3252–3259 Nucleic Acids Research, 1995, Vol. 23, No. 16

Isolation and characterization of the chicken homeodomain protein AKR

Aimee K. Ryan+, Max L. Tejada, Donna L. May, Marianna Dubaova and Roger G. Deeley*

Cancer Research Laboratories, Queen's University, Kingston, Ontario K7L 3N6, Canada

Received April 24, 1995; Revised and Accepted July 4, 1995

Genbank accession no. U25353

ABSTRACT

Expression of the avian apoVLDLII gene is liver specific and completely dependent on estrogen. Previous analyses of protein binding sites in the apoVLDLII promoter revealed interactions between liver-enriched and ubiquitous factors at a location, site F′, between nucleotides –229 and –260 relative to the major transcriptional start site. Site-directed mutagenesis of G residues contacted by these factors decreased expression from the promoter ~5-fold in the avian hepatoma cell line LMH2A. We have used this site to screen a cDNA expression library constructed from day 9 embryonic liver RNA. One of the two DNA binding factors isolated is a novel homeodomain protein. With the exception of the homeodomain itself, which is atypically located close to the protein N-terminus, the factor displays little similarity to any known DNA binding protein. Its homeodomain is most similar to that of the maize protein Knotted-1, while the most closely related vertebrate domain is that of the human proto-oncoprotein Pbx1. We demonstrate that the DNA binding specificity of the factor is consistent with its involvement in the ubiquitous complex formed with site F′ and that it is capable of suppressing expression from the apoVLDLII promoter in short-term transfection experiments. These studies, combined with its DNA binding specificity, the tissue distribution of its mRNA and its developmental regulation, suggest a role as a negative regulator of gene expression in non-hepatic tissues and in the liver early during embryogenesis.

INTRODUCTION

The homeodomain was originally identified as a conserved region in genes responsible for homeotic mutations in *Drosophila* (1,2) and it has since been found in DNA binding proteins in organisms ranging from plants to vertebrates. It is a 60–61 amino acid domain which generally recognizes and binds to A/T-rich sequences (3) in a manner similar to the helix–turn–helix motif of prokaryotic gene regulatory proteins (4). *In vitro* many homeodomain proteins are capable of binding to the same sequence (5) and a single homeodomain can bind to very degenerate binding sites (6). Since many vertebrate homeodomain proteins have been isolated by virtue of the relatively highly conserved nature of the homeodomain, the specific DNA elements to which they bind and their function with respect to gene regulation remain unknown. In the studies presented here we describe the cloning and characterization of an unusual homeodomain protein that binds to a regulatory element in the promoter of the liver-specific, estrogen-dependent very low density apolipoprotein II (apoVLDLII) gene.

The apoVLDLII gene encodes a major egg yolk apolipoprotein and it is normally active only in the livers of laying hens (7). However, hepatic expression of the gene can be activated in embryos, chicks and roosters by the administration of exogenous estrogen (8,9). A developmental window has been defined, between days 7 and 14 of embryogenesis, during which the liver acquires the ability to express first the apoVLDLII gene and then the vitellogenin genes in response to hormone (9–11). By day 7 of embryogenesis mRNAs encoding other liver-specific proteins, such as serum albumin and apolipoprotein B (apoB), can already be detected (9). Furthermore, expression of the apoB gene also responds to exogenous estrogen at this stage, indicating that the delayed ability to activate the apoVLDLII and vitellogenin genes is not simply attributable to the lack of functional hepatic estrogen receptors (11). These observations suggest that competence to express individual genes or subsets of genes, characteristic of mature liver, may be influenced by the developmental stage at which different *trans*-acting factors are produced.

In addition to its estrogen response elements, binding sites for several ubiquitous and liver enriched *trans*-acting factors have been identified in the apoVLDLII proximal promoter, including sites for the liver-enriched factor C/EBPα (12,13). We have previously demonstrated that the F′ element, which is located between –229 and –260 relative to the major apoVLDLII transcription start site and immediately upstream from a previously identified imperfect estrogen response element, is critical for maximal estrogen inducible expression from the apoVLDLII promoter (14). Deletion or mutation of the F′ site results in a >5-fold decrease in activity of the promoter. We have identified two major, chromatographically separable F′ binding activities in liver nuclear protein extracts (14). One of these is liver enriched and the other is expressed at variable levels in a range of tissues. The abundance of the liver-enriched factor increases dramatically

* To whom correspondence should be addressed

+Present address: University of California at San Diego, CMM 345, 9500 Gilman Drive, 0648, La Jolla, CA 92093, USA

between days 7 and 9 of embryogenesis. In contrast, the factor(s) responsible for the ubiquitous complex is already present at significant levels in the liver by day 7.

Direct screening of a λgt11 cDNA expression library constructed from day 9 embryonic liver RNA was used to isolate cDNA clones encoding two factors capable of interacting with the F′ binding site. One of the proteins is the homolog of C/EBPα. The other factor, which we describe here, contains a homeodomain, but has structural features distinct from the previously described classes of these proteins. We demonstrate that the specificity with which the protein binds to normal and mutated F′ sites is consistent with its involvement in the ubiquitous complex detected in nuclear protein extracts. We also describe the structure of the protein, the tissue specificity and developmental profile of expression of its mRNA and its ability to negatively regulate expression from the apoVLDLII promoter. Since the DNA binding domain of the protein is most similar to that of the maize protein Knotted-1, we have named it Avian Knotted-Related (AKR) protein (15).

MATERIALS AND METHODS

Electrophoretic mobility shift assay (EMSA)

EMSA assays were performed using synthetic oligonucleotide binding sites which were end-labeled and incubated on ice for 20–30 min with liver nuclear protein extracts (14) or the products of *in vitro* transcription/translation. Incubations were carried out in 25 μl containing 25 mM HEPES, pH 8.0, 12.5 mM $MgCl_2$, 50 mM KCl, 20% glycerol, 2 mM DTT and 100 ng poly(dI·dC). Typically 2–4 fmol binding site and 1–2 μg extract were used per assay. Free and bound oligonucleotides were separated by electrophoresis (80 V for 3–4 h) through 6% polyacrylamide gels in 0.5× TAE or 0.25× TBE. The gels were dried and subjected to autoradiography.

