Nuclear Matrix Protein ARBP Recognizes a Novel DNA Sequence Motif with High Affinity[†,‡]

Hartmut Buhrmester, Jens P. von Kries,[§] and Wolf H. Strätling*

Institut für Physiologische Chemie, Universitäts-Krankenhaus Eppendorf, Martinistrasse 52, 20246 Hamburg, FR Germany

Received July 21, 1994; Revised Manuscript Received December 13, 1994[⊗]

ABSTRACT: ARBP is a nuclear protein that specifically binds to matrix/scaffold attachment regions (MARs/SARs). Here we characterize by DNase I footprinting, dimethyl sulfate protection, and mobility shift assays two binding sites for ARBP within a chicken lysozyme MAR fragment. Our results indicate that ARBP recognizes a novel DNA sequence motif containing the central sequence 5′-GGTGT-3′ and flanking AT-rich sequences. Binding occurs through major groove contacts to two guanines of the central sequence. Collective and single-base substitutions in the 5′-GGTGT-3′ core motif result in loss or significant reductions of ARBP binding, underscoring the importance of the GC-rich core sequence. Structural elements of the sequence motif are probably also recognized. The affinity of ARBP to both binding sites is surprisingly high [$K_D = (2-6) \times 10^{-10}$ M]. High-affinity recognition of the identified DNA motif in MARs/SARs by ARBP is likely an important feature in the domain organization of chromatin.

In eukaryotes, gene expression is regulated by *cis*-acting elements, named enhancers and silencers. These exert their effects over considerable distances (from 0.2 to at least 10 kb), and, moreover, they are rather promiscuous in their interaction with promoters, modulating the proper promoter as well as heterologous ones in appropriate constructs. For example, the −6.1 kb enhancer of the chicken lysozyme gene can increase transcription from the lysozyme promoter but also from the herpes simplex virus thymidine kinase gene promoter (Grewal et al., 1992). These properties raise the question of how the action of enhancers is directed to a specific gene. For many genes, it has been shown that when they are "active" they are embedded together with their *cis*-acting regulatory elements in a chromatin domain of generally elevated DNase I sensitivity (Weintraub & Groudine, 1976; Strätling et al., 1986; Levy-Wilson & Fortier, 1990). Thus, domains would correspond to units of gene regulation. Moreover, it is thought that domains also define units of chromatin compaction and that these are generated by attachment of specific sequences at domain boundaries to a proteinaceous framework structure, called matrix in interphase nuclei and scaffold in metaphase chromosomes [for reviews, see Gasser and Laemmli (1987), Phi-Van and Strätling (1990), and Cook (1991)]. For many genes, such sequences (named MARs or SARs) have been localized by in vitro studies analyzing the interaction of genomic DNA or cloned fragments with nuclear matrix (scaffold) (Mirkovitch et al., 1984; Cockerill & Garrard, 1986; Phi-Van & Strätling, 1988). MARs/SARs have been characterized as AT-rich sequences of much greater length than enhancers and to exhibit a sequence diversity that does not allow any cross-hybridization between different MARs. A significant step in the elucidation of the functional role of MARs was the use of synthetic domains containing a reporter gene and flanking MARs in stable transfection experiments. In stably integrated copies of such constructs flanking chicken lysozyme 5′ MARs elevated gene expression about 10-fold and rendered expression less sensitive to effects of the chromatin structure at the site of integration (Stief et al., 1989; Phi-Van et al., 1990). Very similar results have been reported for a MAR from the human β-interferon gene domain (Klehr et al., 1991). In transgenic mice, adjacent lysozyme 5′ MARs conferred position-independent, proper hormonally and developmentally regulated expression to the whey acidic protein transgene (McKnight et al., 1992). A MAR from tobacco has also been shown to insulate reporter gene expression from chromatin effects at the site of integration (Breyne et al., 1992).

Very likely the formation of loop-domains in chromatin requires proteins that specifically interact with MARs and other functionally important DNA sequences at domain boundaries (Udvardy et al., 1985; Kellum et al., 1991). Using a Southwestern blotting assay, we have identified two chicken nuclear proteins with apparent molecular masses of 95 and 120 kDa that specifically bind to MAR fragments of the chicken lysozyme gene and to heterologous MARs (von Kries et al., 1991, 1994a). The 95 kDa protein, named ARBP (attachment region binding protein), was purified to homogeneity and characterized as a novel protein that binds to multiple sequences within the lysozyme MAR fragment H1-*Hae*II. ARBP is solubilized from chromatin with 200−300 mM NaCl, thus at a salt concentration definitely lower than that required to remove histones from chromatin. Since ARBP was found to have an unusually low sedimentation coefficient (2.1 S) but a high molecular mass in gel filtration, it was concluded that the protein has an elongated shape (von Kries et al., 1994b). Other proteins that bind to MARs and might have a function in loop-domain formation include DNA topoisomerase II, SAT-B1, lamin B1, histone H1

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

[‡] The complete nucleotide sequence of the fragment H1-*Hae*II is deposited in the EMBL, GenBank, and DDBJ Nucleotide Sequence Databases under Accession Number X84223.

* Corresponding author Telephone: 49-40-4717-2392. Fax: 49-40-4717-4862.

[§] J.P.v.K. contributed to this paper by the preparation of purified ARBP.

[⊗] Abstract published in *Advance ACS Abstracts,* March 1, 1995.

0006-2960/95/0434-4108$09.00/0

variants, and transcription factors (Adachi et al., 1989; Dickinson et al., 1992; Ludérus et al., 1992; Käs et al., 1993; Bidwell et al., 1993).

Here we characterize two binding sites for ARBP in a chicken lysozyme MAR fragment. We show that ARBP recognizes a novel DNA sequence motif with surprisingly high affinity. This motif contains the central sequence 5′-GGTGT-3′ and flanking AT-rich sequences. This property of ARBP likely serves an important function in the formation of chromatin loop-domains.

MATERIALS AND METHODS

Materials. ARBP was purified to homogeneity as previously described (von Kries et al., 1991). Purity was assessed by SDS−polyacrylamide gel electrophoresis, staining, and Southwestern blotting assays. Protein concentrations were also determined by SDS−polyacrylamide gel electrophoresis and staining with Coomassie blue using bovine serum albumin as standard. The chicken 5′ MAR fragments used are *Sau*3A-*Sau*3A (266 bp), *Sau*3A-*Hae*II (235 bp), and a 420-bp fragment containing a direct repeat of the sequence *Sau*3A-*Hae*II (Figure 7) (von Kries et al., 1991). In Figure 6, a chemically synthesized 41-bp site I sequence (coordinates 1986−2026) with attached *Xba*I and *Sal*I sites was cloned into *Xba*I/*Sal*I-cut vector pBend2 (Kim et al., 1989) and subsequently recovered into a 163-bp fragment by cleavage with *Xho*I. Two mutated 41-bp site I sequences (mutants 1 and 3; see Results) were cloned similarly. The correct nucleotide sequences were verified by sequencing using the dideoxy nucleotide chain termination method. A map indicating the positions of the sequences *Sau*3A-*Sau*3A and *Sau*3A-*Hae*II within fragment H1-*Hae*II is shown in Figure 1 (bottom panel). DNA fragments were radioactively 3′-end-labeled with [α-^{32}P]dATP using the Klenow fragment of DNA polymerase I (Sambrook et al., 1989).

Electrophoretic Mobility Shift Assay. Binding reaction mixtures (total volume, 20 μl) contained 1 ng of end-labeled DNA fragment [approximately 10 000 cpm (Cerenkov counting)], 1 μg of poly(dG-dC)·poly(dG-dC) (Hirano et al., 1990), and 0.25−2 ng of purified ARBP in binding buffer. Ten-fold binding buffer was prepared by suspending dry milk powder at 0.5% (w/v) in 100 mM Tris-HCl, pH 7.5, 500 mM NaCl, 10 mM EDTA, 10 mM dithiothreitol, and 50% (v/v) glycerol, heating the suspension at 95 °C for 20 min, and removing the insoluble material by centrifugation in an Eppendorf centrifuge. Binding reactions were carried out at room temperature for 30 min, and then the samples were run on native 3.5−6% polyacrylamide gels (preelectrophoresed for 1 h) in 10 mM Tris-HCl, pH 7.5, 1 mM EDTA at 20 mA for 120 min. After electrophoresis, gels were dried onto Whatman 3MM and examined by autoradiography.

