Chapter 3

MODIFYING THE BINDING SPECIFICITY OF THE HOMEODOMAIN PROTEIN, AKR

3.1 Statement of Authorship

This section was written for submission to DNA and Cell Biology. The authors are M.L. Tejada and R.G. Deeley. I performed all of the experiments and wrote the initial draft of the manuscript as well as participating in its subsequent editing. Special thanks to Mr. Vitali Petrounevitch for his tremendous assistance with all of the modelling presented in this chapter.

3.2 Abstract

Homeodomain-containing proteins are key regulators of development and differentiation. Although the overall structures of homeodomain-containing proteins vary significantly, structural studies of a number of homeodomains, free or complexed to DNA indicate that they bind to their recognition sequences in a highly conserved manner. Avian Knotted-Related (AKR) is a highly unusual vertebrate homeodomain protein with an Ile at position 50 of the DNA recognition helix, helix 3. Presently, along with several other Ile50-containing homeodomain proteins that include Meis1, the Meis-related group (MRG) of proteins and PKNOX/Prep1, AKR is classified on the basis of a three amino acid loop extension (TALE) inserted between helix 1 and 2. The Ile50-containing homeodomain proteins share a significant degree of identity within their helix 3 regions and consequently recognize a binding element 5'-TGACAG-3', designated the MPRE for Meis/Prep recognition element. However, AKR differs from the other Ile50-containing TALE proteins in that its homeodomain is located near the NH_2-terminus of the protein and it displays high affinity

binding to a site, Opt-1, whose sequence matches that of the MPRE, in the absence of cofactor(s). In addition, studies thus far have indicated that AKR does not interact with any of the cofactors demonstrated for Meis1 or Prep1. To determine the underlying factors responsible for the differences observed in binding to the MPRE sequence by AKR or Meis1 and Prep1, we mutated specific nucleotides within the hexanucleotide core as well as the 5' and 3' flanking regions. These studies revealed that AKR makes specific contacts within an area of approximately 13 nucleotides. The potentially greater number of contacts available may explain, in part, the higher binding affinity observed for AKR. We also demonstrate that the amino acid residues that determine the DNA-binding specificity and affinity of AKR reside within the homeodomain. To examine the relative contributions of specific amino acids within the NH_2-terminal arm or helix 3 regions to the DNA-binding affinity or specificity of AKR, the amino acids at positions 4, 6, 47, 50, 52 and 54 of the homeodomain were mutated. PCR-mediated target site selection was performed to determine the optimal binding sites of these mutant proteins. The result of these studies indicate that increasing the number of basic residues in the NH_2-terminal arm results in mutant proteins that exhibit an increased variability in their nucleotide preferences throughout the hexanucleotide core. In contrast, the helix 3 mutants displayed changes in their preferences for nucleotides at specific positons within the hexanucleotide core of the AKR binding site. Lastly, we have refined our molecular model of the AKR/DNA complex and used it to rationalize the impact of the mutations introduced into AKR, upon the binding specificity of these mutant proteins.

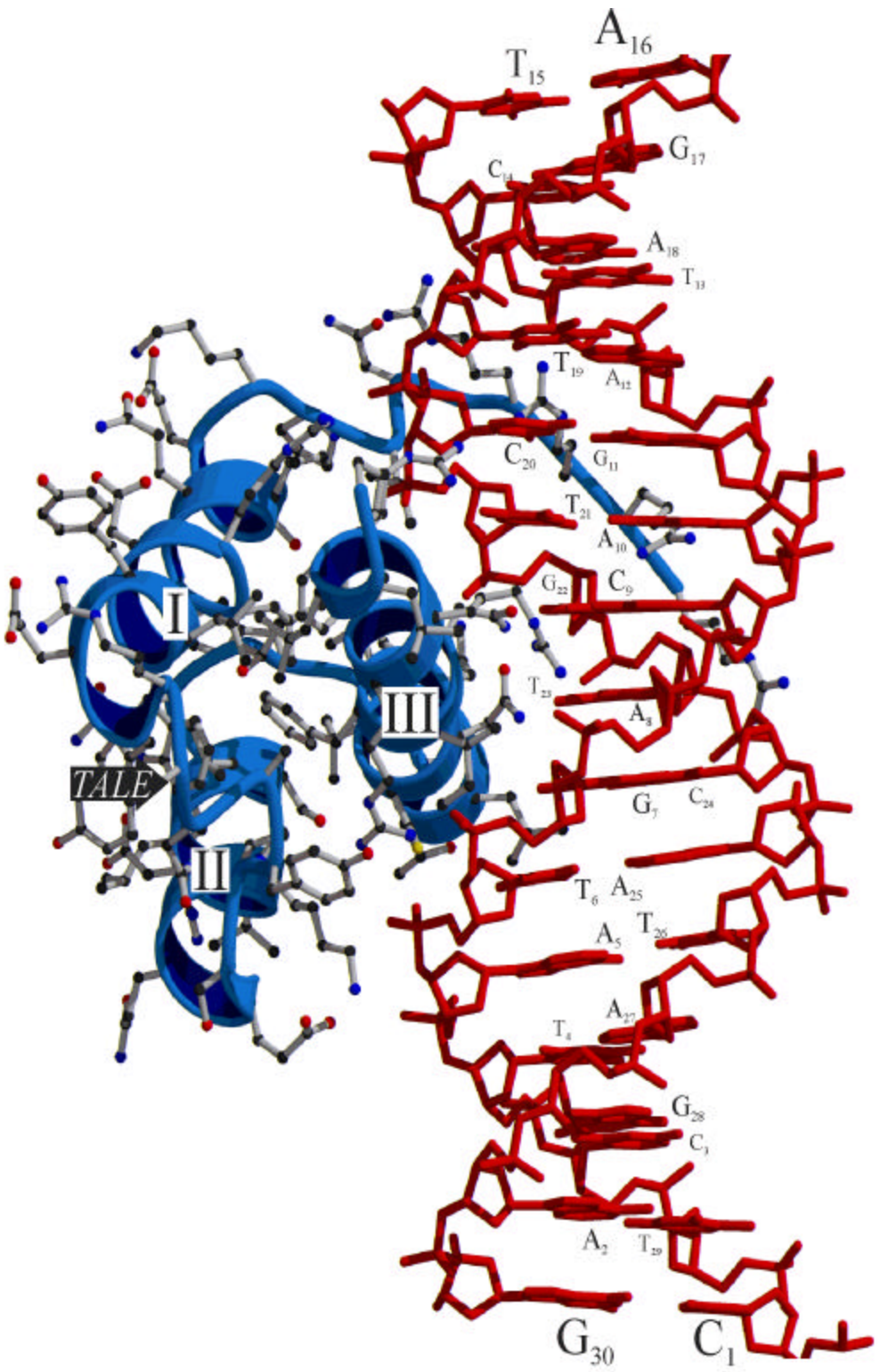
3.3 Introduction

The avian Very Low Density apolipoprotein II (apoVLDLII) is a major egg-yolk protein produced exclusively in hen liver during vitellogenesis. Hepatic expression of this gene can also be induced in embryos, chicks and roosters by administration of estrogen [295; 298]. All of the elements essential for the efficient hormone-responsive expression of the apoVLDLII gene can be found

within a 300-nucleotide region of the proximal promoter [327-330]. Deletion and mutational analyses demonstrated that a region spanning -229 to -247 from the major transcriptional start site is critical to the estrogen-responsiveness of the gene [313]. The sequence of this site, designated F', bore no resemblance to any previously characterized factor binding site. EMSA performed using nuclear extracts from a variety of tissues and an oligonucleotide binding site representing the F' sequence, indicated the existence of a slower migrating complex that was abundant in all tissues examined except blood [317]. The ubiquitously expressed factor responsible for this activity was cloned from a day 9 embryonic liver cDNA library using labelled concatamers of F' [314]. At the time of its characterization, the sequence of this factor was most similar to that of the maize homeodomain protein Knotted-1, hence it was designated Avian Knotted-Related (AKR).

The homeodomain is one of the most widely studied, highly conserved DNA-binding motifs and hundreds of homeodomain-containing proteins have presently been identified. The amino acids that dictate the secondary structure of the homeodomain are highly conserved across species. Therefore, it is not surprising that proteins containing these motifs bind to their regulatory elements in a highly analogous manner (Figure 3.1) [259-262]. The recognition helix, helix 3 of the homeodomain is inserted into the major groove of the DNA where it establishes the majority of base-specific contacts. Within this helix, varying levels of importance have been ascribed to amino acids at positions 47, 50, 51 and 54. Of these, the residues at positions 50 and 54 have been suggested to be the most critical determinants of binding specificity [259; 263-265].

Figure 3.1. **A ribbon diagram of the AKR homeodomain bound to its DNA recognition element.** Amino acids 2 to 61 of AKR were threaded using the data generated from the crystal structure of yeast TALE homeodomain protein MATá2 complexed to its recognition element [279]. The resultant structure was coupled to a 15-base pair fragment of DNA was generated with Sybil v6.4 using the sequence of the optimal binding site of AKR, Opt-1. The DNA/protein complex underwent energy minimisation and Procheck analysis. This structure was then rendered using MolScript and Raster3D software [352; 353]. The positions of helices 1 to 3 are shown along with the position of the TALE region. The sequence of the double-stranded fragment of DNA is presented numbered from 1 to 30, along with the positions of the major and minor grooves.


A16
T15
G17
C14
A18
T13
T19
A12
C20
G11
T21
A10
G22
C9
T23
A8
G7
C24
T6
A25
A5
T26
A27
T4
G28
C3
A2
T29
G30
C1
I
II
III
TALE
Major
Minor
5′-C 1A 2C 3T 4A 5T 6G 7A 8C 9A 10G 11A 12T 13C 14T 15-3′
3′-G30T29G28A27T26A25C24T23G22T21C20T19A18G17A16-5′

Although the amino acids that determine the DNA binding specificity of the homeodomain are predominantly located within helix 3, it appears that different combinations of amino acids at positions 3 to 7 in the NH_2-terminal arm also contribute to the preference of a specific homeodomain protein for particular nucleotides within its binding site core [260; 266-268]. In these cases, it has been suggested that differences in the amino acid composition of the NH_2-terminal arm of these proteins, specifically the number of basic residues present, can result in small changes in the orientation of the amino acids side chains in the NH_2-terminus. This is then transmitted through the protein resulting in minor adjustments of side chains in helix 3. This significantly influences the stability, and consequently the specificity with which the proteins bind to their recognition elements. Subtle repositioning of the protein has been used to explain why the Ubx and Dfd homeodomains with nearly identical recognition helices select binding sites that differ in the region presumably contacted by helix 3 [269].

There are a number of puzzling observations concerning the DNA-binding specificity described for homeodomain proteins. The binding motifs involved are typically only 5-6 base pairs in length and as such are encountered frequently in the genome. In addition, homeobox proteins that specify entirely different biological functions are capable of recognizing very similar, if not identical, DNA sequences *in vitro* [257; 258]. Therefore, it is difficult to reconcile the biological diversity displayed by these proteins *in vivo* with the apparent redundancy and lack of specificity suggested by *in vitro* studies.

In order to explain the molecular basis of homeodomain/DNA recognition it is necessary to integrate three types of information: i) the structure of the homeodomain of interest, ii) the identification of the amino acid-base pair contacts that influence the specificity of DNA/protein interactions and iii) the mode of binding to DNA (monomeric binding, homo/heterodimerization).

