Chapter 2

DETERMINANTS OF THE DNA BINDING SPECIFICITY OF THE AVIAN HOMEODOMAIN PROTEIN, AKR.

2.1 Statement of Co-authorship

This paper was published in DNA and Cell Biology in 1999 [**18**(10):791-804]. The authors are M.L. Tejada, Zhongchao Jia, D. May and R.G. Deeley. Dr. Jia provided the technical and instrumental assistance that facilitated the generation of the AKR/DNA complex model shown in figure 3.3. Donna May performed the original experiment which demonstrated that concatamers of the F' site confer estrogen responsiveness on a minimal promoter. This provided the means with which to compare differences in the activities of AKR and the AKR mutants. I performed all of the experiments for the manuscript. I also wrote the initial draft of the manuscript and participated in subsequent editing.

2.2 Abstract

AKR (Avian Knotted-Related) was the first example of a vertebrate homeodomain protein with a highly divergent Ile residue at position 50 of the DNA-recognition helix. The protein was cloned from a day-9 chick embryo, liver cDNA expression library by virtue of its ability to bind to the F' site in the proximal promoter of the avian apoVLDLII gene. Expression of the apoVLDLII gene is completely estrogen-dependent and mutation or deletion of the F' site decreases estrogen inducibility 5- to 10-fold. Subsequent data indicated that AKR was capable of repressing the hormone responsiveness of the apoVLDLII promoter, specifically through binding to F'.

Involvement of the F' site in the hormone-dependent activation of apoVLDLII gene expression, as well as AKR mediated repression, strongly suggests that both positive and negative regulatory factors interact with this site. Although several mammalian proteins have now been isolated whose homeodomains share of the many structural features of AKR, including the Ile at position 50, little is known of their functions *in vivo* or the identities of the genes they regulate. Consequently, the elements through which they exert their effects and the structural determinants of their binding specificities remain largely uncharacterized. In this study, we have defined the sequence specificity of binding by AKR using PCR assisted optimal site selection and determined the affinity with which the protein binds to both the optimized site as well as the F' site. Additionally, we have generated a three-dimensional model of the AKR homeodomain binding to its optimized site and have probed the validity of the model by examining the consequences of mutating amino acid residues in recognition helix 3 and the NH_2-terminal arm on the binding specificity of the homeodomain. Finally, we present evidence that the F' site itself may act as an ERE when in the vicinity of imperfect or canonical EREs and that AKR can repress hormone inducibility mediated via this site.

2.3 Introduction

The chicken homeobox protein, Avian Knotted Related (AKR), contains an unusual homeodomain with an atypical DNA binding specificity. AKR is widely expressed very early during embryogenesis and its levels remain relatively high in several adult tissues including the intestine, kidney, spleen and gizzard. However, in other tissues such as the liver, expression of the protein decreases markedly during development. AKR was originally cloned from a day 9 embryonic liver cDNA library during a search for proteins expressed early during embryogenesis that were capable of binding to *cis*-regulatory elements in the promoter of the estrogen dependent, egg-yolk protein gene encoding apo Very Low Density Lipoprotein II (apoVLDLII).

Some yolk protein genes, including the apoVLDLII gene, are normally active only in the livers of laying hens [297; 298]. However, high levels of expression can be induced in chicks and roosters by a single treatment of estrogen [295; 298]. The apoVLDLII gene can also be activated in embryonic liver but competence to express the gene does not develop until sometime between days 7-9 [325; 326]. During this two-day period, the levels of AKR in the liver decline several fold, suggesting that the protein may be involved in suppressing the apoVLDLII gene early during embryogenesis. This suggestion is supported by transient transfection assays which indicate that AKR decreases estrogen-inducible expression from the apoVLDLII promoter [314]. Many of the elements required for efficient hormone-dependent activation of the apoVLDLII gene are clustered within a proximal promoter region of approximately 300 nucleotides. These include both perfect and imperfect estrogen response elements (EREs) as well as recognition sites for LF-A1, COUP-TF and various liver-enriched factors such as C/EBP, LAP and DBP [327-330]. The element to which AKR binds (designated F') is located between -247 and -230 relative to the major transcriptional start site and is situated between two imperfect EREs. In previous studies, we demonstrated that mutation of the F' site results in a greater than 5-fold decrease in the estrogen-inducible activity of the promoter [313]. The importance of the F' site for hormone inducibility of the apoVLDLII gene combined with the AKR-mediated repression suggests that both positive as well as negative regulatory factors may interact with this site. Electrophoretic mobility shift assays (EMSAs) using nuclear extracts from a variety of tissues revealed the presence of both liver-enriched and ubiquitously-expressed proteins that bound specifically to F'. Methylation interference analysis also identified specific G residues, which could be used to distinguish between the binding specificities of the liver-enriched and ubiquitous proteins. Based on this information, AKR was tentatively identified as the ubiquitous factor detected in nuclear extracts.

At the time of its characterization, the AKR homeodomain most closely resembled that of the maize protein, Knotted-1 [331]. However, it differed from that of any known vertebrate protein in several

aspects: i) the homeodomain of AKR is atypically positioned at the NH_2-terminus of the protein; ii) it contains the unusual Ile at position 50 of the recognition helix of its DNA-binding domain; and iii) the sequence of the F' site bound by AKR represented a novel homeodomain recognition element. Presently, AKR is classified as a member of a growing family of homeodomain proteins typified by a three amino acid loop extension (TALE) between helices 1 and 2 [248; 332]. Within this family, AKR represents a group of vertebrate homeobox proteins that contain an Ile at position 50 of helix 3. This group includes the murine orthologue of AKR, TGIF [319]; as well as the Meis1 [321]; and Prep-1 proteins [288; 289; 333; 334]. Relatively little is known of the mechanisms governing the binding specificities of these unusual homeodomain proteins, their regulatory functions or the identities of the genes they regulate.

In the studies described here, we have defined an optimal binding element for AKR and characterized changes in the binding specificity resulting from mutation of critical amino acid residues within helix 3 and the NH_2-terminal arm of the protein's homeodomain. We have integrated these data with structural information derived from X-ray crystallographic data of the TALE homeodomain protein, MATá2, to generate a molecular model of the interactions between AKR and its optimized binding site [279]. Finally, we address the importance of F' to the estrogen-dependent activation of the apoVLDLII gene by demonstrating that concatamers of the F' element can confer weak estrogen responsiveness on a minimal promoter and that AKR and high affinity binding mutants can act as potent repressors of hormone inducibility. These results suggest that homeodomain proteins such as AKR may attenuate the hormone responsiveness of some genes regulated by ER and possibly other members of the nuclear receptor superfamily through antagonistic interactions at elements such as those represented by the F' region.

2.4 Materials and Methods

2.4.1 *Generation of GST-AKR$_{1-178}$ fusion protein*

A vector encoding a Glutathione-S-Transferase fusion protein containing the NH_2-proximal 178 amino acids of AKR was constructed by PCR-amplification of an AKR cDNA using the primers RGD981 and 5319 (Table 3.1, Appendix A). The 534 base-pair product was subcloned between the *Bam*HI and *Sma*I sites of the pGEX2T vector (Pharmacia Biotech) using the *Bam*HI site added to primer A and the blunt 3'-end of the PCR product. This construct was used to transform XL1 blue strain of *Escherichia coli* cells that had been made electrocompetent.

2.4.2 *Site-directed mutagenesis*

GST-AKR mutants were generated using a recombinant PCR strategy. The wild-type GST-AKR$_{1-178}$ cDNA served as a template for PCR-amplification of NH_2- or COOH-terminal fragments of AKR which included a varied set of point mutations using invariant NH_2- or COOH-terminal primers RGD981 and 5319, respectively along with the respective set of complementary mutant primers: R4K (RGD983/RGD982), G6R (RGD1019/RGD1020), N47I (RGD1085/1084), I50K (RGD1088/RGD1087), R54M (RGD1570/RGD1571) (Table 3.1, Appendix A).

A typical PCR reaction contained 20 mM Tris-HCl pH 8.2, 10 mM KCl, 6 mM $(NH_4)_2SO_4$, 20 mM $MgCl_2$, 0.1% Triton X-100, 10 ng/ìl nuclease-free BSA, 0.3 mM dNTPs, 10 ng of template, 1 ìM each primer and 2.5 U of Pfu™ DNA polymerase (Stratagene) in a final volume of 100 ìl. The resultant NH_2- and COOH-terminal fragments were gel purified and ethanol precipitated. To generate each full-length mutated cDNA fragment a portion (10-20%) of the matching NH_2- and COOH-terminal PCR products were mixed in PCR reactions, which excluded the primers for the first round of denaturation, annealing and extension was included and primers were excluded. After

allowing the polymerase to synthesize the double-stranded template from the two half-molecules, the invariant NH_2- and COOH-terminal primers were added and amplification was allowed to proceed. The resultant products were subcloned into the PGEX2T vector in the manner outlined above. All GST-fusion proteins were expressed and isolated using the pGEX system (Pharmacia Biotech) according to the manufacturer's suggestions.

2.4.3 *Binding Site Selection Procedure*

A random oligonucleotide (RGD783, Table 3.2 Appendix A) was designed to contain *EcoR*I, *BamH*I and *Hind*III restriction sites in order to facilitate PCR-amplification of selected sequences using primers RGD784 and RGD785 (Table 3.2, Appendix A) and allow cloning of the selected PCR-amplified products. Site selection was carried out essentially as described by Margalit *et al.* [335]. Either 5 ìg of purified GST-AKR_{1-178} fusion protein (2 ìg/ìl) or 6 ìl of *in vitro* transcribed/translated AKR (rabbit reticulosyte lysate, Promega Corporation) was incubated on ice with approximately 400 ng of the randomer binding site pool for 1 hour in a 30 ìl reaction volume (25 mM HEPES, pH 8.0, 12.5 mM $MgCl_2$, 50 mM KCl, 10% glycerol, 1 ìg BSA, 2 mM dithiothreitiol, 1 ìg of poly dIdC). Protein/DNA complexes were then precipitated using 5 ìl of the protein A-purified fraction of rabbit polyclonal anti-AKR antiserum. The reaction was allowed to proceed with rotation for 30 minutes at 4ºC at which point 15 ìl of a 10% (v/v) slurry of GammaBind® Plus Sepharose® (Pharmacia Biotech) in NET (50 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM EDTA, 0.02% (w/v) sodium azide, 1% (v/v) NP40, 0.1% (v/v) Triton X-100, 0.1% (w/v) SDS, 0.1% (v/v) Tween 20) was added and the mixture was incubated a further 30 minutes. After pelleting at 12 000 rpm for 30 s, the resultant complex was washed in cold wash buffer (25 mM HEPES, pH 8.0, 12.5 mM $MgCl_2$, 50 mM KCl, 10% glycerol, 1 ìg BSA, 2 mM dithiothreitiol). After 5 washes, the bound DNA was released by incubation at 55ºC in elution buffer (50 mM Tris, pH 7.5, 5 mM EDTA, 0.5% SDS, 300 mM NaOAc) and subsequently ethanol

precipitated in the presence of 1 ìg of glycogen. The selected products were resuspended in 10 ìl of water of which 5 ìl was re-amplified by PCR to be used in a subsequent round of selection. After 6 rounds of enrichment, the selected products were subcloned into pBluescript SK^+ (Stratagene). Fifty-six independent clones were sequenced using the dideoxy chain termination method [336] and Sequenase Version 2.0 (US Biochemicals). The resultant nucleotide sequences of these clones were then visually aligned.

2.4.4 *Electrophoretic Mobility Shift Assays (EMSAs)*

Semi-quantitative EMSAs were used to assess the relative binding affinity of the wild-type GST-AKR_{1-178} fusion protein to the wild-type or mutated AKR binding sites. For these assays, the double-stranded binding site oligonucleotides were labelled to equivalent specific activities. An oligonucleotide (3'-CCAACGGACTTTTACATC-5') complementary to the 3'-end of each site was labelled using T_4 polynucleotide kinase. A portion of the labelled oligonucleotide was then annealed to the single-stranded binding sites and the second strand synthesized using klenow fragment. All other oligonucleotide binding sites used were end-labelled with ^{32}P using polynucleotide kinase or filled-in using Klenow fragment [336]. EMSAs were carried out as described previously [314]. Protein/DNA complexes were separated using 6% polyacrylamide gels (29:1) in 1 x Tris borate EDTA buffer at 140 V for 3 hours. The gels were dried and the bands visualized by autoradiography.