In some experiments anti-AKR rabbit polyclonal antiserum, which was raised against a GST–AKR bacterial fusion protein, was added to the binding reaction. A 459 bp PCR fragment spanning the homeobox sequence was cloned into *Bam*HI/*Sma*I-digested expression vector pGEX-2T. The fusion protein was purified from lysates of transformed *Escherichia coli* DH5α using standard protocols recommended by Pharmacia. To generate polyclonal antibodies against GST–AKR rabbits were injected with 100 μg purified fusion protein in Freund's complete adjuvant and boosted twice with 40 μg protein. The immune serum was affinity purified on a protein A–agarose column.

Isolation of site F′ binding protein cDNA clones

A randomly primed λgt11 cDNA expression library was constructed from mRNA isolated from the livers of 9-day-old chick embryos (16). An aliquot of the unamplified library (representing ~2×10^6 independent clones) was screened with a concatenated binding site as described (16,17). Binding site monomers were end-labeled using polynucleotide kinase and concatenated using T4 DNA ligase. The remaining 5′ overhangs were filled in with [α-^{32}P]dATP, [α-^{32}P]dCTP, dTTP and dGTP and the Klenow fragment of DNA polymerase. The cDNA inserts were subcloned into the *Eco*RI site of pBluescript SK$^+$ (Stratagene). Each strand was sequenced independently using double-stranded DNA template, synthetic oligonucleotide primers and the Sequenase II sequencing kit (US Biochemical). Additional overlapping clones were isolated using standard DNA screening techniques (18).

Northern analysis

Poly(A)$^+$ RNA was isolated from various adult chicken tissues and embryonic livers using a FastTrack mRNA Isolation kit (Invitrogen). This method consistently yielded RNA preparations that contained ~50% by weight poly(A)$^+$ RNA. Poly(A)$^+$ RNA (4 μg) was size fractionated on a 1.2% agarose–formaldehyde gel, transferred to Zetaprobe nylon membrane (BioRad) and cross-linked to the membrane using a UV Stratalinker 1800 (Stratagene). To ensure that approximately equal amounts of RNA were loaded onto each lane ethidium bromide was added to the samples to visualize the RNA following electrophoresis (19). In some instances the RNA was size fractionated using a glyoxal gel system (18). Northern blots were probed with cloned cDNA fragments which were labeled with [α-^{32}P]dATP by random priming to a specific activity of >5×10^8 (18). Hybridization and washing conditions were as previously described (16).

DNA transfection

Transfection experiments have been described elsewhere (14). Three apoVLDLII–CAT reporter constructs were used in the experiments described here. The F′wt construct contains the proximal promoter and first exon of the apoVLDLII gene from nucleotides (nt) –307 to +38, relative to the major transcription start site. A Transformer Site-Directed Mutagenesis kit (Clontech) was used to mutate the F′ site in the F′wt–CAT reporter construct to produce the Fm1 and Fm2 constructs (14). The F′ sites in the three constructs are:

F′wt, GAAAGGGGCTCTATGACATGGTTGCCTGAA (wild-type);
Fm1, GAAAGGGGCTCTAT**TAA**ATGGTTGCCTGAA (ubiquitous G contacts mutated);
Fm2, GAAAGGGGCTCTATGACAT**TT**TTGA**A**TGAA (liver-enriched G contacts mutated).

The AKR expression vector used in these studies was constructed in pRC/CMV (Invitrogen) and includes nt 196–1214 of the AKR cDNA sequence.

RESULTS

Isolation of a cDNA clone encoding a novel chicken homeobox protein

We have demonstrated previously that the F′ element, located 235 nt upstream from the apoVLDLII transcription start site, is critical for expression from this promoter (Fig. 1A). Ubiquitous and liver-enriched factors which interact with this element have been detected in crude nuclear protein extracts and can be separated from each other by heparin–agarose chromatography. Methylation interference studies have demonstrated that these factors contact overlapping, but not identical, G residues. EMSA using oligonucleotides mutated at these contact residues confirm that the factors can be distinguished by the specificity with which they bind to mutant F′ sites. As described in Materials and Methods, two mutant binding sites, designed on the basis of methylation interference studies (14), were synthesized in which either the G residue on each strand that is contacted by both complexes or the two pairs of downstream G residues on each strand that are

contacted in the liver-enriched complex were converted to T residues (F′m1 and F′m2 respectively). Consistent with previous data from methylation interference experiments, formation of the ubiquitous complex was essentially eliminated by mutation of the middle G nucleotide on each strand (F′m1). The liver-enriched factor bound well to this oligonucleotide and decreased formation of the ubiquitous complex was accompanied by an increase in abundance of the liver-enriched complex. However, unlike the ubiquitous complex, the liver-enriched complex did not form with the binding site in which the four downstream G contacts (two on each strand) were altered (Fig. 1B). We used this information to screen an unamplified day 9 embryonic liver λgt11 cDNA expression library for clones encoding proteins with appropriate binding specificity, as described in Materials and Methods. Approximately 2×10^6 plaques were screened and 40 potential positives were detected. Following secondary and tertiary screening it was found that protein produced by only one of the initially selected clones bound specifically to site F′ and failed to bind to the F′m1 mutant binding site while retaining the ability to bind to the mutant site F′m2 (Fig. 1C). Thus its binding characteristics mimicked those of the factor(s) involved in formation of the ubiquitous F′ complex.