To determine the dissociation constants by electrophoretic mobility shift assays, 2 ng of purified ARBP was incubated with increasing amounts of labeled DNA at 20 °C. In separate experiments, 0.56 ng of DNA was incubated with increasing amounts of ARBP. After electrophoresis on native polyacrylamide gels, radioactivity separated into free DNA and bound DNA was determined by Cerenkov counting.

DNase I Footprint Analysis. End-labeled DNA probes (approximately 10 000 cpm Cerenkov) were preincubated with 0,10, 30, or 60 ng of ARBP for 30 min as in the mobility shift assays. Then DNase I (Cooper Biomedical) was added to 20 or 40 ng/sample in 200 μL of binding buffer containing 2 mM $MgCl_2$ and 0.1 mM (instead of 1 mM) EDTA, and incubated for 1 min at room temperature. Reactions were stopped by adding 50 μL of 50 mM EDTA, 1250 mM NaCl, and 50 μg of *Escherichia coli* DNA per milliliter. This was followed by extraction with phenol/chloroform/isoamyl alcohol (12:12:1) and precipitation with ethanol. After the pellets were dried, 4−10 μl of loading buffer containing 90% deionized formamide, 10 mM EDTA, 0.2% bromophenol blue, and 0.2% xylene cyanol was added. The suspended samples were heated for 5 min at 95 °C, chilled on ice, and loaded onto 5% denaturing sequencing gels according to Maxam and Gilbert (1980). After electrophoresis, gels were fixed, dried, and exposed to Kodak XAR-5 film at −80 °C. Changes in the DNase I cutting pattern were monitored by scanning with a Bio-Rad Model 620 video densitometer.

Dimethyl Sulfate (DMS)[1] *Methylation Protection Assay.* End-labeled DNA probes were preincubated with 30 or 50 ng of ARBP as in the mobility shift and DNase I footprinting assays. Then 180 μL of reaction buffer, that contained 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, and 1 mM EDTA, and 1 μL of DMS were added. Reactions were incubated for 1 or 2 min at room temperature, followed by the addition of 50 μL of stop solution containing 1.5 M sodium acetate, pH 7.0, 1 M 2-mercaptoethanol, and 100 μg of yeast tRNA per milliliter. DNA was precipitated with ethanol, collected by centrifugation, and washed once with 0.075 M sodium acetate, pH 7.0, 75% ethanol and once with ethanol. Chemical G and G+A reactions were carried out as described (Maxam & Gilbert, 1980) and followed by separation on sequencing gels and autoradiography.

RESULTS

We have previously shown by Southwestern blotting assays and deletion mutational analysis that multiple sequences within the MAR fragment H1-*Hae*II, which is located at the 5′ boundary of the chicken lysozyme gene domain, can lead to efficient binding of the nuclear matrix protein ARBP (von Kries et al., 1991). Here we used electrophoretic mobility shift assays, DNase I footprinting, and DMS methylation protection analyses to characterize two binding sites for this protein. First, we performed electrophoretic mobility shift assays with purified ARBP and two DNA fragments, *Sau*3A-*Sau*3A and *Sau*3A-*Hae*II, that contain an upstream and a downstream segment, respectively, of fragment H1-*Hae*II [see map in Figure 1C (bottom panel)]. Figure 1A (top panel), lanes 1−5, and 1B (middle panel) show that each of the fragments formed a single main complex with ARBP. These complexes are specific, since their formation was competed by a 50-fold excess of unlabeled fragment but not of *Hin*fI-digested Bluescript (Figure 1A, lanes 6 and 8). To verify that these complexes were indeed generated by ARBP, proteins bound to the retarded DNA fragment were recovered by electroelution of the retarded band, displayed on an SDS−polyacrylamide gel, blotted, and visualized by incubation with probe H1-*Hae*II under conditions of low stringency (in the presence of only 50 μg/mL of *E. coli* competitor DNA). The only DNA binding protein detected had the apparent molecular mass

[1] Abbreviation: DMS, dimethyl sulfate.

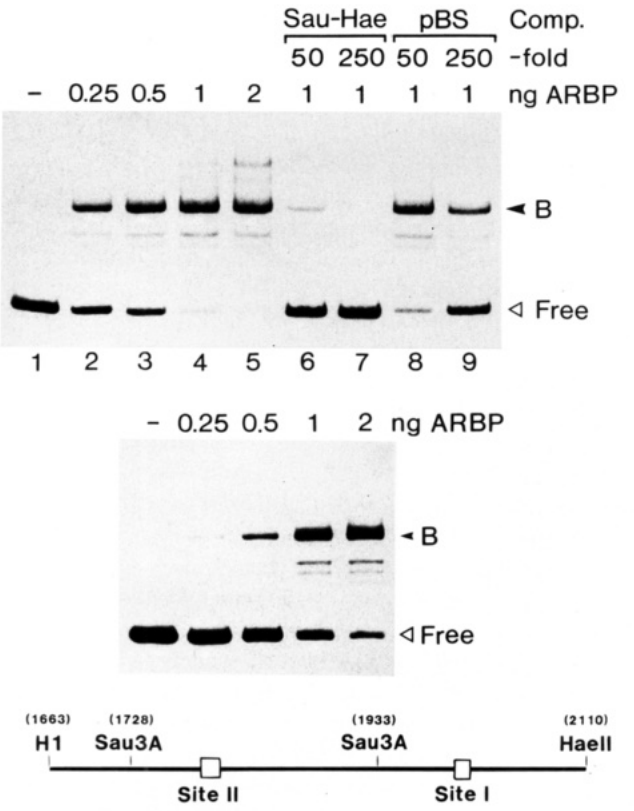

FIGURE 1: An electrophoretic mobility shift assay resolves defined complexes of ARBP with two MAR fragments. Labeled fragments *Sau*3A-*Hae*II (A, top panel) and *Sau*3A-*Sau*3A (B, middle panel) were incubated with the indicated amounts of purified ARBP and subjected to native 5% polyacrylamide gel electrophoresis. Where indicated, a 50- or 250-fold molar excess of unlabeled *Sau*3A-*Hae*II (*Sau-Hae*) (A, lanes 6 and 7) or of *Hin*fI-digested Bluescript (pBS) (A, lanes 8 and 9) was present during incubation. The positions of free DNA (Free) and those of the main ARBP−MAR fragment complexes (B) are marked. (C, bottom panel) Map of the 5′ MAR fragment H1-*Hae*II showing the positions of relevant restriction sites and of ARBP binding sites I and II. The numbers in parentheses refer to positions within the 5′ MAR fragment B-1-X1 defining the first nucleotide in the *Bam*HI site B-1 as position 1.

of intact ARBP (95 kDa) (data not shown) (Buhrmester, 1994). The fainter complexes visible in our mobility shift assays are thought to be formed with proteolytic fragments of ARBP, since their dissociation constants were identical to those of the main complexes (see below).

We then used electrophoretic mobility shift assays to determine the dissociation constants of the complexes between ARBP and each of the two MAR fragments. By titration of a fixed (2 ng) amount of ARBP with increasing amounts of labeled *Sau*3A-*Hae*II fragment and determination of the ^{32}P radioactivity in the position of free DNA and that of the main complex (Figure 2A, top), we calculated by a Scatchard analysis a dissociation constant (K_D) of the ARBP−*Sau*3A-*Hae*II DNA complex of 2.1×10^{-10} M (Figure 2A, bottom). Conversely, by varying the concentration of ARBP with a fixed amount of *Sau*3A-*Hae*II fragment (see Figure 1A), the K_D was estimated to be 2.7×10^{-10} M, which agrees closely with the previous experiment. The K_D of the ARBP−*Sau*3A-*Sau*3A DNA complex was estimated to be 6.7×10^{-10} and 4.2×10^{-10} M by using a fixed amount of ARBP (Figure 2B) and a fixed amount of the DNA fragment (see Figure 1B), respectively. Thus, ARBP binds to both sequences with surprisingly high affinity. The dissociation constants are in the same range as those of several transcription factors, such as nuclear factor I ($K_D = 2.4 \times 10^{-11}$ M) and the homeodomains of the *Drosophila* proteins Antennapedia ($K_D = 1.6 \times 10^{-9}$ M) and engrailed [$K_D = (1-2) \times 10^{-9}$ M] (Meisterernst et al., 1988; Affolter et al., 1990; Kissinger et al., 1990).