2.4.5 *Kinetic Analysis of Binding*

The equilibrium dissociation constants (K_d) for both the wild-type F' site (wt F') as well as the optimized site (Opt1) were determined by EMSA. Binding reactions were carried out at a constant GST-AKR concentration (0.25 ìg) and decreasing amounts of binding site (initial concentration of binding sites was 1 pM). Following autoradiography of the dried gels, bound and free

oligonucleotides fractions were excised and counted by liquid scintillation. Apparent K_d's were determined by Scatchard analyses. The results of three separate experiments were plotted and a line of best fit was calculated using all of the data on the plots.

2.4.6 *Transient Transfections*

The chicken hepatoma cell line LMH/2A, which has been stably transfected to express chicken estrogen receptor (ER), was used in short-term transfections [337]. LMH/2A cells were cultured in Waymouth's medium (Life Technologies™) supplemented with 10% (v/v) L-Glutamine, 10% (v/v) fetal calf serum. Four luciferase reporter constructs were generated which had concatamers of the F' (6 copies), FM1 (4 copies), FM2 (6 copies) and Opt-1 (5 copies) binding sites placed upstream of an SV40 minimal promoter in the pGL2p vector (Promega Corporation). These reporter constructs were transfected alone or were co-transfected along with 0, 5, 10 ng of either wild-type or NH_2-terminal mutant expression vectors. As controls the reporter constructs were also transfected along with corresponding amounts of the parental expression vector pcDNA1/Amp (Invitrogen®). Transfections were carried out using the FuGENE™ 6 transfection reagent (Boehringer Mannheim). Following the first 24 hours of expression, the transfectants were incubated in the presence of 1ìM Diethylstilbestrol (DES) or, in the case of the controls, an equivalent volume of ethanol. After a further 24 hour incubation the cells were harvested and assayed for luciferase activity using the Dual-Luciferase Reporter Assay System developed by Promega (Dual-Luciferase™ Reporter Assay System, Promega Corporation). The DLR Assay System utilizes two different luciferase reporter enzymes, the firefly luciferase and the sea pansy (Renilla reniformis) luciferase, that can be measured in the same sample. The luciferase activities were detected using an microplate luminometer (EG+G Berthold MicroLumat Plus, model LB96V). The firefly luciferase activity of the reporter constructs was normalized to the activity of a control Renilla luciferase expression vector, pRL-TK, that served as an internal control for transfection efficiency

or cell viability. These values were then expressed either relative to the activities of the respective reporters in the absence of AKR/DES or in the presence of DES alone.

2.5 Results

2.5.1 *Identification of an optimized AKR DNA-binding element*

AKR was cloned by virtue of its ability to bind an oligonucleotide corresponding to the F' element from the apoVLDLII promoter (F' binding site, Appendix A Table 2.1) [314]. At that time, we used methylation interference experiments to identify G residues in the site, mutation of which either abolished or enhanced AKR binding (FM1 or FM2 binding sites, respectively, Appendix A Table 2.1). Additional mutagenesis studies also revealed the importance of other bases within the F' site, but failed to identify the prototypical homeodomain core-binding element, ATTA. To further define the binding specificity of AKR, a PCR-assisted binding site selection protocol was used [335]. Both *in vitro* translated full-length AKR, as well as a GST-fusion protein containing the first 178 amino acids of the protein which includes the homeodomain (GST/AKR_{1-178}) were used to screen a library of binding sites containing 24 base pair long cores of random sequence. Limiting amounts of the library were used to enhance the likelihood of selecting high affinity binding sites. Bound sites were selected using a rabbit polyclonal antiserum that recognized both the *in vitro* translated and GST-fusion proteins. The isolated sites were then amplified as described in the Material and Methods section and another round of selection was initiated. Following 6 rounds of enrichment, the selected oligonucleotides were cloned and sequenced. A visual alignment of the binding site pools selected by either the *in vitro* translated protein or GST/AKR_{1-178} did not detect any differences in the binding specificity of either protein. Therefore, the results of the selection were pooled and are presented in the form of a consensogram depicting the frequency with which a particular nucleotide appeared at a specific position within a given sequence for the 56 selected sequences analysed (Figure 2.1). The consensus sequence defined for AKR matches the F' site at 6

of 10 residues and includes the two central G residues determined previously to be critical for AKR binding [314]. A comparison of site F' with the sequences selected by AKR and its murine orthologue TGIF revealed a common pentanucleotide core (5'-TGACA-3'). The selection protocol performed with TGIF consisted of three successive rounds of selection by EMSA using three different concentrations of a GST-fusion protein containing the TGIF homeodomain. The consensus sequence was derived from computer analysis of forty-two sequences. The results of this strategy provided no clear indication of the nucleotide preferences for TGIF outside the core pentanucleotide. The greater number of rounds of selection carried out with AKR may have contributed to the identification of an additional five 3' nucleotides (5'-TGACAGATCT-3') where the G residue was found in greater than 85% of the selected sequences and the remaining four nucleotides were each present in 50% of the selected sequences (Figure 2.1).

2.5.2 *Comparison of the binding of AKR to Site F' and the optimized site Opt-1*

EMSAs using equimolar concentrations of the F' site core (5'-TGACAT-3'), a site where the 3' T of the F' site core was replaced with G (F'(G)) as found in the optimized site, and the entire optimized site (Opt-1) that included all five additional nucleotides identified by our selection protocol were carried out with the GST-AKR fusion protein to provide a preliminary indication of relative binding affinities.

Figure 2.1. **Definition and characterization of a hexanucleotide consensus binding element for AKR.** The binding site selection carried out using either a GST-AKR_{1-178} or *in vitro* translated full-length AKR defined a hexanucleotide consensus site matching the wt F' site at 5/6 nucleotides. There are also additional bases 3'- to this core which were selected for and may act to enhance the binding affinity of AKR.

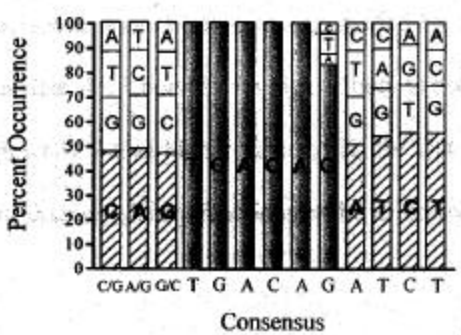
A

```
GTCATCCGTCTGACAGGTCGGGGT
  AGACGAATGACAGATCTGATCTGG
 GCGACCATTGACAGTTCTGGTTTG
GACAGGCCCATTGACAGGCCTTAT
       ATATGACAGGCCAGTTGACAGCCC
  TGGCTAGGAGTGACAGGAAGTGAC
AGGTTAATGACTCGTGACAGAGCA
  CAATGCCTAGTGACAGATCATGAC
 GTCAGTGCCTTTGACAGACGGTAT
       CGGTGACAGATCTTCAACGGACCT
   GCGACAGTTGACAGCTCTGGTTTG
    GCCAAGTGACAGATGTATGTTGGC
    CCACAGTGACAGCTGTAAAGTATT
   AACATTGATGACAGACTTTGTCTG
    AGGCTGTGACAGGTAGTATCTACC
     GGGAGTGACAGTAGTCGTCAGCCC
     GCCAATGACAGATCTTTGCACTCC
ACTAATCAAGGACGTGACAGTGTT
   GGGAGGCAATGACAGTGCCAGCCC
 CTGCCATTACGGTGACAGATCTCC
   GTGAGGTCGTGACAGCGCGTTCTT
 GTCAATGTCGACTGACAGCACGTT
   CAAGCACACTGACAGATCTATGAC
       CGACTGACACAGCCATATAAAGTG
    TACCTCTGACAGGCCTGTGACAAC
   TGGACAATGACAGCGCTCATTGAC
     CCCATTGACAGATCTCAGTTTGCT
  GACGTATCGCTGACAGTGTAGGGT
 TGAGAGGCACATTGACAGCGCAAC
ATATAACGTGTGCGTGACAGATCG
       ATTCTGACAGATTCCCTGCAACTC
CTAGAGTGTCTGGGTGACAGATCT
   TGACGTCCGCTGACAGTCCTTGCC
       AACGTGACAGGAGTGTCATCCTAG
       CCCTGACAGACTTCTGACACCGCT
  AGGCCAACTAGTGACAGTGTTGGC
 AGGAGGGGGGCGGTGACAGCTCTG
 TCATGTATGTCGTGACAGCAGACG
    TGTCACACTGACAGATCCTGTCC
   TAAACACATTGACAGATCGACGTC
       GACTGACAGATCTAATTACCGGGT
    TCACAAATTGACAGGGATACGTCT
AATGTCAAAGGTACTGACAGATCT
 GGATGGATGTGGTGACAGATCTAG
  ACACCAGCTTGTGACAGGACTGTC
GCAGCTCAAAAGGGTGACAGATTG
       TCACTGACAGATGTTATTGTATGG
     ATACGGTGACAGATCTATGGGTAA
```

B

```
          TGCGTGACAAATCCCAFCTAAAA
TACCGCACGAGAGTGACAAATTC
          CGCTGACACATCTGTAACACTGCC
  ACGGGGGACAGTGACATGACAGGT
       TTTCGATGACATATCGATAATCGT
ATATCTCTGTCCGTGACATATTTG
          CATTGACATATCTAAAGCTGCCAG
 GAAGGGCAGGACTGACATCATCTT
```

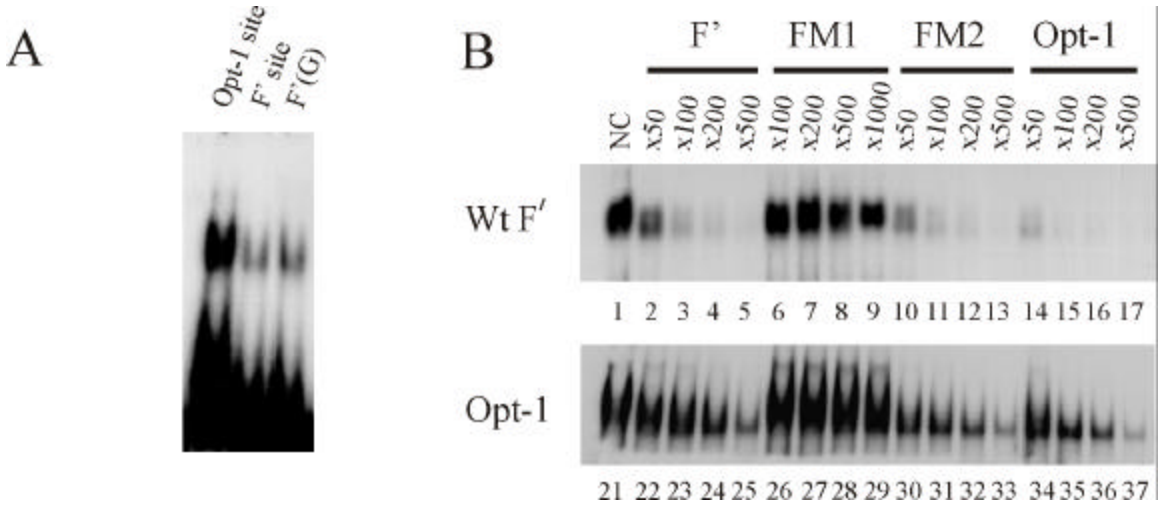
C



The sites were labelled to equivalent specific activities, allowing direct comparison of the extent of complex formation with AKR. Densitometry of the autoradiograph indicated that AKR binding to Opt-1 was approximately three-fold greater than to either the F' or F'(G) sites (Figure 2.2A). Competition experiments were also carried out in which labelled F' and Opt-1 sites were bound by AKR in the presence of increasing molar concentrations of unlabelled competitor binding sites (Figure 2.2B). Under the conditions of the assay, a 50-fold molar excess of unlabelled Opt-1 greatly diminished binding to F' (Figure 2.2B, compare lanes 1 and 14), while a 500-fold molar excess of F' was required for an equivalent decrease in binding to Opt-1 (Figure 2.2B, compare lanes 21 and 25). The specificity of binding to both F' and Opt-1 was verified by using the oligonucleotides mutated at G residues previously demonstrated to abrogate or enhance binding by AKR (FM1 and FM2, respectively, Appendix B Table 2.1) [314].