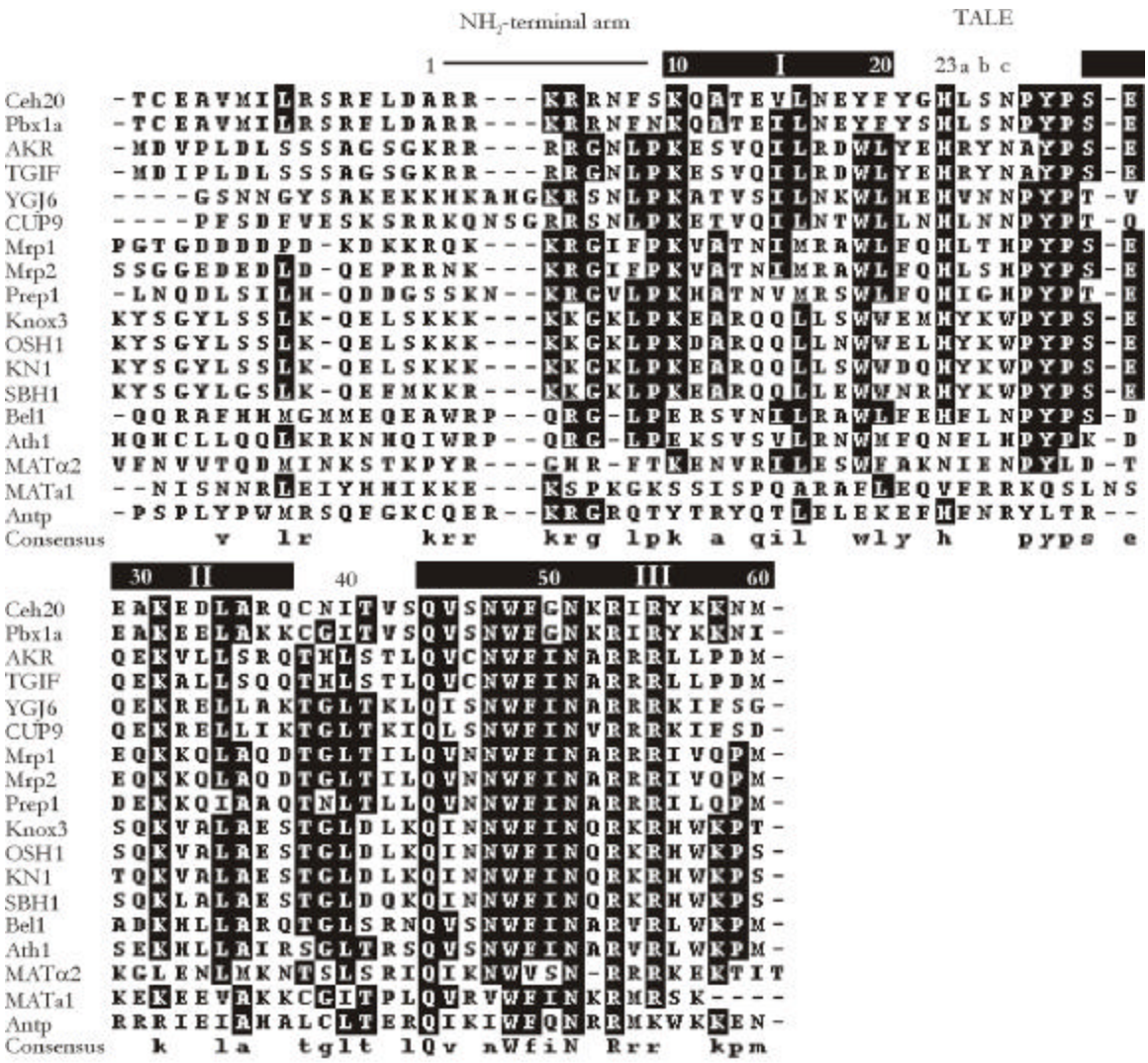
The binding specificities of several families of homeodomain proteins are influenced by the presence of an extra domain with DNA-binding properties, as in the case of the CUT homeodomain proteins, the POU domain proteins, the LIM homeodomain proteins and the paired domain groups. This extra domain augments the stringency with which these proteins recognize their response elements [270; 354; 355]. Although, sequence analysis of AKR revealed no other recognized DNA binding motifs, the homeodomain does contain a three amino acid insertion between helices 1 and 2. This feature has been used to categorize AKR as a member of the TALE (three amino acid loop extension) group of homeodomain proteins [248; 332].

The involvement of a homeodomain protein partner can also raise the binding specificity and affinity of homeodomain/DNA interactions. This is exemplified by the interactions of the yeast mating type homeodomain proteins MATa1/MATá2. In the diploid a/á cell type, the two homeodomain proteins form heterodimers that binds to sites upstream of haploid-specific genes (hsg) [278]. In the absence of á2, the a1 protein exhibits no detectable specific binding to DNA.. However, the presence of a1 in solution dramatically raises the affinity of á2 for *hsg* operators. [279]. In a manner analogous to the interactions of MATa1 and MATá2, the members of the PBC family of proteins that include the mammalian oncoproteins Pbx1-3 and their various isoforms, the *Caenorhabditis elegans* ceh-20 and their *Drosophila* orthologue, extradenticle (Exd), have been demonstrated to act as co-factors for a number of Hox proteins [280-285]. PBC/Hox dimers bind to elements that are composites of PBC and Hox recognition elements. Furthermore, the choice of Hox protein partner depends upon subtle differences within the DNA binding site. Recently, Pbx1 has also been shown to interact with the Ile50-containing members of the TALE group of homeodomain proteins, which include Meis1, the Meis-related proteins Mrg1 (Meis2) and Mrg2 (Meis3) and Prep1/pKnox1 [263; 288-290]. These proteins share a high degree of identity throughout the DNA recognition helix and as a result bind a common recognition motif, 5'-TGACAG-3', designated the MPRE or Meis1/Prep1 Recognition Element. The sequence of the

MPRE is quite divergent from the canonical homeodomain 5'-TAAT-3' motifs. Meis1 and Prep1 bind to DNA as heterodimers with members of the PBC class of proteins [263; 288-290]. More recently, homothorax (Hth), the *drosophila* orthologue of Meis1, has been demonstrated to be essential to the nuclear localization of exd [291; 356; 357]. Meis1 has also been shown to dimerize with members of the AbdB-like family of homeodomain proteins in manner that increases the stability of Meis1 binding to its recognition element [292]. Finally, Jacobs *et al*. (1999) have shown that trimers comprised of Pbx/Meis and Hoxb1 regulate the transcriptional activity of the *Hoxb2* enhancer [293].

AKR was the first vertebrate homeobox protein to be identified with an Ile at position 50 of helix 3. It binds to an atypical homeodomain recognition element within the F' site of the form 5'-TGACAT-3', which exhibits a 5/6 match to the MPRE [313; 314]. PCR-assisted target site selection indicated that AKR bound optimally to a site with a sequence identical to the MPRE [320]. Presently, AKR is classified as a member of the Ile50-containing group of TALE homeodomain proteins. However, a number of characteristics distinguish AKR from the other members of this group. The homeodomain of AKR is situated near the NH_2-terminus of the protein, in contrast to the more traditional COOH-terminal location. The composition of the NH_2-terminal arm of the AKR homeodomain also differs from that of the other Ile50-containing TALE proteins (Figure 3.2). Additionally, the amino acid residues comprising the TALE region participate in the interactions between Exd/Ubx, MATa1/MATá2, and Pbx1/HoxB1. These residues are highly conserved within groups of proteins. However, a high degree of variability exists across groups These three residues are LSN for members of the PBC group, LTH for Meis1, MRG1 and MRG2, IGH for Prep1 and RYN for AKR (Figure 3.2).

Figure 3.2. **A comparison of the amino acid sequences of a variety of homeodomain proteins to that of AKR.** A BLAST search for sequences similar to the homeodomain region of AKR was carried out using the services available at the NCBI [358]. Of these, only the sequences from a representative of each class of homeodomain proteins were subjected to CLUSTAL alignment available at http://www.hgsc.bcm.tmc.edu/tools. Subsequently, the homeodomain sequences of the *Drosophila* protein Antennapedia (Antp) and the yeast protein MATá2 were included. Homologies are indicated at the bottom of the alignment as similarities (shown in lowercase) and identities (shown in uppercase). The region of the NH_2-terminal arm, the helices of the homeodomain and the TALE region are indicated above the alignment.


NH2-terminal arm
TALE
1
10
I
20
23a b c
Ceh20
Pbx1a
AKR
TGIF
YGJ6
CUP9
Mrp1
Mrp2
Prep1
Knox3
OSH1
KN1
SBH1
Bel1
Ath1
MATα2
MATa1
Antp
Consensus
30
II
40
50
III
60

AKR does not contain any of the other protein interaction motifs that have been demonstrated to exist in the members of the PBC or Meis family of proteins. TGIF, the murine orthologue of AKR, has been shown to participate in the smad2-mediated downregulation of several TGFâ-responsive genes through its recruitment of histone deacetylase 1 (HDAC1) [324]. This indicates that TGIF, and by analogy AKR, may employ a different set of co-factors than those recruited by the PBC or MEIS proteins. AKR and TGIF both exhibit high affinity binding to their recognition elements *in vitro*. Additionally, AKR displays high affinity binding to sites (F' and Opt-1) that share a high degree of identity to the MPRE (5'-TGACAG-3') as well as a site, designated G, whose sequence is completely diverged from the MPRE [314; 320]. In contrast, the high affinity binding of the other members of this particular group of TALE homeodomain proteins is dependent upon interactions with the members of the PBC class or with other Hox proteins [263; 288-290]. Studies thus far, have indicated that both AKR and TGIF act as potent repressors of genes that are responsive to nuclear hormone receptors [314; 319; 320].

To begin to explain the variability in the binding affinities displayed by AKR and the Ile50-containing TALE proteins for their common consensus element, we introduced changes to various nucleotides outside of the hexanucleotide core binding element of the F' site. The results indicate that AKR establishes contacts to a 13 base pair region of DNA that encompasses this hexanucleotide core element. We have also generated truncations of AKR and demonstrate that regions outside of the homeodomain do not influence the specificity and affinity of binding. Structural and chemical information regarding the role of a particular residue in dictating the DNA binding specificity can be obtained in part, by systematic mutation of key predicted residues of the homeodomain. Here, we describe the effects of specific mutations, generated within the NH_2-terminal arm and the recognition helix of the homeodomain, on the DNA binding affinity and specificity of AKR. We have also refined our previous model to enable us to rationalize the effects that these mutations exert upon the binding specificity of this unusual homeodomain protein. This

model provides insights into the causes of the differences in binding affinity displayed by AKR and other Ile50-containing TALE proteins.

3.4 Materials and Methods

3.4.1 *Construction and expression of DNA-binding mutants of AKR*

DNA-binding mutants of AKR were generated using a recombinant PCR strategy. A GST-AKR_{1-178} construct was used as a template for all PCR reactions. This construct encodes the first 178 amino acids of the protein, which includes the first 34 NH_2-terminal amino acids, the 63 amino acid homeodomain and 81 COOH-terminal amino acids. In the first step of each reaction, NH_2-terminal as well as COOH-terminal fragments of AKR were PCR-amplified to include various point mutations using the sets of primers described in Table 3.1 (Appendix A) along with the invariant primers RGD981 and 5319. The primers were designed such that the two products overlapped by a region of approximately 15 nucleotides containing the mutated site. 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 dNTP, 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. In the second step, the two PCR products from the first reaction were mixed (approximately 10-20% of each), denatured, annealed and the recombinant product made double-stranded in the initial step of the PCR reaction, which contained no primers. The common NH_2- and COOH-terminal invariant primers (RGD981 and 5319) were then added and the PCR reaction was allowed to proceed as described above. The resultant product was subcloned into the PGEX2T vector using the *BamH*I and *Sma*I sites on the vector. All GST-fusion proteins used were expressed and isolated using the pGEX system (Pharmacia) according to the manufacturer's suggestions.

