Journal of Cellular Biochemistry 55:513–529 (1994)

Transcription Factor Binding Sites in the Matrix Attachment Region (MAR) of the Chicken α-Globin Gene

Teni Boulikas

Institute of Molecular Medical Sciences, Palo Alto, California 94306

Abstract Nuclear matrix is a nuclear protein–DNA superstructure believed to be the exclusive site of DNA replication, transcription, repair, and recombination. The attachment regions of chromatin loops to the nuclear matrix, called MARs, nest origins of replication, have transcriptional enhancer activity, and via their interaction with protein transcription factors may govern gene switch during development and tissue-specific gene expression. In this study the 967 bp MAR of the chicken α-globin gene is analyzed for the presence of hexanucleotides from a number (83 in total) of vertebrate protein transcription factors and core origins of replication. A total number of 760 hexanucleotides from factor sites or origins of replication were used for this search. We found that: (1) The occurrence of protein transcription factor binding sites overall on the MAR fragment as well as on the enhancer and promoter regions of other genes is only about 1.2–1.5 times higher than in random DNA, something consistent for all MAR and enhancer sequences examined. However, a high concentration (up to 2.7 times over random sequences) of hexanucleotide factor sites is observed on small stretches of the α-globin gene MAR. (2) Some regulatory protein binding sites are underrepresented whereas others are overrepresented, giving to an MAR a particular transcription factor flavor. (3) The DNA curvature map of the MAR sequence and the potential sites of positioned nucleosomes suggest the sites where a competition between core histone octamers and protein transcription factors for DNA might be found. This approach might provide a novel technique to diagnose for the regulatory or nonregulatory function of a stretch of DNA. Furthermore, MARs are proposed to constitute important regulatory elements of genes in addition to enhancers, promoters, silencers, locus control regions, and origins of replication. Additional parameters such as interaction of a transcription factor with other transcription factors fixed at vicinal sites, DNA methylation, intrinsic DNA curvature torsional strain, and nucleosome positioning might also determine the high-affinity binding of a transcription factor to its functional sites and its exclusion from or low affinity binding to other nonregulatory regions.

Key words: transcription factors, nuclear matrix, MARs, transcriptional enhancers, curved DNA, positioned nucleosomes, origins of replication

Polynucleosomes are constrained into loops or domains by the formation of crosscomplexes between certain classes of DNA sequences spaced every 5 to 200 kb with specific chromosomal proteins leading to the formation of a distinct nuclear superstructural entity termed nuclear matrix [see Hancock and Boulikas, 1982; Verheijen et al., 1988; Gasser et al., 1989; de Jong et al., 1990; Bonifer et al., 1991; van Driel et al., 1991; Pienta et al., 1991; Berezney, 1991; Fey et al., 1991]. Nuclear matrix is believed to be implicated in the differential gene expression during development [Dworetzky et al., 1990; Boulikas, 1992a], with topoisomerase II orchestrating the topological looped organization of chromatin domains [Razin et al., 1991].

MAR sequences have been demonstrated to have transcriptional enhancer activity [Xu et al., 1989; Blasquez et al., 1989; Stief et al., 1989; Phi-Van et al., 1990; Klehr et al., 1991]. Since transcription takes place on the nuclear matrix [e.g., Ciejek et al., 1983; see Boulikas, 1987 for a review], understanding the nature of DNA sequences that are attached to the nuclear matrix is important for a comprehensive view of the regulation of gene expression in eukaryotic cells.

We have proposed [Boulikas, 1992c] that the promoter region of the genes might be anchored to the nuclear matrix via interactions between nonhistone proteins bound to the enhancer considered to be a constitutive MAR element and

Abbreviations used: nt, nucleotides; MAR, matrix attached region; ORI, origin of replication.

Received March 8, 1994; accepted March 9, 1994.

Send correspondence to Teni Boulikas, Institute of Molecular Medical Sciences, 460 Page Mill Road, Palo Alto, CA 94306.

promoter elements (facultative MAR elements), with looping out of the DNA. In this model the transcriptional enhancers of the various genes are MARs. Electron micrographs of the matrix protein SAF-A mixed with DNA show the formation of loops [Romig et al., 1992]. Usually one constitutive MAR activity present in all cell types is found in a 5′ far upstream site of a gene and a second constitutive MAR activity in a 3′ downstream site, delineating the boundaries of a chromatin domain such as in the 19 kb locus of the chicken lysozyme gene [Phi-Van and Strätling, 1988] and in the 49 kb locus of the human apolipoprotein B gene [Levy-Wilson and Fortier, 1989]. On the other hand, the facultative class of MAR activities is believed to depend on the transcriptional and repair activity of the gene or stage of development. Thus, the facultative MAR activity of a given gene will be present only in *cells expressing* this gene.

Origins of replication (ORIs) are nested in the nuclear matrix, and their differential activation is intimately associated with cell type formation and differential gene expression during development [see Boulikas, 1992a,b] and during the cell cycle [Mah et al., 1993]. A multitude of ORIs function in early stages of development with a suppression of specific ORIs taking place concomitantly with the establishment of cell memory, that is, cell type formation [e.g., Spradling and Orr-Weaver, 1987]. The ORIs whose function is suppressed during development may be related to the facultative class of MAR activities. Indeed, some MARs share common sequence characteristics with some origins of replication, including the presence of ATTA and ATTTA motifs [Boulikas, 1992a] and the presence of potential cruciform structures [Boulikas and Kong, 1993a,b]. A facultative type of MAR activity, detected only in cell types harboring the particular gene in an active chromatin structure, is located between nucleotides −2.7 and −1.8 kb to the 5′ side of the human apolipoprotein B gene [Levy-Wilson and Fortier, 1989].

Today only a small fraction of nuclear matrix anchorage regions (MARs) for polynucleosomes are known, most particularly the AT-rich class of MARs [e.g., Mirkovitch et al., 1984; Cockerill and Garrard, 1986; Levy-Wilson and Fortier, 1989]. We have recently identified a non-AT-rich type of MARs containing CT-, and GA-rich *blocks [Boulikas* and Kong, 1993a,b].

The view of MARs as mass binding sites for protein transcription factors [Dworetzky et al., 1992; Isomura et al., 1992; Boulikas, 1992a,b,c, 1993a,b; Bidwell et al., 1993; van Wijnen et al., 1993] explains both the transcriptional enhancer activity associated with MARs and the fact that the nuclear matrix is the site of initiation of transcription/replication. Studies aimed at theoretically predicting protein binding sites along DNA have appeared [e.g., Frech et al., 1993]. We have applied a very simple method based on the search of hexanucleotides derived from the recognition sites of protein transcription factors along DNA. Application of this method to the MAR residing to the 3′ flanking region of the chicken α-globin gene domain [Farache et al., 1990] as well as to other MAR sequences and to known promoters and enhancer regions establishes some simple rules for predicting protein sites on a DNA stretch of unknown function.

MATERIALS AND METHODS

Factor Hexanucleotide Searches

A list of vertebrate transcription factor binding sites has been compiled [Faisst and Meyer, 1992] and classified into AT-, GC-, GA-, TG-rich, and mixed motifs [Boulikas, in press]. To this list we have added the core regions from the origins of replication of SV40, bovine papilloma virus, the c-*myc* gene origin of replication, the adenoviruses 2 and 4 core origin (ORI sites), and the *Xenopus* Hbox1 and human HOX4D homeodomain protein sites (HOMEO sites) (Table I). The α-globin gene MAR [Farache et al., 1990] was screened for the presence of hexanucleotide sequences from the factor, ORI, and HOMEO sites using a Macintosh computer and the Microsoft Word 5.0 program. The hexanucleotides for each particular factor binding sequence were changed into a different font using combinations of sizes 10, 12, 14, and 18; with type Helvetica, Chicago, Palatino, and Times; with formats normal, bold face, outlined, italics, and shadowed; and with formats nonunderlined, underlined, dotted underlined, or double underlined. This combination gives $4 \times 4 \times 5 \times 4 = 320$ distinct letter characters. This way a different font was used for each factor site. This permits the visual inspection of a MAR sequence for clusters of factor sites.