The original clone contained a 702 bp insert with a single extended open reading frame. This insert was used to re-screen the cDNA library to isolate overlapping clones. Screening of ~10^5 plaques yielded 15 additional clones, one of which contained a larger fragment of 1335 base pairs (bp) with 197 and 436 nt of additional 5′ and 3′ sequence, respectively. 5′ RACE experiments (20) were also carried out using day 9 embryonic liver RNA to search for evidence of additional 5′ untranslated sequence. However, the longest product isolated was found to end within 3 nt of the cloned fragment. An in-frame stop codon is present at nt 120, suggesting that the ATG at nt 249 is the translation start site. The sequence around this ATG matches the Kozak consensus sequence at 7 of 10 positions (21). The complete nucleotide sequence has been deposited in the GenBank database (accession no. U25353). The open reading frame encodes a protein of 269 amino acids with a predicted M_r of 29442.

Comparison of the sequence to the GenBank database (performed at the NCBI using the BLAST network service; 22) revealed no homologs of the predicted protein. However, the analyses identified a region located close to the N-terminus of the protein that was related to the homeodomains found in the maize protein Knotted-1 (15), the *C.elegans* protein CEH-20 (23) and the human proto-oncogene product Pbx-1 (24,25). The similarity between the F′ binding protein and these other proteins extended only through their homeodomains. An alignment of the homeodomain of the protein with the homeodomains of several plant, mammalian, nematode, yeast and *Drosophila* proteins is shown in Figure 2. Since the homeodomain of the F′ binding protein was most similar to that of Knotted-1 (45% identity), we have designated the protein Avian Knotted-Related (AKR) protein.

The AKR homeodomain, like those of Knotted-1, CEH-20, Pbx-1, -2 and -3 and yeast MATα2, contains a three amino acid insertion between helices 1 and 2 that is not present in the homeodomain consensus. Four invariant amino acids that are present in helix 3 of essentially all homeodomains characterized to date are present in AKR. In addition, AKR also contains eight matches (seven identical and one conserved) to the 14 amino acids that have been found to be most highly conserved in homeodomains. As indicated in Figure 2, we have also noted six

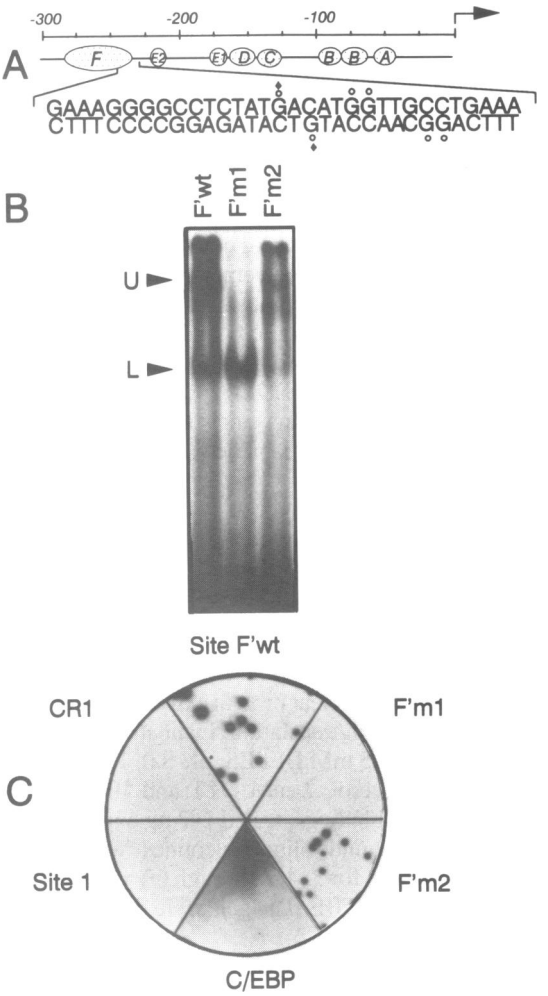

Figure 1. The promoter of the apoVLDLII gene and a comparison of protein binding to native and mutant F′ binding sites. (**A**) The locations of eight protein binding sites in the apoVLDLII proximal promoter are depicted. These sites were identified by *in vivo* and *in vitro* footprinting (37). Footprint A contains a recognition sequence for the liver-enriched factor LF-A1/HNF4 and the ubiquitous factor COUP-TF. A second LF-A1/HNF4 binding site is present at footprint C. Footprints B and D contain binding sites for the liver-enriched factors C/EBP and DBP. E1 and E2 contain a canonical estrogen response element and an imperfect response element respectively. Footprint F does not contain significant sequence matches to any previously characterized DNA binding protein. The sequence and location of the F′ element is indicated. Sites of methylation interference with protein binding by the ubiquitous and liver-enriched F′ binding activities are indicated by ♦ and ○ respectively. (**B**) Binding of liver nuclear protein extracts to native and mutant F′ binding sites is shown. The liver-enriched and ubiquitous F′ binding activities are indicated by L and U respectively. (**C**) A λgt11 cDNA expression library constructed from poly(A)$^+$ RNA from livers of 9 day embryos was screened with oligonucleotide F′wt. Positive plaques were then re-screened to determine their binding specificity using oligonucleotides F′m1 and F′m2. The figure shows a composite of autoradiographs from plaque lifts of plate-amplified isolates that displayed binding specificity similar to that of the ubiquitous complex formed with site F′. In addition to being screened with normal and mutant F′ oligonucleotides the plaques were also screened with an optimized binding site for mammalian C/EBPα (42) and two additional protein binding sites from the apoVLDLII 5′ flanking region: site 1 (43) and CR1 (14).