As expected, a 1000-fold molar excess of FM1 was ineffective in competing for binding to either the F' or Opt-1 sites (Figure 2.2B, lanes 6-9 and 26-29). However, a 50-fold molar excess of FM2 was sufficient to compete completely for binding to the F' site (Figure 2.2B, lane 10). Consistent with the higher affinity of AKR for Opt-1 a 50-fold molar excess of FM2 only reduced binding to Opt-1 approximately three-fold (Figure 2.2B, lane 30). Equilibrium dissociation constants were determined for the binding of AKR to the F' and Opt-1 site, by carrying out EMSAs under conditions of constant protein and serially diluted concentrations of labelled Opt-1 or F' oligonucleotides. After drying and autoradiography of the gel, the bound/unbound fractions of oligonucleotide were excised and quantified by liquid scintillation counting. Scatchard analyses indicated that AKR bound with high affinity to both sites, although the K_d for F' (3.7×10^{-10} M) was approximately 7-fold higher than the K_d for Opt-1 (5.2×10^{-11} M).

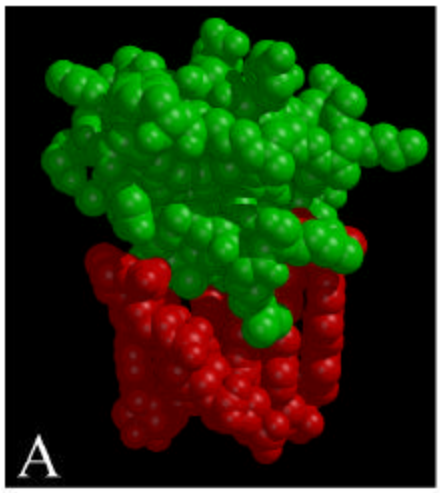
Figure 2.2. **Characterizing the binding of F' and Opt-1 by AKR.** A, 0.5 ìg of GST-AKR_{1-178} fusion protein was incubated for 20 minutes on ice along with 10 fmoles of either the Opt-1, F' or F'(G) binding sites which had been labelled to equivalent specific activities. B, competition EMSA were also carried out using 0.5 ìg of GST-AKR_{1-178} and 5 fmoles of the F' or Opt-1 binding sites as described previously [314]. ^{32}P-labelled F' or Opt-1 binding sites were bound by AKR in the presence of increasing molar concentrations of various unlabelled competitor sites including F' (lanes 2-5, 22-25), FM1 (lanes 6-9, 26-29), FM2 (lanes 10-13, 30-33) or Opt-1 (lanes 14-17, 34-37).

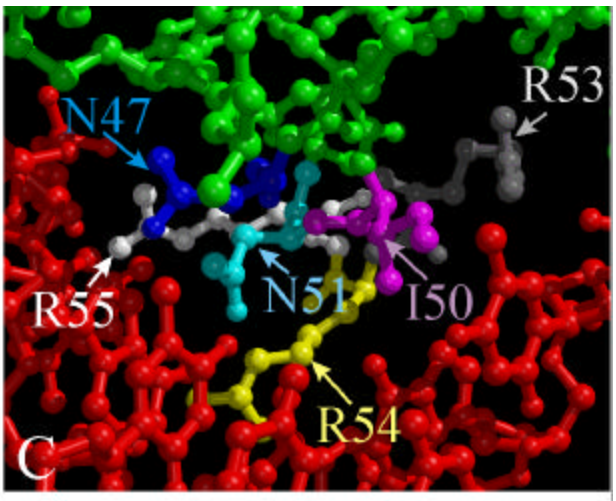

A
Opt-1 site
F' site
F'(G)
B
F'
FM1
FM2
Opt-1
NC
x50
x100
x200
x500
x100
x200
x500
x1000
x50
x100
x200
x500
x50
x100
x200
x500
Wt F'
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
Opt-1
21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37

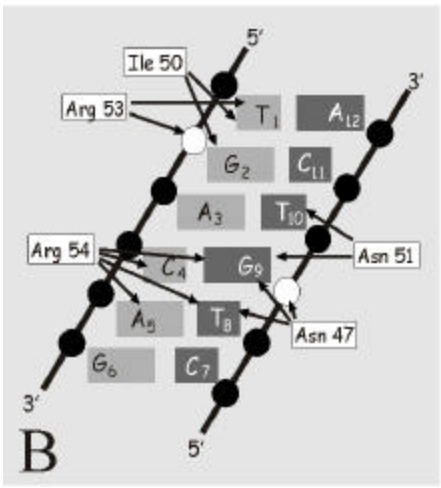
2.5.3 *Homology modelling of the AKR homeodomain protein*

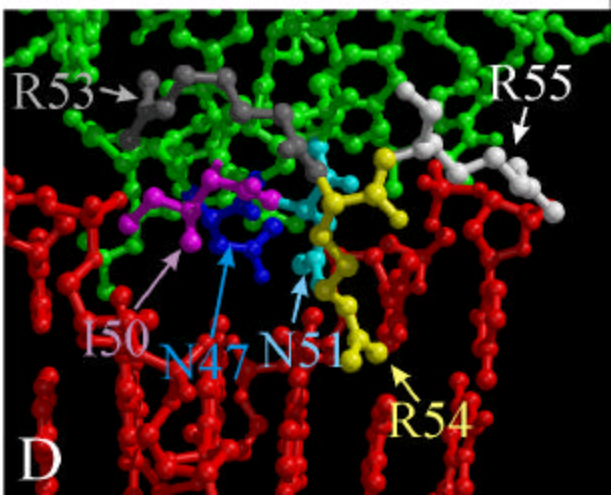
NMR and X-ray crystallographic studies have revealed that the tertiary structure of the homeodomain, its interaction with DNA and its binding mechanism are evolutionarily well-conserved [260-262; 338]. However, the homeodomain of AKR differs from previously characterized homeodomains both with respect to the presence of atypical amino acid residues at key locations and to its binding specificity. To obtain some insight into structural features that may determine the DNA-binding specificity of AKR, we modelled the recognition helix of its homeodomain complexed with the Opt-1 core element. The model was based on the assumption that AKR bound the Opt-1 core hexanucleotide as a monomer. This assumption was made because no differences in site selection were observed when either full-length *in vitro* translated AKR or a fragment, GST-AKR_{1-178}, was used and because only a single complex was observed in EMSAs using the complete Opt-1 sequence, over a wide range of protein concentrations. The model was generated using coordinates derived from studies of the crystal structure of the MATá2/MATa1 dimer complexed to DNA [279]. AKR and MATá2 are sufficiently homologous through their homeodomains (32% identity, 41% similarity) to permit use of the SWISS-3D modeling service (http://expasy.hcuge.ch/swissmod/SWISS-MODEL.html) to generate a preliminary model of an AKR homeodomain/DNA complex. The DNA sequence was then changed to match the hexanucleotide core of the Opt-1 sequence and the docking position was derived from the MATá2/MATa1/DNA complex structure. The model was then subjected to energy minimization to generate the three-dimensional structure illustrated in Figure 2.3. Figure 2.3B is a schematic representation of the predicted contacts made by the AKR homeodomain with its binding site. Nucleotides have been assigned the numbers 1-12 to facilitate the interpretation and subsequent discussion. The predicted interactions between residues of helix 3 and the Opt-1 hexanucleotide are summarized in Table 2.2 (Appendix A).

Figure 2.3. **Modelling the AKR/Opt-1 complex.** A space-filling representation of the AKR/Opt-1 complex model. A, modelling was carried out using the program FRODO [339] and energy minimization was carried out using the X-PLOR [340]. The protein is shown in green and DNA in red, respectively. B, cartoon diagram of contacts established by key residues within the AKR recognition helix to Opt-1. Nucleotides are numbered from 1-12 to facilitate discussion. Phosphates are represented as circles and those phosphates contacted by the homeodomain are represented by open circles. C, ball-and-stick representation of the contact area in the AKR/Opt-1 complex model. Key residues in AKR are highlighted and labelled in various colours corresponding to the residue colour. D, rear view of figure 2.3 C) with the same colour notation.


A


N47
R53
R55
N51
I50
R54
C


5'
Ile 50
3'
Arg 53
T1
A12
G2
C11
A3
T10
Arg 54
C4
G9
Asn 51
A5
T8
Asn 47
G6
C7
3'
5'
B

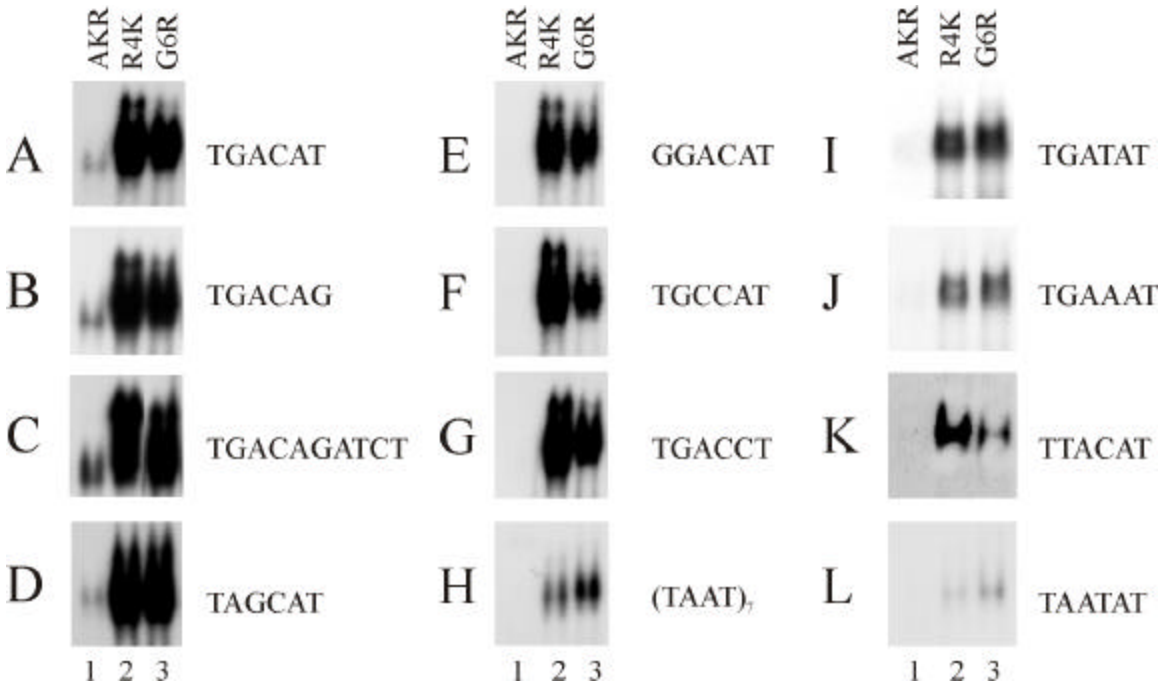

R53
R55
I50
N47
N51
R54
D

Most notable are the number of predicted contacts between Asn47, Asn51 and Arg54 and the last four base positions of the hexanucleotide element defined by the site selection strategy (5'TGACAG-3'). All but four of the contacts postulated to occur within the Opt-1 hexanucleotide involve at least one of the eight bases (A3 to T10) that make up this tetranucleotide core.

2.5.4 *Mutations in helix 3 influence the DNA binding specificity of AKR*

To test the model, amino acids within the helix 3 region that were predicted to be important for determining DNA-binding specificity or affinity were changed to residues found more frequently in other homeodomain proteins. The effects of these mutations on either the affinity or the specificity of AKR binding were then assessed by EMSAs using the wt F' site, or sites where nucleotides within the hexanucleotide core had been systematically modified. Nucleotide changes at positions 1 through 5 essentially abolished binding by the wild-type AKR homeodomain, confirming its unusually high specificity for the hexanucleotide recognition element (Figure 2.4, E-I and Figure 2.5, E-L). Varying levels of importance have been ascribed to amino acids at positions 47, 50, 51 and 54 of helix 3. Of these, position 50 has been suggested to be the most critical determinant of binding specificity [338; 341; 342]. However, it is accepted that contacts established by residue 50 to the two nucleotides 5'- of the homeodomain tetranucleotide core are reinforced through the contacts mediated to the adjacent nucleotides by the residue at positon 47. In this manner, the residue at position 47 is able to contribute to but does not necessarily define, the binding specificity of the homeodomain. Within this helix, Asn 47 and Ile 50 were substituted with either Ile or Lys, respectively. Ile 47 is most commonly found in members of groups 1-11 and 13 of the HOX family, where it makes contacts to thymine [261; 342]. Lys 50 is present most frequently in the paired family of homeodomain proteins where it specifies a GG dinucleotide at the first two positions of the recognition element [338; 343].