We searched for a total of 83 transcription factor, ORI, and HOMEO sites together. The size of the factor sites varied from pentanucleotides for CTCF (CCCTC) to 23 nucleotides for Pax-1 (CACCGTTCCGCTCTAGATATCTC) and

TABLE I. Vertebrate Homeodomain Protein Binding Sites and Origins of Replication Known to Interact With Protein Replication/Transcription Factors Used in the Searches of This Work

Sequence	Protein or repl. origin	Species	Reference
5′-AACAATTACAAA-3′ 5′-TTTGTAATTGTT-3′	Hbox 1	Frog	Cho et al., 1988
CGTTTTATTAGG CCTAATAAAACG	Hox 4D	Human	Zappavigna et al., 1991
CATTAATC GATTAATG			
TATAATC GATTATA			
AGGCATAATATCATTAC GTAATGATATTATGCCT			
CATAAAATTTTTATTG CAATAAAAATTTTATG			
TGAATAATTGTTGTTAACAATAATC GATTATTGTTAACAACAATTATTCA	BPV-ORI	Bovine	Ustav et al., 1991
CATCATCAATAATATACC GGTATATTATTGATGATG	Ad 2,4-ORI	Human	Hay, 1985
TTGGATTGAAGCCAATA TATTGGCTTCAATCCAA	Ad 2,4-ORI		
AGAGGCCGAGGCGGCCTCGGCCTCTG CAGAGGCCGAGGCCGCCTCGGGCCTCT	T antigen site II on SV40 ORI	Monkey	Hay and De Pamphilis, 1982
TCTCTTATGCGGTTGAATAGT ACTATTCAACCGCATAAGAGA	c-myc-ORI	Human	Kumano et al., 1992

to 25 and 26 nt for BPV and SV40 ORI (Table I). Both strand factor motifs were used in this search. Thus, for hexanucleotide factor sites, like Sp1, two motifs were used, GGGCGG and its complementary CCGCCC; for heptanucleotide factor sites such as the TCTCTTA site of Myc, four hexanucleotide motifs (TCTCTT, CTCTTA, and the complementary TAAGAG and AAGAGA) were used. From the 18 nt CBF factor site, 26 total hexanucleotides were extracted, and for the 23 nt Pax-1 factor site 36 total hexanucleotides were extracted. Thus, the total number of hexanucleotides extracted from a factor binding site was 2 (n − 5), where n is the number of nt in the recognition and binding factor site. A total of 760 hexanucleotides were derived.

The probability of finding a given hexanucleotide is $(1/4)^6$ in sequences in which all four nucleotides are equally represented, or one in 4,096 nt of DNA. Ninety-four out of 760 hexanucleotides were found to be shared by at least two factor sites in our searches (see Table II). The occurrence of many common hexanucleotides among transcription factors, when only 83 factor sites were screened with an average length of 10 nt, demonstrates the common language proteins use when they talk to their cognate DNA motifs. Octanucleotide matches between a factor site and a stretch of the MAR sequence were taken as three hexanucleotide matches and heptanucleotide matches as two hexanucleotide matches.

Sequences Searched

These include MAR sequences known today [Boulikas, 1992a], other MAR sequences identified in our laboratory [e.g., Boulikas and Kong,

TABLE II. Hexanucleotides That Occur in Two or More Transcription Factor Binding Sites*

Hexanucleotides (complementary not shown)	Protein transcription factor		
AAACAC	H1TF1	TGT3	
GACCTT	ELP	TR	
CTCCCC	H4TF1	NTF	
AAGAGA	ISGF2	c-myc-ORI	
AAAGAA	LSF	CD28	
GAATTT	NFκB	CD28	Oct-5
TTCCTC	Pu .1	NTF	
CAGTTT	IREBF	ISGF3	NF-AT
CTTTCC	PCF	EBP-1	
ATTTTT	YB-1	HOX4D	
GAAAAA	NF-AT	YB-1	
TGTTTG	H1TF1	TGT3	
TTTCCT	PEA3	NF-AT	
CTTTGG	LIT-1	YB-1	
GGAATC	MBP-1	NFκB	
GGTCAT	NFE3	VDR	NF-μE3
GGGAGG	NTF	H4TF1	
TTATCT	GATA	RVF	
GTCAAG	ARP-1	TR	
AGGTCA	ER	TR	RAR/H2RII
AACTGA	ISGF3	MyoD	
ATGAAA	NF-AT	AAF	
TGAAGG	ISGF3	Ad 2,4-ORI	
ATGTGG	TFE3	NF-μE3	
GGAAAA	NF-AT	TR	
AAAAAC	NF-AT	TR	
GACTTT	EBP-1	NFκB	
TTATTG	HOX4D	Ad 2,4-ORI	
ATTGTT	Hbox1	BPV-ORI	
GGGAAA	IREBF	PCF	
GGGGGA	H4TF1	NTF	
CTTTCT	ISGF2	PCF	
TTAATG	HNF-1	HOX4D	
GTCATG	VDR	NF-μE3	
TAATAT	SRF	Ad 2,4-ORI	
ACATGA	TFE3	NF-μE3	
AATATG	CBF	AAF	
AAAAAT	YB-1	HOX4D	
CTAATA	SRF	HOX4D	
ATAATT	HNF1	BPV-ORI	
TAAAAT	Oct-5	HOX4D	
CATATT	CBF	SRF	
ATGCAA	NF-W	Oct-3	
TTGACC	ARP-1	TR	
CCCTTT	PCF	LIT-1	
TAAAAC	RVF	HOX4D	

*Hexanucleotides occurring in two or more transcription factor binding sites. Only one strand is shown. This table lists 46 such hexanucleotides. Others may be found. The total number of theoretically possible hexanucleotides (considering only one strand) is 2,048.

1993a,b], and several known enhancers and promoters; random DNA of desired A + T content was generated using a program written by Jerzy Jurka (Linus Pauling Institute).

DNA Curvature Maps and Positioned Nucleosomes Along the MAR Sequence

These maps are based on programs developed by Ed Trifonov and collaborators [Ioshikhes et al., 1992; Shpigelman et al., 1993]. This type of analysis permits positioned nucleosomes and intrinsically curved sites to be superimposed on the protein transcription site map and thus to theoretically predict possible nucleosome-free regions occupied by nonhistones.

RESULTS

Figure 1 shows the hexanucleotide sites for protein transcription factors along the MAR in the 3′ flanking region of the chicken α-globin gene. This MAR was identified by Farache and coworkers [1990]. Our search reveals that the 967 bp MAR sequence possesses hexanucleotide motifs that are present in the binding sites of transcription factors. Most factors interact with a 6–14 nucleotide consensus binding sequence, with a range of 4–23 [Wingender, 1988; Faisst and Meyer, 1992; Boulikas, in press], and stretches of contact of transcription factors with DNA can be from 8–30 nt long from DNase I footprinting studies [e.g., Lichtsteiner et al., 1987] or up to 42 bp for the unusual TFIIIA that contains nine zinc fingers [Clemens et al., 1992].

Figure 1 suggests that the chicken α-globin gene MAR sequence is a mosaic of six nt binding sites of a great number of transcription factors, including the mammalian Oct-3, μ-enhancer binding protein, CD28, NFκB, Ets-1, SRF, CBF, and the Pu.1 factor related to Ets-1 oncoprotein. In addition, several six nt motifs contained in the c-*myc* and bovine papilloma virus origins of replication and from the core sequence of the adenovirus two and four origin of replication, which is known to interact with protein replication/transcription factors (Table I), were found.

The probability of finding an intact 8–30 nt factor binding site on the 967 bp fragment is very small. A given seven nt motif occurs once every 4^7 = 16,384 nt and an eight nt motif once every 4^8 = 65,536 nt. For this reason we have broken the consensus sequence of the transcription factor binding site into all possible hexanucleotide motifs. The probability of finding a given hexanucleotide on a stretch of DNA is

$(1/4)^6$, or one such hexanucleotide every 4,096 nt of DNA. We have found that 55 out of 83 factor sites examined are represented with at least one hexanucleotide in the α-globin gene MAR. Out of the 760 factor site hexanucleotides searched, 222 were found in the MAR sequence (Fig. 1). Whereas 179.4 (760 × 967:4096) are expected; the χ^2 when the observed is compared with the expected is $(222\text{-}179.4)^2/179.4 = 10.11$; analysis of ten independent MAR sequences gave similar χ^2 values which gives P value below 0.001.