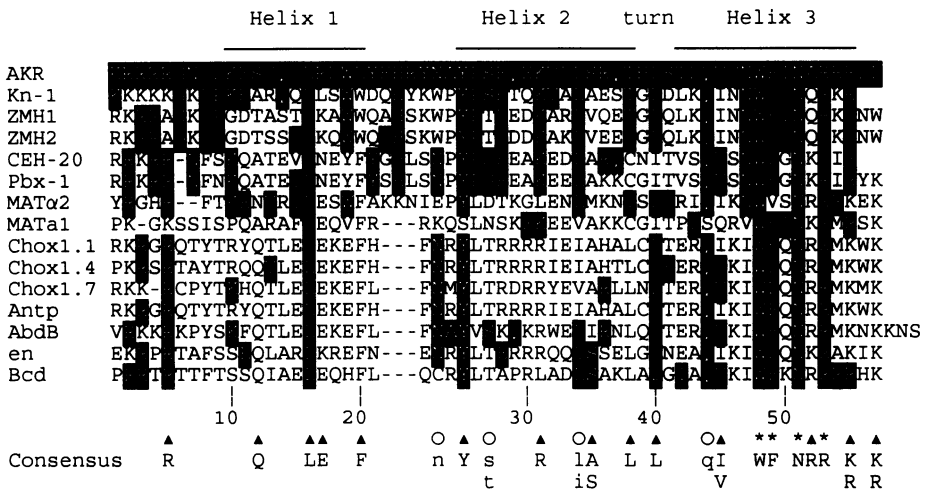

Figure 2. Comparison of the chicken AKR homeodomain with other homeodomain sequences. The deduced amino acid sequence of the AKR homeodomain was aligned with the homeodomains of: Knotted-1, ZMH1 and ZMH2 maize proteins (15); CEH-20 from *C.elegans* (23); the human proto-oncoprotein Pbx-1 (24,25); yeast MATa1 and MATα2 (36); Chox-1.1, Chox-1.4 and Chox-1.7 from chicken (sequences from 44); Antennapedia (Antp), Abdominal B (AbdB), bicoid (Bcd) and engrailed (en) from *Drosophila* (sequences from 45). Conventional numbering for the homeodomain consensus sequence is given such that the three amino acid insertion found in AKR and in some of the other homeodomain proteins is not numbered. Amino acid identities to AKR are highlighted. The four invariant amino acids and the 14 highly conserved amino acids are indicated by * and ▲, respectively. The additional four positions which are well conserved across several homeodomain families are indicated by ○ and the lower case sequence in the consensus sequence. Gaps introduced to maximize the alignment are indicated by –. The regions which form the three main α-helices are indicated.

other residues which are relatively well conserved. The amino acid residue corresponding to position 50 of the homeodomain consensus (position 9 of helix 3) is known to be important in determining DNA binding specificity (26,27). Unlike the majority of other homeodomains described to date, which have lysine or glutamine at this location, AKR, Kn-1, ZMH1, ZMH2 and MATa1 contain an Ile. Thus AKR is the first animal homeodomain protein to be identified with Ile at this position of helix 3.

Tissue and developmental expression profile of AKR mRNA

Poly(A)$^+$ RNA samples isolated from various adult tissues and from embryonic liver at different stages of development were assayed for AKR mRNA expression by Northern blot analysis. The AKR cDNA hybridized to a single size class of mRNA of 1.7–1.8 kb that was present in all tissues examined (Fig. 3A). In adult birds it was most abundant in kidney, intestine, lung, spleen and gizzard and least abundant in liver, brain, heart and testes. The abundance in hen and rooster liver was approximately the same when standardized to 28S rRNA levels. Expression of AKR mRNA was also examined in the liver during embryogenesis and immediately following hatching (Fig. 3B and C). AKR mRNA was most abundant early during embryogenesis and between days 7 and 20 its levels decreased by ~80%, to levels similar to those in adult birds. Changes in post-hatching levels were minor.

DNA binding specificity of AKR

To confirm the DNA binding specificity of AKR a full-length protein was produced by *in vitro* transcription/translation and its ability to bind to normal and mutated F′ oligonucleotides was determined by EMSA. That the binding observed was attributable to AKR was confirmed by using a rabbit antiserum raised against an AKR fusion protein to supershift the complex formed with the *in vitro* translation mix. The entire complex was supershifted by the antiserum and was unaffected by pre-immune serum (Fig. 4). *In vitro* synthesized AKR displayed the same specificity for the F′ oligonucleotides as that observed with the ubiquitous complex formed with nuclear protein extracts (Fig. 4). No binding was detected when the two G residues implicated in binding of the ubiquitous complex by methylation interference assays were mutated (oligonucleotide F′m1). Binding was actually enhanced when the downstream pairs of G residues contacted in the liver-enriched complex were converted to T residues, suggesting that AKR may also interact with this region of the site.

The F′ binding site does not contain TAAT or TTAT motifs that form the core recognition sequences of many, but not all, homeodomain proteins (28,29). However, it does contain the hexanucleotide sequence 5′-TCTATG-3′. Methylation interference and mutagenesis studies have implicated the 3′-terminal G residue of this sequence in AKR binding (14). This hexanucleotide matches a known binding site for the homeodomain protein caudal, 5′-TTTATG-3′, with the exception of a C at a position corresponding to the first nucleotide of the tetranucleotide core (30). Consequently, we examined the ability of AKR to bind to an oligonucleotide consisting of a triplet tandem repeat of this known caudal binding sequence and the ability of the oligonucleotide to compete with a site F′ oligonucleotide. No competition for site F′ binding was observed and only extremely weak direct binding to the caudal site was detected. An oligonucleotide consisting of a repeated TAAT sequence also competed very poorly (data not shown).