Figure 2.4. **Mutations within the NH_2-terminal arm of AKR exhibit increased binding activities for wild- type and mutant binding sites.** EMSAs were carried out using GST fusions of the wild-type AKR or the NH_2-terminal mutant proteins along with a variety of binding sites containing mutations within the hexanucleotide core. A-D, both the G4R and R6K NH_2-terminal mutants were incubated in the presence of wt F', F'(G) and Opt-1 binding sites as well as a site where the G and A at positions 2 and 3 had been exchanged. E-G, binding of these mutant proteins was also examined in the presence of sites containing mutations at positions 1, 3 and 5 that abrogated binding by the wild-type protein. H, the NH_2-terminal mutants were also examined for alterations to their binding specificity by using a site with a hepatmer repeat of a typical homeodomain 5'-TAAT-3' core that is not recognized by the wild-type protein. The critical nature of the G residues at positions 2 and 9 to the binding of the two NH_2-terminal mutants was also examined. I, EMSA was carried out using a binding site where the G at position 9 was changed to T. J, a binding site was also used which contained an A at this position. K, alternatively binding of these mutants to a site containing a mutation of G2 to T was examined. L, finally, binding was examined to a site containing a mutation of both G2 and G9 residues, as found in the FM1 mutant site which was shown previously to be essential for binding of wild-type AKR.


AKR
R4K
G6R
A
TGACAT
B
TGACAG
C
TGACAGATCT
D
TAGCAT
1 2 3
E
GGACAT
F
TGCCAT
G
TGACCT
H
(TAAT)7
1 2 3
I
TGATAT
J
TGAAAT
K
TTACAT
L
TAATAT
1 2 3

The N47I mutant demonstrated marked increases in binding to F' as well as the F'(G) and Opt-1 binding sites relative to the binding activity of the wild-type protein (Figure 2.4, A-C). These results were consistent with those predicted by the model that indicated an increase in the number of hydrophobic contacts made by the Ile residue to the nucleotides at positions 9 and 10 of the hexanucleotide element relative to the wild-type Asn.

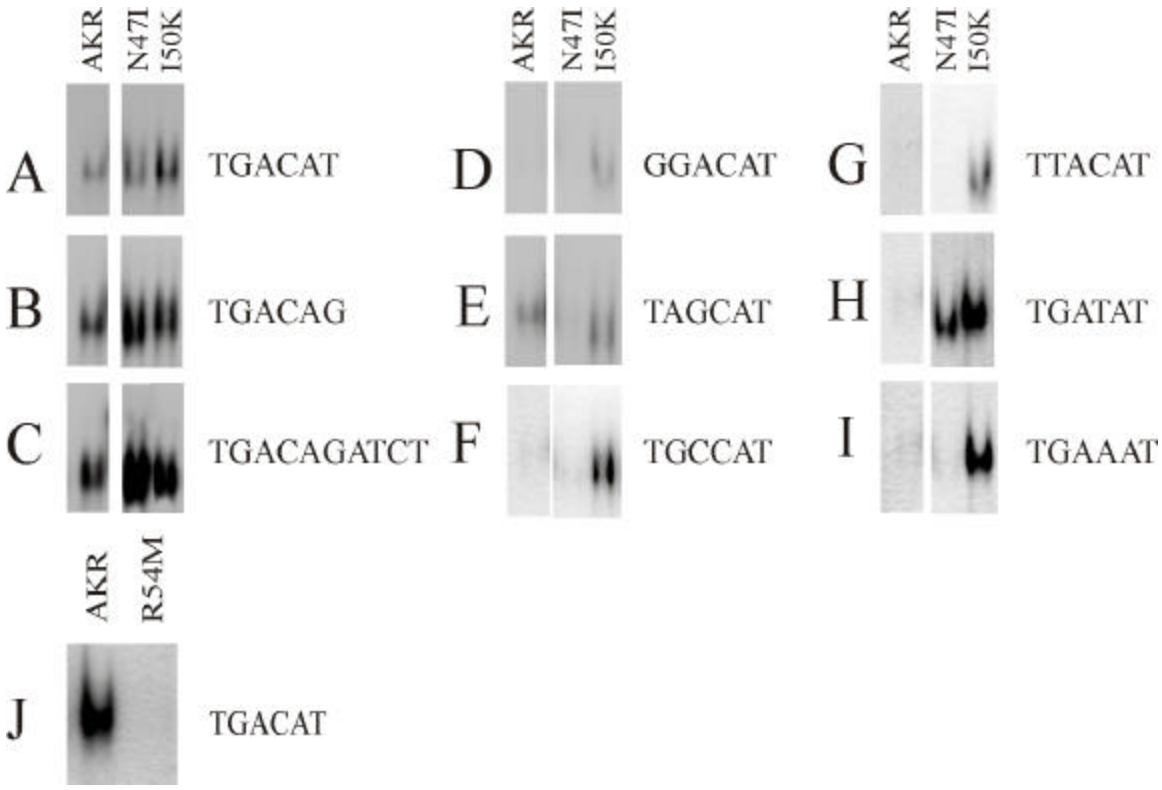
This prediction is supported by the observation that the N47I mutant was capable of compensating for a G to A transition but not a G to T transversion at position 9 of the hexanucleotide (Figure 2.4, H and I), suggesting that this residue is positioned to establish contacts with nucleotides at this location. Similarly to the wild-type protein, the N47I mutant failed to bind to the FM1 mutant binding site or to a binding site containing the prototypical homeobox consensus 5'-TAAT-3'. These data are consistent with the predicted increase in the number of contacts established by the N47I mutation to the nucleotides at position 9 and 10. The increased number of contacts apparently enables the N47I mutant to compensate for conservative changes at position 9 of the hexanucleotide core without altering the more general binding characteristics of the wild-type protein. The proposed model predicts that the I50K mutation should increase the number of contacts made to the TG dinucleotide 5' to the ACA core including formation of an H-bond to G2 (Figure 2.3B and Table 2, respectively). As predicted, the I50K mutation noticeably increased the affinity of the protein for F' as well as the F'(G) and Opt-1 binding sites relative to the wild-type protein (Figure 2.4, A-C). Substitution of T1 with a G or exchange of G2 and A3 severely affected binding by this mutant (Figure 2.4, D-E). This indicates a preference for a G versus an A at position 2 of the hexanucleotide by this mutant, as opposed to the wild-type protein which binds both similarly. Mutation of the A to C at position 4 of the hexanucleotide core had no significant effects on binding by this mutant (Figure 2.4F). The I50K mutant exhibited a moderate decrease in binding to a site containing a G to T mutation at position 2 (Figure 2.4G). In contrast, no such effects were observed when sites containing either a G to A transition or a G to T transversion were used (Figure 2.4 H-I).

However, a site containing both of the mutations at positions 2 and 9, represented by the FM1 mutant binding site, completely abolished binding (data not shown). Finally, the I50K mutant failed to recognize the typical homeobox consensus binding element suggesting that the increased binding affinity had not compromised its binding specificity (data not shown). The residue at position 54 of the homeodomain is predicted to establish base-specific contacts. There is also considerable variability in the identity of the residue at this position within groups of homeodomains that provides a potential source of variation in sequence specificity [338]. For example, in the case of the vnd/NK-2 homeodomain, Tyr 54 has been demonstrated to be the major determinant of binding specificity to the uncommon recognition element (5'-CAAGTG-3') and mutation of Tyr to Met lowered binding affinity by an order of magnitude, while a Y54M mutation changed the binding specificity of the thyroid transcription factor to that of Antp [260]. Our model predicts that Arg 54 of AKR makes the largest number of contacts with the optimized site of any amino acid. To determine the effect of mutating this residue on binding affinity and/or specificity, it was mutated to Met, a residue that is found predominantly at this position in members of the Antp family of homeodomain proteins. Although Met 54 is positioned to make contacts to the T at position 4 of the tetramer core (5'-TAAT-3') in the Antp/DNA complex, it is not a major determinant of binding specificity since the En homeodomain which contains an alanine residue at this position also recognizes TAAT [260]. EMSAs demonstrated that the R54M mutation eliminated binding to the F' site (Figure 2.4J) as well as to all other oligonucleotides tested (data not shown), supporting the predicted importance of this residue for binding by the AKR homeodomain. We did not mutate the residue at position 51 since it is conserved among most homeodomain proteins including AKR and previous studies have demonstrated its importance in DNA binding [338; 344; 345].

2.5.5 *Mutations within the NH_2-terminal arm influence DNA binding by AKR*

To complement the mutational studies carried out on helix 3, additional mutations were made that modified the number of charged residues in the NH_2-terminal arm of AKR. Within the NH_2-terminal arm, homeodomain proteins vary in the number of basic residues present in this region of the protein. This difference in composition significantly influences the stability and the specificity with which the protein binds to its recognition element. Residues at positions 3 through 7 have been implicated as having the most direct effects on protein binding [260; 266]. In order to determine the effects that changes to the number of charged residues within the NH_2-terminal arm of AKR had on the protein's binding specificity, the Arg and Gly residues at positions 4 and 6 were mutated to Lys and Arg, respectively. These residues are found predominantly among members of the PBC class of homeodomain proteins (Pbx, ceh20, exd), members of the PKNOX and MEIS classes, as well as members of the BEL group of homeodomain proteins [248; 260; 332; 338]. With the exception of BEL1, whose recognition element is yet undetermined, all of these homeodomain proteins recognize elements containing a 5'-TGAC/T-3' core. However, the affinity with which they bind to these elements is fairly weak and in most cases requires heterodimerization to become detectable. Both of the NH_2-terminal mutations displayed marked increases in their binding affinities for the F' site over that observed with wild-type protein (Figure 2.5A-G, I, J and K). The increased affinity, predicted by modelling to result from augmenting charged contacts with the DNA phosphate backbone, also enabled the mutated homeodomains to tolerate a variety of nucleotide changes within the hexanucleotide element that abolished binding of the wild-type protein (Figure 2.5, E-K). Notably, both mutants bound to a half ERE that matches the wt F' site at 5/6 nucleotides (5'-TGACCT-3' and 5'-TGACAT-3' respectively). In contrast, wild-type AKR fails to bind the former (Figure 2.5, A and G).

Figure 2.5. **Helix 3 mutations validate the AKR/DNA complex model.** EMSAs were carried out using GST fusion proteins of either the wild-type or helix 3 mutants along with a variety of binding sites containing mutations within the hexanucleotide core. A-C, the N47I and I50K mutants were incubated with the wt F', F'(G) and Opt-1 binding sites. D, binding was examined using a site containing a G to T mutation at position 1. E and F, sites were also used which contained mutations at positions 1 and 2 or 3. G, the binding of these mutants was tested using a site that contained a G at position 2 or alternatively; H-I, sites containing G to A or G to T mutations at position 9. J, the R54M mutant was incapable of recognizing any of the sites tested, supporting the predicted importance of this residue for binding by the AKR homeodomain. Presented is a comparison of the binding of wild-type AKR and the R54M mutant to the wild-type F' site.


AKR
N471
I50K
A
TGACAT
B
TGACAG
C
TGACAGATCT
D
GGACAT
E
TAGCAT
F
TGCCAT
G
TTACAT
H
TGATAT
I
TGAAAT
AKR
R54M
J
TGACAT

Mutation of either G residue shown previously to be essential for binding of wild-type AKR diminished binding by both mutant proteins (Figure 2.5, I-K). Mutation of G2 to T (Figure 2.5K) as well as G9 to T or A reduced binding by these mutants significantly (Figure 2.5, I and J). Although very weak binding of the mutant proteins to a concatamer of 5'-TAAT-3' could be detected (Figure 2.5, H), mutation of both G2 and G9 of F' to create the sequence 5'-TAATAT-3' abolished binding (Figure 2.5L). Thus, these mutant proteins were significantly less sensitive to single nucleotide changes within the hexanucleotide core than the wild-type protein.