Figure 1 reveals that many hexanucleotide factor sites overlap on the α-globin gene MAR. These stretches are shown in detail in Figure 2. Since 760 hexanucleotides in total from factor sites and origins of replication were used to search the α-globin gene MAR, we expect to find 760 sites every 4,096 nucleotides or approximately one site every 5.4 nucleotides. Thus the nine-nt-long sequence at position 75 of the chicken α-globin gene MAR (Fig. 2) is expected to have 9/5.4 = 1.7 factor sites. Four factor hexanucleotide recognition sequences, those of CD28, AGIE, DBP, and HNF-1 [see Faisst and Meyer, 1992], were found to overlap on this nine nt stretch; therefore the density of factor hexanucleotides is 4 (found)/1.7 (predicted) = 2.4 times higher than random sequences on this nine nt sequence of the MAR (Fig. 2).

The vertebrate factor sites used in this search of the chicken sequence come from species as diverse as frogs, chickens, and mammals. The chicken counterpart of a human, frog, mouse, or rat factor site is expected to be identical in cases of high evolutionary conservation or to differ in one or more nucleotide sites in cases of lower evolutionary conservation.

Due to the common language regulatory proteins use when talking to "their DNA," the presence of overlapping factor hexanucleotide recognition sequences on a MAR stretch might signal the binding sites of several transcription factors that interact with one another. However, since most factors possess a longer than hexanucleotide recognition sequence, this could indicate the binding site of a single factor. This regulatory protein factor might be the species-specific counterpart of one of the proteins whose binding sites overlap on a particular stretch of the MAR fragment, a variant of this protein transcription factor expressed at a different stage during development, or a different, as yet undiscovered, transcription regulating protein.

Some factor sites share common hexanucleotides. For example, AAACAC occurs in both H1TF1 and TGT3 sites, the GACCTT in ELP and TR sites, and the CTCCCC in H4TF1 and NTF. Table II lists hexanucleotides identified as common to two or more factor binding sites. The identification of more than 47 such hexanucleotides, or 2 × 47 = 94, with their complementary sequences out of 4,096 possible as being represented at least twice in factor binding sites (Table II) indeed supports the concept that DNA sequence motifs use a common language to talk to "their proteins."

An important feature of the factor hexanucleotides on the chicken α-globin gene MAR, supposed to have a transcriptional enhancer activity (see introductory remarks), is that some factor sites are underrepresented while other factor sites are overrepresented in this MAR sequence (Table III). The most striking examples of underrepresented factor/ORI sites are those of SV40-ORI, ETS-1, VDR, TFEB, RAR, and HOX4D. The most overrepresented factor sites are those of CD28, Pu.1, PEA3, and AGIE (Table III).

Intrinsically Curved DNA Stretches and Positioned Nucleosomes in the α-Globin Gene MAR

One major class of the nuclear matrix DNA harbors AT-rich sequences [e.g., Mirkovitch et al., 1984; see Boulikas, 1992a]; however, a non AT-rich class of MARs has been described [Boulikas and Kong, 1993a,b]. MARs are characterized by the presence of intrinsically curved DNA [Anderson, 1986; Homberger, 1989]. Figure 3 shows the curvature map of the chicken α-globin gene MAR sequence. Bendable DNA arising by low-level periodic signals along the DNA may in part determine the positioning of nucleosomes, which have a tendency to place such intrinsically curved motifs at certain locations within the core particle [Trifonov and Sussman, 1980; reviewed by Trifonov, 1991; Thoma, 1992]. There is a delicate balance between stability of nucleosomes and their functionality; the natural nucleosomes are rather metastable (i.e., they unfold easily by a small energy input facilitating replication and transcription) as opposed to nucleosomes formed on strongly curved DNA [Shrader and Crothers, 1989] that can act like immovable objects (E.N. Trifonov, personal communication). Many protein transcription factors bind to intrinsically curved DNA or cause DNA

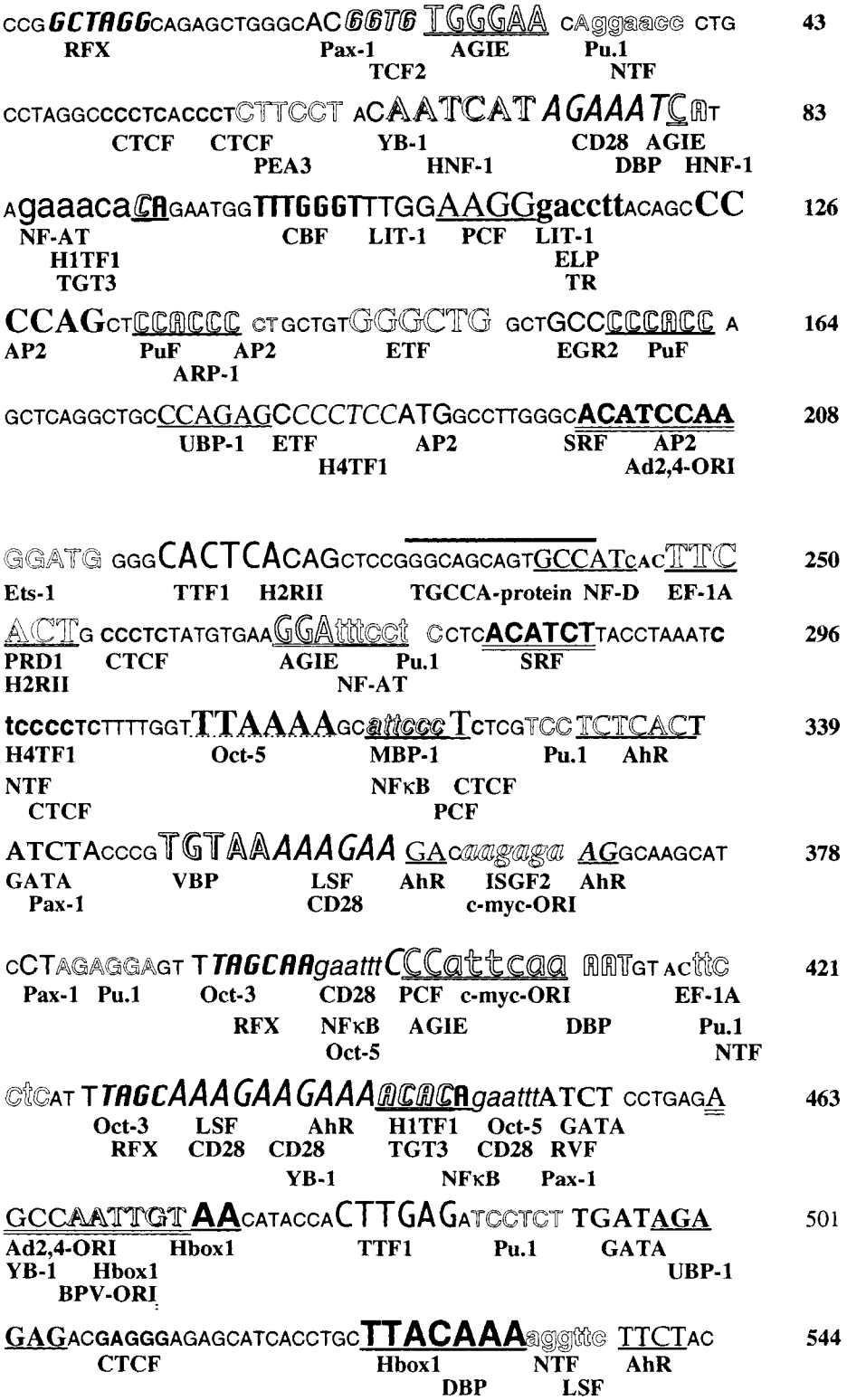

Fig. 1. *Protein transcription* factor binding sites in the chicken α-globin gene matrix attachment region. Different fonts and formats of all possible hexanucleotides from the consensus binding sites of vertebrate protein transcription factors [Faisst and Meyer, 1992; Boulikas, in press], homeodomain protein sites, as well as from vertebrate core sites of origins of replication known to interact with transcription/replication protein factors (Table I) were used to screen this MAR sequence. Protein factor acronyms are aligned below the hexanucleotide motifs. Four topoisomerase II recognition sites predicted by Farache and coworkers [1990] are overlined; the region from nucleotides 50–334 is the CRI repeat containing a site for the TGGCA binding protein (overlined), the chicken analog of NF-1 determined by DNase I footprinting [Farache et al., 1990]. (Continued on next page.)