To identify the specific sequences involved in binding of ARBP, we performed DNase I footprinting experiments. Purified ARBP was incubated with fragment *Sau*3A-*Hae*II and digested briefly with DNase I, and footprints were analyzed on denaturing polyacrylamide gels. Using varying amounts of ARBP and varying concentrations of DNase I, we reproducibly found that a single site, named site I, was partially protected. The bars in Figure 3A (left panel) mark the footprinted regions in either strand. The degree of protection was quantitated by comparison of scans of ARBP−MAR DNA complex band patterns (lanes 4−7 and 11−14) to corresponding scans of naked DNA band patterns (lanes 2, 3, and 10) as 40%; one example for either strand is shown in Figure 3B (right panel). Though the footprints are thus relatively weak, we are nevertheless confident that they represent true binding sites of ARBP, since (i) the binding was shown to be of high affinity (see Figure 2), (ii) the band shift results in Figure 1 were remarkably clear, and (iii) we used a purified ARBP preparation (not a nuclear extract). Furthermore, identical results were obtained with a fragment containing a direct repeat of the sequence *Sau*3a-*Hae*II (see also Figure 6) (von Kries et al., 1991). This fragment contains two binding sites, and both were equally protected (data not shown) (Buhrmester, 1994). Since the failure to obtain more complete DNase I footprints may result from an inefficient formation of MAR DNA−ARBP complexes under the footprinting conditions, we performed control mobility shift assays with aliquots taken after DNase I digestion. At both ARBP concentrations used, the complexes contained 100% of the labeled fragments and had sustained DNase I digestion (data not shown). Thus, other explanations, e.g., a rapid on−off rate of protein binding, have to be envisaged for the failure to obtain more complete DNase I footprints.

DNase I footprinting experiments with fragment *Sau*3A-*Sau*3A also revealed a single partially protected site, named site II (Figure 4A, left panel). In site II, the protected regions, marked by bars, were slightly better visible than in site I. By densitometric scanning of ARBP−MAR DNA complex lanes with free DNA lanes the degree of protection in either strand was determined as 61% (Figure 4B, right panel). Identical results were obtained with a fragment containing a direct repeat of the sequence *Sau*3A-*Sau*3A (von Kries et al., 1991). Both binding sites in this fragment were equally protected (data not shown) (Buhrmester, 1994). An interesting observation is that, in the naked fragments, sites I and II each contain a very prominent DNase I cleavage site (marked by filled arrowheads in Figures 3A and 4A). In fact, the cleaved phosphoester bond in site II is the most prominent cutting site in the whole fragment. This suggests similarities in the mode of recognition of sites I and II by ARBP and in the mode of binding of DNase I to distinct DNA structures (see also Discussion).

In Figure 8, the footprinted sequence of site I is aligned with the 180°-rotated sequence of site II. The comparison reveals that both sites share two significant characteristics: a central GC-rich motif, 5′-GGTGT-3′ (shown in boldface), and flanking runs of AT-containing sequences. Further,

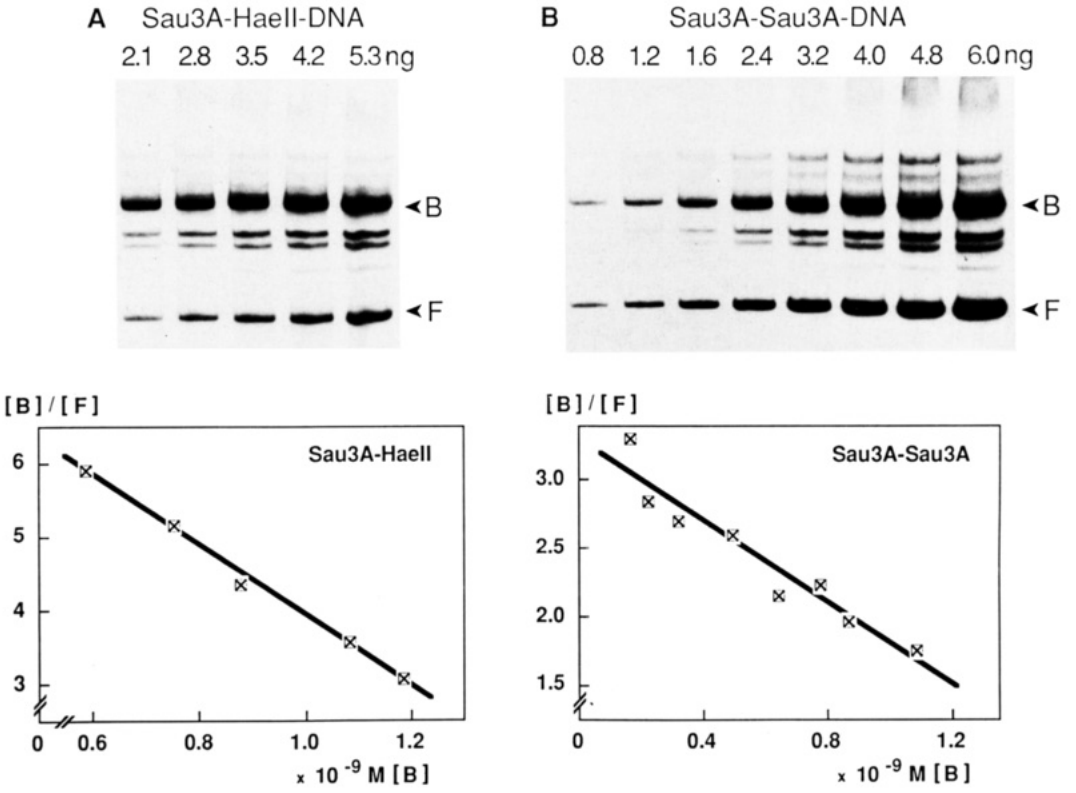

FIGURE 2: Dissociation constants of ARBP−MAR complexes. Purified ARBP (2 ng) was incubated with the indicated amounts of labeled fragment *Sau*3A-*Hae*II (A) or *Sau*3A-*Sau*3A (B) at 20 °C. Free (F) DNA and bound (B) DNA were separated on native 5% polyacrylamide gels (upper panels), and ^{32}P radioactivity in both types of DNA was determined by Cerenkov counting and used for Scatchard analysis (lower panels).

arrowheads mark the preferred DNase I cutting sites in naked site I and site II DNA fragments. These map to identical phosphoester bonds in the C-containing strand of the shared GC-rich motif. This common feature emphasizes a similarity in DNA helix structure of the two sites. Slightly DNase I-hypersensitive sites and further changes in the DNase I cleavage pattern on the flanks of footprinted sites I and II (indicated by stars and brackets, respectively) are indicative of alterations in the DNA structure about 20−25 bp remote from the binding site centers (Figures 3, 4, and 8).