To further delineate the binding requirements of AKR the binding site was split into two halves so that the 5′ half contained the sequence preceding and including the putative core sequence 5′-TCTATGA-3′, but lacked the 3′ proximal CG base pair

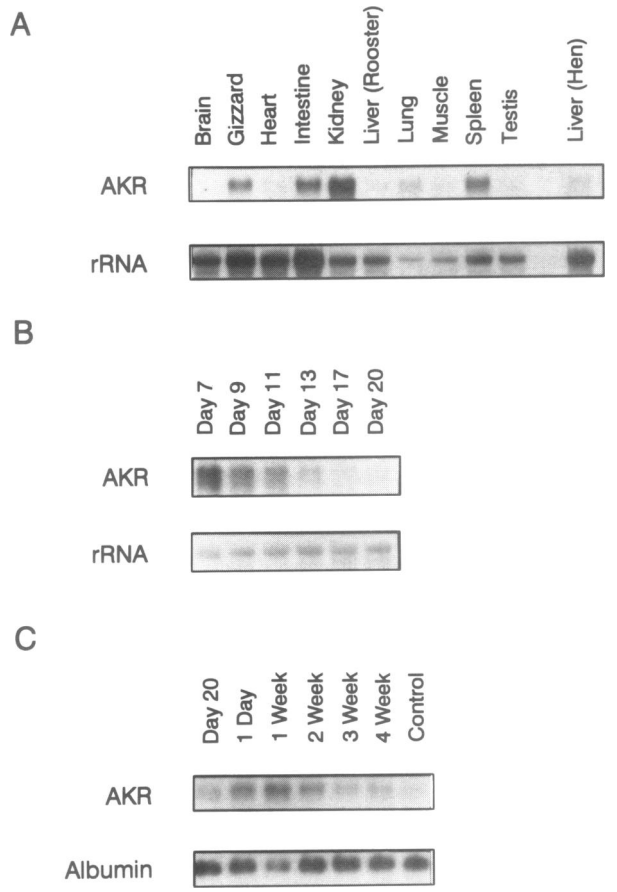

Figure 3. Expression pattern of AKR. The level of expression of AKR mRNA in selected tissues from adult roosters (**A**) and in the livers of embryos (**B**) and chicks post-hatching (**C**) were determined by probing Northern blots of poly(A)$^+$ RNA with a fragment of the AKR cDNA (AKR). The amount of RNA in each lane was monitored by hybridization with a 500 bp DNA fragment derived from a chicken 28S rRNA DNA fragment (rRNA) or a 520 bp DNA fragment derived from chicken serum albumin mRNA (46). Control indicates RNA isolated from the livers of roosters.

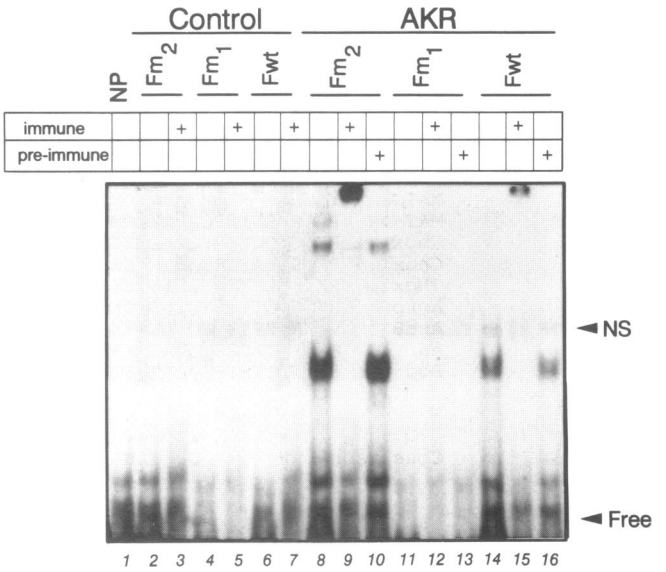

Figure 4. Electrophoretic mobility shift analysis of AKR binding to native and mutant F′ binding sites. Each binding reaction was performed at 4°C using 1 µl *in vitro* translated AKR or vector control, 5 fmol (20 000 c.p.m.) ^{32}P-labeled binding sites Fm2, Fm1 or Fwt in the presence of either pre-immune or immune sera. Lane 1, no protein; lanes 2–7, 1 µl *in vitro* translated vector control in the presence of Fm2 (lanes 2 and 3), Fm1 (lanes 4 and 5) and Fwt (lanes 6 and 7) and 3 µl polyclonal anti-AKR antibody (lanes 3, 5 and 7); lanes 8–16, 1 µl *in vitro* translated AKR in the presence of Fm2 (lanes 8–10), Fm1 (lanes 11–13) and wild-type F site (Fwt) (lanes 14–16) and 3 µl either pre-immune (lanes 10, 13 and 16) or immune sera (lanes 9, 12 and 15) as indicated.

contacted in the ubiquitous complex. AKR failed to bind to either half site, suggesting that the CG base pair was essential for binding and that it was not capable of binding independently to elements in the downstream half of the site (data not shown). We next examined the requirement for residues within the putative core element. Conversion of the sequence from 5′-CTATGAC-3′ to 5′-CTTAATC-3′ completely abolished binding, indicating that residues within the motif ATGA are essential for binding (Fig. 5). The site was then converted so that it more closely matched the optimal binding sites for Antennapedia (5′-TAATGAC-3′), Abdominal B (5′-TTATGAC-3′) and Pbx1 (5′-CAATGAC-3′). AKR bound equally well to the latter two sites and displayed the highest level of binding to 5′-TAATGAC-3′ (Fig. 5). The first two nucleotides of the core were also mutated from CT to AG or GC. These mutations diminished, but did not abolish, binding, indicating considerable flexibility with respect to sequence preference at these two locations (Fig. 5).

Effect of AKR on expression from the apoVLDLII promoter

To determine the ability of AKR to influence expression from the apoVLDLII promoter we constructed a pRC/CMV-AKR vector for use in short-term transfection assays in which synthesis of AKR was under the control of the human cytomegalovirus promoter. Various amounts of the AKR expression vector were co-transfected into the human hepatoma cell line Hep3B with apoVLDLII–CAT reporter constructs containing a normal site F′ sequence or the F′m1 or F′m2 mutations. Since expression from the apoVLDLII promoter is extremely estrogen dependent, a fixed amount of an estrogen receptor expression vector was included in all transfections. The results demonstrate a dose-dependent inhibition of expression from the apoVLDLII constructs containing either a normal site F′ sequence or the F′m2 mutation, but not from the construct containing the F′m1 mutation (Fig. 6). Thus the profile of inhibition observed is consistent with the DNA binding specificity of AKR as determined by EMSA and with the possibility that AKR is a negative regulator of the gene.