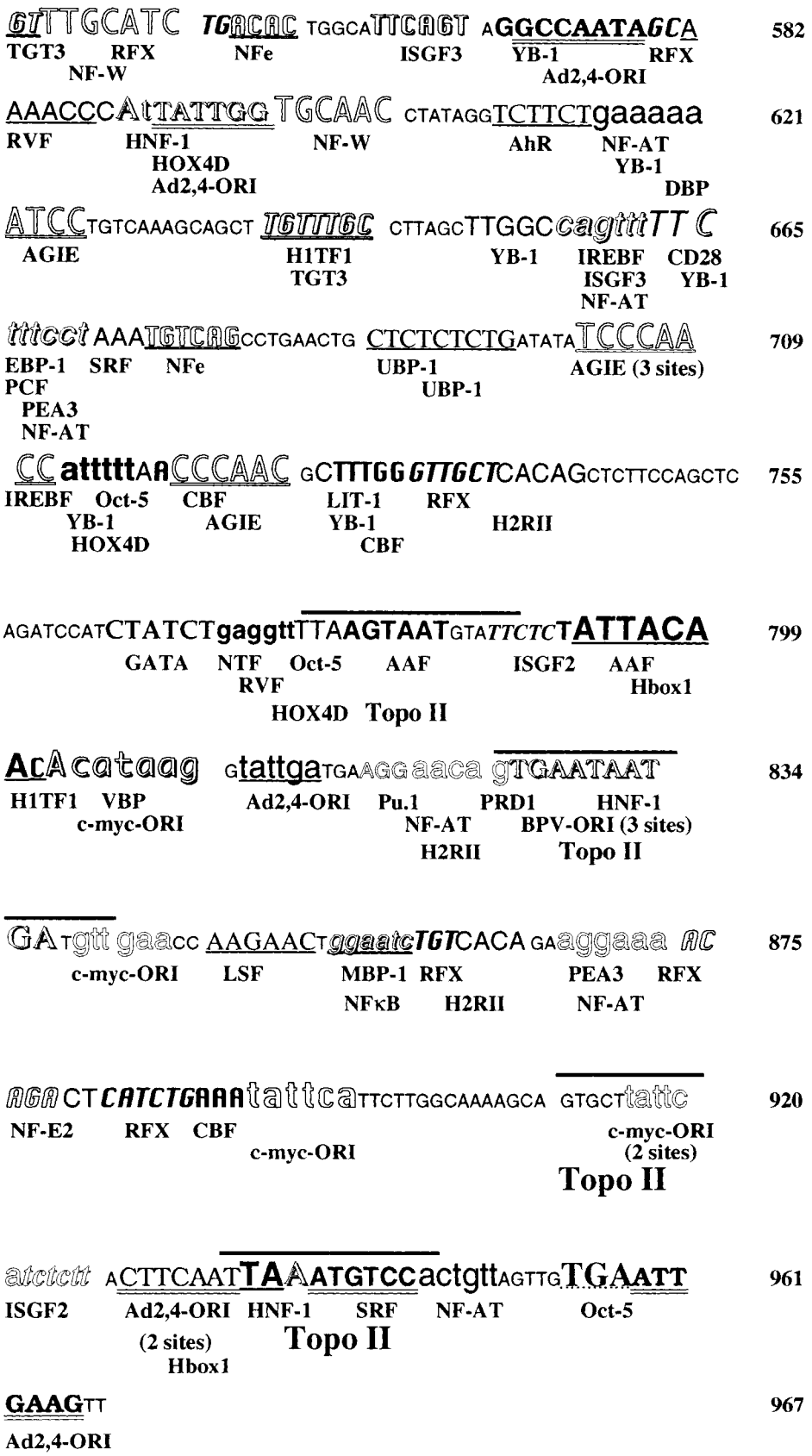

Figure 1. (Continued.)

75 AGAAATCAT (9 nt, Predicted sites: 1.7, Found: 4, Found/Predicted: 2.4)
AGAAAT
CD28

GAAATC
AGIE

AAATCA
DBP

AATCAT
HNF-1

99 TTTGGGTTTGGAAGGGACCTT (21 nt, Predicted sites: 3.9, Found: 7 , Found/Predicted: 1.8)
TTTGGGT GGACCT
CBF LIT-1

TTTGGA GACCTT
LIT-1 ELP
TR

AAGGGA
PCF

201 ACATCCAAGGATG (13 nt, Predicted sites: 2.4, Found: 4, Found/Predicted: 1.7)
ACATCC AGGATG
SRF Ets-I

ATCCAA
Ad 2,4-ORI

CCAAGG
AP2

240 GCCATCACTTCACTG (15 nt, Predicted sites: 2.8, Found: 4, Found/Predicted: 1.4)
GCCATC TTCACT
NF-D PRD1

CACTTC
EF-IA

TCACTG
H2RII

319 ATTCCCTC (8 nt, Predicted sites: 1.5, Found: 4, Found/Predicted: 2.7)
ATTCCC
MBP-1
NFκB

TTCCCT
PCF

CCCTC
CTCF

330 TCCTCTCACTATCTA (15 nt, Predicted sites: 2.8, Found: 4, Found/Predicted: 1.4)
TCCTCT TATCTA
Pu .1 Pax-1

TCTCAC
AhR

CTATCT
GATA

349 TGTAAAAAGAAGACAAGAGAAG (22 nt, Predicted sites: 4.1, Found: 6, Found/Predicted: 1.7)
TGTAAA AAGAGA
VBP ISGF2
c-myc-ORI

AAAGAA GAGAAG
CD28 AhR

AGAAGA
AhR

390 TTAGCAAGAATTTCCCATTCAAAAT (25 nt, Predicted sites: 4.6 Found: 10, Found/Predicted: 2.2)
TTAGCA GAATTT CAAAAT
Oct-3 Oct-5 DBP
NFκB

TAGCAA ATTTCC
RFX IREBF

GAATTTC ATTCAA
CD28 c-myc-ORI

TTTCCC
PCF

TTCCCA
AGIE

427 TTAGCAAAGAAGAAAACACAGAATTTATCTC (31 nt, Predicted sites: 5.8, Found: 14, Found/Predicted: 2.4)
TTAGCA AGAAAA GAATTT
Oct-3 YB-1 Oct-5
CD28
NFκB

TAGCAA AAACACA TTTATC
RFX H1TF1 RVF

AAAGAA AAACAC TTATCT
LSF TGT3 GATA

AGAAGA TATCTC
AhR Pax-1

463 AGCCAATTGTAA (12 nt, Predicted sites: 2.2, Found: 5, Found/Predicted: 2.3)
AGCCAA
Ad 2,4-ORI

GCCAAT
YB-1

AATTGT
BPV-ORI

AATTGTAA
Hbox1

526 TTACAAAAGGTTCTTCT (17 nt, Predicted sites: 3.1 Found: 6, Found/Predicted: 1.9)
TTACAAA GGTTCT
Hbox1 LSF

ACAAAA TCTTCT
DBP AhR

AGGTTC
NTF

Fig. 2. Overlapping protein transcription factor sites on the chicken α-globin gene MAR. Due to the common language of protein transcription factor sites, such motifs are proposed to be enucleation sites of protein transcription factors in vivo, provided that this is permitted by positioned nucleosomes (see Fig. 3). The complete prototype sequence motif is shown on top, whereas the hexa- or longer nucleotide motifs and the transcription factor whose recognition site represent are properly aligned below. The nucleotide position of the motif in the chicken α-globin gene MAR sequence is shown to the far left. (Figure 2 continued on next page.)