2.5.6 *The NH_2-terminal mutations augment repression mediated by the homeodomain*

The use of bacterially-derived fusion proteins in the EMSAs described could be potentially complicated by the possibility that differences exist in the proportion of the purified protein that is correctly folded and functional. Although this would not necessarily affect binding specificity studies, it would influence the extent of the binding observed. This would be most evident in the NH_2-terminal mutations where the results of EMSAs clearly demonstrated substantial increases in the binding activities of these two mutant homeodomain proteins over that of wild-type AKR. Consequently, we also examined the effects that these mutations had upon the activity of AKR in vivo using reporter constructs in which concatamers of potential AKR binding sites were placed upstream of a minimal promoter.

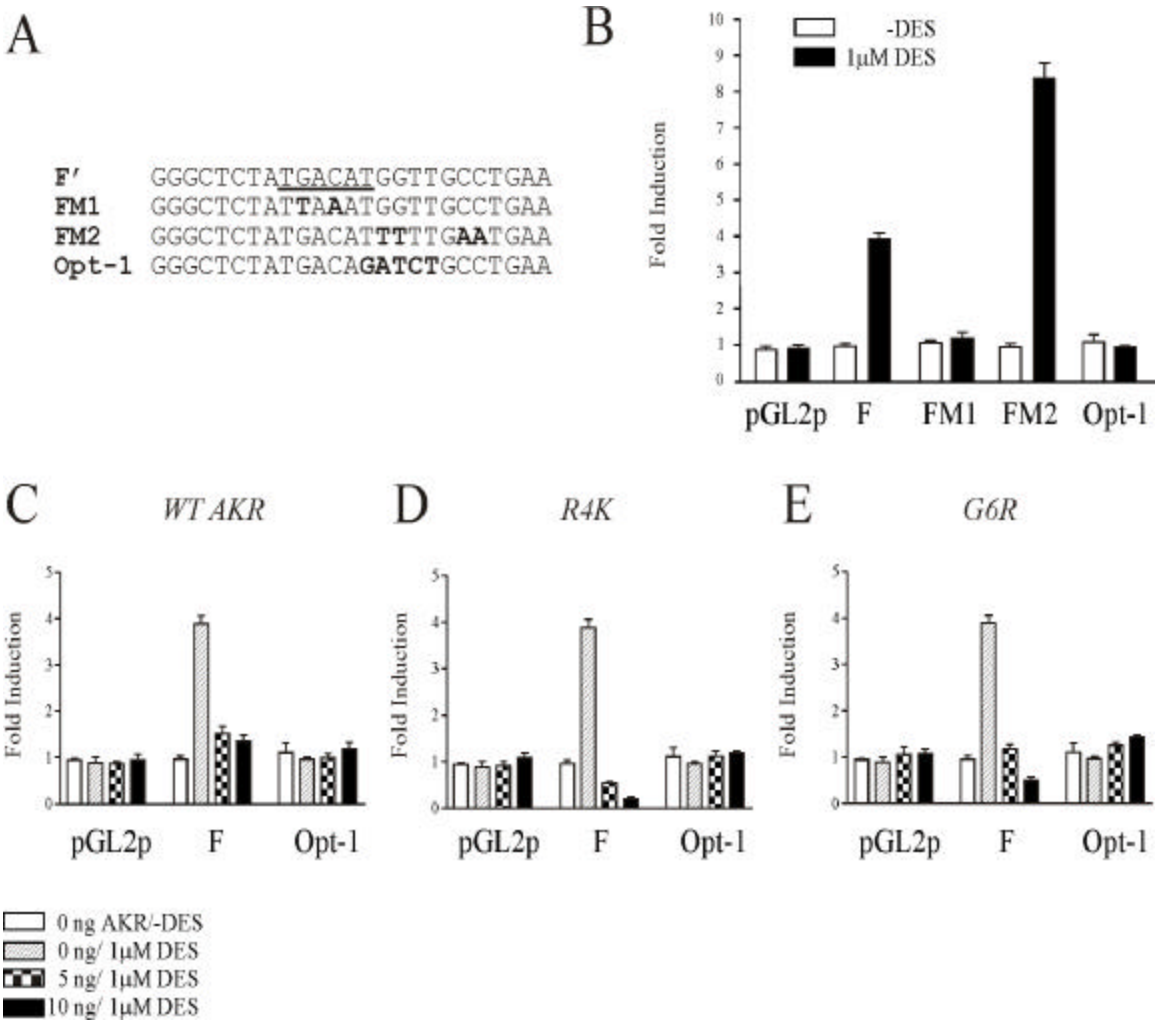
Previously, we demonstrated that the hexanucleotide core of F' was essential for optimal estrogen-dependent expression of the apoVLDLII gene and that the ability of AKR to repress hormone-dependent activation required the same hexanucleotide element. The F' core sequence matches that of a half ERE at 5 of 6 residues including 2 of the 3 nucleotides contacted by the ER. This suggests that ER and AKR may compete for binding to the F' core and that F' in the appropriate context might itself act as an ERE [313; 314; 346].

To assess the potential hormone-responsiveness of the F' element, transient transfections were carried out using the chicken hepatoma cell line LMH/2A and luciferase reporter constructs containing concatamers of the F', FM1, FM2 and Opt-1 binding sites (Figure 2.6A) upstream of an SV40 minimal promoter element. As shown, DES treatment of transfected cells in the absence of AKR results in a 4- to 8-fold increase in the activities of the F' and FM2 reporter constructs, respectively (Figure 2.6B). Neither the FM1 or the Opt-1 reporter constructs displayed any estrogenic responsiveness.

The estrogen-inducibility of the F' luciferase reporter construct provided a functional assay to corroborate the effects that the NH_2-terminal mutations may have upon the DNA-binding affinity of the wild-type protein. Transfection of an expression vector for wild-type AKR resulted in the coincident decrease in expression from the F' reporter construct (Figure 2.6C). No AKR-dependent effects were observed with the control vector pGL2p or the Opt-1 reporter construct, to which AKR binds most strongly. Consistent with the results of EMSA, co-transfection of the vectors encoding the two NH_2-terminal mutant proteins were more effective at repressing the activity of the F' luciferase reporter construct (Figure 2.6 D,E). With 10 ng of expression vector for wild-type AKR, DES responsiveness was markedly diminished but still detectable. The same concentration of expression vector for either of the mutant proteins completely abolished the response to DES and repressed expression from the F' reporter construct below basal levels.

The effect on basal expression was specific for the F' construct and was not observed with either the parental pGL2p vector or the construct containing the optimized binding site for AKR. The mechanism by which the mutant proteins decrease basal expression is presently not known.

Figure 2.6. **The NH_2-terminal mutants suppress the estrogen-responsiveness of the F' and FM2 reporter constructs more actively than the wild-type factor.** A, luciferase reporter constructs were generated which had concatamers of the F', FM1, FM2 or Opt-1 binding site placed upstream of a minimal SV40 promoter. An alignment of the single binding site sequences are shown. B, These reporter constructs were transfected into the avian liver cell line, LMH/2A [337]. After a 24-hour incubation, the transfectants were treated with 1 ìM DES or an equivalent concentration of ethanol. Following a further 24-hour incubation, the cells were harvested and luciferase activity assayed (described under Materials and Methods). The luciferase reporter constructs (pGL2p, F', Opt-1) were also transiently co-transfected along with 0, 5 or 10 ng of; C, wild-type AKR; D, the R4K mutant or; E, the G6R mutant expression vectors. After a 24-hour incubation the transfectants were treated with 1 ìM DES and processed as described previously. These values were then expressed relative to the activities of the respective reporters in the absence of AKR and DES.


A
F′ GGGCTCTATGACATGGTTGCCTGAA
FM1 GGGCTCTATTAAATGGTTGCCTGAA
FM2 GGGCTCTATGACATTTTTGAATGAA
Opt-1 GGGCTCTATGACAGATCTGCCTGAA
B
-DES
1μM DES
Fold Induction
pGL2p F FM1 FM2 Opt-1
C
WT AKR
D
R4K
E
G6R
Fold Induction
pGL2p F Opt-1
0 ng AKR/-DES
0 ng/ 1μM DES
5 ng/ 1μM DES
10 ng/ 1μM DES

2.6 Discussion

2.6.1 *AKR belongs to the TALE family of Homeodomain proteins that recognize the atypical consensus element 5'-TGACAG-3'*

The binding site selection carried out with either the homeodomain of AKR or the entire protein confirmed its ability to bind with high affinity to the sequence 5'-TGACAG-3'. Thus, the binding specificity of AKR is similar to that of several recently identified TALE homeodomain proteins that contain Ile and Arg at positions 50 and 54 of their recognition helices, respectively. Presently, this group includes the members of the Meis and Prep family of homeodomain proteins which share a high degree of identity within their helix 3 regions (Figure 2.1)[332]. Discounting the highly invariant Asn51, residues at positions 50 and 54 are proposed to be two of the most critical determinants of binding specificity because of the direct contacts they establish with specific nucleotides within a given consensus element [260; 338; 342; 347]. However, despite the common presence of Ile50 and Arg54, members of this group differ considerably in affinity for their common binding site, referred to here as the Meis/Prep Recognition element (MPRE). Both Meis1 and Prep1 bind very weakly as monomers to the MPRE but heterodimerization with Pbx significantly increases the affinity of both proteins for a compound binding site consisting of a Pbx half-site 5'-TGAC-3' immediately preceding the MPRE [263; 290]. The Meis1 partner of the Pbx/Meis1 heterodimer can also recognize other unrelated elements, such as the cAMP regulatory sequence (CRS1) of the CYP17 gene promoter (5'-TGATGGACAG-3') and three conserved repeat elements in the Hoxb1 rhombomere (r4) enhancer (5'-AGATGGACAG-3', 5'-TGATTGAAGT-3', 5'-TGATGGATGG-3'). Prep1 appears to heterodimerize with Pbx1 in the absence of DNA to form a complex capable of binding with higher affinity [289]. AKR and its ortholog TGIF, are 81% identical (87.5% similar) to Meis and Prep1 through helix 3 but unlike the other proteins, their homeodomains are very close to their NH_2-termini. They also lack a motif bearing homology to the

Pbx heterodimerization domain. All of the proteins contain Asn47, Ile50 and Arg54, and TGIF recognizes an element on the rat CRBPII gene promoter with a sequence identical to that of the MPRE [319]. However, unlike Meis1 or Prep1, TGIF can bind this element as a monomer with sufficiently high affinity to compete for binding with RXRá. In addition, TGIF does not bind DNA cooperatively with Pbx1 [263]. These observations suggest that TGIF differs from Meis and Prep 1 in the manner with which it recognizes this common target element.

2.6.2 *AKR recognizes 5'-TGACA(T/G)-3' hexanucleotide elements with high affinity*

We showed previously that the F' region of the apoVLDLII promoter is essential for the optimal estrogen-dependent activation of the gene [313]. The F' region contains a hexanucleotide element, 5'-TGACAT-3', which matches the MPRE (5'-TGACAG-3') at 5/6 residues. In the studies described here, we have confirmed that the MPRE is an optimal binding site for AKR and that substitution of the 3'-proximal G of the MPRE by T, as it occurs in the apoVLDLII promoter, resulted in a 7-fold decrease in binding affinity. Nevertheless, the affinity with which the homeodomain of AKR binds to both sites is high, with K_d values for F' and the optimized site, Opt-1, of 3.7×10^{-10} M and 5.2×10^{-11} M, respectively.

2.6.3 *Molecular modelling of the AKR homeodomain*

The critical determinants of binding specificity in the AKR homeodomain are predicted from our model to be Asn47, Ile50, Asn51 and Arg54. The model indicates that Asn47, Asn51 and Arg54 establish extensive contacts with C4, T8, G9 and T10 defining a trimeric core 5'-TGACAG-3' within the Opt-1 hexanucleotide (Table 2.2). The highly invariant Asn51 was postulated to establish contacts to such a core (ACA/TGT) by Bertolino et al [319] and this is consistent with our model. The wild-type AKR homeodomain displays a high degree of specificity for the MPRE and essentially all mutations of the hexanucleotide at positions 1 to 5 abrogate binding. The additional

contacts made by Asn47, Ile50 and Arg54 are predicted to contribute to the stringent binding requirements of AKR.