DISCUSSION

The predicted amino acid sequence of AKR indicates that it is not a homolog of any previously characterized homeodomain protein. The AKR homeodomain is most closely related to those found in a subgroup of plant proteins, which includes the maize proteins Knotted-1 (15), ZMH1 and ZMH2 (15). The most

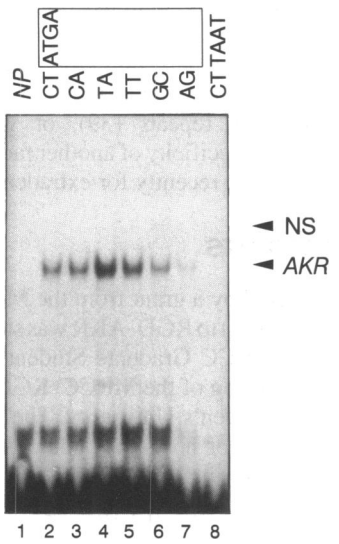

Figure 5. EMSA of AKR binding to native site F′ and F mutant core sites. Each binding reaction was performed at 4°C using 1 µl *in vitro* translated AKR, 5 fmol (20,000 c.p.m.) ^{32}P-labeled binding sites Fwt, FMC1 (5′-CCTCTTAAT-CATGGTTGCCTGAAA-3′), FMC2 (5′-CCTAGATGACATGGTTGCCTG-AAA-3′), FMC3 (5′-CCTTAATGACATGGTTGCCTGAAA-3′), FMC4 (5′-CCTTTATGACATGGTTGCCTGAAA-3′), FMC5 (5′-CCTCAATGA-CATGGTTGCCTGAAA-3′) or FMC6 (5′-CCTGCATGACATGGTTGCCT-GAAA-3′). Only core changes appear in the lane labels. Boxed regions define areas of the core that are present in both the wild-type and mutant binding sites. Lane 1, no protein; lane 2, 1 µl *in vitro* translated AKR in the presence of Fwt; lane 3, 1 µl *in vitro* translated AKR in the presence of FMC5; lane 4, 1 µl *in vitro* translated AKR in the presence of FMC3; lane 5, 1 µl *in vitro* translated AKR in the presence of FMC4; lane 6, 1 µl *in vitro* translated AKR in the presence of FMC6; lane 7, 1 µl *in vitro* translated AKR in the presence of FMC2; lane 8, 1 µl *in vitro* translated AKR in the presence of FMC1.

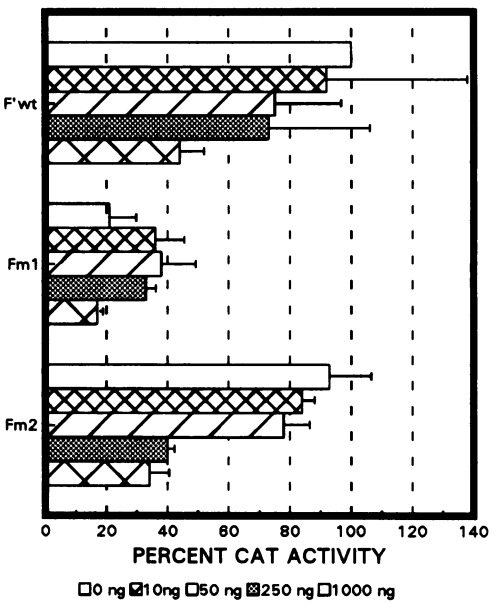

Figure 6. AKR suppresses a wild-type site F′ (F′wt) apoVLDLII reporter construct, but not a construct in which the AKR binding residues have been changed (Fm1). The human hepatoma cell line Hep 3B was used in short-term transfections. Three apoVLDLII–CAT reporter constructs F′wt, Fm1 and Fm2 (4 µg) were co-transfected with an ER expresion vector, pV16CHER (0.4 µg), and various amounts of an AKR expression vector, pRC/CMV-AKR, (0, 10, 50, 250 and 1000 ng). During the expression period the transfections were incubated in the presence of 1 µM DES. CAT activity is expressed relative to F′wt co-transfected with pV16CHER in the presence of DES. The data shown are based on two independent transfection experiments. In each experiment duplicate transfections were carried out and assayed independently. Results of duplicate assays were averaged and normalized to the activity obtained from the wild-type apoVLDLII reporter construct in the absence of AKR, which was arbitrarily set at 100%. The data shown are the means and ranges of the two experiments.

similar homeodomains in vertebrates are present in the human Pbx family of proto-oncogenes (24,25,31), which are homologs of the *Drosophila* protein extradenticle. The homeodomains of all of these proteins contain a three amino acid insertion between helices 1 and 2 that is not present in the consensus homeodomain sequence and which includes a highly conserved histidine in the middle position. Outside the homeodomain AKR has no similarity to any of these proteins. It also differs from these proteins with respect to the location of the homeodomain, which is atypically very close to the N-terminus of the protein. In addition, the protein contains no regions related to other previously identified DNA binding domains that are also found in some homeodomain proteins, such as POU, Paired and CUT repeat sequences (26,32,33).