545 GTTTGCATCTGACAC (15 nt, Predicted sites: 2.8, Found: 7, Found/Predicted: 2.5)
GTTTGC CTGACAC
TGT3 NFe

TTTGCA
Oct-3

TTGCATC
NF-W

CATCTG
RFX

572 GGCCAATAGCAAAACCCATTATTGG (25 nt, Predicted sites: 4.6, Found: 9, Found/Predicted: 2.0)
GGCCAA AAAACC ATTATT
YB-1 RVF HNF-1

GCCAATA TTATTG
Ad 2,4-ORI HOX4D

TAGCAA ATTATTGG
RFX Ad 2,4-ORI

610 TCTTCTGAAAAAATCC (16 nt, Predicted sites: 3.0, Found: 5, Found/Predicted: 1.7)
TCTTCT AAAATC
AhR DBP

GAAAAA
NF-AT
YB-1

AAATCC
AGIE

652 TTGGCCAGTTTTTCTTTCCTAA (22 nt, Predicted sites: 4.1, Found: 11, Found/Predicted: 2.7)
TTGGCC TTTTTC TCCTAA
YB-1 CD28 SRF

CAGTTT CTTTCC
IREBF EBP-1
ISGF3 PCF
NF-AT

TTTCTT
YB1

TTTCCT
PEA3
NF-AT

704 TCCCAACCATTTTTAACCCAAC (22 nt, Predicted sites: 4.1, Found: 9, Found/Predicted: 2.2)
TCCCAACC ACCCAA
AGIE CBF

CCATTT CCCAAC
IREBF AGIE

ATTTTT
YB-1
HOX4D

TTTTAA
Oct-5

727 CTTTGGGTTGCTCACAG (17 nt, Predicted sites: 3.2, Found: 6, Found/Predicted: 1.9)
CTTTGG TCACAG
LIT-1 H2RII
YB-1

TTTGGGT
CBF

GTTGCT
RFX

770 GAGGTTTTAAGTAAT (15 nt, Predicted sites: 2.8, Found: 5, Found/Predicted: 1.8
GAGGTT AGTAAT
NTF AAF

GGTTTT
RVF

GTTTTA
HOX4D

TTTTAA
Oct-5

788 TTCTCTATTACAACACATAAG (21 nt, Predicted sites: 3.9, Found: 7, Found/Predicted: 1.8)
TTCTCT AACACA
ISGF2 H1TF1

TATTAC ACATAA
AAF VBP

ATTACAA CATAAG
Hbox1 c-myc-ORI

819 AGGAACAGTGAATAATGA (18 nt, Predicted sites: 3.3, Found: 8, Found/Predicted: 2.4)
AGGAAC TGAATAAT
Pu .1 BPV-ORI
(3 sites)

AACAGT TAATGA
NF-AT HNF-1

ACAGTG
H2RII

AGTGAA
PRD1

853 GGAATCTGTCACA (13 nt, Predicted sites: 2.4, Found: 4, Found/Predicted: 1.7)
GGAATC GTCACA
MBP-1 H2RII
NFκB

ATCTGT
RFX

868 AGGAAAACAGACTCATCTGAAATATTCA (28nt, Predicted sites: 5.2, Found: 7, Found/Predicted: 1.3)
AGGAAA GACTCA AAATAT
PEA3 NF-E2 CBF

AACAGA CATCTG TATTCA
RFX RFX c-myc-ORI

929 CTTCAATTAA (10 nt, Predicted sites: 1.9, Found: 4, Found/Predicted: 2.1)
CTTCAAT
Ad 2,4-ORI

CAATTA
Hbox1

AATTAA
HNF-1

Figure 2. (Continued.)

TABLE III. Unusual Occurrence of Factor Sites on the Chicken α-Globin Gene MAR*

Underrepresented				
Factor	Recognition sequence	Number of 6 nt motifs	Number of hexanucleotides: Expected	Found
SV40-ORI	See Table I	20	4.7	0
Ets-1	5′-$^{GA}_{CC}$GGA$^{A}_{T}$GY-3′ 5′-RC$^{T}_{A}$TCC$^{TC}_{GG}$-3′	16	3.8	0
VDR	AGGTCATGACCT AGGTCATGACCT	14	3.3	0
TFEB	GGCCACGTGACC GGTCACGTGGCC	14	3.3	0
PAX-1	CACCGTTCCGCTCTAGATATCTC GAGATATCTAGAGCGGAACGGTG	36	8.5	4
HOX4D	See Table I	50	11.8	3
RAR	AGGTCATGACCT AGGTCATGACCT	14	3.3	0
SRF	GGATGTCCATATTAGGACATCT AGATGTCCTAATATGGACATCC	34	8.0	4
Oct-6	ATGCAAAT ATTTGCAT	6	1.4	0
Oct-R	ATGCAAAY RTTTGCAT	6	1.4	0
C/EBP	GTGG$^{TTT}_{AAA}$G C$^{AAA}_{TTT}$CCAC	8	1.9	0
Myb	YAAC$^{G}_{T}$G C$^{C}_{A}$GTTR	8	1.9	0
MyoD	CAACTGAC GTCAGTTG	6	1.4	0
NF-μE3	GCCACATGACC GGTCATGTGGC	12	2.8	0
EGR-2	CCGCCCCCGC GCGGGGGCGG	10	2.4	0
ETF	CAGCCCCCGCGCAGC GCTGCGCGGGGGCTG	20	4.7	2
WT-ZPF	CGCCCCCGC GCGGGGGCG	8	1.9	0
Ker1	GCCTGCAGGC GCCTGCAGGC	10	2.4	0
ARP-1	TGAnCCCTTGACCCCT AGGGGTCAAGGGnTCA	14	3.3	1
CTCF	CCCTC GAGGG	6	1.4	0
TFE3	GCCACATGACC GGTCATGTGGC	12	2.8	0

Table III continued on next page.

TABLE III. Unusual Occurrence of Factor Sites on the Chicken α-Globin Gene MAR* (*continued*)

Factor	Recognition sequence	Number of 6 nt motifs	Number of hexanucleotides	
			Expected	Found
Underrepresented				
αA-CRYBP	GGGAAATCCC GGGATTTCCC	10	2.4	0
TEF-1	AAGYATGCA TGCATRCTT	6	1.4	0
TR	AGGTCAAGGTCA TGACCTTGACCA	14	3.3	1
Overrepresented				
CD28RC	5′-AAAGAAATTCC-3′ 5′-GGAATTTCTTT-3′	12	2.8	8
Pu.1	AGAGGAACT	8	1.9	6
PEA3	AGTTCCTCT AGGAAR YTTCCT	4	0.9	3
AGIE BP	GGTTGGGAAATCCC GGGATTTCCCAACC	18	4.2	11
Oct-3,4	ATGC$^{T}_{A}$AAT ATT$^{A}_{T}$GCAT	6	1.4	3
Oct-5	TTAAAATTCA TGAATTTTAA	10	2.4	6
Hbox	AACAATTACAAA TTTGTAATTGTT	14	3.3	7
UBP-1	CTCTCTGG CCAGAGAG	6	1.4	4
H1TF1	AACAAACACAAA TTTGTGTTTGTT	14	3.3	6
PuF	GGGTGGG CCCACCC	4	0.9	2
RFX	CCCCTAGCAACAGATG CATCTGTTGCTAGGGG	22	5.2	9
TGT3	AAGTGTTTGC GCAAACACTT	10	2.4	5
NF-W	GTTGCATC GATGCAAC	6	1.4	3

*Protein transcription factors whose some hexanucleotides extracted from their DNA binding sequence are under- or overrepresented in the 967 bp chicken α-globin gene MAR. Their recognition sequences are from Faisst and Meyer [1992], Boulikas [in press], and Table I. Both strands of the recognition sequence are shown in their 5′ to 3′ orientation.

bending upon binding as the CAP protein on the lac promoter of *E. coli* [Wu and Crothers, 1984]. Peaks above about 0.3 curvature units on the map (Fig. 3A) are significant curvature points [Shpigelman et al., 1993]. The highest curved synthetic or naturally occurring DNA fragments display curvature values of up to 0.6 units in our program. The curvature map shows intrinsically curved DNA points in decreasing order at positions 622 (0.47 units), 429, 447, 576, 583, 490, 459, 409, 662, 875, and 353 (0.35 units).

We have scanned the 967 bp MAR sequence for the preferred sites of positioned nucleosomes