We further used DMS methylation protection assays to identify the contact points of ARBP. Fragments *Sau*3A-*Hae*II or *Sau*3A-*Sau*3A were incubated with ARBP and then reacted with DMS, and the purified DNA fragments were subjected to Maxam−Gilbert G or G+A sequencing reactions and analyzed on sequencing gels. The left side of Figure 5A (left panel) shows representative results for the upper (G-containing) strand of site I in fragment *Sau*3A-*Hae*II. Comparison of the G or G+A band pattern in ARBP−MAR DNA complexes (lanes 2 and 4) with that in naked DNA (lanes 1 and 3) clearly shows that the two central guanines, G-2004 and G-2006, in the motif 5′-GGTGT-3′ were protected from methylation by DMS. This was confirmed by densitometric scanning (Figure 5B (right panel, top)]. Careful comparison of the band intensities at the 5′-GGTGT-3′ motif with those in the upstream and downstream flanking sequences further indicates that also G-2003 is slightly protected. The right side of Figure 5A shows an analysis of the lower (G-containing) strand of site II in fragment *Sau*3A-*Sau*3A. Comparison of the G ladder in ARBP−MAR DNA complexes (lanes 3 and 4) with that in naked DNA (lanes 1 and 2) demonstrates that similarly the two central guanines, G-1801 and G-1799, in the shared motif 5′-GGTGT-3′ were protected from methylation. Densitometric scanning verified this conclusion (Figure 5B, bottom). In additional DMS methylation interference assays, we obtained very similar results for site I as well as for site II (data not shown). Thus, we conclude that in each ARBP binding site identical guanines in the shared sequence 5′-GGTGT-3′ were identified as ARBP contact points, as shown by direct comparison in Figure 8.

To show directly that binding site I as defined by the above DNase I footprinting and methylation protection experiments is sufficient for binding of ARBP, we cloned a 41-bp synthetic fragment corresponding to the sequence from coordinates 1986−2026 into the vector pBend2 and recovered the binding site together with flanking vector sequences in a 163-bp fragment. This fragment bound ARBP as efficiently as fragment *Sau*3A-*Hae*II (Figure 6A), giving definite proof that the 41-bp site I is sufficient for binding of ARBP. Furthermore, this result allows specification of the above dissociation constant measurements by stating that ARBP binds with high affinity to site I rather than solely to fragment *Sau*3A-*Hae*II. The 41-bp site I sequence was further used in a mutational analysis to explore in more detail the importance of the GC-rich core sequence for ARBP binding. First, the central three bases (5′-GTG-3′) were substituted by 5′-TCT-3′, creating an AT-rich mutant sequence (mutant 1). A 163-bp fragment containing this mutated sequence showed only marginal binding of ARBP (Figure 6B). To underscore this result, we also performed competition band shift experiments. Figure 6D, lane 2, shows that the nonmutated 41-bp fragment (at a 300-fold molar excess) competed efficiently with the 163-bp fragment, which contains the 41-bp sequence. As expected, a mutant-1 41-bp fragment was unable to compete (Figure 6D, lane 3),

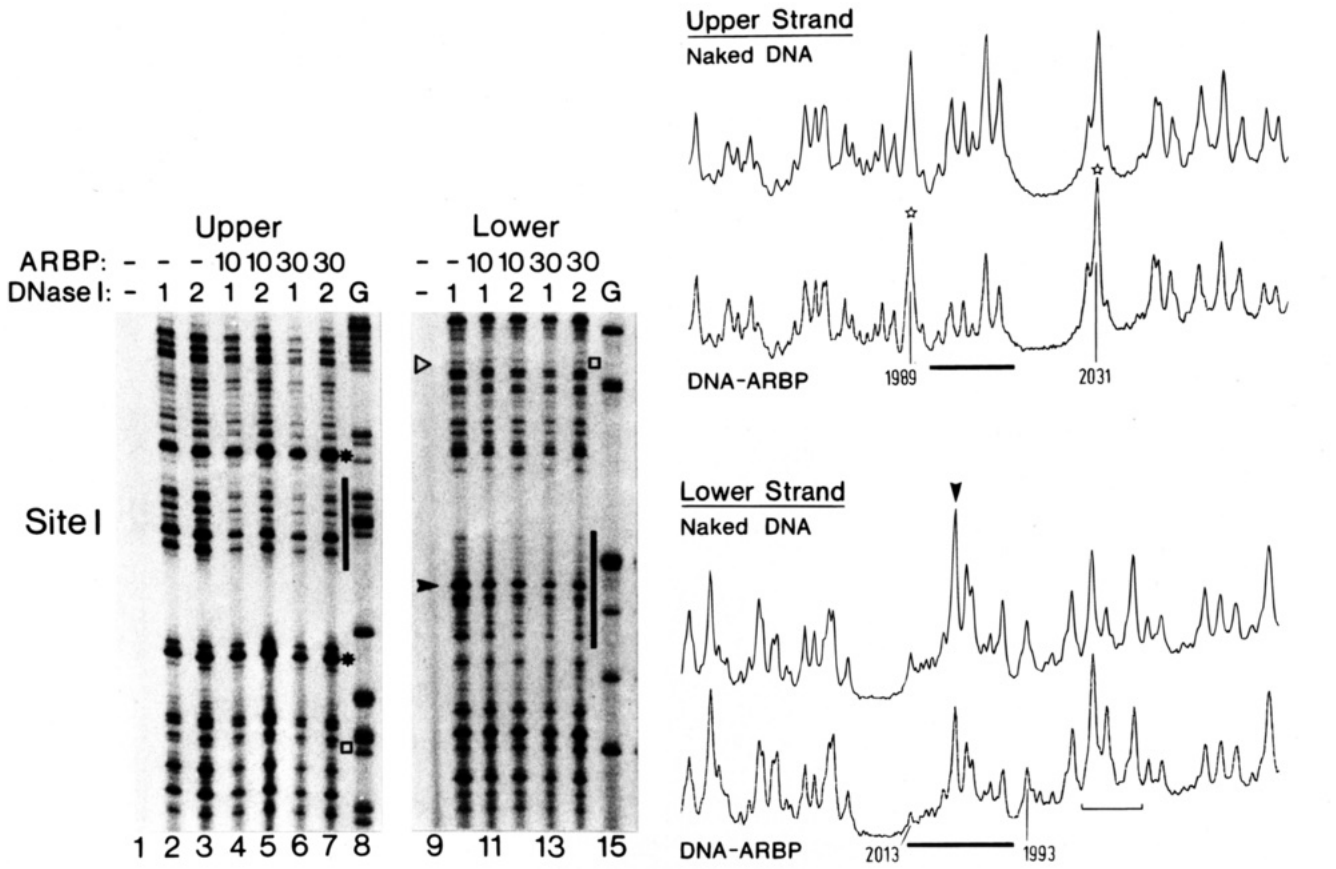

FIGURE 3: DNase I footprinting of ARBP binding site I in *Sau*3A-*Hae*II MAR DNA. (A, left panel) End-labeled *Sau*3A-*Hae*II DNA was incubated with 10 or 30 ng of purified ARBP and digested with 1×10^{-7} or 2×10^{-7} g of DNase I/mL. For comparison, naked DNA was digested with 1×10^{-7} or 2×10^{-7} g of enzyme/mL (lanes 2, 3, and 10). Lanes 1 and 9 (controls) contain DNA, to which no ARBP and no DNase I was added. Lanes 8 and 15 contain chemically G-specific cleaved DNA. Upper and lower strands are displayed in the left and right panels, respectively, of panel A. Footprinted regions (site I) are indicated by bars and two DNase I hypersensitive sites by stars. A preferred DNase I cleavage site in the lower strand of naked DNA (lane 10) is marked by a filled arrowhead. The open squares indicate a site containing the sequence 5′-GGTGT-3′ that is not recognized by ARBP and not cleaved by DNase I (lane 10, open triangle). (B, right panel) Representative DNase I cleavage patterns of naked DNA (scans of lanes 2 and 10) were compared with corresponding ones of ARBP−MAR DNA complexes (scans of lanes 4 and 14). Top, upper strand; bottom, lower strand. The bracket indicates a region where the DNase I cleavage pattern is altered by ARBP binding. The numbering system is as in Figure 1C.

emphasizing that the central sequence, 5′-GTG-3′, in the GC-rich core motif is critical for binding of ARBP. We further synthesized mutated 41-bp fragments substituted individually at each of the three G's of the core binding sequence (5′-GGTGT-3′). Substitution of the first and the third G resulted in a reduced competence to compete, indicating that these are critical for binding of ARBP (Figure 6D, lanes 4 and 6). A 41-bp fragment substituted in the second G competed as efficiently as the nonmutated fragment, suggesting that this G might be less critical for binding. However, since this assay requires a 300-fold excess of the 41-bp fragment over the 163-bp fragment, we thought that small changes in the binding affinity might be undetectable in this assay. We therefore directly analyzed binding of the mutant-3 sequence with attached vector sequences as a 163-bp fragment. Comparison of Figure 6C with Figure 6A shows that substitution of the second G reduced binding of ARBP about 4-fold. We conclude that each individual guanine of the central 5′-GGTGT-3′, though to a different degree, is important for ARBP binding.