3.4.2 *Binding Site Selection*

The random oligonucleotide used previously to determine the optimal binding site of wild-type AKR was used in these experiments (RGD783, Table 3.1 Appendix A). This oligonucleotide consisted of 24 base pairs of random sequence flanked by elements of non-random sequence that were used as priming sites to PCR-amplify the selected sequences using primers RGD784 and 785 (Table 3.1, Appendix A). These non-random flanking sequences also contained *EcoR*I, *BamH*I and *Hind*III restriction sites in order to facilitate cloning of the selected PCR-amplified products. Site selection was carried out as described previously but with a few modifications [320]. Briefly, 5 ìg of the purified wild-type full-length AKR or the mutant GST-AKR_{1-178} fusion proteins, was incubated along with approximately 400 ng of the randomer binding site pool, on ice 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 dI· dC). Protein/DNA complexes were then precipitated using 5 ìl of a protein A-purified fraction of rabbit polyclonal anti-AKR antiserum generated by a former member of this laboratory [359]. This antibody displays specific binding to recombinant wild-type AKR and has never been observed to disrupt DNA binding by the wild-type protein in EMSA, suggesting that it would not affect the DNA binding of the mutant proteins. 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 centrifuging at 12 000 rpm for 30 s in a microcentrifuge (Eppendorf), the resulting pellets were 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 and 5 ìl was re-amplified by PCR to be used in a subsequent round of selection. Following 6 rounds of enrichment the GST-AKR_{1-178} mutant and their respective selected products were tested by EMSA, under conditions of binding site excess, in order to determine whether all of the mutants had retained DNA-binding activity. Positive selected sites, verified in this manner, were cloned into pBluescript SK^+ (Stratagene). An average of 30 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.

3.4.3 *Electrophoretic mobility Shift Assays (EMSAs)*

Semi-quantitative EMSAs were used to assess the relative binding affinity of the GST-AKR_{1-178} binding site mutant proteins for the wild-type F' as well as the various selected binding sites (Table 3.2 and Table 3.3, Appendix A). For these assays, the double-stranded binding site oligonucleotides were labelled to equivalent specific activities. An oligonucleotide complementary to the 3'-end of each site was labelled with ^{32}P-dATP using T_4 polynucleotide kinase (primer 7919, Table 3.1). A portion of the labelled oligonucleotide was then annealed to the single-stranded binding sites and the second strand synthesized using klenow fragment [336]. The resultant labelled double-stranded products were then purified using a 10% polyacrylamide gel (19:1 acrylamide:bis-acrylamide) run under non-denaturing conditions. The binding sites where then extracted from the gel by passive diffusion using elution buffer (50 mM Tris-HCl, 5 mM EDTA, 100 mM NaOAc, 0.5% SDS), and ethanol precipitated along with 1ìg of glycogen (Boehringer Mannheim). EMSAs were carried out as described previously [314]. Protein/DNA complexes were separated using 6% polyacrylamide gels (29:1 poly:bis-acrylamide) in 1 x Tris borate EDTA buffer at 140 V for 3 hours. The gels were dried and the bands visualized by autoradiography. For determination of equilibrium dissociation constants (K_d) a constant amount of protein (0.25 ìg)

from each of the AKR binding mutants was incubated with various amounts of labelled DNA (4 x 10^{-11} M to 2.0 x 10^{-9} M) for 30 minutes at 4ºC to allow protein/DNA complexes to reach equilibrium in EMSA binding buffer minus poly dI· dC. Electrophoresis conditions were as described above. Bound and Free DNAs were quantitated using the InstantImager™ (Electronic Autoradiography, Packard, Meridan, CT 06450). The apparent K_d of the GST-AKR_{1-178} binding mutants for their various selected sites was determined by Scatchard analyses.

3.4.4 *Western Blotting*

All of the steps described were performed at room temperature unless otherwise specified. Protein samples (0.5 to 1 ìg) were boiled in 1 X SDS sample-loading buffer (25 mM Tris-HCl pH 6.8, 4% (w/v) SDS, 20% (v/v) glycerol, 10% (v/v) 2-mercaptoethanol, 0.002% (w/v) bromophenol blue) for 5 to 10 minutes with vortexing. Particulate matter was then removed by centrifugation at 12000 rpm for 10 minutes. The various samples were subjected to SDS-PAGE (4% stacking gel, 12% separating gel) with a pre-stained broad-range protein standard (New England Biolabs). The gel was run at 140 V for 3 hours in 1 x protein gel running buffer (25 mM Tris, 250 mM glycine pH 8.3, 0.1% (w/v) SDS). The separated samples were subsequently transferred from the gel onto polyvinylidene difluoride (PVDF) membrane (Immobilon™-P, MILLIPORE). This membrane had been pre-cut to the dimensions of the gel, soaked in methanol for 10 sec and then rinsed in water. The membrane, as well as the protein gel, was soaked in 1 x transfer buffer (39 mM glycine, 48 mM Tris base, 0.037% (w/v) SDS, 20% (v/v) methanol) for 10 minutes. Transfer was carried out overnight at 120 mA using a TE Series Transphor Electrophoresis unit (HOEFER Scientific Instruments, San Francisco). The fidelity of the transfer was verified by staining the membrane in amido black (0.1% (w/v) Amido Black 10% (v/v) 2-propanol 2% (v/v) glacial acetic acid) for 30 s followed by four rounds of destaining in 100 mL of destain solution (25% (v/v) isopropanol, 10% (v/v) glacial acetic acid). The blots were air dried for 5-10 minutes and stored at 4ºC for up to 2

months. Alternatively, immediately following destaining the blot was washed 3 times, for 5 minutes each time, in 1 X TBS-T (20 mM Tris pH 7.5, 150 mM NaCl, 0.05% (v/v) Tween 20). Blocking was carried out in 30 mL blocking solution (4% (w/v) Carnation brand skim milk powder in 1 x TBS-T) for 1 hour. At this point the primary antibody (1/500 dilution) was added to 5 mL of fresh blocking solution and this was incubated for 4 hours with gentle shaking [359]. The blot was washed 5 times, for 5 minutes each time in 1 x TBS-T and subsequently incubated with a horseradish peroxidase-conjugated goat anti-rabbit monoclonal antibody (1/10000 dilution in 10 mL of blocking buffer) (PIERCE ImmunoPure® No. 31462, Lot 97061216). Following a 2 hour incubation the blot was again washed 5 times, for 5 minutes each time, in 1 x TBS-T. The Renaissance® system (NEN™ LifeScience Products) was used to develop the blot and the bands were visualized by autoradiography on X-ray film.

3.4.5 *Molecular modelling of the AKR homeodomain*

The amino acid sequence from position 2 to 61 of the AKR homeodomain (GenBank accession no. U25353) was submitted to SWISSPROT server at http://expasy.hcuge.ch/swissmod/SWISS-MODEL.html in FASTA format. This server threaded the á-carbons of AKR using the coordinates derived from the crystal structure of the MATá2 homeodomain complexed to DNA (MMDB ID 3857, PDB ID 1YRN). The resultant structure was then docked to a fragment of double-stranded DNA that was generated using Sybyl v6.4 software (Tripos Inc., St. Louis), as guided by the MATá2/DNA complex structure. The DNA sequence used was that of the optimal binding element of AKR, Opt-1 (5'-CTCTATGACAGATCT-3', the hexanucleotide core binding element is underlined). Following energy minimization the homeodomain/DNA complex was analysed using Procheck, a program that checks the stereochemical quality of protein structures [360]. The figures derived from this model were generated using the Molscript [352] and Raster3D programs [353]. The surface model depicting the putative hydrophobic pocket of AKR was generated using GRASP [361].

3.5 Results

3.5.1 *AKR recognizes DNA in a manner that is distinct from the Ile50-containing members of the TALE family of homeodomain proteins.*

Previous results indicated that AKR was capable of establishing contacts to nucleotides 3'- of its hexanucleotide response element. As shown in Figure 3.3, AKR exhibits an increased binding to the FM2 mutant site in which the GG dinucleotide 3' to the hexanucleotide core binding element, 5'-TGACAG-3', is mutated to AA (shown underlined in Figure 3.3A). In addition, a site selection protocol performed using *in vitro* translated AKR also indicated a preference for an ATCT tetranucleotide 3' to the 5'-TGACAG-3' core [320]. To examine whether AKR established an extended set of contacts to nucleotides 5' of the core binding element, a number of mutations were introduced to nucleotides in this region (binding sites FMC1 to FMC6, Table 3.3 Appendix A). The ability of AKR to bind these 6 sites was then tested in EMSAs. Full-length AKR displays a binding preference for these mutant binding sites in the order FMC3>FMC4>FMC5>FMC6/wt F'>FMC2 (Figure 3.4B). As expected AKR fails to recognize the FMC1 binding site since this particular mutation disrupts the core-binding element. These results demonstrate AKR binding preferences for TA>TT>CA>CT/GC>AT at positions 3 and 4, 5' of the hexanucleotide core binding element. The results indicate that AKR establishes contacts to a 13-base pair region of DNA, which is appreciably larger than the typical tetranucleotide elements bound by the hox proteins or the hexanucleotide elements bound by Meis1/Prep1 and the Meis-related homeodomain proteins. The greater number of contacts established by AKR to its recognition element may be responsible for the greater DNA binding activity observed for this protein. The differences in binding affinities observed for Meis1 and AKR could also result from structural differences between the two proteins, particularly in regions outside of their homeodomains.

Figure 3.3 **Electrophoretic mobility shift analysis of AKR binding to native and mutant F' binding sites.** *In vitro* translated full-length AKR was incubated with ^{32}P-labelled F', FM1 and FM2 oligonucleotide binding sites in the presence of pre-immune or immune antisera. A, the sequences of the F', FM1 and FM2 binding sites are provided. The hexanucleotide core of the F' site is underlined. The nucleotides confirmed by methylation interference analyses to mediate the binding of AKR to F' are designated with solid squares. B, Lane 1, no protein; lanes 2-7, 1ìL unprogrammed lysate in the presence of FM2 (lanes 2 and 3), FM1 (lanes 4 and 5) and F' (lanes 6 and 7) and 3 ìL of a rabbit polyclonal anti-AKR antibody (lanes 3, 5 and 7); lanes 8-16, 1 ìL of *in vitro* translated full-length AKR in the presence of FM2 (lanes 8-10), FM1 (lanes 11-13), and F' (lanes 14-16) and 3 ìL of pre-immune (lanes 10, 13 and 16) or immune sera (lanes 9, 12 and 15) as indicated [adapted from *Ryan et al*. (1995) [314]].