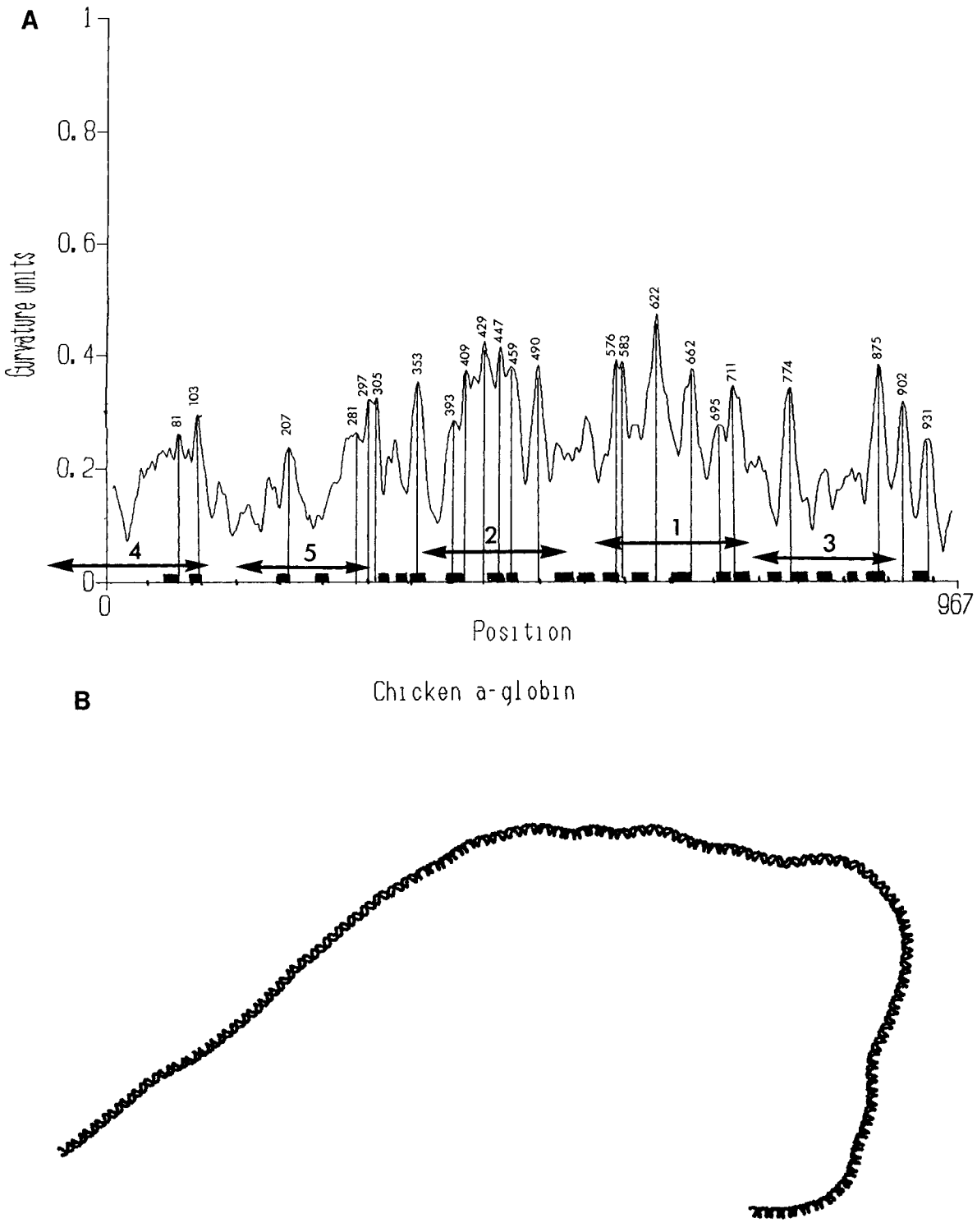

Fig. 3. Intrinsic DNA curvature map of the chicken α-globin gene MAR. **A:** A 15 nt window was moved along the sequence. Peaks above 0.2 curvature units represent significant positions of curvature. Numbers denote the nucleotide positions of major peaks on the map. Filled rectangles on the abscissa show the overlapping factor sites (Fig. 2). Arrows above the abscissa show possible nucleosome locations (Table IV). **B:** Three-dimensional projection of curvature along the 967 bp fragment.

using a program that takes into consideration all AA and TT dinucleotides and compares their positions against a given matrix of strongly positioned nucleosomes [Ioshikhes et al., 1992]. Nucleosome positioning is determined by several factors: preference of core histone octamers for features on DNA, interactions between neighboring nucleosomes, by transcription/replication regulatory proteins acting as boundaries for positioned nucleosomes, and the higher order folding of chromatin [reviewed by Thoma, 1992]. This analysis shows a strongly positioned nucleosome (score 0.304) around nucleotide 648, occupying the 564–731 bp region of the MAR fragment (Table IV). The most probable locations of five nucleosomes on the 967 bp MAR fragment are shown as arrows in Figure 3A. This map does not take into account changes in positioned nucleosomes arising from strongly bound transcription regulatory proteins at precise positions on the MAR fragment that could act as boundaries to position two nucleosomes at their flanking regions [Thoma, 1992]. This information can only arise from footprint analysis.

Figure 3A also shows the locations of overlapping protein factor sites (Fig. 2) on the curvature map (shown on the abscissa as solid rectangles). Only five stretches of clustered factor sites out of 23 total clusters fall within predicted linkers, those at the flanking region of nucleosome 2. The occupancy of a portion of the MAR region by a nucleosome with strong regulatory protein sites in the nucleosome flanking regions might bring into juxtaposition protein factors bound to the flanking regions at the entry and exit points of the DNA to and from the nucleosome. Binding sites of regulatory proteins along the DNA also determine nucleosome positioning [Roth et al., 1990]. The interplay between protein transcription factors and nucleosomes on the chicken α-globin gene 3′ MAR should await footprinting of protein factors in vivo.

DISCUSSION

MARs as Mass Binding Sites for Protein Transcription Factors

A critical mass of activator protein seems to be necessary to effectively stimulate the basic transcriptional machinery [Carey et al., 1990]. Creating polymers of single factor binding sites, which by themselves may have weak or no stimulatory activity, can lead to strong transcriptional stimulation [Ondek et al., 1987; Schirm et al., 1987], especially when one single high-affinity binding site is present in the complex [Scheuermann, 1991; Zuckerkandl and Villet, 1988]. Synergistic effects between transcription factors can best occur when regulatory proteins saturate their binding sites on DNA. However, precise positioning of two transcription protein factors on vicinal sites on DNA is an absolute necessity for formation of stable complexes and provides a model to understand the dramatic activation of the albumin gene in hepatic compared with other cell types [Milos and Zaret, 1992].

Given our incomplete knowledge of the principles used by nuclei for placing protein factors along the DNA, we have made an attempt to predict the positions of transcription factors and nucleosomes by screening the chicken α-globin gene MAR fragment for hexanucleotide motifs from core recognition sites of transcription/replication factors. The data (Figs. 1, 2) show that the 967 bp MAR fragment has 1.25 times more six nt factor sites than random DNA. A computer-generated 5 kb random DNA fragment showed 0.97 times the theoretically-predicted density of hexanucleotide factor sites whereas all other MAR and enhancer sequences

TABLE IV. Estimation of the Most Preferred Nucleosome Locations in the Chicken α-Globin Gene MAR*

Nucleosome	Nucleotide position for nucleosome center	Nucleotide positions of 167 bp nucleosome	Score for positioned nucleosome	Linker length to the 3′ side
1	648	564–731	0.304	14
2	451	367–534	0.262	30
3	829	745–912	0.248	Not determined
4	35	−49–118	0.204	32
5	234	150–317	0.200	50

*Nucleosomes are listed in order of decreasing probability of formation from top to bottom. Linker lengths between a nucleosome and its neighbor to the 3′ site are given in the last column.

examined consistently showed a 1.2 to 1.5 times higher concentration of such hexanucleotides. However, some stretches of the chicken α-globin gene MAR fragment (Fig. 2) have a high density of six nt factor sites that could be up to 2.7 times higher than the number expected from probability calculations.

We have proposed that a major class of regulatory chromosomal nonhistones (transcription protein factors) are components of the nuclear matrix [Boulikas, 1992a,b,c, 1993a,b]. Growing evidence independently supports this notion [Dworetzky et al., 1992; Isomura et al., 1992; Bidwell et al., 1993; van Wijnen et al., 1993]. Thus nuclear matrix proteins might number in the thousands. Only a few have been identified and well characterized, to date. Among these are topoisomerase II [Berrios et al., 1985; Razin et al., 1991]; the chicken ARBP protein (attachment region binding protein), which recognizes the consensus sequence ATTTCA$^{C}_{G}$TTGTAAAA in the MAR of the chicken lysozyme gene locus [von Kries et al., 1991]; the NMP-2 localized exclusively in the nuclear matrix [Bidwell et al., 1993]; the Sp1, ATF, CCAAT, C/EBP, and AP-1 transcription factors [van Wijnen et al., 1993]; the yeast ACBP protein (ARS consensus binding protein), which interacts with the ARS element [Hofmann and Gasser, 1991]; SAF-A, able to form loops with naked DNA as observed by electron microscopy [Romig et al., 1992]; the tissue-specific human SATB1 protein expressed predominantly in thymus that binds to the minor groove of a special class of AT-rich MARs with A, T, or C but not G on one strand [Dickinson et al., 1992]; matrin 3, an acidic protein of the internal nuclear matrix network of human and rat cells [Belgrader et al., 1991]; matrin F/G [Hakes and Berezney, 1991]; the transcription protein factor RFP involved in the activation of the *ret* proto-oncogene [Isomura et al., 1992]; DNA polymerase α [Smith and Berenzney, 1980]; c-myc [Waitz and Loidl, 1991] and Myb proteins [Klempnauer, 1988]; corticosteroid receptors [van Steensel et al., 1991]; H1 [Izaurralde et al., 1989]; HMG 1,2 [Ivachenko and Avramova, 1992]; lamins from the peripheral nuclear matrix [Boulikas, 1986; Ludèrus et al., 1992]; and several others. The recognition sequences of some of these nuclear matrix proteins are known.