2.6.4 *Mutations within the recognition helix of AKR modify the binding specificity of the homeodomain*

The residue at position 47 has been postulated to have moderate effects on the binding specificity of some other homeodomain proteins [348]. In the proposed model, Asn47 contacts T8, G9 and T10 of Opt-1. A N47I mutation is predicted to increase the number of hydrophobic contacts to G9 and T10 while eliminating the hydrogen bond with T8. However, the effect of losing the H-bond should be more than offset by the increase in hydrophobic contacts. Consistent with this prediction, the N47I mutation increases binding to sites recognized by the wild-type protein. Additionally, the N47I mutant tolerates a G to A transition, but not a G to T transversion at position 9 of the hexanucleotide. This implies that the N47I mutant homeodomain is able to maintain contacts provided a purine is present at this position. Finally, the mutant protein fails to recognize a binding site containing a T to G mutation at position 10 of the hexanucleotide core validating the importance of the contacts that were indicated by our model.

The residues at positions 50 and 54 have been noted to be the primary determinants of specificity for the three nucleotides 5'- to the binding site core [261; 341; 342; 349]. For example, a Gln50/Met54 combination specifies the first three nucleotides of the Antp binding site (5'-CAATTA-3'). Mutation of these residues to Lys50/Tyr54 changes the binding specificity of the homeodomain to 5'GGCTTA-3' [349]. The AKR model predicts that contacts established by Ile50 of AKR specify the TG pair 5' to the ACA core. The I50K mutation results in an increased number of contacts made to this dinucleotide and augments binding to sites recognized by the wild-type protein. The specificity displayed by this mutant for the nucleotides at positions 1 and 2 indicates a

preference for TT>TA>GG at these positions. Unlike wild-type AKR, the I50K mutant recognizes a binding site with GG at the 5' end, although binding is relatively weak. This is consistent with the binding specificity of Bicoid which recognizes 5'-GGATTA-3' where the combination of Lys50 and Arg54 establishes contacts with the GG dinucleotide [338].

In AKR, Arg54 is predicted to make extensive contacts to C4, A5 and T8. Modification of this residue to Lys or Met is predicted to severely decrease the number of contacts established to the Opt-1 core. Consistent with the prediction, the R54M mutant failed to bind to any of the binding sites used. Immunoblot analyses carried out using a polyclonal anti-AKR antibody raised against the NH_2-terminal portion of the homeodomain verified that the mutant protein was produced efficiently and remained intact during purification (data not shown). The NK family of homeodomains contains Tyr54 and recognizes 5'-CACTTG-3', a mutation of this residue to methionine lowers affinity for the NK element by an order of magnitude [347].

Variability in the identity of the amino acid at position 54 may be responsible for differences in the binding specificity of the TALE homeodomain proteins that share an isoleucine at position 50. Among TALE proteins, Asn47 is usually paired with Ile50 (BEL,CUP,Kn, HAc-ATYP, KNOX, MEIS families) but position 54 may be lysine (KNOX), valine (BEL) or arginine (TGIF/AKR, MEIS) [248; 332]. This suggests that, along with the highly invariant Asn51, Ile50 and Asn47 may account for recognition of TGAC, while the residue at position 54 contributes to determining the identity of the two 3' proximal nucleotides and may significantly influence the overall strength of binding. For example, the helix 3 region of the a hooded barley homeodomain protein, HvH21, differs from that of AKR only at Gln52 and Lys54 and only the latter residue is predicted to make base-specific contacts [350]. HvH21 recognizes a 5'-TGACNN-3' motif suggesting that Lys at position 54 introduces a degree of ambiguity for nucleotides at these positions. Our model predicts that an R54K substitution would essentially abrogate all contacts made to C4 and A5, and more

importantly also to T8 which may explain the decreased specificity of HvH21 for the 3' dinucleotide [350].

2.6.5 *Mutations in the NH_2-terminal arm affect AKR binding affinity*

Both R4K and G6R mutations increased binding to sites recognized by AKR and allowed binding to sites to which we could not detect binding by the wild-type protein. The NH_2-terminal mutations of AKR allowed binding to sites resembling a number of half-sites that are recognized by Pbx/Meis1 heterodimers (5'-TGATGGACAG-3' ,CYP17 promoter and 5'-AGATGGACAG-3' on the Hoxb1 rhombomere (r4) enhancer) as well as a site resembling a half ERE (5'-TGACCT-3'). These observations are similar to those obtained using HoxA-1 mutant proteins, which demonstrated that mutations A3R and N2K act synergistically to increase the ability of HoxA-1 to transactivate the HoxD-4 promoter 23-fold. The mutant HoxA-1 protein was also found to bind to a HoxD-4 recognition element with the same affinity as HoxD-4 itself [351]. The increased binding activities of the NH_2-terminal mutants did not eliminate the binding specificity of the homeodomain since these proteins failed to recognize a typical homeodomain consensus element or a binding site where the G residues directly contacted by wild-type AKR had been mutated (FM1).

2.6.6 *The NH_2-terminal mutations augment AKR mediated repression*

We presented evidence previously that AKR could partially repress estrogen-inducibility of the apoVLDLII promoter and that this apparently involved binding of AKR to F'. We also demonstrated that mutations such as FM1 that abolished AKR binding markedly decreased estrogen-responsiveness. This suggested that the F' element in the context of the apoVLDLII promoter where it is flanked by known EREs might be acting as a response element itself and that AKR might compete with the ER for binding. We have now confirmed that concatamers of the F' element can confer estrogen-responsiveness on a minimal promoter when appropriate reporter

constructs are transfected into avian hepatoma cells that express estrogen receptor. Furthermore, the FM1 mutation, which abrogates AKR binding also abolished estrogen responsiveness while the FM2 mutation, which increases AKR binding, enhanced inducibility. Thus, the two central G residues in the F' core that are essential for binding by AKR, are also critical for estrogen-responsiveness, as would be expected if the estrogen receptor was binding directly to the F' core. The Opt-1 element, which binds AKR with high affinity and which also contains the two central G residues, did not confer estrogen responsiveness. However, unlike the two other functional AKR binding sites, it matches a canonical half ERE at only 4, rather than 5, out of six positions

Kato *et al.* had previously shown that native or artificial widely-spaced PuGGTCA repeats conferred hormone responsiveness when placed upstream of CAT reporter constructs [171]. Similarly, our results also indicate that naturally occurring elements recognized by AKR, and presumably TGIF and related homeodomain proteins, may contribute to the hormone-responsiveness of the promoters in which they are located. The estrogen dependence of the F' luciferase expression vector allowed us to compare the ability of wild-type AKR and the NH_2-terminal mutants to function as repressors of hormone inducibility. Co-transfection experiments carried out in LMH/2A cells showed that the R4K and G6R mutants were more potent repressors than the wild-type protein. Thus, they support the predictions of the proposed model of the AKR homeodomain and the results of the EMSAs, which indicated a higher binding affinity for these mutant proteins.

It has been proposed that TGIF exerts its negative regulatory effects on the CRBPII gene by binding initially with high affinity to its consensus recognition element, the MPRE, which then permits subsequent cooperative interactions with downstream DR1 elements within the proximal promoter [319]. Our results indicate that AKR negatively regulates the estrogen-dependent expression of the apoVLDLII gene by binding to the 5'-TGACAT-3' element of the F' region which lies between two

perfect half EREs containing the TGAC motif [313; 314]. Although wild-type AKR does not display detectable binding to an individual estrogen-response element by EMSA we have shown that a single mutation in the NH_2-terminal arm of the homeodomain, which is predicted to increase charged interactions with the DNA phosphate backbone, results in readily detectable binding to a single half-ERE. Thus, it is possible that weak interactions with EREs may be stabilized by the high affinity binding of AKR to the F' site via protein/protein interactions as proposed for Meis and Pbx heterodimers. These data and those derived from studies on TGIF, suggest that AKR and its mammalian orthologues may contribute to the regulation of both retinoid- and estrogen-regulated genes [319].

Table 2.1. Binding Site Oligonucleotides used in EMSAs presented in this study.

Binding Site	Function	Nucleotide Sequence[a,b]
WT F'	EMSA Kinetic Analyses	5'-AAAGGGGCCTCTATGACATGGTTGCCTGAA-3 3'-CGGAGATACTGTACCAACGGACTTTTACAT-5'
FM1	EMSA	5'-AAAGGGGCCTCTATTAAATGGTTGCCTGAA-3' 3'-CGGAGATAATTTACCAACGGACTTTTACAT-5'
FM2	EMSA	5'-AAAGGGGCCTCTATGACATTTTTGAATGAA-3= 3'-CGGAGATACTGTAAAAACTTACTTTTACAT-5'
Opt-1	EMSA	5'-GTTTATGAAAGGGGCCTCTATGACAGATCTGCCTGAAAATGTAGG-3' 3'-CAAATACTTTCCCCGGAGATACTGTCTAGACGGACTTTTACATCC-5'
WT F' Core	EMSA Kinetic Analyses	5'-GAAAGGGGCCTCTATGACATGGTTGCCTGAAAATGTAG-3' 3'-CCAACGGACTTTTACATC-5'
Core Mutant 1	EMSA	5'-GAAAGGGGCCTCTAGGACATGGTTGCCTGAAAATGTAG-3'
Core Mutant 2	EMSA	5'-GAAAGGGGCCTCTATGCCATGGTTGCCTGAAAATGTAG-3'
Core Mutant 3	EMSA	5'-GAAAGGGGCCTCTATGACCTGGTTGCCTGAAAATGTAG-3'
Core Mutant 4	EMSA Kinetic Analyses	5'-GAAAGGGGCCTCTATGACAGGGTTGCCTGAAAATGTAG-3'
Core Mutant 5	EMSA	5'-GAAAGGGGCCTCTATTACATGGTTGCCTGAAAATGTAG-3'
Core Mutant 6	EMSA	5'-GAAAGGGGCCTCTATGAAATGGTTGCCTGAAAATGTAG-3'
Core Mutant 7	EMSA	5'-GAAAGGGGCCTCTATGATATGGTTGCCTGAAAATGTAG-3'

[a] All oligonucleotides were synthesized using an Oligo 1000 DNA Synthesizer (Beckman).

[b] The regions representing the hexanucleotide core are underlined.

Table 2.2 Summary of Predicted Contacts derived from the AKR/Opt-1 Complex Model.

Residue	Atom	H-bond	Van der Waals
N47	(Oä1) (Nä2)	T8 G9 O^2P	T10 C5A, O4, C5 G9 C2', N9, N7
I50	(Cγ) (Cδ)		G2 N7, C3 T1 C5, C6 G2 N7
N51	(Nä2) (Oä1) Câ Cã	G9 N7	G9 C5, N7, O6 T8 O4, C5A; G9 O6; C4 N4 T8 C5A T8 C5A; G9 O6
R53	(Nç2) (C2)	G2 O^2P	T1 C2', C3' G2 O^2P
R54	(Nç1) (Nç2) (Cî) (Nå)	T8 O4; A5 N6; C4 N4	C4 N4; C4; C5 C4 C4; G9 O6; A5 C6 T8 O4; A5 N6; C4 C4, N4 C4 N4
R55	(Nç1) (Nç2) (Cî)	C7 O5' C7 O5' C7 O5'	C7 C5' C7 C5' C7 C5'
I47	Câ Cã1 Cã2 Cä		T10 C5A T10 C5A T10 C5A, O4 T10 C5A; G9 C2', O^2P; C8
K50	(Nî) Cå Cä	G2 O6	T1 O4, C4, C5; G2 N7, C6 G2 O6 N7 T1 C5, C6; G2 N7
K54			C4 C5
M54			C4 N4

In order to facilitate reading of this table, an example is provided. The Oä1 of Asn47 forms a hydrogen bond with T8 of Opt-1. Oä1 also establishes van der waals interactions with C5A, O4 and C5 at distances of T10 as well as C2', N9 and N7 of G9. Lastly, Nä2 of Asn47 establishes a hydrogen bond to G9.

REFERENCES

244. Scott MP, Weiner AJ: Structural relationships among genes that control development: sequence homology between the *Antennapedia*, *Ultrabithorax*, and *fushi tarazu* loci of *Drosophila*. *Proc.Natl.Acad.Sci.U.S.A.* 1984, **81**:4115-4119.

245. Gehring WJ, Affolter M, Bürglin T: Homeodomain proteins. *Ann.Rev.Biochem.* 1994, **63**:487-526.

246. Stein S, Fritsch R, Lemaire L, Kessel M: Checklist: vertebrate homeobox genes. *Mech.Devel.* 1996, **55**:91-108.

247. Kappen C, Ruddle FH: Evolution of a regulatory gene family: HOM/HOX genes. *Curr.Opin.Genet.Dev.* 1993, **3**:931-938.