To analyze the significance of the AT-containing runs in the flanks of ARBP binding sites I and II, we performed mobility shift assays using as competitors the fluorochrom Hoechst 33258 and the antiviral drug netropsin, that specifically bind to AT-rich regions (Zimmer & Wähnert, 1986). The DNA fragment employed in this experiment harbors two copies of the sequence *Sau*3A-*Hae*II and consequently two ARBP binding sites I (von Kries et al., 1991). At low concentrations of ARBP, it forms a complex (I) containing one molecule of ARBP and, at higher concentrations, a complex (II) containing two molecules of ARBP (data not shown). This demonstrates that each binding site in the dimer has the potential to bind ARBP. In Figure 7, we used an intermediate concentration (1 ng) that generated both complexes (lanes 2). As shown in Figure 7A (top panel), 7.5−30 nM Hoechst 33258 effectively competed with binding of ARBP to 1 ng (75 nM of base pairs) of labeled fragment. These concentrations of the dye have previously been demonstrated to efficiently compete for binding of HMG proteins I and Y and of peptide S6 from the sea urchin spermatogenous histone H1 (Suzuki, 1989; Reeves & Nissen, 1990). Binding of Hoechst 33258 (or netropsin) alone did not produce any change in mobility of the DNA fragment (data not shown). Figure 7B (bottom panel), lane 3, shows that netropsin completely inhibited binding of ARBP at a netropsin-to-base-pair-tetramer ratio of <0.2. Since the drug specifically recognizes tetrameric runs of AT-containing sequences with high affinity (Zimmer & Wähnert, 1986), the low ratio convincingly shows that netropsin inhibits

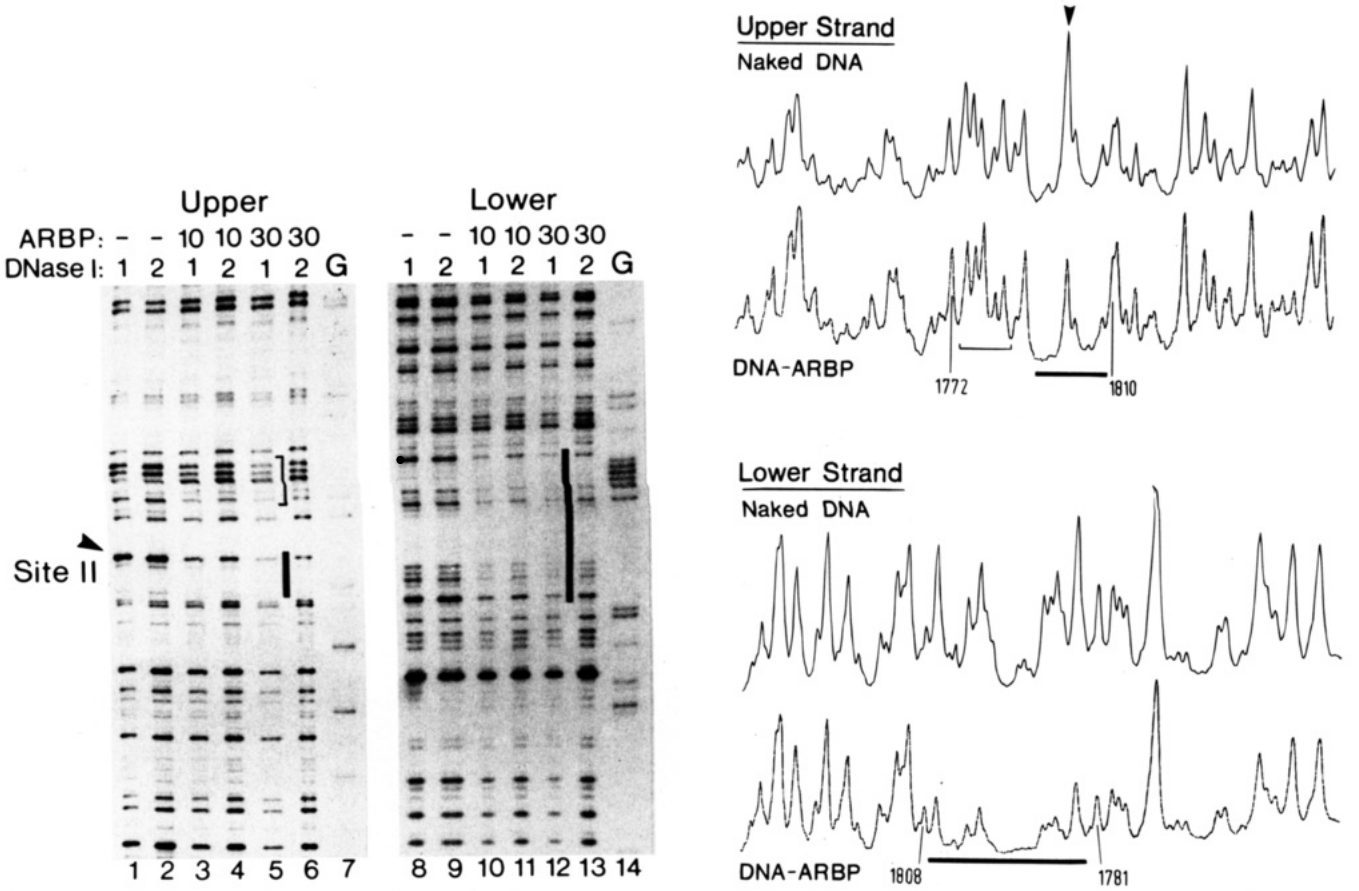

FIGURE 4: DNase I footprinting of ARBP binding site II in *Sau*3A-*Sau*3A MAR DNA. (A, left panel) End-labeled *Sau*3A-*Sau*3A DNA was incubated with 10 or 30 ng of purified ARBP and digested with 1×10^{-7} or 2×10^{-7} g of DNase I/mL. For comparison, naked DNA was digested with 1×10^{-7} or 2×10^{-7} g of enzyme/mL (lanes 1, 2, 8, and 9). Lanes 7 and 14 contain chemically G-specific cleaved DNA. Footprinted regions (site II) are indicated by bars. A preferred DNase I cleavage site in the upper strand of naked DNA (lanes 1 and 2) is marked by an arrowhead. The bracket indicates a region of altered DNase I cleavage pattern due to ARBP binding. (B, right panel) DNase I cleavage patterns of naked DNA (scans of lanes 1 and 8) were compared with corresponding ones of ARBP−MAR DNA complexes (scans of lanes 5 and 12). Top, upper strand; bottom, lower strand.