A


Wt F' 5'-GGGGCCTCTATGACATGGTTGCCTGAAAA-3'
FM1 5'-GGGGCCTCTATTAAATGGTTGCCTGAAAA-3'
FM2 5'-GGGGCCTCTATGACATTTTTGAATGAAAA-3'

B

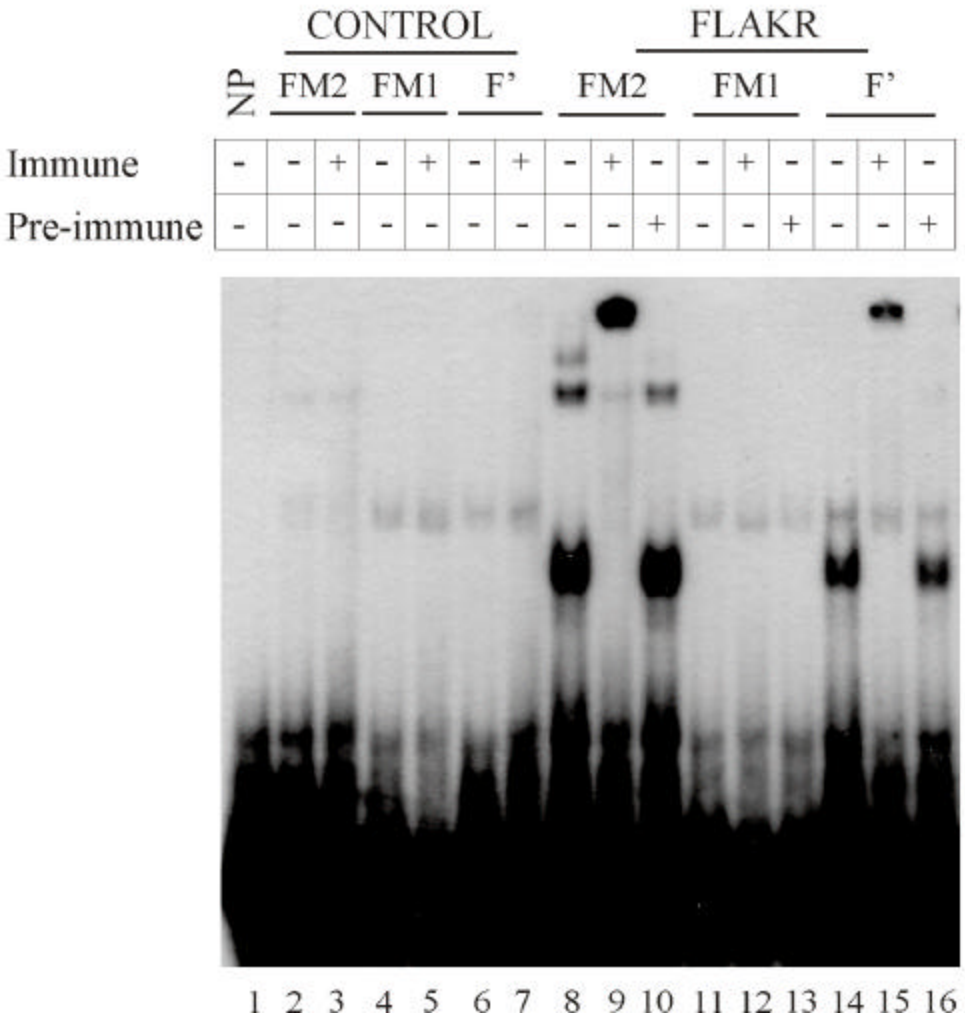
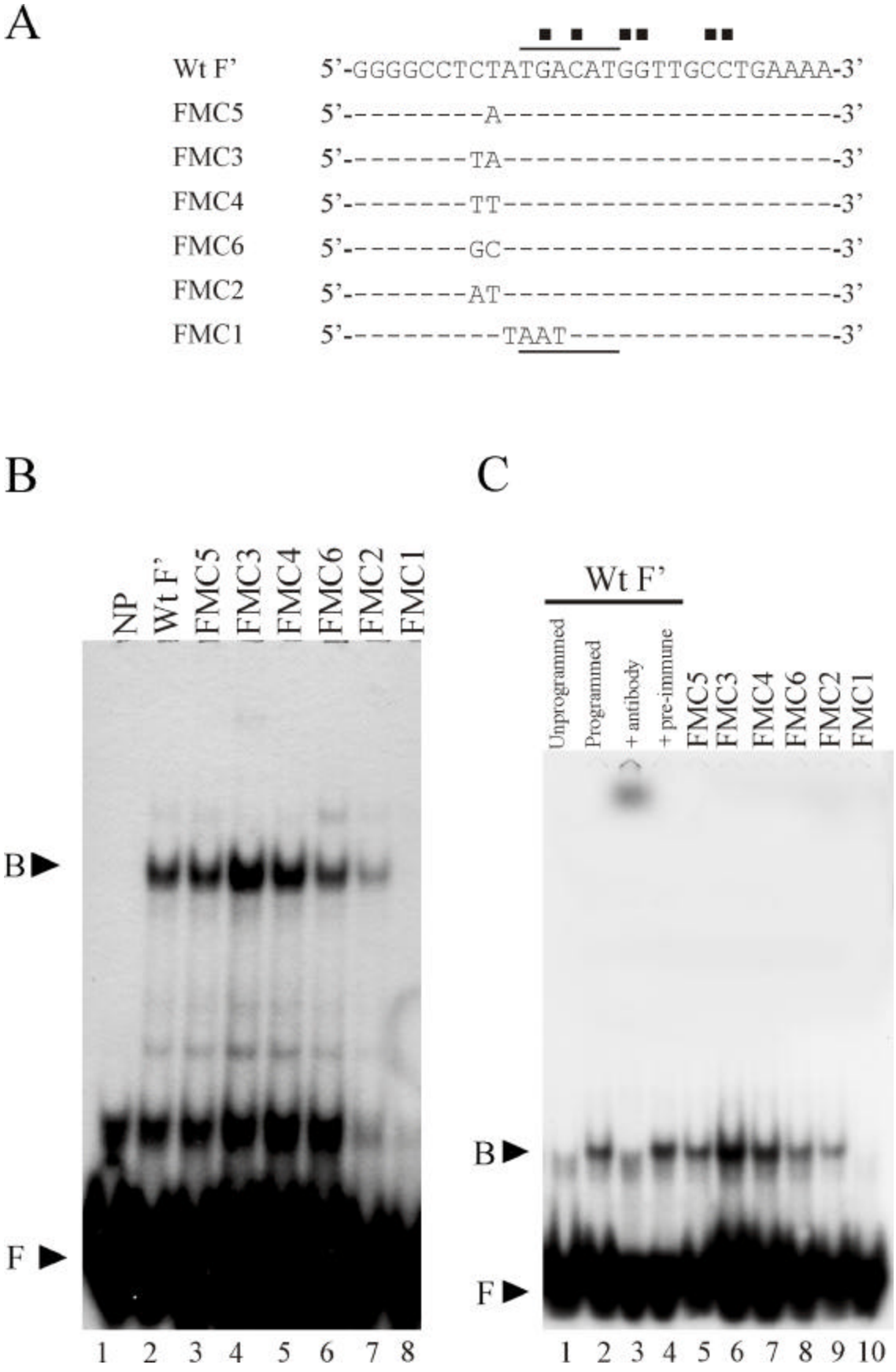
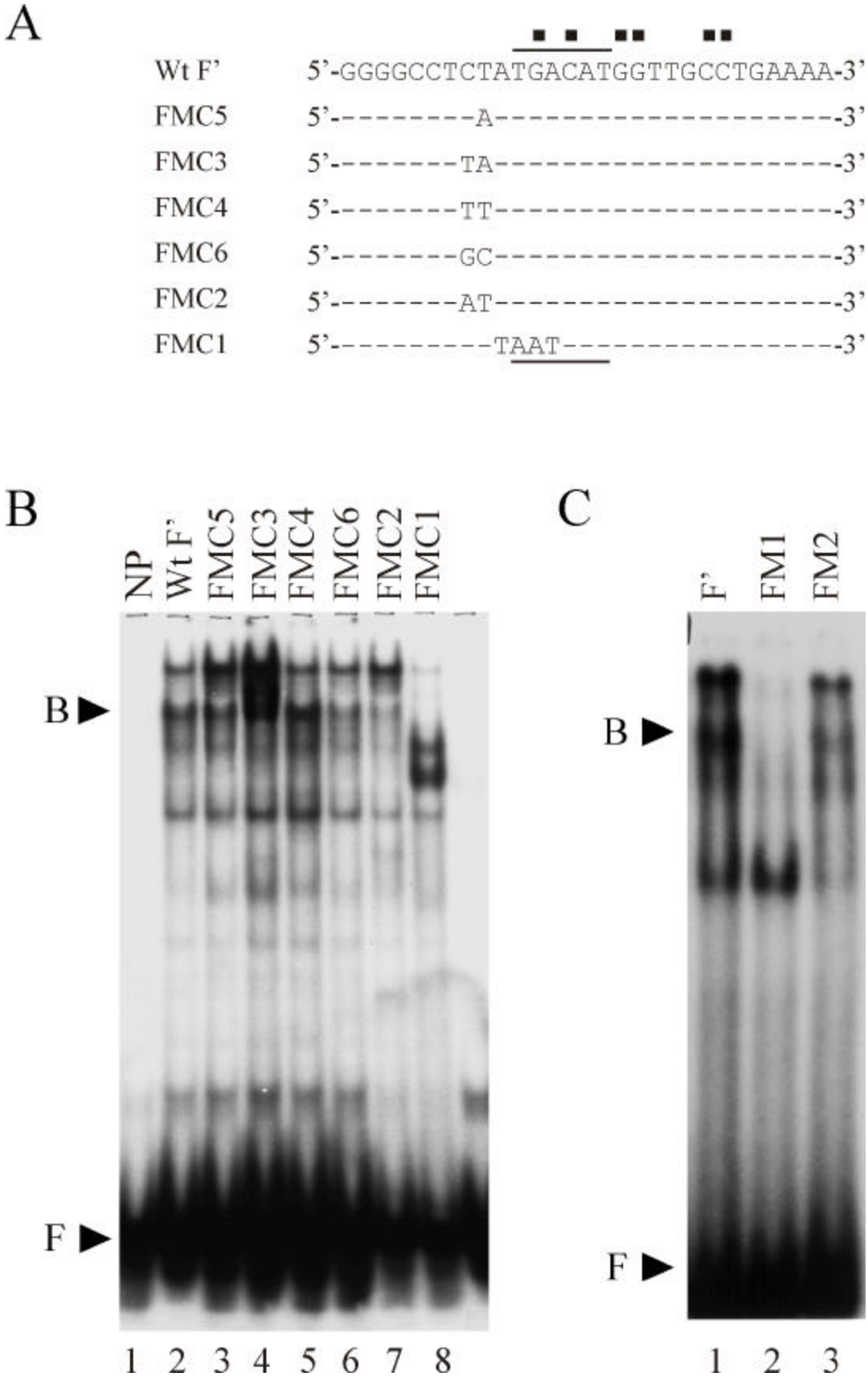

CONTROL
FLAKR
NP
FM2
FM1
F'
FM2
FM1
F'
Immune
Pre-immune
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Figure 3.4. **Electrophoretic mobility shift analysis of AKR binding to native and mutant F' binding sites.** *In vitro* translated full-length AKR was incubated with ^{32}P-labelled F', FMC1 to FMC6 oligonucleotide binding sites. A, the sequences of the respective binding sites are presented. Apart from the sequence of F', only the nucleotide differences are listed. The hexanucleotide core of the binding sites is demarkated with a black bar. The positions of the residues determined by methylation interference to be contacted by AKR are denoted by black squares. B, lane 1, no protein; lanes 2-8, 1 ìL *in vitro* translated full-length AKR incubated in the presence of wild-type F' (lane 2), FMC5 (lane 3) and FMC3 (lane 4), FMC4 (lane 5), FMC6 (lane 6), FMC2 (lane 7), FMC1 (lane 8). C, lanes 1-4, 1 ìL of unprogrammed rabbit reticulocyte lysate (lane 1) or 1 ìL of *in vitro* translated AKR_{21-114} (programmed) (lanes 2-4) incubated in the presence of wild-type F'. The construct AKR_{21-114} encodes the homeodomain at position 35 to 99 as well as 20 NH_2-terminal and 15 COOH-terminal amino acids. Lane 2, 1 ìL of *in vitro* translated $AKR_{21-114.}$ Lane 3, 3 ìL of a polyclonal anti-AKR antibody; lane 4, 3 ìL of preimmune serum. Lanes 5-10, 1 ìL of *in vitro* translated AKR_{21-114} incubated in the presence of FMC5 (lane 3), FMC3 (lane 4), FMC4 (lane 5), FMC6 (lane 6), FMC2 (lane 7), FMC1 (lane 8). Bound and Free binding sites are designated by B and F, respectively.


