nuclease-hypersensitive element of the human c-*myc* promoter interacts with a transcription initiation factor. Mol Cell Biol 9:5123–5133.

Prives C, Murakami Y, Kern FG, Folk W, Basilico C, Hurwitz J (1987): DNA sequence requirements for replication of polyomavirus DNA in vivo and in vitro. Mol Cell Biol 7:3694–3704.

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

Reisman D, Yates J, Sugden B (1985): A putative origin of replication of plasmids derived from Epstein-Barr virus is composed of two cis-acting components. Mol Cell Biol 5:1822–1832.

Rose SM, Garrard WT (1984): Differentiation-dependent chromatin alterations precede and accompany transcription of immunoglobulin light chain genes. J Biol Chem 259:8534–8544.

Schmid CW, Manning JE, Davidson N (1975): Inverted repeat sequences in the *Drosophila* genome. Cell 5:159–172.

Smith GR, Kunes SM, Schultz DW, Taylor A, Triman KL (1981): Structure of chi hotspots of generalized recombination. Cell 24:429–436.

Soeda E, Kimura G, Miura K-I (1978): Similarity of nucleotide sequences around the origin of DNA replication in mouse polyoma virus and simian virus 40. Proc Natl Acad Sci USA 75:162–166.

Sollner-Webb B, Melchior W Jr, Felsenfeld G (1978): DNA-ase I, DNAase II and staphylococcal nuclease cut at differ-

ent, yet symmetrically located, sites in the nucleosome core. Cell 14:611–627.

Stow ND, McMonagle EC (1983): Characterization of the TR_S/IR_S origin of DNA replication of Herpes Simplex Virus Type 1. Virology 130:427–438.

Subramanian KN, Dhar R, Weissman SM (1977): Nucleotide sequence of a fragment of SV40 DNA that contains the origin of DNA replication and specifies the 5′ end of "early" and "late" viral RNA. III. Construction of the total sequence of *Eco*R II-G fragment of SV40 DNA. J Biol Chem 252:355–367.

Thornell A, Hallberg B, Grundström T (1988): Differential protein binding in lymphocytes to a sequence in the enhancer of the mouse retrovirus SL3-3. Mol Cell Biol 8:1625–1637.

van Wijnen AJ, Wright KL, Massung RF, Gerretsen M, Stein JL, Stein GS (1988): Two target sites for protein binding in the promoter region of a cell cycle regulated human H1 histone gene. Nucl Acids Res 16:571–592.

Vaughn JP, Dijkwel PA, Mullenders LHF, Hamlin JL (1990): Replication forks are associated with the nuclear matrix. Nucl Acids Res 18:1965–1969.

Vrieling H, Van Rooijen ML, Groen NA, Zdzienicka MZ, Simons JWIM, Lohman PHM, van Zeeland AA (1989): DNA strand specificity for UV-induced mutations in mammalian cells. Mol Cell Biol 9:1277–1283.

Walsh K (1989): Cross-binding of factors to functionally different promoter elements in c-*fos* and skeletal actin genes. Mol Cell Biol 9:2191–2201.

Ward GK, McKenzie R, Zannis-Hadjopoulos M, Price GB (1990): The dynamic distribution and quantification of DNA cruciforms in eukaryotic nuclei. Exp Cell Res 188: 235–246.

Ward GK, Shihab-el-Deen A, Zannis-Hadjopoulos M, Price GB (1991): DNA cruciforms and the nuclear supporting structure. Exp Cell Res 195:92–98.

Weller SK, Spadaro A, Schaffer JE, Murray AW, Maxam AM, Schaffer PA (1985): Cloning, sequencing, and functional analysis of ori_L, a Herpes Simplex Virus Type 1 origin of DNA synthesis. Mol Cell Biol 5:930–942.

Wu R, Wu T, Ray A (1987): Adaptors, linkers and methylation. Methods Enzymol 152:343–349.

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.

Yan D-H, Hung M-C (1991): Identification and characterization of a novel enhancer for the rat *neu* promoter. Mol Cell Biol 11:1875–1882.

Zannis-Hadjopoulos M, Kaufmann G, Wang S-S, Lechner RL, Karawya E, Hesse J, Martin RG (1985): Properties of some monkey DNA sequences obtained by a procedure that enriches for DNA replication origins. Mol Cell Biol 5:1621–1629.

Zannis-Hadjopoulos M, Frappier L, Khoury M, Price GB (1988): Effect of anti-cruciform DNA monoclonal antibodies on DNA replication. EMBO J 7:1837–1844.