The DNA binding specificities of the homeodomains most closely related to that in AKR, i.e. those in Knotted-1, ZMH1 and ZMH2, have not been defined. The maize proteins were isolated on the basis of primary sequence conservation, rather than by virtue of their ability to bind to known regulatory elements. Our studies on the DNA binding characteristics of AKR indicate essential sequence requirements other than the simple presence of TAAT or TTAT core motifs. The normal F′ site to which AKR binds does not contain either tetranucleotide, while the F′m1 mutant site to which it does not bind contains the sequence 5′-ATTAAAT-3′, mutated from 5′-ATGACAT-3′ (Figs 1B and 4). Conversion of the normal sequence to 5′-**TAAT**CAT-3′ also completely abolishes binding. In addition, the caudal site oligonucleotide to which AKR binds very poorly contains a 5′-TTATG-3′ sequence (data not shown). An avian homeobox protein, CdxA, has been identified that binds very well to the caudal consensus sequence, which includes the motif TTTATG. However, CdxA clearly belongs to the *caudal*-type subfamily of homeodomain proteins, while AKR does not (30). Creation of a half-site which lacks only the most 3′ of the CG base pairs contacted in the ubiquitous complex formed with nuclear protein extracts also completely abolishes binding. The results of these analyses indicate that the motif 5′-ATGAC-3′ is extremely important in determining binding specificity.

Position 51 of the homeodomain is a highly conserved asparagine residue in the DNA recognition helix of the domain. It normally contacts the A residue at the third position of the common tetranucleotide cores TAAT and TTAT. N51 is conserved in AKR, which, combined with mutagenesis studies described above, suggested that its more extended core recognition sequence in site F′ may be 5′-TCTATGAC-3′, where A is the residue contacted by N51 (29). Residues R3, R5, I47 and N51 have been shown to be highly conserved in homeobox proteins that preferentially recognize sites with a TAAT core tetranucleotide (29). In AKR these residues are conserved with the exception of I47, which is asparagine. The 2- to 3-fold increase in the extent

of binding observed when the sequence in site F′, 5′-TCTATGAC-3′, is converted to 5′-TTAATGAC-3′ is consistent with the predicted preference of R5 and R3 for the 5′-TA of the TAAT core. However, recent studies on optimal site selection using the human Pbx1 homeodomain, which also contains R5, defined a preferred sequence with C in the first position of the tetranucleotide, as it is in the site F′ sequence (35). Our studies indicate that binding by AKR is slightly reduced with C at the first position. Similarly, a slight reduction in extent of binding was observed when the second position of the TAAT tetranucleotide was converted to T, as it is in the normal F′ binding sequence. Preference for T at this position is a feature of homologs of the *Drosophila* protein Abdominal B (34). Lys and Pro at positions 3 and 7 of the homeodomains of these proteins have been implicated in the preference for T at the second position of the tetranucleotide core. However, neither of these residues are conserved in AKR (29). The extent of AKR binding was significantly diminished when the first two nucleotides of the TAAT sequence were converted to AG or GC, but it was not abolished. These preliminary analyses indicate that binding by AKR is relatively tolerant of substitutions in the 5′ 2 nt of the tetranucleotide core, particularly with respect to T or C at the first position and T or A at the second position.

One of the unusual features of AKR is the presence of Ile at position 9 of helix 3 of the homeodomain (Fig. 2). The only other homeodomain proteins known to have Ile at this position are the three maize proteins Knotted-1 (15), ZMH1 and ZMH2 (15) and the yeast mating factor MATa1 (36). Residue 9 in the DNA recognition helix (position 50 in the homeodomain) has been shown to be an extremely important determinant of DNA binding specificity (26,27). This region of AKR and the maize proteins contains the most extensive stretch of contiguous identical residues within the homeodomain, 47-NWFIN-51. With the exception of I50 this sequence is also present in Pbx1 and the *C.elegans* protein CEH-20 (Fig. 2). Residues in this region of the homeodomain interact not only with positions 3 and 4 of the tetranucleotide core, but also with the 3′ proximal dinucleotide pair (27). Our data indicate that a 3′ proximal G residue is extremely important for strong interaction with AKR. However, at present the specific importance of I50 for this interaction is not known.

Analyses of adult tissues indicate that the liver has one of the lowest levels of AKR mRNA and that it is more abundant in kidney, intestine, gizzard and spleen (Fig. 3A). The mRNA for the liver-enriched factor C/EBPα, which is known to be an important positive regulator of apoVLDLII expression (12,37), is also detectable in these tissues, although expression of apoVLDLII is not (A.K.Ryan, D.L.May, C.E.Grant and R.G.Deeley, unpublished results). In the embryo levels of hepatic AKR mRNA are highest in the liver at times during development when the apoVLDLII gene cannot be activated in response to exogenous estrogen or is expressed relatively poorly (7,8,38) (Fig. 3B). Thus in adult birds AKR mRNA is most abundant in tissues which do not express the apoVLDLII gene, but which may also contain known positive regulatory factors. The expression profile of AKR mRNA suggests a negative, rather than positive, role for the factor with respect to apoVLDLII expression. This suggestion is consistent with co-transfection experiments in which estrogen-dependent expression of an apoVLDLII reporter construct containing either a normal site F′ or a mutated site which retains the ability to bind AKR was repressed by the protein. The AKR dependence of the response was further supported by the lack of repression observed with the mutant site F′m1, which does not bind the protein or form the ubiquitous complex observed with nuclear protein extracts. We are currently investigating whether AKR may act by directly preventing binding of a positive regulator, as in the case of homeodomain CCAAT displacement proteins containing CUT repeats (39), or whether it may contribute to the binding specificity of another factor, as proposed for Pbx1 and demonstrated recently for extradenticle (40,41).

ACKNOWLEDGEMENTS

This work was supported by a grant from the Medical Research Council of Canada (MRCC) to RGD. AKR was supported for part of these studies by a MRCC Graduate Studentship. DLM is a clinician-scientist in-training of the MRCC. RGD is the Stauffer Research Professor of Queen's University. The authors wish to thank Dr Caroline Grant for her invaluable contribution to the preparation of this manuscript.