One can ask whether or not the presence of a hexanucleotide density for a protein transcription factor on the α-globin gene MAR of 1.25 times that of random DNA is sufficient to give to this sequence its expected transcriptional enhancer properties, that is, acting as a strong enucleation site for protein transcription factors. Since known transcriptional enhancers and promoters, including those of *Drosophila engrailed* gene, human c-*myc* gene, rat albumin gene, human histone gene, human growth hormone gene, and SV40 72 bp repeat, display a density comparable to MARs for hexanucleotides from regulatory protein recognition sites [Boulikas, in preparation], it is concluded that this small difference in density of hexanucleotide sites between MARs and random DNA can be diagnostic of the presence of a regulatory sequence of DNA.

We feel that additional features other than DNA sequence itself, notably intrinsic curvature of DNA [Trifonov, 1991], are important in directing regulatory proteins to a DNA site; even straight DNA can be curved by the binding of some transcriptional regulatory proteins [reviewed by van der Vliet and Verrijzer, 1993]. Also, the cooperative binding of some transcription-regulating proteins and the strong heterologous protein–protein interactions on the regulatory regions of genes are important determinants of the formation of a transcription-promoting ternary complex. A more exciting phenomenon is the overrepresentation of some unique factor sites on this MAR fragment. A short (50–100 nt) segment of an enhancer germane to a particular gene needs to interact with probably about five or even more protein factors. Which segments of the 967 bp MAR are crosscomplexed with protein factors and which are wrapped in nucleosomes is not known. Thus, understanding the spacing between protein factor sites and which particular factors are likely to be present in a given cell type is of utmost importance for the successful theoretical prediction of which stretch of an MAR will be occupied by which protein factors. It is interesting to note that Farache and coworkers (1990) have observed four topoisomerase II recognition sites (by sequence analysis) and one site for the chicken TGGCA binding protein (by DNase I footprinting), a relative of the mammalian nuclear factor 1 (NF-1), within the CR1 repeat (overlined in Fig. 1). Fifty-five different sequence-specific proteins out of 85 tested have hexanucleotides from their core binding site on the chicken α-globin gene MAR. A total number of 222 such factor nucleotides were found on this MAR (Figs. 1, 2). Most of these proteins may have their counterparts in

chickens with a slightly modified or identical sequence specificity.

Several protein transcription factors interact with one another on a transcriptional enhancer sequence germane to a particular gene. The interdigitation and crosstalk between transcription protein factor sites on the regulatory regions of genes allow for the complex process of cell type speciation to be manifested during development. Some of these recognition sites may be 23 nt long but most of them are of about eight nt. The strong specificity, giving a dissociation constant of a protein factor from the DNA in the order of 10^{-7}–10^{-15}, may, therefore, depend on the precise interaction of protein factors with one another after locking onto precisely positioned sites on the DNA. The possibility of the occupancy of a stretch of 167–200 bp of this MAR by a positioned nucleosome, and therefore the competition of protein factors with core histone octamers for interaction with a stretch of DNA, needs to be taken into consideration.

On account of the common preference of protein binding sites for certain di-, tri-, tetra-, penta-, and hexanucleotide motifs (Table II), we believe that the approach used here, although giving only a 1.2–2.7-fold higher hexanucleotide factor site concentration on sequences believed to have a regulatory function, nevertheless can be diagnostic of regulatory regions on DNA and can be used to distinguish between functional and nonfunctional DNA.

Despite our incomplete knowledge of all mammalian transcription factor consensus sequences, the results of this search may have an important impact on our understanding of the function of stretches of MARs as transcriptional enhancers and origins of replication. It will be interesting to correlate the cellular function of a gene with the type of transcription factor sites theoretically predicted to be present on its MAR sequence. Such data can further be treated taking into account the possible involvement of an MAR at the origin of replication and/or transcriptional enhancer for the gene with which it is associated. It is proposed that MARs constitute important regulatory sequences of genes of an importance equal to that of promoters, enhancers, locus control regions, silencers, and origins of replication.

PERSPECTIVES

Since regulatory regions on DNA display a 1.2–2.7 higher concentration of hexanucleotides derived from regulatory protein binding sites, the cell has probably invented additional mechanisms such as protein factor–protein factor interaction, intrinsic curvature of DNA, nucleosome exclusion, cruciform formation, Z-DNA formation, DNA methylation, and others to direct protein *trans*-acting factors to the regulatory regions of genes. Studies on the interaction of protein transcriptional regulators with the omnipresent nucleosomes rather than naked DNA which compose the physiological milieu of DNA sequences in nuclei are likely to produce a different landscape of knowledge in this field and to advance our understanding of the relationship between negative and positive regulatory protein–DNA complex elements in development, active and inactive chromatin loops, and active versus inactive matrix-attached region activities. Such data can arise only from the elucidation of interactions of regulatory proteins with one another, with DNA, with nucleosomes, and with the nuclear matrix in proximal and distal regulatory sequences of genes. A further understanding of the networks that regulate transcription initiation will arise from studies on protein factor phosphorylation tightly coupled to up- or downregulation of *cis*-acting factors at precise intervals of the cell cycle and development. Finally, knowledge of the mechanisms of differential activation of origins of replication during development tightly connected to the differential gene expression [see Boulikas and Kong, 1993b, for references] and their relationship to MARs and enhancers will fit the bulk of knowledge on protein transcription factor–regulated genes into their context of chromatin structure, carcinogenesis, and development.

ACKNOWLEDGMENTS

Special thanks and gratitude to Ed Trifonov, Dawn Brooks, Irina Gribanovsky, Jolanta Walichiewicz, Emile Zuckerkandl, Jerzy Jurka, S. Karlin, and B.E. Blaisdell for their valuable input.

REFERENCES

Anderson JN (1986): Detection, sequence patterns and function of unusual DNA structures. Nucleic 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.

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.

Bidwell JP, van Wijnen AJ, Fey EG, Dworetzky S, Penman S, Stein JL, Lian JB, Stein GS (1993): Osteocalcin gene promoter–binding factors are tissue-specific nuclear matrix components. Proc Natl Acad Sci 90:3162–3166.

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

Bonifer C, Hecht A, Saueressig H, Winter DM, Sippel AE (1991): Dynamic chromatin: The regulatory domain organization of eukaryotic gene loci. J Cell Biochem 47:99–108.

Boulikas T (1986): Protein–protein and protein–DNA interactions in calf thymus nuclear matrix using crosslinking with UV. Biochem Cell Biol 64:474–484.

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

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): Chromatin and nuclear matrix in development and in carcinogenesis: A theory. Int J Oncol 1:357–372.

Boulikas T (1992c): Poly(ADP-ribosyl)ation, repair, chromatin and cancer. In Spandidos DA (ed): "Current Perspectives in Molecular and Cellular Oncology, Vol 1." London: JAI Press, pp 1–109.

Boulikas T (1993a): Nature of DNA sequences at the attachment regions of genes to the nuclear matrix. J Cell Biochem 52:14–22.

Boulikas T (1993b): Homeodomain protein binding sites, inverted repeats, and nuclear matrix attachment regions along the human β-globin gene complex. J Cell Biochem 52:23–36.

Boulikas T (in press): A compilation and classification of DNA binding sites for protein transcription factors from vertebrates. Crit Rev Eukaryotic Gene Expression.

Boulikas T, Kong CF (1993a): A novel class of matrix-attached regions (MARs) identified by random cloning and their implications in differentiation and carcinogenesis. Int J Oncol 2:325–330.

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

Boulikas T (in preparation): A theoretical prediction of transcription factor binding sites on promoters, enhancers, origins of replication, and MARS.

Carey M, Lin Y-S, Green MR, Ptashne M (1990): A mechanism for synergistic activation of a mammlian gene by GAL4 derivatives. Nature 345:361–365.