A
Wt F'
5'-GGGGCCTCTATGACATGGTTGCCTGAAAA-3'
FMC5
5'---------A----------3'
FMC3
5'--------TA----------3'
FMC4
5'--------TT----------3'
FMC6
5'--------GC----------3'
FMC2
5'--------AT----------3'
FMC1
5'----------TAAT----------3'
B
NP
Wt F'
FMC5
FMC3
FMC4
FMC6
FMC2
FMC1
B
F
1 2 3 4 5 6 7 8
C
Wt F'
Unprogrammed
Programmed
+ antibody
+ pre-immune
FMC5
FMC3
FMC4
FMC6
FMC2
FMC1
B
F
1 2 3 4 5 6 7 8 9 10

To determine whether regions outside the homeodomain are important for the binding specificity or activity of AKR, various truncations of the protein were generated, *in vitro* translated and used in EMSA along with the mutant binding sites FMC1 to FMC6. The results of EMSA performed using these truncations along with the various mutant binding sites, indicate that the binding specificity of AKR resides in the homeodomain itself. Therefore, only the results obtained with $AKR_{21\text{-}114}$ are shown (Figure 3.4C). To determine if AKR displayed similar binding *in vivo*, EMSAs were performed using liver nuclear extracts along with the various mutant binding sites (FMC1 to FMC6). As shown in figure 3.5, EMSAs using these nuclear extracts and the mutant binding sites resulted in the formation of four complexes. Using mutant binding sites that abrogate or enhance AKR binding (FM1 and FM2 respectively) we have shown previously that the DNA/protein complex designated B contains AKR (Figure 3.5C) [314]. The relative binding of this complex to the mutant sites was similar to that obtained using the *in vitro* translated full-length AKR or $AKR_{21\text{-}114}$ indicating that AKR dictates the specificity of DNA binding within this presumed complex.

Figure 3.5. **Electrophoretic mobility shift analysis of AKR binding to native and mutant F' binding sites.** Nuclear extracts (0.5ìg) isolated from estrogen-treated rooster liver were incubated with ^{32}P-labelled F', FMC1 to FMC6 oligonucleotide binding sites. A, the sequences of the respective binding sites are presented. Apart from the sequence of F', only the nucleotide differences are listed. The hexanucleotide core of the binding sites is demarked with a black bar. The positions of the residues determined by methylation interference to be contacted by AKR are denoted by hollow circles. B, lane 1, no protein; lanes 2-7, 2 ìL rooster liver nuclear extracts (0.25ìg/ ìL) incubated in the presence of wild-type F' (lane 2), FMC5 (lane 3) and FMC3 (lane 4), FMC4 (lane 5), FMC6 (lane 6), FMC2 (lane 7), FMC1 (lane 8). C, lane 2, 2 ìL rooster liver nuclear extracts (0.25ìg/ ìL) incubated in the presence of wild-type F' (lane 1), FM1 (lane 2) and FM2 (lane 3). The complex presumably containing AKR is designated with a B. The unbound sites are designated by F (free).


A
Wt F'
5'-GGGGCCTCTATGACATGGTTGCCTGAAAA-3'
FMC5
5'---------A----------3'
FMC3
5'--------TA----------3'
FMC4
5'--------TT----------3'
FMC6
5'--------GC----------3'
FMC2
5'--------AT----------3'
FMC1
5'----------TAAT----------3'
B
NP
Wt F'
FMC5
FMC3
FMC4
FMC6
FMC2
FMC1
B
F
1 2 3 4 5 6 7 8
C
F'
FM1
FM2
B
F
1 2 3

3.5.2 *Mutations in the homeodomain of AKR produce subtle differences in the manner in which the protein recognizes DNA.*

Previously, we introduced specific mutations in the NH_2-terminal arm and the helix 3 regions of AKR [320]. These AKR mutants exhibited levels of binding activity that were quite different from those observed for the wild-type protein. Furthermore, the mutations allowed these proteins to bind to sites not recognized by wild-type AKR. This suggested that the mutations had modified the target site requirements of the homeodomain. To determine the specific effects of these changes upon the binding site preferences of the variant AKR proteins, we performed PCR-assisted binding site selection using a protocol that we have used previously to determine the optimal binding site of wild-type AKR [320]. AKR recombinant proteins were used to screen a library of binding sites containing 24 base pair cores of random sequence. These 24 base pair cores are flanked by 20 base pairs of non-random sequence that could be used to PCR-amplify the selected products. In addition, the flanking sequences contained restriction sites to allow the subsequent subcloning of the final pool of selected products. Limiting amounts of the library were used to enhance the likelihood of selecting high affinity binding sites. The mutants used in the selection protocol were generated in the GST-AKR_{1-178} construct, which encodes a protein that is missing 111 COOH-terminal amino acids. Single amino acid mutations were introduced at positions in the NH_2-terminal arm and the recognition helix, helix 3. Recombinant full-length AKR was included as a positive control since the results produced with this control could be compared to our previous results in order to gauge the relative effectiveness of the protocol. In addition, GST alone was used as a negative control. Bound sites were immunoprecipitated using a rabbit polyclonal anti-AKR antibody generated by a previous member of the laboratory [359]. In order to verify that the mutations introduced had not modified the epitopes recognized by this antiserum, western blotting was performed using all of the mutant fusion proteins generated. Wild-type GST-AKR_{1-178} was included as a positive control. Figure 3.6 demonstrates that this antibody is capable of recognizing

the wild-type and mutant GST-fusion proteins. The sizes of the bands recognized by the antibody indicated that the GST fusions were expressed properly. The isolated sites were amplified as described in the *Material and Methods* section and another round of selection was initiated. Following 6 rounds of enrichment, EMSAs were performed using the various GST-AKR_{1-178} mutants and their respective selected site pools, in order to confirm the DNA-binding activity of the mutant proteins prior to cloning and sequencing of the selected sites. The results of a number of studies have indicated that some combinations of amino acids in the homeodomain are inconsistent with DNA binding [261; 268; 349]. This may explain the lack of DNA binding activity exhibited by the R54M mutant, since this mutant did not bind either to F' or a number of binding sites recognized by the wild-type protein [320]. Similar results were obtained with the NH_2-terminal double mutant (R4K, G6R) and the homeodomain mutant I50G (Figure 3.7). The data gathered from visually alignment of the binding site pools selected by the remaining GST-AKR_{1-178} mutant proteins are presented in Figure 3.8 in the form of a consensogram. The numbering of the nucleotide positions is given in Figure 3.9 to facilitate the description and subsequent discussion of the results. The numbering is such that the sequence of the hexanucleotide core binding element 5'-TGACAG-3' spans positions 1 to 6 and sequence 5' and 3' of the core are described as -1 to -3 and +1 to +4, respectively. Generally, the sequences of the selected binding sites resembled the sequence of the F' site. However, the mutations caused minor variations to the nucleotides specified by these proteins, particularly within the hexanucleotide core. The consensus sequence determined for the selected binding site pool is presented using the Staden letter code of ambiguity provided in Table 3.4 (Appendix A). This table may also be found at http://www.angis.su.oz.au/Documents/Tables/.

Figure 3.6 **The protein-A purified rabbit polyclonal anti-AKR antibody is able to recognize all of the GSTAKR$_{1\text{-}178}$ mutants generated.** The various mutants were expressed and purified as described in the *Materials and Methods* section prior to being separated on a 12% SDS-polyacrylamide gel (4% stacking), transferred onto Immobilon™-P membrane and probed using a rabbit polyclonal anti-AKR antibody generated by a former member of this laboratory [359]. Molecular weight markers are delineated in kilodaltons (kDa).

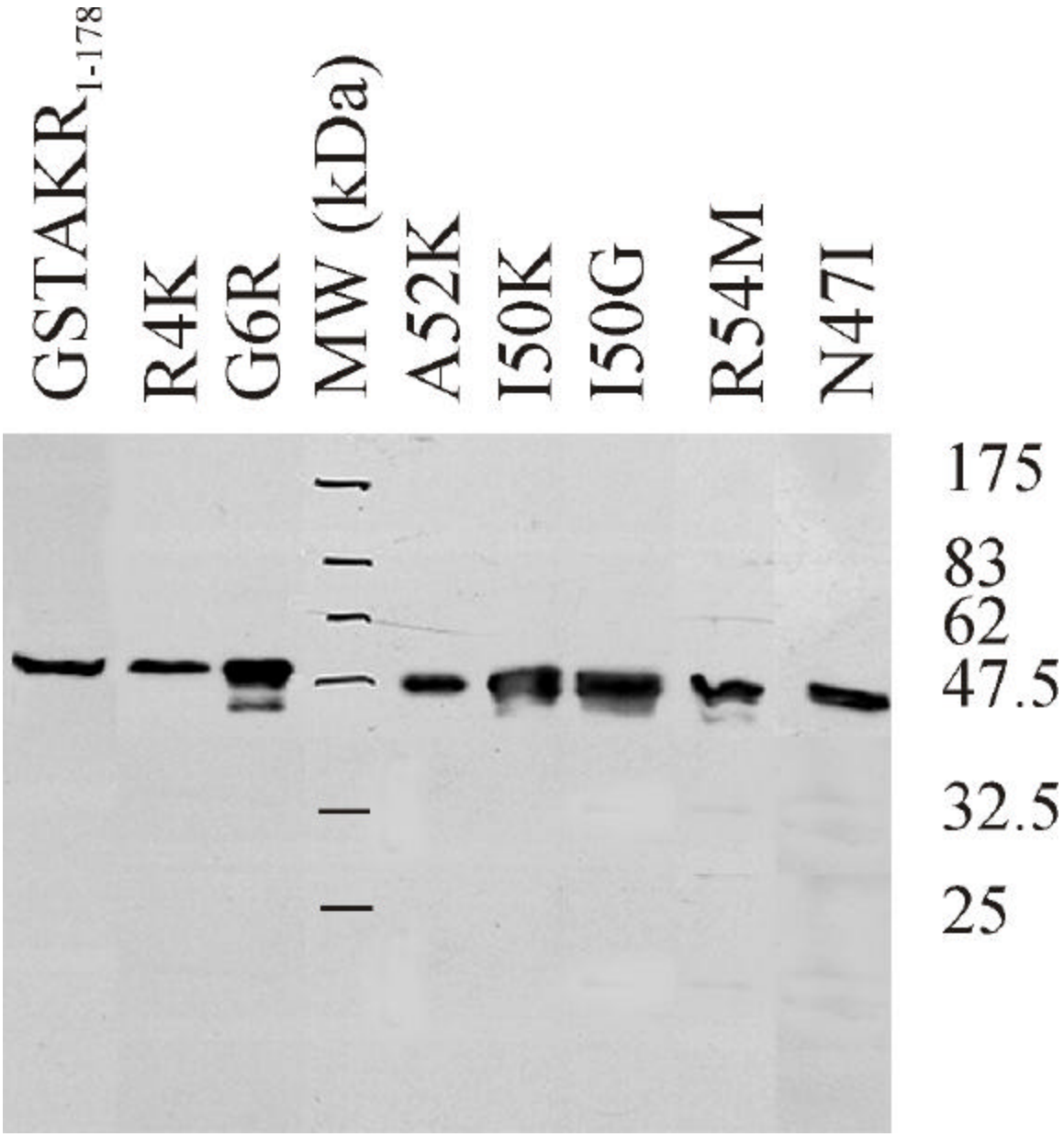

GSTAKR1-178
R4K
G6R
MW (kDa)
A52K
I50K
I50G
R54M
N47I
175
83
62
47.5
32.5
25

Figure 3.7 **The different mutants display varying degrees of binding to their respective selected binding site pools.** EMSA analysis was performed using the respective binding mutants (0.25 ìg) and a portion of the amplified pool of selected sites (1 to 2 ìL). After a 20 minute incubation on ice in the presence of 500 ng of the double-stranded non-specific competitor poly dI· dC, the DNA/protein complexes were then loaded onto a 6% non-denaturing PAGE in 1 x TBE and electrophoresed at 30 mA for 3 hours. The gel was then dried and the bands visualized by autoradiography on X-ray film. Lane 1, no protein (NP); lane 2, GST only; lane 3, I50K; lane 4, I50G; lane 5, GST only; lane 6, A52K; lane 7, NP, lane 8, N47I; lane 9, GST only; lane 10, R4K; lane 11, G6R; lane 12 R4K/G6R; lane 13, R54M; lane 14, wt GST-AKR_{1-178}; lane 15, thrombin-cleaved recombinant full-length AKR.