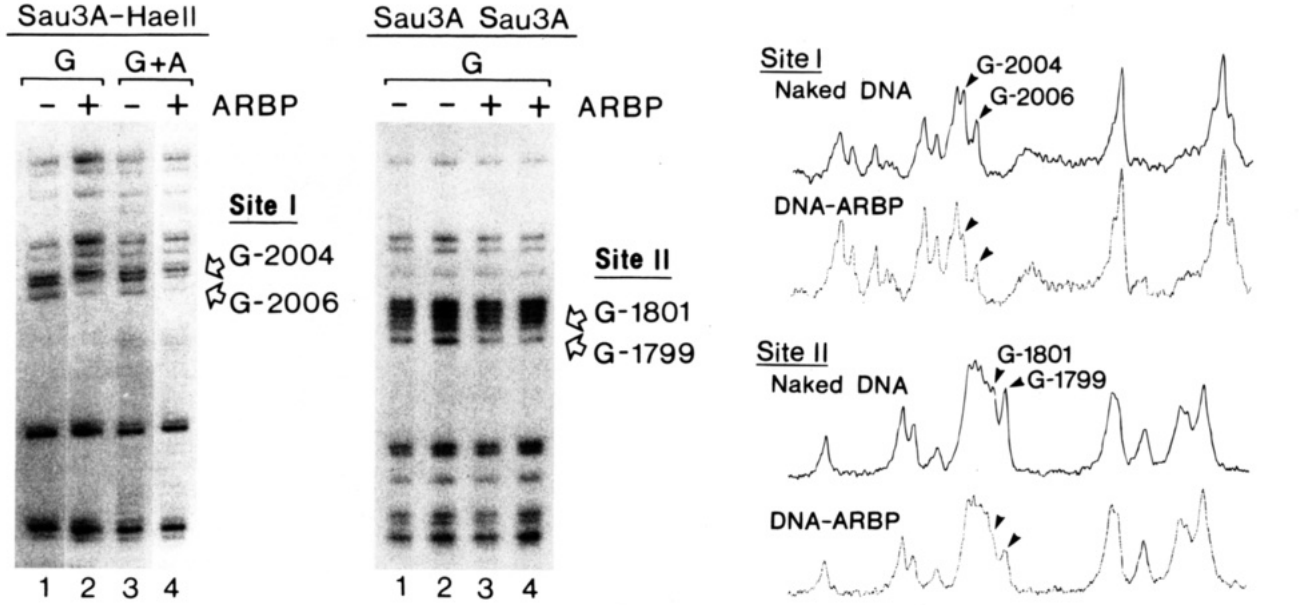

FIGURE 5: Methylation protection assay of ARBP binding sites in MAR DNA. (A, left panel, left side) *Sau*3A-*Hae*II DNA, end-labeled in the upper strand, was not (lanes 1 and 3) or was incubated with 30 ng of purified ARBP (lanes 2 and 4), reacted with DMS for 2 min, cleaved following the G (lanes 1 and 2) or G+A (lanes 3 and 4) protocol of Maxam and Gilbert (1980), and analyzed on a sequencing gel. (Right side of panel A) *Sau*3A-*Sau*3A DNA, end-labeled in the lower strand, was not (lanes 1 and 2) or was incubated with 50 ng of ARBP (lanes 3 and 4), reacted with DMS for 1 min (lanes 1 and 3) or 2 min (lanes 2 and 4), cleaved following the G protocol, and analyzed on a sequencing gel. Guanines protected from methylation by ARBP binding are indicated. (B, right panel) Sequencing ladders without and with prior incubation with ARBP were compared. Top, site I (scans of lanes 3 and 4); bottom, site II (scans of lanes 2 and 4).

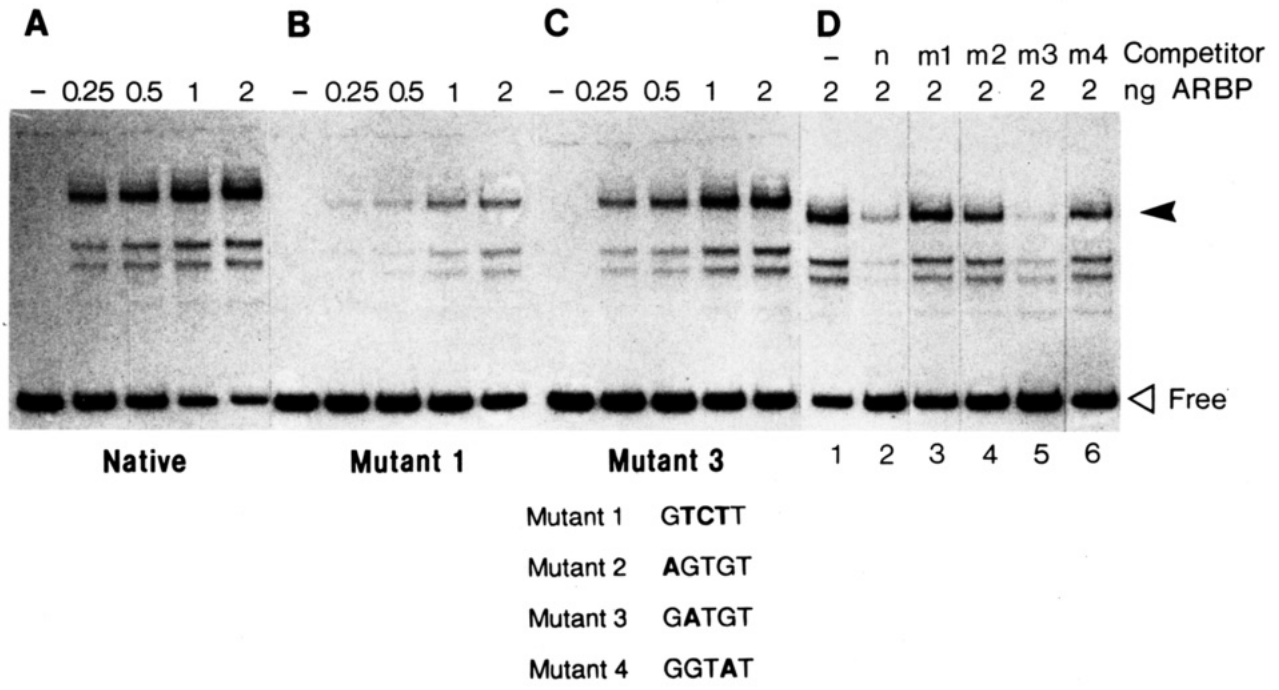

FIGURE 6: Mutational analysis of the ARBP binding core motif. (A) A labeled 163-bp fragment containing the 41-bp site I sequence from coordinates 1986−2026 with attached sequences of the vector pBend2 was incubated with the indicated amounts of purified ARBP and electrophoresed on native 6% polyacrylamide gels. In (B) and (C), the 163-bp fragments used contain mutated ARBP binding sites I, in which the core sequence (5′-GGTGT-3′) is substituted by the sequences 5′-GTCTT-3′ and 5′-GATGT-3′, respectively (Mutants 1 and 3). (D) In competition band shift experiments, labeled 163-bp fragment containing the native 41-bp site I sequence was incubated in the absence (lane 1) or presence of a 300-fold molar excess of the native (n) 41-bp fragment (lane 2) or 41-bp fragments containing the indicated mutations 1−4 (m1−m4) of the core binding sequence (5′-GGTGT-3′) (lanes 3−6).

binding of ARBP by interacting with AT-clusters flanking site I.

Since many transcription factors and structural proteins have been described to bend target DNA fragments [as an example, see Kerppola and Curran (1991)], we employed a circular permutation analysis to explore a possible bending of the site I DNA sequence by ARBP (Wu & Crothers, 1984). The 41-bp site I sequence from coordinates 1986 to 2026 cloned into the circularly permutated sequence of vector pBend2 was recovered by use of different restriction enzymes into 163-bp fragments containing binding site I at different positions of the fragment ends. All fragments formed high-affinity complexes with ARBP, but of identical electrophoretic mobilities, showing that this assay could not detect a measurable distortion in DNA structure affected by ARBP binding (Figure 7A and data not shown) (Buhrmester, 1994).

DISCUSSION

ARBP Recognizes a Novel DNA Sequence Motif. The experiments reported in this paper show that ARBP binds to the MAR fragment H1-*Hae*II at two sites. These share several sequence and structural features that are summarized in Figure 8. First, the centers of both sites contain the sequence motif 5′-GGTGT-3′. The DMS methylation protection experiments show that specifically the two central guanines of the motif are recognized through contacts in the major groove. Substitution of the central three bases of the 5′-GGTGT-3′ motif by an AT-rich sequence abolished binding of ARBP. After single-base substitutions of each individual guanine of the motif, the mutated sequences showed significant reductions in the binding affinity of ARBP. These results underscore the importance of the central GC-rich 5′-GGTGT-3′ motif for ARBP binding. A specific phosphodiester bond in the ACACC-containing strand of both binding sites is a highly preferred cleavage site for DNase I. DNase I cleavage is known to be very sensitive to local variations of the DNA secondary structure, and cleavage rates can vary as much as 500-fold. One critical parameter recognized is the width of the minor groove. The highest frequency of cutting occurs within regions where the minor groove width is closest to the intermediate or even wider values for B-DNA (Drew & Travers, 1984; Suck & Oefner, 1986; Suck et al., 1988). A second critical parameter is the DNA stiffness which resists bending toward the major groove, i.e., the energy required to form a bent DNA−enzyme complex (Hogan et al., 1989; Brukner et al., 1990). The structure of a nicked octanucleotide complex with DNase I revealed widening of the minor groove by 3 Å and a 21.5° bend of the DNA, away from the bound enzyme toward the major groove (Suck & Oefner, 1986). This suggests that distinct structural features, e.g., an elevated flexibility of the DNA helix, are shared by both ARBP binding sites and, furthermore, are commonly recognized by DNase I and ARBP.