Cho KWY, Goetz J, Wright CVE, Fritz A, Hardwicke J, De Robertis EM (1988): Differential utilization of the same *reading frame in a Xenopus* homeobox gene encodes two related proteins sharing the same DNA-binding specificity. EMBO J 7:2139–2149.

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

Clemens KR, Liao X, Wolf V, Wright PE, Gottesfeld JM (1992): Definition of the binding sites of individual zinc fingers in the transcription factor IIIA-5S RNA gene complex. Proc Natl Acad Sci USA 89:10822–10826.

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.

de Jong L, van Driel R, Stuurman N, Meijne AML, van Renswoude J (1990): Principles of nuclear organization. Cell Biol Int Rep 14:1051–1074.

Dickinson LA, Joh T, Kohwi Y, Kohwi-Shigematsu (1992): A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. Cell 70:631–645.

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.

Dworetzky SI, Wright KL, Fey EG, Penman S, Lian JB, Stein JL, Stein GS (1992): Sequence-specific DNA-binding proteins are components of a nuclear matrix-attachment site. Proc Natl Sci USA 89:4178–4182.

Faisst S, Meyer S (1992): Compilation of vertebrate-encoded transcription factors. Nucleic Acids Res 20:3–26.

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. Nucleic Acids Res 18:401–409.

Fey EG, Bangs P, Sparks C, Odgren P (1991): The nuclear matrix: Defining structural and functional roles. Crit Rev Eukaryotic Gene Expression 1:127–143.

Frech K, Herrmann G, Werner T (1993): Computer-assisted prediction, classification, and delimitation of protein binding sites in nucleic acids. Nucleic Acids Res 21:1655–1664.

Gasser SM, Amati BB, Cardenas ME, Hofmann JF-X (1989): Studies on scaffold attachment sites and their relation to genome function. Int Rev Cytol 119:57–96.

Hakes DJ, Berezney R (1991): Molecular cloning of matrin F/G: A DNA binding protein of the nuclear matrix that contains putative zinc finger motifs. Proc Natl Acad Sci USA 88:6186–6190.

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

Hay RT (1985): Origin of adenovirus DNA replication. Role of the Nuclear Factor I binding site in vivo. J Mol Biol 186:129–136.

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.

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.

Ioshikhes I, Bolshoy A, Trifonov EN (1992): Preferred positions of AA and TT dinucleotides in aligned nucleosomal DNA sequences. J Biomol Struct Dyn 9:1111–1117.

Isomura T, Tamiya-Koizumi K, Suzuki M, Yoshida S, Taniguchi M, Matsuyama M, Ishigaki T, Sakuma S, Takahashi M (1992): RFP is a binding protein associated with the nuclear matrix. Nucleic Acids Res 20:5305–5310.

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

Izaurralde E, Käs E, Laemmli UK (1989): Highly preferential nucleation of histone H1 assembly on scaffold-associated regions. J Mol Biol 210:573–585.

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.

Klempnauer K-H (1988): Interaction of myb proteins with the nuclear matrix in vitro. Oncogene 2:545–551.

Kumano M, Nakagawa T, Imamura Y, Galli I, Ariga H, Iguchi-Ariga SMM (1992): Stimulation of SV40 DNA replication by the human c-*myc* enhancer. FEBS Lett 309:146–152.

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.

Ludérus MEE, de Graaf A, Mattia E, den Blaauwen JL, Grande MA, de Jong L, van Driel R (1992): Binding of matrix attachment regions to lamin B_1. Cell 70:949–959.

Mah DCW, Dijkwel PA, Todd A, Klein V, Price GB, Zannis-Hadjopoulos M (1993): ors12, a mammalian autonomously replicating DNA sequence, associates with the nuclear matrix in a cell cycle–dependent manner. J Cell Sci 105:807–818.

Milos PM, Zaret KS (1992): A ubiquitous factor is required for C/EBP-related proteins to form stable transcription complexes on an albumin promoter segment in vitro. Genes Dev 6:991–1004.

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.

Ondek B, Shepard A, Herr W (1987): Discrete elements within the SV40 enhancer region display different cell-specific enhancer activities. EMBO J 6:1017–1025.

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.

Phi-Van L, von Kries JP, Ostertag W, Strätling 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.

Pienta KJ, Getzenberg RH, Coffey DS (1991): Cell structure and DNA organization. Crit Rev Eukaryotic Gene Expression 1:355–385.

Razin SV, Petrov P, Hancock R (1991): Precise localization of the α-globin gene cluster within one of the 20- to 300-kilobase DNA fragments released by cleavage of chicken chromosomal DNA at topoisomerase II sites in vivo: Evidence that the fragments are DNA loops or domains. Proc Natl Acad Sci USA 88:8515–8519.

Romig J, Fackelmayer FO, Renz A, Ramsperger U, Richter A (1992): Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements. EMBO J 11:3431–3440.

Roth SY, Dean A, Simpson RT (1990): Yeast a2 repressor positions nucleosomes in TRP1/ARS1 chromatin. Mol Cell Biol 10:2247–2260.

Scheuermann RH (1991): The tetrameric structure of NF-μNR provides a mechanism for cooperative binding to the immunoglobulin heavy chain μ enhancer. J Biol Chem 267:624–634.

Schirm S, Jiricny J, Schaffner W (1987): The SV40 enhancer can be dissected into multiple segments each with a different cell type specificity. Genes Dev 1:65–74.

Shpigelman ES, Trifonov EN, Bolshoy A (1993): Curvature: Software for the analysis of curved DNA. Comp Appl Biol Sci 9:435–440.

Shrader TE, Crothers DM (1989): Artificial nucleosome positioning sequences. Proc Natl Acad Sci USA 86:7418–7422.

Smith HC, Berezney R (1980): DNA polymerase α is tightly bound to the nuclear matrix of actively replicating liver. Biochem Biophys Res Commun 97:1541–1547.

Spradling A, Orr-Weaver T (1987): Regulation of DNA replication during *Drosophila* development. Annu Rev Genet 21:373–403.

Stief A, Winter DM, Strätling WH, Sippel AE (1989): A nuclear DNA attachment element mediates elevated and position-independent gene activity. Nature 341:343–345.

Thoma F (1992): Nucleosome positioning. Biochim Biophys Acta 1130:1–19.

Trifonov EN (1991): DNA in profile. Trends Biochem Sci 16:467–470.

Trifonov EN, Sussman JL (1980): The pitch of chromatin DNA is reflected in its nucleotide sequence. Proc Natl Acad Sci USA 77:3816–3820.

Ustav M, Ustav E, Szymanski P, Stenlund A (1991): Identification of the origin of replication of bovine papillomavirus and characterization of the viral origin recognition factor E1. EMBO J 10:4321–4329.

van der Vliet PC, Verrijzer CP (1993): Bending of DNA by transcription factors. BioEssays 14:25–32.

van Driel R, Humbel B, de Jong L (1991): The nucleus: A black box being opened. J Cell Biochem 47:311–316.

van Steensel B, van Haarst AD, de Kloet ER, van Driel R (1991): Binding of corticosteroid receptors to rat hippocampus nuclear matrix. FEBS Lett 292:229–231.

van Wijnen AJ, Bidwell JP, Fey EG, Penman S, Lian JB, Stein JL, Stein GS (1993): Nuclear matrix association of multiple sequence-specific DNA binding activities related to SP-1, ATF, CCAAT, C/EBP, OCT-1, and AP-1. Biochemistry 32:8397–8402.

Verheijen R, Van Venrooij W, Ramaekers F (1988): The nuclear matrix: Structure and composition. J Cell Sci 90:11–39.

von Kries JP, Buhrmester H, Strätling WH (1991): A matrix/scaffold attachment region binding protein: Identification, purification and mode of binding. Cell 64:123–135.

Waitz W, Loidl P (1991): Cell cycle dependent association of c-*myc* protein with the nuclear matrix. Oncogene 6:29–35.

Wingender E (1988): Compilation of transcription regulating proteins. Nucleic Acids Res 16:1879–1902.

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.

Zappavigna V, Renucci A, Izpisúa-Belmonte J-C, Urier G, Peschle C, Duboule D (1991): *HOX4* genes encode transcription factors with potential auto- and cross-regulatory capacities. EMBO J 10:4177–4187.

Zuckerkandl E, Villet R (1988): Generation of high specificity of effect through low-specificity binding of proteins to DNA. FEBS Lett 231:291–298.