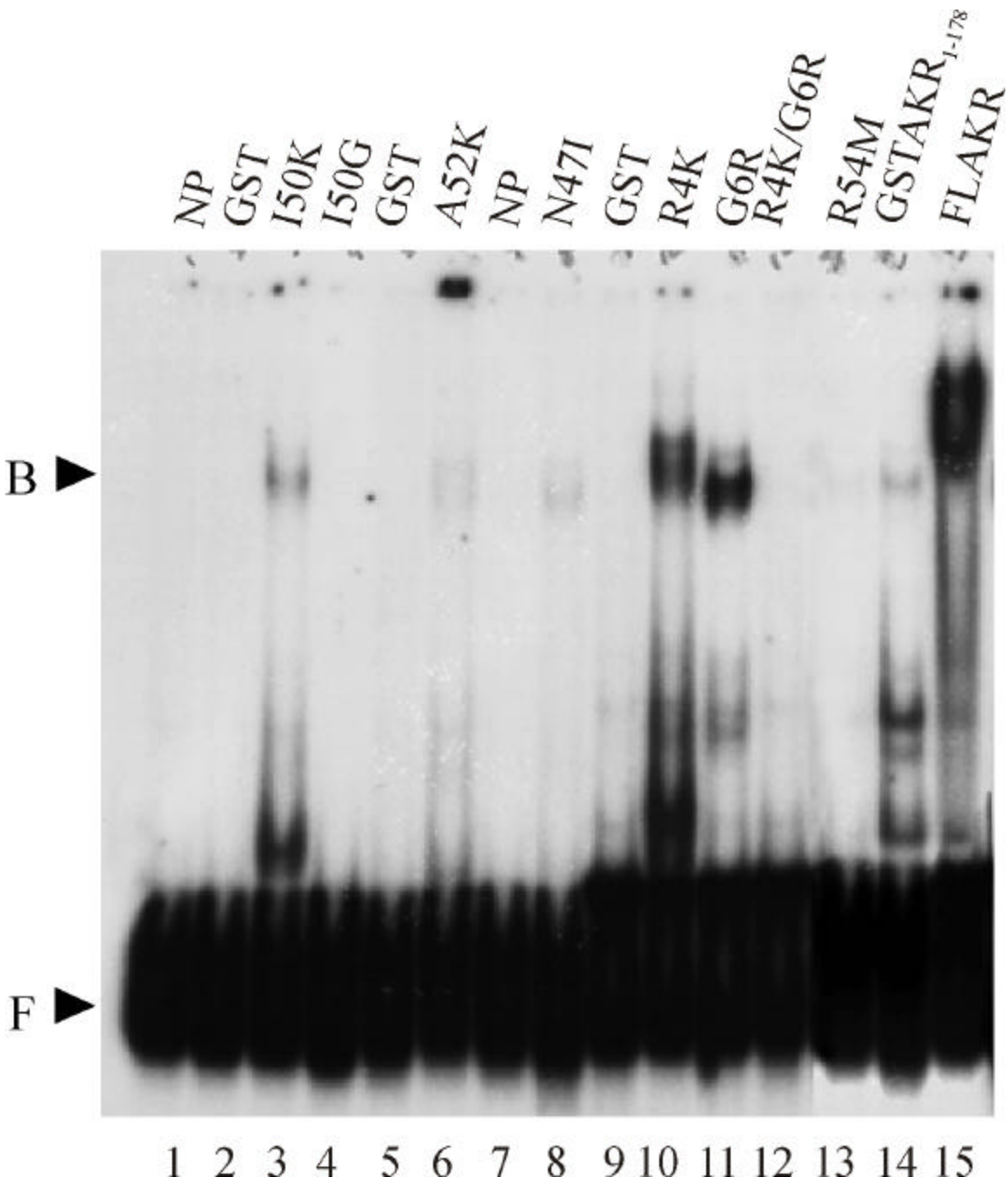
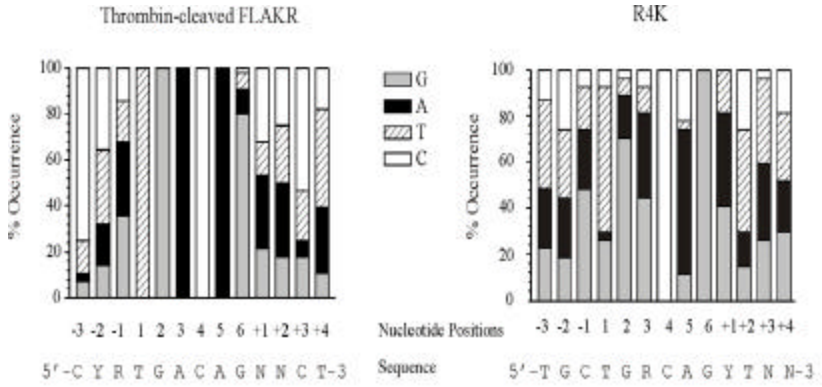
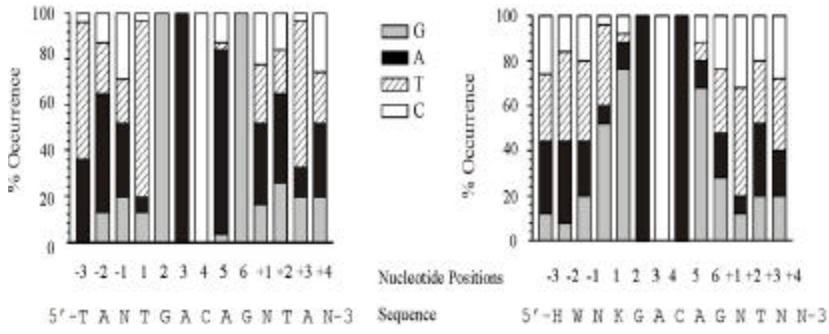

NP
GST
I50K
I50G
GST
A52K
NP
N47I
GST
R4K
G6R
R4K/G6R
R54M
GSTAKR1-178
FLAKR
B
F
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

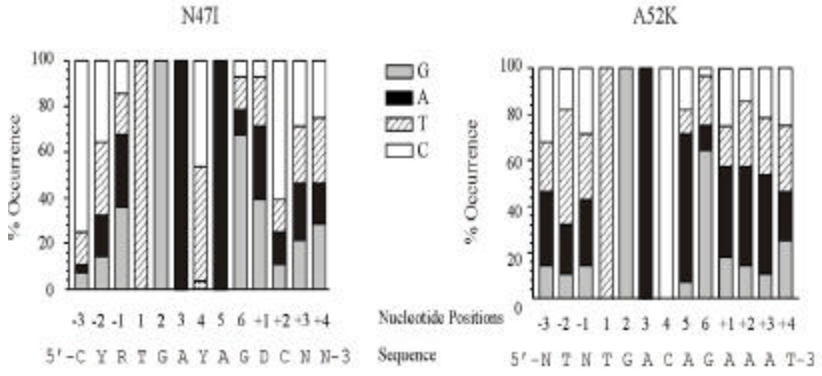
Figure 3.8 **Defining optimal binding sites for the GSTAKR$_{1\text{-}178}$ mutants.** Binding site selection was performed, as described under *Materials and Methods*, using the various mutant proteins. The sequences of the selected sites were then visually aligned and the results are presented in the form of a consensogram that depicts the percent occurrence of a given nucleotide at a particular position for A, recombinant full-length AKR; B, the R4K mutant; C, the G6R mutant; D, the N47I mutant; E, the I50K mutant and F, the A52K mutant. The consensus sequences displayed beneath each consensogram were derived in the following manner. In the flanking region a specific nucleotide is considered "preferred", if it appears with a frequency ≥40% at a particular position and this value is two-fold higher than the next highest level. Within the hexanucleotide core binding element a nucleotide is considered to be preferred, if it appears with a frequency ≥50% and if this value is two-fold greater that the next higher level. Two and three nucleotides are listed in the same position using the Staden code of ambiguity (http://www.angis.su.oz.au.) if they appear equivalently at this position at a frequency ≥30%. An N depicts regions were any of the 4 bases were found to occur at equivalent frequencies.


Thrombin-cleaved FLAKR
% Occurrence
100
80
60
40
20
0
-3 -2 -1 1 2 3 4 5 6 +1 +2 +3 +4
5′-C Y R T G A C A G N N C T-3
G
A
T
C
Nucleotide Positions
Sequence

R4K
% Occurrence
100
80
60
40
20
0
-3 -2 -1 1 2 3 4 5 6 +1 +2 +3 +4
5′-T G C T G R C A G Y T N N-3


G6R
% Occurrence
100
80
60
40
20
0
-3 -2 -1 1 2 3 4 5 6 +1 +2 +3 +4
5′-T A N T G A C A G N T A N-3
G
A
T
C
Nucleotide Positions
Sequence

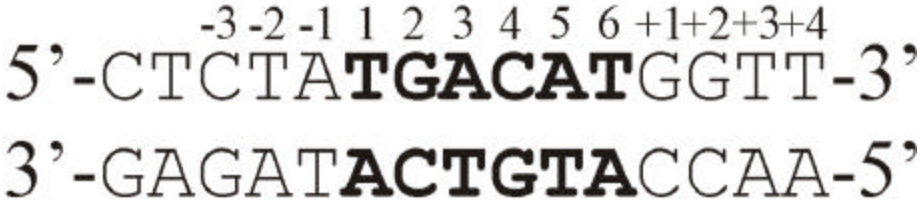
I50K
% Occurrence
100
80
60
40
20
0
-3 -2 -1 1 2 3 4 5 6 +1 +2 +3 +4
5′-H W N K G A C A G N T N N-3


N47I
% Occurrence
100
80
60
40
20
0
-3 -2 -1 1 2 3 4 5 6 +1 +2 +3 +4
5′-C Y R T G A Y A G D C N N-3
G
A
T
C
Nucleotide Positions
Sequence

A52K
% Occurrence
100
80
60
40
20
0
-3 -2 -1 1 2 3 4 5 6 +1 +2 +3 +4
5′-N T N T G A C A G A A A T-3

Figure 3.9 **Sequence and numbering of the F' binding site.** *A*, The sequence of the F' site is shown in along with the numbering scheme used to express and to discuss the results of PCR-assisted binding site selection performed with the various mutants. *B*, Alignment of the binding sites included in the *Results* and *Discussion* sections. The numbering scheme used to describe these binding sites is presented. The hexanucleotide core binding sites are aligned and underlined.