A second feature shared by both binding sites are flanking AT-rich clusters that are critical for binding of ARBP. Since homopolymers of dA·dT tetramers, such as AATT and AAAA, rank as the highest affinity binding sites of netropsin (Zimmer & Wähnert, 1986), the efficient competition of ARBP binding at the low ratio of 0.2 mol of netropsin to mole of base-pair-tetramers indicates that AT-homopolymers play a major role for binding of ARBP. Since netropsin and Hoechst 33258 are minor groove binders (Zimmer & Wähnert, 1986; Teng et al., 1988), our competition experiments may indicate that ARBP binds through additional contacts in the minor groove. Transcription factors from

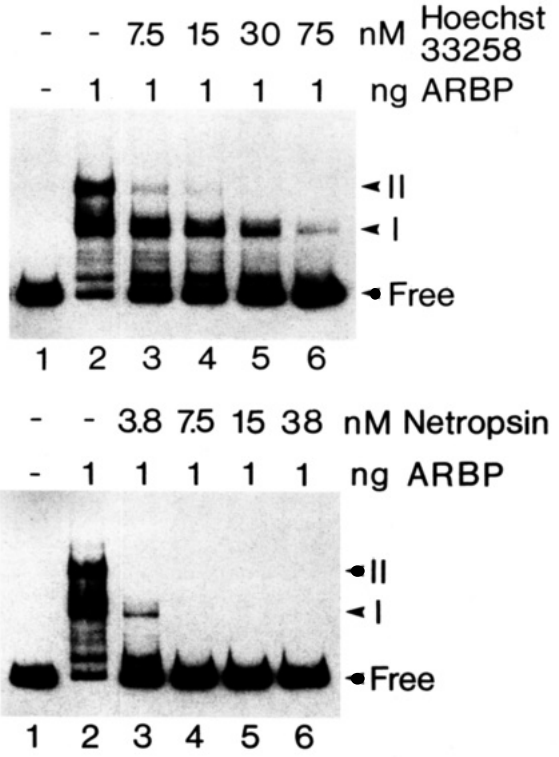

FIGURE 7: Inhibition of binding of ARBP by Hoechst 33258 and netropsin. Labeled fragments containing the dimerized sequence *Sau*3A-*Hae*II (1 ng, equivalent to 75 nM of base pairs) were incubated without (lanes 1) or with 1 ng of purified ARBP (lanes 2−6) in the presence of the indicated concentrations of Hoechst 33258 (A, top panel) or netropsin (B, bottom panel) and subjected to native 3.5% polyacrylamide gel electrophoresis. Complexes containing one (I) or two (II) molecules of ARBP are marked.

diverse classes (e.g., homeodomain proteins and HNF-3) have previously been shown to make major as well as minor groove interactions (Kissinger et al., 1990; Clark et al., 1993). However, the alternative explanation that competition by Hoechst 33258 and netropsin might be caused by an induced conformational change of the DNA cannot be excluded (Dorn et al., 1992). X-ray crystal analysis of complexes of netropsin and Hoechst 33258 with oligomers has shown that these ligands induce a widening of the minor groove and bending of the helix away from the attachment site (Kopka et al., 1985; Pjura et al., 1987). It is possible that the ligand-induced structural changes are even stronger in solution due to a higher flexibility (Triebel et al., 1994). The sequence 5′-GGTGT-3′ occurs twice more in the fragment H1-*Hae*II, and two additional sequences are located shortly upstream of the H1-site. However, none of these sites was recognized by ARBP, and furthermore, they are not distinguished by flanking AT-containing homopolymers and were not preferentially cleaved by DNase I. As an example, the open squares in Figure 3 mark one of these sequences, while the open triangle indicates the absence of DNase I hypersensitivity in the naked DNA fragment (lane 10). These results clearly show that distinct, AT-rich sequences flanking the central GGTGT-motif are important for binding of ARBP. The absence of a more extended sequence specificity furthermore indicates that structural features of the binding site are significant elements for recognition. These results agree with the characterization of MARs/SARs as AT-rich sequences exhibiting very little sequence similarity (Gasser & Laemmli, 1987; Phi-Van & Strätling, 1990).

The sequence 5′-GGTGT-3′ flanked by AT-rich sequences is (to our knowledge) a novel recognition motif, not previously found as a motif recognized by transcription factors or nuclear proteins with structural functions. This would correlate with amino acid sequence data of the DNA binding domain of ARBP and secondary structural prediction that gave no sign for a known DNA binding motif (J. Weitzel and W. H. Strätling, to be published). These data also indicate that ARBP, unlike the 120 kDa chicken MAR binding protein, is not related to a known heterogeneous nuclear ribonucleoprotein (von Kries et al., 1994a).

MARs have been generally characterized as AT-rich sequences. For a trivial reason, however, AT-richness per se is not a sufficient criterium for specific sequence recognition of MAR binding proteins. While MARs are relatively extended sequences (300 to several thousand base pairs in length), MAR binding proteins (in the form of monomers or oligomers) can bind only to much shorter sequences, and short AT-rich sequences also occur in other types of genomic sequences. In line with this reasoning, we found that ARBP reconizes a GC-rich central core sequence. It was also reported that DNA topoisomerase II, a major component of the metaphase scaffold, cleaves the *Drosophila* histone SAR in vivo at sites where the AT/GC ratio is approximately

FIGURE 8: Summary of the footprinting data. The sequence of binding site I is aligned with that of the 180°-rotated site II according to the central sequence motif 5′-GGTGT-3′ (shown in boldface). Bases protected from DNase I digestion are indicated by bars above and below the sequence. Brackets mark regions of altered DNase I cleavage pattern, and stars mark DNase I hypersensitive sites. Arrowheads indicate phosphoester bonds that are highly sensitive to cleavage by DNase I in naked DNA. Bases protected from DMS methylation are marked by open arrows.

balanced, but that are flanked by AT-containing sequences (Käs & Laemmli, 1992). Furthermore, the human MAR binding protein SATB1 recognizes a special AT-rich sequence, where A's and T's are mixed with C's (Dickinson et al., 1992).

How Does ARBP Function in the Formation of Chromatin Loop-Domains? For a protein with supposedly structural functions, ARBP has several unusual properties. First, ARBP is solubilized from interphase nuclei with 200−300 mM NaCl, i.e., at significantly lower salt concentrations than histones, in particular histone H1 (von Kries et al., 1991). Second, ARBP recognizes a relatively precisely defined DNA sequence (Figure 8). Third, ARBP binds to its cognate binding sites with high affinity (Figure 2). These properties are routinely known as features of transcription factors. Transcription factors that bind to remote enhancer sequences are thought to act by protein−protein interactions with general transcription factors in the transcription initiation complex by looping out the DNA in between (Ptashne, 1988; Tjian & Maniatis, 1994). Presently, it is certainly difficult to detect chromatin loops at defined genetic loci. In view of this difficulty, it is extremely valuable that the juxtaposition of the enhancer and the promoter region of the rat prolactin gene, though separated by several hundred base pairs, could be demonstrated by use of a nuclear ligation assay (Cullen at al., 1993). From the biochemical properties shared by ARBP and transcriptional factors, it may be expected that they act in an analogous manner. ARBP might form loop-domains through protein−protein interactions involving an as yet unidentified effector domain. The nuclear protein RAP-1, a component of the yeast chromosomal scaffold, that binds to diverse genomic loci, represents an example for a nuclear matrix/scaffold protein, where loop formation between different silencers and between silencer and promoter sequences of the mating type locus HML has been demonstrated already in vitro (Hofmann et al., 1989).