248. Bürglin TR: The Evolution of Homeobox Genes. In *Biodiversity and Evolution,* Edited by Araj R, Kato M, Doi Y. Tokyo: The National Science Museum Foundation.; 1995:291-336.

249. McGinnis W, Krumlauf R: Homeobox genes and axial patterning. *Cell* 1992, **68**:283-302.

250. Mann RS: The specificity of homeotic gene function. *BioEssays* 1995, **17**:855-863.

251. Maconochie M, Nonchev S, Morrison A, Krumlauf R: Paralogous Hox genes: function and regulation. *Annu.Rev.Genet.* 1996, **30**:529-556.

252. Struhl G: A homoeotic mutation transforming leg to antenna in *Drosophila*. *Nature* 1981, **292**:635-638.

253. Schneuwly S, Klemenz R, Gehring WJ: Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia. *Nature* 1987, **325**:816-818.

254. Garcia-Bellido A: The development of concepts on development--a dialogue with Antonio Garcia-Bellido [interview by Enrique Cerda-Olmedo]. *Int.J.Dev.Biol.* 1998, **42**:233-236.

255. Lewis EB: A gene complex controlling segmentation in *Drosophila*. *Nature* 1978, **276**:565-570.

256. Akam M: Hox genes and the evolution of diverse body plans. *Philos.Trans.R.Soc.Lond.B.Biol.Sci.* 1995, **349**:313-319.

257. Desplan C, Theis J, O'Farrell PH: The sequence specificity of homeodomain-DNA interaction. *Cell* 1988, **54**:1081-1090.

258. Hoey T, Levine M: Divergent homeo box proteins recognize similar DNA sequences in Drosophila. *Nature* 1988, 332:858-861.

259. Wolberger C, Vershon AK, Liu B, Johnson AD, Pabo CO: Crystal structure of a MATá2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell* 1991, **67**:517-528.

260. Damante G, Pellizzari L, Esposito G, Fogolari F, Viglino P, Fabbro D, Tell G, Formisano S, Di Lauro R: A molecular code dictates sequence-specific DNA recognition by homeodomains. *EMBO J.* 1996, **15**:4992-5000.

261. Pomerantz JL, Sharp PA: Homeodomain Determinants of Major Groove Recognition. *Biochemistry* 1994, **33**:10851-10858.

262. Gehring WJ, Qian YQ, Billeter M, Furukubo-Tokunaga K, Schier AF, Resendez-Perez D, Affolter M, Otting G, Wüthrich K: Homeodomain-DNA recognition. *Cell* 1994, 7**8**:211-223.

263. Chang CP, Jacobs Y, Nakamura T, Jenkins NA, Copeland NG, Cleary ML: Meis proteins are major *in vivo* DNA binding partners for wild-type but not chimeric Pbx proteins. *Mol.Cell Biol.* 1997, **17**:5679-5687.

264. Gruschus Jm, Tsao Dh, Wang Lh, Nirenberg M, Ferretti Ja: Interactions of the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance spectroscopy: basis of binding specificity. *Biochemistry* 1997, **36**:5372-5380.

265. Ekker SC, Young KE, Von Kessler DP, Beachy PA: Optimal DNA sequence recognition by the Ultrabithorax homeodomain of *Drosophila*. *EMBO J.* 1991, **10**:1179-1186.

266. Riechmann JL, Krizek BA, Meyerowitz EM: Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proc.Natl.Acad.Sci.U.S.A.* 1996, **93**:4793-4798.

267. Vershon AK, Jin Y, Johnson AD: A homeodomain protein lacking specific side chains of helix 3 can still bind DNA and direct transcriptional repression. *Genes Dev.* 1995, **9**:182-192.

268. Clarke ND: Covariation of residues in the homeodomain sequence family. *Protein Sci.* 1995, **4**:2269-2278.

269. Ekker, S. C., von Kessler, D. P., and Beachy, P. A. Differential DNA sequence recognition is a determinant of specificity in homeotic gene action. *EMBO J.* 1992, **11**: 4059-4072.

270. Harada R, Bérubé G, Tamplin OJ, Denis-Larose C, Nepveu A: DNA Binding Specificity of the Cut Repeats from the Human Cut-Like Protein. *Mol. Cell. Biol.* 1995, **15**:129-140.

271. Andres V, Chiara MD, Mahdavi V: A new bipartite DNA-binding domain: cooperative interaction between the cut repeat and homeo domain of the cut homeo proteins. *Genes Dev.* 1994, **8**:245-257.

272. Verrijzer CP, Alkema MJ, Van Weperen WW, Van Leeuwen HC, Stratling MJJ, Van Der Vliet PC: The DNA binding specificity of the bipartite POU domain and its subdomains. *EMBO J* 1992, **11**:4993-5003.

273. Verrijzer CP, Van Der Vliet PC: POU domain transcription factors. *Biochim.Biophys.Acta* 1993, **1173**:1-21.

274. Fujioka M, Miskiewicz P, Raj L, Gulledge AA, Weir M, Goto T: *Drosophila* Paired regulates late *even*-skipped expression through a composite binding site for the paired domain and the homeodomain. *Development* 1996, **122**:2697-2707.

275. Jun S, Desplan C: Cooperative interactions between paired domain and homeodomain. *Development* 1996, **122**:2639-2650.

276. Wolberger C: Homeodomain interactions. *Curr.Opin.Struct.Biol.* 1996, **6**:62-68.

277. Piper DE, Batchelor AH, Chang CP, Cleary ML, Wolberger C: Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell* 1999, **96**:587-597.

278. Goutte C, Johnson AD: a1 protein alters the DNA binding specificity of á2 repressor. *Cell* 1988, **52**:875-882.

279. Li T, Stark MR, Johnson AD, Wolberger C: Crystal Structure of the MATa1/MATá2 Homeodomain Heterodimer Bound to DNA. *Science* 1995, **270**:262-293.

280. Carson-Jurica MA, Lee AT, Dobson AW, Conneely OM, Schrader WT, O'Malley BW: Interaction of the chicken progesterone receptor with heat shock protein (HSP) 90. *J.Steroid Biochem.* 1989, **34**:1-9.

281. Smeal T, Angel P, Meek J, Karin M: Different requirements for formation of jun:jun and jun:fos complexes. *Genes Dev.* 1989, **3**:2091-2100.

282. Phelan ML, Featherstone MS: Distinct HOX N-terminal Arm Residues are Responsible for Specificity of DNA Recognition by HOX monomers and HOX-PBX Heterodimers. *J.Biol.Chem.* 1997, **272**:8635-8643.

283. Shen WF, Rozenfeld S, Lawrence HJ, Largman C: The Abdb-like Hox homeodomain proteins can be subdivided by the ability to form complexes with PBX1a on a novel DNA target. *J.Biol.Chem.* 1997, **272**:8198-8206.

284. Neuteboom, S. T. and Murre, C. Pbx raises the DNA-binding specificity but not the selectivity of the *Antennapedia* Hox proteins. *Mol.Cell.Biol* 1997, **17**:4696-4706.

285. Chan SK, Ryoo HD, Gould A, Krumlauf R, Mann RS: Switching the *in vivo* specificity of a minimal Hox-responsive element. *Development* 1997, **124**:2007-2014.

286. Peltenburg LTC, Murre C: Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. *EMBO J.* 1996, **15**:3385-3393.

287. Passner JM, Ryoo HD, Shen L, Mann RS, Aggarwal AK: Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature* 1999, **397**:714-719.

288. Berthelsen J, Zappavigna V, Ferretti E, Mavilio F, Blasi F: The novel homeoprotein Prep1 modulates Pbx-Hox protein cooperativity. *EMBO J.* 1998, **17**:1434-1445.

289. Berthelsen J, Zappavigna V, Mavilio F, Blasi F: Prep1, a novel functional partner of Pbx proteins. *EMBO J.* 1998, **17**:1423-1433.

290. Knoepfler PS, Calvo KR, Chen H, Antonarakis SE, Kamps MP: Meis1 and pKnox1 bind DNA cooperatively with pbx1 utilizing an interaction surface disrupted in oncoprotein E2a-pbx1. *Proc.Natl.Acad.Sci.U.S.A.* 1997, **94**:14553-14558.

291. Rieckhof GE, Casares F, Ryoo HD, Abu-Shaar M, Mann RS: Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein. *Cell* 1997, **91**:171-183.

292. Shen WF, Montgomery JC, Rozenfeld S, Moskow JJ, Lawrence HJ, Buchberg AM, Largman C: AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. *Mol.Cell Biol.* 1997, **17**:6448-6458.

293. Jacobs Y, Schnabel CA, Cleary ML: Trimeric association of Hox and TALE homeodomain proteins mediates Hoxb2 hindbrain enhancer activity. *Mol.Cell Biol.* 1999, **19**:5134-5142.

294. Jackson RL Lin H-Y, Chan L, Means AR: Amino acid sequence of a major apoprotein from hen plasma very low density lipoproteins. *J.Biol.Chem.* 1977, **252**:250-253.

295. Chan L, Jackson RL, O'malley BW, Means AR: Synthesis of very low density lipoproteins in the cockerel. *J Clin Inv* 1976, **58**:368-379.

296. Hillyard LA, White HM, Pangburn SA: Characterization of apolipoproteins in chicken serum and egg yolk. *Biochemistry* 1972, **11**:511-518.

297. Colgan V, Elbrecht A, Goldman P, Lazier CB, Deeley RG: The avian apoprotein II very low density lipoprotein gene: methylation patterns of 5' and 3' flanking regions during development and following induction by estrogen. *J.Biol.Chem.* 1982, **257**:14453-14460.

298. Wiskocil R, Bensky P, Dower W, Goldberger RF, Gordon JI, Deeley RG: Coordinate regulation of two estrogen-dependent genes in avian liver. *Proc.Natl.Acad.Sci.USA* 1980, **77**:4474-4478.

299. Gordon DA, Shelness GS, Nicosia M, Williams DL: Estrogen-induced destabilization of yolk precursor protein mRNA in avian liver. *J.Biol.Chem.* 1988, **263**:2625-2631.

300. Cochrane AW, Deeley RG: Estrogen-dependent activation of the avian very low density apolipoprotein II and vitellogenin genes - transient alterations in mRNA polyadenylation and stability early during induction. *J.Mol.Biol.* 1988, **203**:555-567.

301. Margot JB, Williams DL: Estrogen induces the assembly of a multiprotein messenger ribonucleoprotein complex on the 3'-untranslated region of chicken apolipoprotein II mRNA. *J.Biol.Chem.* 1996, **271**:4452-4460.

302. Ito Y, Azrolan N, O'connell A, Walsh A, Breslow JL: Hypertriglyceridemia as a result of human apo CIII gene expression in transgenic mice. *Science* 1990, **249**:790-793.

303. St Clair RW: The contribution of avian models to our understanding of atherosclerosis and their promise for the future. *Lab.Anim.Sci.* 1998, **48**:565-568.

304. Luskey KL, Brown MS, Goldstein JL: Stimulation of the synthesis of very low density lipoproteins in rooster liver by estradiol. *J.Biol.Chem.* 1974, **249**:5939-5947.

305. Haché RJG, Wiskocil R, Vasa M, Roy RN, Lau PCK, Deeley RG: The 5' noncoding and flanking regions of the avian very low density apolipoprotein II and serum albumin genes. *J.Biol.Chem.* 1983, **258**:4556-4564.

306. Haché RJG, Deeley RG: Organization, sequence and nuclease hypersensitivity of repetitive elements flanking the chicken apoVLDLII gene: extended sequence similarity to elements flanking the chicken vitellogenin gene. *Nucleic Acids Res.* 1988, **16**:97-113.

307. Hoodless PA, Roy RN, Ryan AK, Haché RJ, Vasa MZ, Deeley RG: Developmental regulation of specific protein interactions with an enhancerlike binding site far upstream from the avian very-low- density apolipoprotein II gene. *Mol.Cell.Biol.* 1990, **10**:154-164.

308. Hoodless PA, Ryan AK, Schrader TJ, Deeley RG: Characterization of liver-enriched proteins binding to a developmentally demethylated site flanking the avian apoVLDLII gene. *DNA Cell Biol.* 1992, **11**:755-765.