REFERENCES

1 McGinnis,W., Levine,M.S., Hafen,E., Kuroiwa,A. and Gehring,W.J. (1984) *Nature*, **308**, 428–433.
2 McGinnis,W., Garber,R.L., Wirz,J., Kuroiwa,A. and Gehring,W.J. (1984) *Cell*, **37**, 403–408.
3 Levine,M. and Hoey,T. (1988) *Cell*, **55**, 537–540.
4 Brennan,R.G. and Matthews,B.W. (1989) *J. Biol. Chem.*, **264**, 1903–1906.
5 Hoey,T. and Levine,M. (1988) *Nature*, **332**, 858–861.
6 Desplan,C., Theis,J. and O'Farrell,P.H. (1988) *Cell*, **54**, 1081–1090.
7 Wiskocil,R., Bensky,P., Dower,W., Goldberger,R.F., Gordon,J.I. and Deeley,R.G. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 4474–4478.
8 Chan,L., Jackson,R.L., O'Malley,B.W. and Means,A.R. (1976) *J. Clin. Invest.*, **58**, 368–379.
9 Elbrecht,A., Lazier,C.B., Protter,A.A. and Williams,D.L. (1984) *Science*, **225**, 639–641.
10 Colgan,V., Elbrecht,A., Goldman,P., Lazier,C.B. and Deeley,R.G. (1982) *J. Biol. Chem.*, **257**, 14453–14460.
11 Evans,M.I., Silva,R. and Burch,J.B.E. (1988) *Genes Dev.*, **2**, 116–124.
12 Beekman,J.M., Wijnholds,J., Schippers,I.J., Pot,W., Gruber,M. and Ab,G. (1991) *Nucleic Acids Res.*, **19**, 5371–5377.
13 Berkowitz,E.A. and Evans,M.I. (1992) *J. Biol. Chem.*, **267**, 7134–7138.
14 Ryan,A.K., Schrader,T.J., Burtch Wright,R., Buchanan,L. and Deeley,R.G. (1994) *DNA Cell Biol.*, **13**, 987–999.
15 Vollbrecht,E., Veit,B., Sinha,N. and Hake,S. (1991) *Nature*, **350**, 241–243.
16 Grant,C.E. and Deeley,R.G. (1993) *Mol. Cell. Biol.*, **13**, 4186–4196.
17 Singh,H., Clerc,R.G. and LeBowitz,J.H. (1989) *BioTechiques*, **7**, 252–261.
18 Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Plainview, NY.
19 Rosen,K.M. and Villa-Komaroff,L. (1990) *Focus*, **12**, 23–24.
20 Frohman,M.A., Dush,M.K. and Martin,G.R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8998–9002.
21 Kozak,M. (1987) *Nucleic Acids Res.*, **15**, 8125–8148.
22 Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
23 Burglin,T.R. and Ruvkun,G. (1992) *Nature Genet.*, **1**, 319–320.
24 Kamps,M.P., Murre,C., Sun,X.-H. and Baltimore,D. (1990) *Cell*, **60**, 547–555.
25 Nourse,J., Mellentin,J.D., Galili,N., Wilkinson,J., Stanbridge,E., Smith,S.D. and Cleary,M.L. (1990) *Cell*, **60**, 535–545.
26 Treisman,J., Gonczy,P., Vashishtha,M., Harris,E. and Desplan,C. (1989) *Cell*, **59**, 553–562.
27 Hanes,S.D. and Brent,R. (1991) *Science*, **251**, 426–430.
28 Ekker,S.C., von Kessler,D.P. and Beachy,P.A. (1992) *EMBO J.*, **11**, 4059–4072.
29 Ekker,S.C., Jackson,D.G., von Kessler,D.P., Sun,B.I., Young,K.E. and Beachy,P.A. (1994) *EMBO J.*, **13**, 3551–3560.
30 Margalit,Y., Yarus,S., Shapira,E., Gruenbaum,Y. and Fainsod,A. (1993) *Nucleic Acids Res.*, **21**, 4915–4922.
31 Monica,K., Galili,N., Nourse,J., Saltman,D. and Cleary,M.L. (1991) *Mol. Cell. Biol.*, **11**, 6149–6157.
32 Aurora,R. and Herr,W. (1992) *Mol. Cell. Biol.*, **12**, 455–467.

33 Harada,R., Dufort,D., Denis-Larose,C. and Nepveu,A. (1994) *J. Biol. Chem.*, **269**, 2062–2067.
34 Benson,G.V., Nguyen,T.-H.E. and Maas,R.L. (1995) *Mol. Cell. Biol.*, **15**, 1591–1601.
35 van Dijk,M.A., Voorhoeve,P.M. and Murre,C. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 6061–6065.
36 Astell,C.R., Ahlstrom-Jonasson,L. and Smith,M. (1981) *Cell*, **27**, 15–23.
37 Wijnholds,J., Muller,E. and Ab,G. (1991) *Nucleic Acids Res.*, **19**, 33–41.
38 Elbrecht,A., Williams,D.L., Blue,M.-L. and Lazier,C.B. (1981) *Can. J. Biochem.*, **59**, 606–613.
39 Skalnik,D.G., Strauss,E.C. and Orkin,S.H. (1991) *J. Biol. Chem.*, **266**, 16736–16744.
40 Rauskolb,C. and Wieschaus,E. (1994) *EMBO J.*, **13**, 3561–3569.
41 van Dijk,M.A. and Murre,C. (1994) *Cell*, **78**, 617–624.
42 Vinson,C.R., LaMarco,K.L., Johnson,P.F., Landschulz,W.H. and McKnight,S.L. (1988) *Genes Dev.*, **2**, 801–806.
43 Hoodless,P.A., Roy,R.N., Ryan,A.K., Haché,R.J., Vasa,M.Z. and Deeley,R.G. (1990) *Mol. Cell. Biol.*, **10**, 154–164.
44 Sasaki,H., Yamamoto,M. and Kuroiwa,A. (1992) *Mechanisms Dev.*, **37**, 25–36.
45 Scott,M.P., Tamkun,J.W. and Hartzell,G.W. (1989) *Biochim. Biophys. Acta*, **989**, 25–48.
46 Gordon,J.I., Burns,A.T., Christmann,J.L. and Deeley,R.G. (1978) *J. Biol. Chem.*, **253**, 8629–8639.