When nuclei are extracted with high salt (2 M NaCl) and then treated with DNase I, ARBP partitions with the insoluble fraction that has been named nuclear matrix (Berezney & Coffey, 1974). Thus, it appears to be a paradox that ARBP is (partially) extractable from intact nuclei with 400 mM NaCl (von Kries et al., 1991, 1994b). This paradox can be resolved, however, by considering that extraction of the chromatin masses from nuclei by the method of Berezney and Coffey (1974) grossly perturbs the balance of the ionic and nonionic interactions among the remaining components. These perturbations may easily lead to the formation of insoluble aggregates of otherwise soluble components. It should be also kept in mind that ARBP is only partially extracted with 400 mM NaCl and it routinely appears as a doublet in denaturing gels (von Kries et al., 1991). Thus, there may exist multiple forms and/or populations of ARBP that have different chromosomal locations and fulfill different functions.

A specific chromatin structure likely assists the function of ARBP in loop-domain formation. Interestingly, binding sites I and II are separated by 200 bp (see map in Figure 1C). Thus, a positioned nucleosome would allow both sites to be located in accessible linker regions. Such an organization would also bring both binding sites into close proximity, thereby facilitating a cooperative interaction of ARBP with its DNA binding sites or its protein targets (von Kries et al., 1991).

ACKNOWLEDGMENT

We thank B. Huckschlag and K. Weber for technical assistance and T. Pillar for reading the manuscript. This paper is based on the doctoral thesis of H. Buhrmester in the Faculty of Biology, University of Hamburg.

REFERENCES

Adachi, Y., Käs, E., & Laemmli, U. K. (1989) *EMBO J. 8*, 3997−4006.
Affolter, M., Percival-Smith, A., Müller, M., Leupin, W., & Gehring, W. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4093−4097.
Berezney, R., & Coffey, D. S. (1974) *Biochem. Biophys. Res. Commun. 60*, 1410−1417.
Bidwell, J. P., van Wijnen, A. J., Fey, E. G., Dworetzky, S., Penman, S., Stein, J. L., Lian, J. B., & Stein, G. S. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3162−3166.
Breyne, P., Van Montagu, M., Depicker, N., & Gheysen, G. (1992) *Plant Cell 4*, 463−471.
Brukner, I., Jurukovski, V., & Savic, A. (1990) *Nucleic Acids Res. 18*, 891−894.
Buhrmester, H. (1994) Dissertation, University of Hamburg.
Clark, K. L., Halay, E. D., Lai, E., & Burley, S. K. (1993) *Nature 364*, 412−420.
Cockerill, P. N., & Garrard, W. T. (1986) *Cell 44*, 273−282.
Cook, P. R. (1991) *Cell 66*, 627−635.
Cullen, K. E., Kladde, M. P., & Seyfred, M. A. (1993) *Science 261*, 203−206.
Dickinson, L. A., Joh, T., Kohwi, Y., & Kohwi-Shigematsu, T. (1992) *Cell 70*, 631−645.
Dorn, A., Affolter, M., Müller, M., Gehring, W. J., & Leupin, W. (1992) *EMBO J. 11*, 279−286.
Drew, H. R., & Travers, A. A. (1984) *Cell 37*, 491−502.
Gasser, S. M., & Laemmli, U. K. (1987) *Trends Genet. 3*, 16−22.
Grewal, T., Theisen, M., Borgmeyer, U., Grussenmeyer, T., Rupp, R. A. W., Stief, A., Quian, F., Hecht, A., & Sippel, A. E. (1992) *Mol. Cell. Biol. 12*, 2339−2350.
Hirano, T., Kinoshita, N., Morikawa, K., & Yanagida, M. (1990) *Cell 60*, 319−328.
Hofmann, J. F.-X., Laroche, T., Brand, A. H., & Gasser, S. M. (1989) *Cell 57*, 725−737.
Hogan, M. E., Roberson, M. W., & Austin, R. H. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9273−9277.
Käs, E., Poljak, L., Adachi, Y., & Laemmli, U. K. (1993) *EMBO J. 12*, 115−126.
Kellum, R., & Schedl, P. (1991) *Cell 64*, 941−950.
Kerppola, T. K., & Curran, T. (1991) *Science 254*, 1210−1214.
Kim, J., Zwieb, C., Wu, C., & Adhay, S. (1989) *Gene 85*, 15−23.
Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., & Pabo, C. O. (1990) *Cell 63*, 579−590.
Klehr, D., Maass, K., & Bode, J. (1991) *Biochemistry 30*, 1264−1270.
Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., & Dickerson, R. E. (1985) *J. Mol. Biol. 183*, 553−563.
Levy-Wilson, B., & Fortier, C. (1990) *J. Biol. Chem. 264*, 21196−21204.
Ludérus, M. E. E., de Graaf, A., Mattia, E., den Blaauwen, J. L., Grande, M. A., de Jong, L., & van Driel, R. (1992) *Cell 70*, 949−959.
Maxam, A., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499−560.
McKnight, R., Shamay, A., Sankaran, L., Wall, R. J., & Hennighausen, L. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6943−6947.
Meisterernst, M., Gander, I., Rogge, L., & Winnacker, E.-L. (1988) *Nucleic Acids Res. 16*, 4419−4435.
Mirkovitch, J., Mirault, M.-E., & Laemmli, U. K. (1984) *Cell 39*, 223−232.
Phi-Van, L., & Strätling, W. H. (1988) *EMBO J. 7*, 655−664.
Phi-Van, L., & Strätling, W. H. (1990) *Prog. Mol. Subcell. Biol. 11*, 1−11.
Phi-Van, L., von Kries, J. P., Ostertag, W., & Strätling, W. H. (1990) *Mol. Cell. Biol. 10*, 2302−2307.

Pjura, P. E., Grzeskowiak, K., & Dickerson, R. E. (1987) *J. Mol. Biol. 197,* 257−271.
Ptashne, M. (1988) *Nature 335,* 683−689.
Reeves, R., & Nissen, M. S. (1990) *J. Biol. Chem. 265,* 8573−8582.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Stief, A., Winter, D. M., Strätling, W. H., & Sippel, A. E. (1989) *Nature 341,* 343−345.
Strätling, W. H., Dölle, A., & Sippel, A. E. (1986) *Biochemistry 25,* 495−502.
Suck, D., & Oefner, C. (1986) *Nature 321,* 620−625.
Suck, D., Lahm, A., & Oefner, C. (1988) *Nature 332,* 464−468.
Suzuki, M. (1989) *EMBO J. 8,* 797−804.
Teng, M., Usman, N., Frederick, C. A., & Wang, A. H.-J. (1988) *Nucleic Acids Res. 16,* 2671−2690.
Tjian, R., & Maniatis, T. (1994) *Cell 77,* 5−8.
Triebel, H., Bär, H., Walter, A., Burckhardt, G., & Zimmer, C. (1994) *J. Biomol. Struct. Dyn. 11,* 1085−1105.
Udvardy, A., Maine, E., & Schedl, P. (1985) *J. Mol. Biol. 185,* 341−358.
von Kries, J. P., Buhrmester, H., & Strätling, W. H. (1991) *Cell 64,* 123−135.
von Kries, J. P., Buck, F., & Strätling, W. H. (1994a) *Nucleic Acids Res. 22,* 1215−1220.
von Kries, J. P., Rosorius, O., Buhrmester, H., & Strätling W. H. (1994b) *FEBS Lett. 342,* 185−188.
Weintraub, H., & Groudine, M. (1976) *Science 193,* 848−856.
Wu, H.-M., & Crothers, D. M. (1984) *Nature 308,* 509−513.
Zimmer, C., & Wähnert, U. (1986) *Prog. Biophys. Mol. Biol. 47,* 31−112.

BI941663U