A



B

Wt F'	5'-CTATGACATGGTTG-3'
FM1	5'-CTATTATATGGTTG-3'
FM2	5'-CTATGACATTTTTG-3'
Opt-1	5'-CTATGACAGATCTG-3'
R4K/OMBS1	5'-TCGTGGCAGGTTTG-3'
G6R/OMBS2	5'-TAATGACAGGTATG-3'
N47I/OMBS3	5'-CTATGATAGGCTTG-3'
I50K/OMBS4	5'-CTAGGACAGGTTTG-3'

Thrombin-cleaved AKR, used as a positive control, selected sites with a consensus sequence, 5'-C(C/T)(G/A)TGACAGNNCT-3' (hexanucleotide core binding element is underlined), that matches the sequence of the hexanucleotide core binding element of the Opt-1 site selected previously [320]. However, the sequence of the regions flanking this core varied slightly (Figure 3.8A). Position -3 was occupied by C in 76% of the sequences selected, in contrast to 48% in the Opt-1 sequence. In the present results, the identity of the nucleotides at positions +1 and +2 was more unclear compared to the AT residues that were found to occupy these positions in the previous trial. Lastly, a CT dinucleotide was found at positions +3 and +4 in 54% and 43% of the cloned selected sequences, respectively. This was consistent with our previous results.

The Arg4 and Gly6 residues of the NH_2-terminal arm of the homeodomain were mutated to Lys and Arg, respectively. Lys4 and Arg6 are found predominantly among members of the PBC class of homeodomain proteins (Pbx, Ceh-20, Exd), members of the pknox and Meis classes, as well as members of the BEL group of homeodomain proteins (Figure 3.2) [248; 260; 332; 338]. As described previously, these are all classified as members of the TALE family of homeodomain proteins. As shown in Figure 3.8B, the R4K binding mutant selected sites with a consensus sequence 5'-TCGTGRCAG(A/G)TNN-3' (where R denotes an A or G) that resembled the sequence of the F' (5'-TGACAT-3') and Opt-1 (5'-TGACAG-3') hexanucleotide core binding elements (all are underlined). The R4K mutation effectively relaxes the stringent binding requirements of the wild-type homeodomain within the core binding element, specifically at positions 1, 2, 3 and 5. In the flanking regions, the R4K mutant displayed a slight preference for a TC dinucleotide at positions -3 and -2, respectively while a G was found at position 5 in approximately half of the clones sequenced. This DNA-binding mutant also exhibits a preference for a purine at position +1 and a T at position +2. Of the two NH_2-terminal mutations, the G6R mutant appeared to have the least effect on the binding specificity of the protein. This mutant selected sites with a consensus sequence 5'-TANTGACAGNTAN-3' (hexanucleotide core is

underlined). In the sequence selected by the wild type AKR, 5 of the 6 positions in the core binding site 5'-TGACAG-3' were invariant in all of the selected sequences and a G residue occupies the sixth position in 80% of the cloned sequences [320]. In the binding sites selected by the G6R mutant, the nucleotides at positions 2, 3, 4 and 6 are invariant and identical to those selected by wt AKR (Figure 3.8C). However, the nucleotides at positions 1 and 5 match the Opt-1 core binding site only 77% and 81% of the time, respectively. The sequence of the core binding site selected by the R4K mutant 5'-TGRCAG-3', although similar to that of the Opt-1 core binding site, displays a high degree of variability at positions 1, 2, 3 and 5 (63%, 70%, 37% and 63% matches respectively, to the sequence of the Opt-1 core binding element) while positions 4 and 6 remained invariant to the Opt-1 hexanucleotide core sequence.

Within the recognition helix, Ile47 is most commonly found in members of groups 1-11 and 13 of the HOX family, where it makes hydrophobic contacts with the methyl group of thymine [261; 342]. The members of these groups do not contain a TALE between helices 1 and 2 of the homeodomain. The N47I mutant selected sites with a sequence (the core hexanucleotide binding element underlined) of 5'-CYRTGAYAGDCNN-3' (where Y denotes T or C, R denotes A or G, D denotes A, G or T) (Figure 3.8D). Positions 1, 2, 3 and 5 are invariant and identical to those selected by the wild-type protein. The N47I mutant selected a T or C at position 4 of the core hexanucleotide element with frequencies of 50% and 46%, respectively. In the flanking regions, this mutant selected a purine and pyrimidine at positions -2 and -1, respectively. The identity of the nucleotide at position +1 was more variable since the mutant displayed a preference for G>A>T while a C was present at position +2 in 61% of the selected sequences (denoted as D, Figure 3.8D).

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 core binding site [338; 343]. The I50K mutant preferred to bind to a site with the sequence 5'-HWNKGACAGNTNN-3' (Figure 3.8E,

hexanucleotide core is underlined, W denotes a T or A, K denotes T or G). This matched the sequence of the Opt-1 core binding element at positions 3 to 6. However, the I50K mutant displayed more variability for the bases at positions 1 and 2 (52% and 76% matches, respectively, to the Opt-1 sequence). In the flanking regions, the nucleotide preferences of this mutant were unclear at positions -3, -1, +1, +3 and +4. At the -2 position the I50K mutant displayed a preference for an A or T (36% and 40% occurrence respectively, denoted as W) compared to the A selected by the wild-type protein (50% occurrence). The T specified at position +2 appeared in approximately 50% of the clones sequenced and matched the nucleotide at this position in the Opt-1 core hexanucleotide binding site.

Lys is found at position 52 of the helix 3 region in the members of the PBC class of proteins [362]. This residue is also found predominantly at this position in Ceh-33, 34 and 40 as well as the sine oculis family of homeodomain proteins [363]. Previously, we determined that the A52K binding mutant only retained the ability to bind the Opt-1 binding site (MLT, unpublished results). These results suggested that mutating Ala52 to Lys diminished the ability of the protein to bind DNA or possibly increased the selectivity of the protein. In order to examine the effects of the A52K mutation, binding site selection was also performed with this mutant. The cloned sites selected by this mutant displayed a consensus sequence of 5'-NTNTGACAGAAAT-3' (Figure 3.8F). The nucleotides at positions 1, 2, 3 and 4 are invariant and match the sequence of the core hexanucleotide element of Opt-1. Although the nucleotide preference at position 5 was more variable, an A occupied this position in 64% of the cloned sequences. In the flanking regions, the A52K mutant exhibits a preference for T at position -2 compared to the A selected by the wild-type protein. This mutant exhibits a high degree of variability for its nucleotide preferences at position 1 selecting an A, T or C in 87% of the clones sequenced (29% frequency of occurrence for each of the 3 nucleotides). Interestingly, the consensus sequence displayed a preference for A at positions +1,

+2 and +3 (40%, 43% and 43% frequency of occurrence, respectively) while a T was present at position +4 at the same frequency.

3.5.3 *The mutant proteins differ in the binding affinities they exhibit for the wild-type F' site and their selected sites*

We had previously quantified the affinities of wild-type AKR to binding sites, F' and G, present in the apoVLDLII proximal promoter [320]. Opt-1 was determined by binding site selection. The equilibrium dissociation constants (K_d) determined for the F' and G sites were similar and an order of magnitude lower than the K_d value determined for Opt-1. In order to compare the binding affinities exhibited by the mutant proteins for their respective selected sites we determined K_d values for these interactions as outlined in the *Materials and Methods* section. To generate the appropriate optimal binding sites for these mutants only those nucleotides that were selected for with a frequency >50% were incorporated into the native F' sequence. Scatchard analyses of the data generated from EMSA performed using the mutant proteins and their respective selected sites indicates that the R4K, G6R and I50K mutants display K_ds similar to that obtained for wt AKR binding to the F' and G sites (Table 3.5, Appendix A) [320]. The N47I and A52K mutants displayed a similar affinity for their respective selected sites that was an order of magnitude lower than those obtained for the wt or mutant AKRs. The lower values for these mutants may reflect minor perturbations in their DNA contacts or disruptions to their secondary structure caused by these mutations.

3.5.4 *A revised model for AKR/DNA interactions*

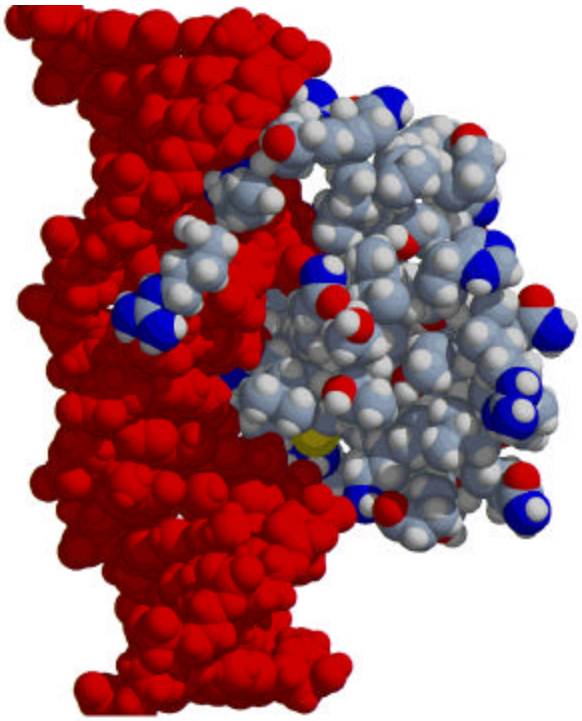
Our previous attempts to model the homeodomain were somewhat limited by the capabilities of the software used which allowed us to include only amino acids 8 to 56 of the homeodomain [320]. Consequently, we were unable to model the NH_2-terminal arm of AKR. Recent advances in software design have allowed us to remodel the homeodomain in its entirety.

To generate the revised model we used a region encompassing amino acids from 2 to 61 of the AKR homeodomain. AKR and MATá2 are sufficiently homologous through their homeodomains (32% identity, 41% similarity) to permit use of the SWISS-3D modelling server (http://expasy.hcuge.ch/swissmod/SWISS-MODEL.html) to generate a preliminary model of an AKR homeodomain/DNA complex. This server threaded the á-carbons of the submitted sequence using the coordinates derived from the structure of the MATá2 complexed to its recognition element [279]. A double-stranded fragment of DNA was generated whose sequence 5'-CACTATGACAGATCT-3' (core binding site is underlined) incorporated the sequence of the Opt-1 binding site shown and numbered in Figure 3.10 [320]. The docking position was derived from the MATá2/DNA complex structure as outlined in the *Materials and Methods* section. The model underwent extensive energy minimization and was evaluated for stereochemical and steric violations using Procheck [360]. The resultant three-dimensional structure is illustrated in Figure 3.10.

As expected, the model confirms that the NH_2-terminal region establishes numerous contacts within a minor groove region that spans A10 to T13 of the binding site core (Figure 3.11). The majority of contacts appear to be made by the Arg residues at positions 3, 4 and 5 of the NH_2-terminal arm. These residues establish numerous contacts with the nucleotides at positions 10-14, 20 and 21 of the DNA. This region includes several bases 3' of the core binding element.