309. Grant CE, Deeley RG: Cloning and characterization of chicken YB-1: Regulation of expression in the liver. *Mol.Cell.Biol.* 1993, **13**:4186-4196.

310. Grant CE, Vasa MZ, Deeley RG: cIRF-3, a new member of the Interferon Regulatory Factor (IRF) family that is rapidly and transiently induced by dsRNA. *Nucleic.Acids Res.* 1995, **23**:2137-2146.

311. Baniahmad A, Muller M, Steiner CH, Renkawitz R: Activity of two different silencer elements of the chicken lysozyme gene can be compensated by enhancer elements. *EMBO J* 1987, **6**:2297-2303.

312. Wijnholds J, Muller E, Ab G: Oestrogen facilitates the binding of ubiquitous and liver-enriched nuclear proteins to the apoVLDL II promoter *in vivo*. *Nucleic Acids Res* 1991, **19**:33-41.

313. Ryan AK, Schrader TJ, Burtch-Wright R, Buchanan L, Deeley RG: Characterization of Protein Interactions with Positive and negative elements regulating the apoVLDLII gene. *DNA Cell Biol.* 1994, **13**:987-999.

314. Ryan AK, Tejada ML, May DL, Dubaova M, Deeley RG: Isolation and characterization of the chicken homeodomain protein, AKR. *Nucleic Acids Res.* 1995, **23**:3252-3259.

315. Beekman JM, Wijnholds J, Schippers IJ, Pot W, Gruber M, Ab G: Regulatory elements and DNA-binding proteins mediating transcription from the chicken very-low-density apoplipoprotein II gene. *Nucleic Acids Res.* 1991, **19**:5371-5377.

316. Wijnholds J, Philipsen JNJ, AB G: Tissue-specific and steroid-dependent interaction of transcription factors with the oestrogen-inducible apoVLDLII promoter *in vivo*. *EMBO J* 1988, **7**:2757-2763.

317. Ryan AK: *Characterization of DNA Binding Proteins which Regulate Expression of the Chicken apoVLDLII Gene. PhD Dissertation.* 1 edn. Kingston, Ontario, Canada: Queen's University; 1994.

318. Bertolino E, Wildt S, Richards G, Clerc RG: Expression of a Novel Murine Homeobox Gene in the Developing Cerebellar External Granular Layer During Its Proliferation. *Developmental Dynamics* 1996, **205**:410-420.

319. Bertolino E, Reimund B, Wildt-Perinic D, Clerc RG: A Novel Homeobox Protein Which Recognizes a TGT Core and Functionally Interferes with a Retinoid-responsive Motif. *J.Biol.Chem.* 1995, **270**:31178-31188.

320. Tejada, ML., May, DL, Jia, Z, And Deeley, RG. Determinants of the DNA-binding specificity of the Avian homeodomain protein, AKR. *DNA Cell.Biol.* 1999, **18**:791-804.

321. Moskow JJ, Bullrich F, Huebner K, Daar IO, Bucher NLR: *Meis1,* a *PBX1*-Related Homeobox Gene Involved in Myeloid Leukemia in BXH-2 Mice. *Mol.Cell.Biol.* 1997, **15**:5434-5443.

322. Nakamura T, Jenkins NA, Copeland NG: Identification of a new family of *Pbx*-related homeobox genes. *Oncogene* 1996, **13**:2235-2242.

323. Steelman S, Moskow JJ, Muzynski K, North C, Druck T, Montgomery JC, Huebner K, Daar IO, Buchberg AM: Identification of a conserved family of Meis1-related homeobox genes. *Genome Res.* 1997, **7**:142-156.

324. Lim DA, Gossen M, Lehman CW, Botchan MR: Competition for DNA binding sites between the short and long forms of E2 dimers underlies repression in bovine papillomavirus type 1 DNA replication control. *J.Virol.* 1998, **72**:1931-1940.

325. Mueller PR, Wold B: *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* 1989, **246**:780-786.

326. Evans MI, Silva R, Burch JBE: Isolations of chicken vitellogenin I and III cDNAs and the developmental regulation of five estrogen-responsive genes in the embryonic liver. *Genes Dev.* 1988, **2**:116-124.

327. Berkowitz EA, Evans MI: Functional analysis of regulatory regions upstream and in the first intron of the estrogen-responsive chicken very low density apolipoprotein II gene. *J.Biol.Chem.* 1992, **267**:7134-7138.

328. Van Den Hoff MJB, Vermeulen JLM, De Boer PAJ, Lamers WH, Moorman AFM: Developmental changes in the expression of the liver-enriched transcription factors LF-B1, C/EBP, DBP and LAP/LIP in relation to the expression of albumin, a-fetoprotein, carbamoylphosphate synthase and lactase mRNA. *Histochem.J.* 1994, **26**:20-31.

329. Cooney AJ, Leng X, Tsai SY, O'Malley BW, Tsai Mj: Multiple mechanisms of chicken ovalbumin upstream promoter transcription factor-dependent repression of transactivation by the vitamin D, thyroid hormone, and retinoic acid receptors. *J.Biol.Chem.* 1993, **268**:4152-4160.

330. Calkhoven, CF, Snippe L, and AB G: Differential stimulation by CCAAT/enhancer-binding protein alpha isoforms of the estrogen-activated promoter of the very-low-density apolipoprotein II gene. *Eur.J.Biochem.* 1997, **249**:113-120.

331. Vollbrecht E, Veit B, Sinha N, Hake S: The developmental gene Knotted-1 is a member of a maize homeobox gene family. *Nature* 1991, **350**:241-243.

332. Bürglin TR: Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic.Acids.Res.* 1997, **25**:4173-4180.

333. Berthelsen J, Vandekerkhove J, Blasi F: Purification and Characterization of UEF3, A Novel Factor Involved in the Regulation of the Urokinase and Other AP-1 Controlled Promoters. *J.Biol.Chem.* 1997, **271**:3822-3830.

334. Chen H, Rossier C, Nakamura Y, Lynn A, Chakravarti A, Antonarakis SE: Cloning of a Novel Homeobox-Conatining Gene, *PKNOX1*, and Mapping to Human Chromosome 21q22.3. *Genomics* 1997, **41**:193-200.

335. Margalit Y, Yarus S, Shapira E, Gruenbaum Y, Fainsod A: Isolation and characterization of target sequences of the chicken CdxA homeobox gene. *Nucleic.Acids.Res.* 1993, **21**:4915-4922.

336. Ausubel FM: Edited by Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG , Smith JA, Struhl K. New York: John Wiley & Sons Inc.; 1997

337. Sensel MG, Binder R, Lazier C, Williams DL: Reactivation of apolipoprotein II gene transcription by cycloheximide reveals two steps in the deactivation of estrogen receptor-mediated transcription. *Mol.Cell.Biol.* 1994, **14**:1733-1742.

338. Laughon, A. DNA binding specificity of homeodomains. *Biochemistry* 1991, **30**:11357-11367.

339. Jones TA, Zhou JY, Cowan SW, Kjeldgaard M: Improved methods for building protein models in electron density maps and the localization of errors in these models. *Acta Crystallogr.* 1991, **47**:110-119.

340. Brünger AT: *X-Plor (Version 3.1) Manual A System for X-ray Crystallography and NMR.*, New Haven: Yale University Press; 1992.

341. Treisman J, Gonczy P, Vashishtha M, Harris E, Desplan C: A single amino acid can determine the DNA binding specificity of homeodomain proteins. *Cell* 1989, **59**:553-562.

342. Wilson DS, Sheng GJ, Jun S, Desplan C: Conservation and diversification in homeodomain-DNA interactions: A comparative genetic analysis. *Proc.Natl.Acad.Sci.USA* 1996, **93**:6886-6891.

343. Sharkey M, Graba Y, Scott MP: Hox genes in evolution: Protein surfaces and paralog groups. *Trends in Genetics* 1997, **13**:145-151.

344. Hanes SD, Brent R: A genetic model for interaction of the homeodomain recognition helix with DNA. *Science* 1991, **251**:426-430.

345. Sawamoto K, Okano H, Kobayakawa Y, Hayashi S, Mikoshiba K, Tanimura T: The function of *argos* in regulating cell fate decisions during *Drosophila* eye and wing vein development. *Developmental Biology* 1994, **164**:267-276.

346. Klein-Hitpass L, Tsai SY, Greene GL, Clark JH, Tsai MJ, O'Malley BW: Specific binding of estrogen receptor to the estrogen response element. *Mol.Cell.Biol.* 1989, **9**:43-49.

347. Weiler S, Gruschus JM, Tsao DH, Yu L, Wang LH, Nirenberg M, Ferretti JA: Site-directed mutations in the vnd/NK-2 homeodomain. Basis Of variations in structure and sequence-specific DNA binding. *J.Biol.Chem.* 1998, **273**:10994-11000.

348. Ades SE, Sauer RT: Specificity of minor-groove and major-groove interactions in a homeodomain-DNA complex. *Biochemistry* 1995, **34**:14601-14608.

349. Pellizzari L, Tell G, Fabbro D, Pucillo C, Damante G: Functional interference between contacting amino acids of homeodomains. *FEBS Letts* 1997, **407**:320-324.

350. Krusell L, Rasmussen I, Gausing K: DNA binding sites recognised *in vitro* by a knotted class 1 homeodomain protein encoded by the hooded gene, k, in barley (*hordeum vulgare*). *FEBS Letts* 1997, **408**:25-29.

351. Phelan ML, Sadoul R, Featherstone MS: Functional differences between HOX proteins conferred by two residues in the homeodomain N-terminal arm. *Mol.Cell Biol.* 1994, **14**:5066-5075.

352. Kraulis, PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr.* 1991, **24**:946-950. 4-16.

353. Merrit, EA and Bacon, DJ Raster3D: photorealistic molecular graphics. *Methods Enzymol.* 1997, **277**:505-524.

354. Herr W, Sturm RA, Clerc RG, Corcoran LM, Baltimore D, Sharp PA, Ingraham HA, Rosenfeld MG, Finney M, Ruvkun G, Horvitz HR: The POU domain: a large conserved region in the mammalian Pit-1, Oct-1, Oct-2, and *Caenorhabditis elegans* unc-86 gene products. *Genes Dev.* 1988, **2**:1513-1516.

355. Robertson M: Homeoboxes, POU proteins and the limits to promiscuity. *Nature* 1988, **336**:522-524.

356. Ryoo HD, Mann RS: The control of trunk Hox specificity and activity by Extradenticle. *Genes Dev.* 1999, **13**:1704-1716.

357. Pai CY, Kuo TS, Jaw TJ, Kurant E, Chen CT, Bessarab DA, Salzberg A, Sun YH: The Homothorax homeoprotein activates the nuclear localization of another homeoprotein, extradenticle, and suppresses eye development in Drosophila. *Genes Dev.* 1998, **12**:435-446.

358. Benson DA, Boguski MS, Lipman DJ, Ostell J, Ouellette BF, Rapp BA, Wheeler DL: GenBank. *Nucleic.Acids.Res.* 1999, **27**:12-17.

359. Duba, M.: Developmental analysis of AKR mRNA expression in chicken embryos, Kingston: Queen's University; 1996.

360. Laskowski, RA, Macarthur, MW, Moss, DS, and Thornton, JM PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* 1993, **26**:283-291.

361. Nicholls, A., Sharp, K. A., and Honig, B. Protein folding and association: insights from interfacial and thermodynamic properties of hydrocarbons. *Proteins* 1991, **11**:281-296.

362. Bürglin Tr, Ruvkun G: New motif in PBX genes. *Nature Genet.* 1992, **1**:319-320.

363. Spitz F, Demignon J, Porteu A, Kahn A, Concordet Jp, Daegelen D, Maire P: Expression of myogenin during embryogenesis is controlled by *Six/sine oculis* homeoproteins through a conserved MEF3 binding site. *Proc.Natl.Acad.Sci.U.S.A.* 1998, **95**:14220-14225.

364. Li H, Tejero R, Monleon D, Bassolinoklimas D, Abateshen C, Bruccoleri RE, Montelione GT: Homology modeling using simulated annealing of restrained molecular dynamics and conformational search calculations with CONGEN: Application in predicting the three dimensional structure of murine homeodomain MSX 1. *Protein Science* 1997, **6**:956-970.

365. Fraenkel E, Pabo CO: Comparison of X-ray and NMR structures for the Antennapedia homeodomain- DNA complex. *Nat.Struct.Biol.* 1998, **5**:692-697.