Figure 3.10 **A space filling model of the AKR/Opt-1 complex.** Modelling and energy minimization was performed using the program Sybil 6.4. A, front view and B, rear view of the contacts established by key residues within the AKR recognition helix to Opt-1. Nucleotides are numbered from 1-12 to facilitate discussion. The residues of the AKR homeodomain are presented in colours that correspond to the respective atoms.

A



Major Minor

5′-C ^{1}A ^{2}C ^{3}T ^{4}A ^{5}T ^{6}G ^{7}A ^{8}C ^{9}A ^{10}G ^{11}A ^{12}T ^{13}C ^{14}T ^{15}-3′

3′-G$_{30}$T$_{29}$G$_{28}$A$_{27}$T$_{26}$A$_{25}$C$_{24}$T$_{23}$G$_{22}$T$_{21}$C$_{20}$T$_{19}$A$_{18}$G$_{17}$A$_{16}$-5′

B

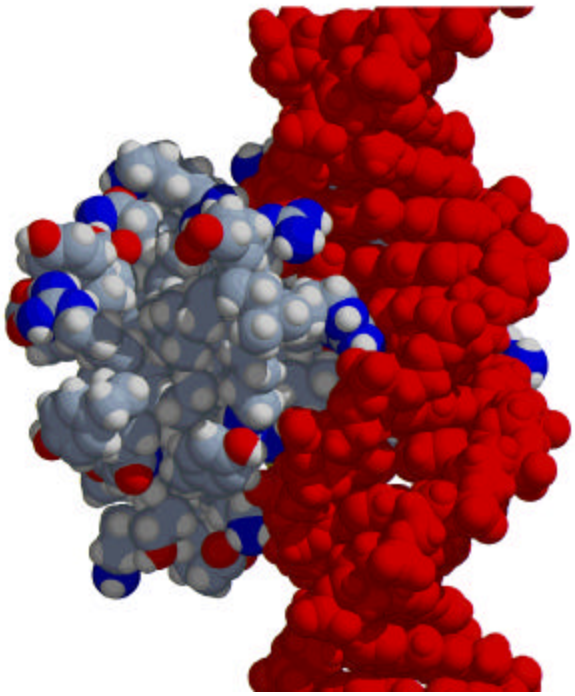
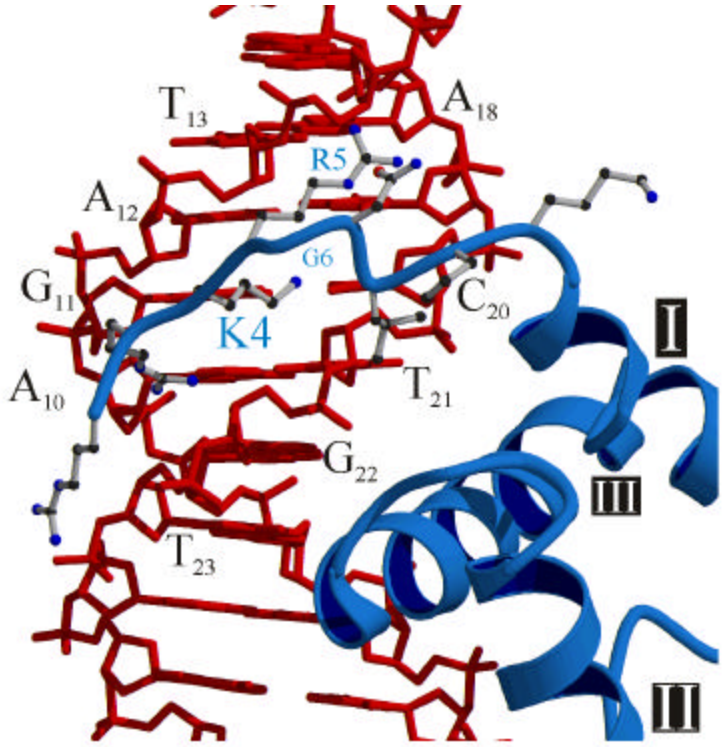


Figure 3.11 **Ribbon diagram of the AKR homeodomain bound to its DNA recognition element.** Amino acids 2 to 61 of AKR were threaded using the data generated from the crystal structure of yeast TALE homeodomain protein MATá2 complexed to its recognition element [279]. The resultant structure was processed and rendered as outlined in the Materials and Methods section. A, ball-and-stick representation of the contact area of the residues in the NH_2-terminal arm in the AKR/Opt-1 complex model. Key residues in AKR are highlighted and labelled. B, rear view of figure 3.11A. The sequence of the double-stranded fragment of DNA is numbered from 1 to 30, along with the positions of the major and minor grooves.

A

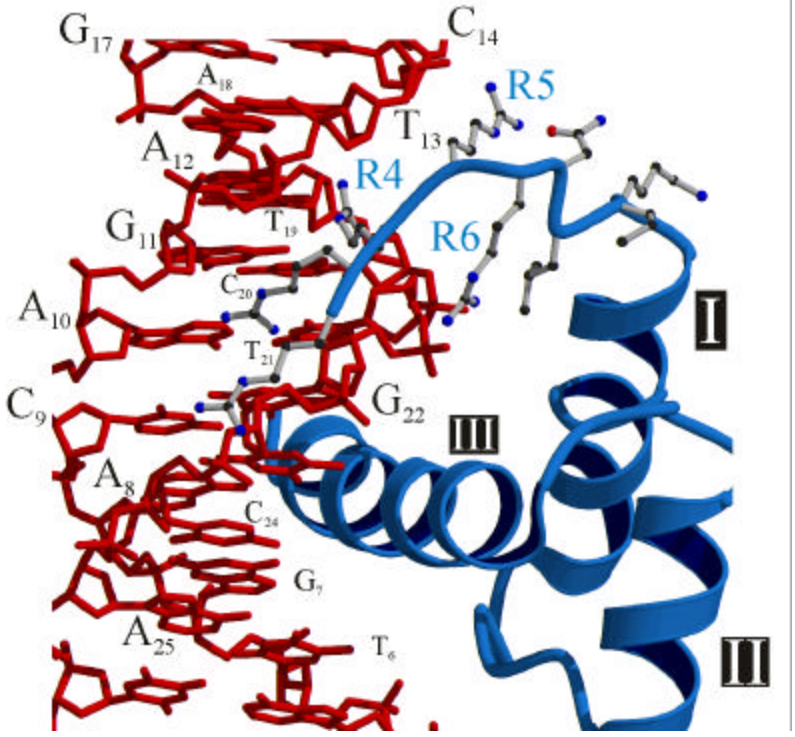


Major Minor

5′-C ^{1}A ^{2}C ^{3}T ^{4}A ^{5}T ^{6}G ^{7}A ^{8}C ^{9}A ^{10}G ^{11}A ^{12}T ^{13}C ^{14}T ^{15}-3′

3′-G$_{30}$T$_{29}$G$_{28}$A$_{27}$T$_{26}$A$_{25}$C$_{24}$T$_{23}$G$_{22}$T$_{21}$C$_{20}$T$_{19}$A$_{18}$G$_{17}$A$_{16}$-5′

B

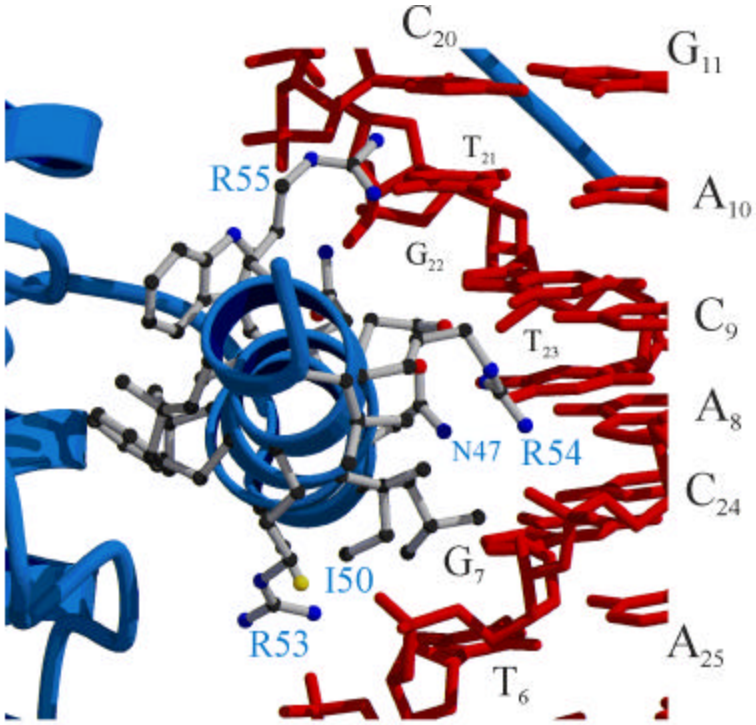


Within the recognition helix the model indicates that residues 47, 50, 54 and 55 establish the majority of base-specific contacts to bases T6 to C9 within the major groove (Figure 3.12). Asn47 is capable of making direct contacts with A, G, and T at positions 8, 22 and 23. Ile50 is in position to make a number of hydrophobic interactions with T6, G7 and possibly A8. Along with Ile50, Arg54 may have the greatest number of major groove contacts to nucleotides A8, T21, G22 or T23. Lastly, Arg55 is positioned to make a number of charged interactions with the nucleotides at the 3'-end of the hexanucleotide core. Specifically, Arg55 may contact a region from C20 to G22.

The model also indicates the presence of a putative hydrophobic pocket in the AKR homeodomain that is comprised of Leu16, Arg17, Leu20, Trp48, Phe49 and Ala52. Figure 3.13 depicts the solvent-accessible surface of AKR, calculated and displayed with GRASP [361]. The double-stranded DNA molecule is displayed as a full atomic model. The possible existence of a hydrophobic pocket is of interest since the interactions of MATá2/MATa1 as well as the interactions of the PBC homeodomain proteins and their various Hox partners occur via the hexapeptide motif of the Hox proteins and a hydrophobic pocket in the PBC partner. In these examples the amino acids that comprise these hydrophobic pockets have been shown to include amino acids in the TALE region that are well conserved in Pbx and Exd [362]. Our model suggests that residues in the TALE region of AKR do not contribute to the structure of the pocket. However, the TALE residues are located in the vicinity of the pocket and as such maintain the potential to contribute in some manner. At present, it is unclear whether amino acid residues in the regions outside of the homeodomain participate in the formation of this pocket since we were unable to model AKR in its entirety because of the lack of an appropriate template with which to perform the modelling.

Figure 3.12 **Ribbon diagram of the recognition helix of the homeodomain of AKR bound to its DNA recognition element.** Amino acids 2 to 61 of AKR were threaded using the data generated from the crystal structure of yeast TALE homeodomain protein MATá2 complexed to its recognition element [279]. The resultant structure was processed and rendered as outlined in the Materials and Methods section. A, ball-and-stick representation of the contact area of the residues in helix 3 in the AKR/Opt-1 complex model. Key residues in AKR are highlighted and labelled. B, rear view of figure 3.12A. The sequence of the double-stranded fragment of DNA is presented numbered from 1 to 30, along with the positions of the major and minor grooves.

A



Major Minor

5′-C ^{1}A ^{2}C ^{3}T ^{4}A ^{5}T ^{6}G ^{7}A ^{8}C ^{9}A ^{10}G ^{11}A ^{12}T ^{13}C ^{14}T ^{15}-3′

3′-G$_{30}$T$_{29}$G$_{28}$A$_{27}$T$_{26}$A$_{25}$C$_{24}$T$_{23}$G$_{22}$T$_{21}$C$_{20}$T$_{19}$A$_{18}$G$_{17}$A$_{16}$-5′

B

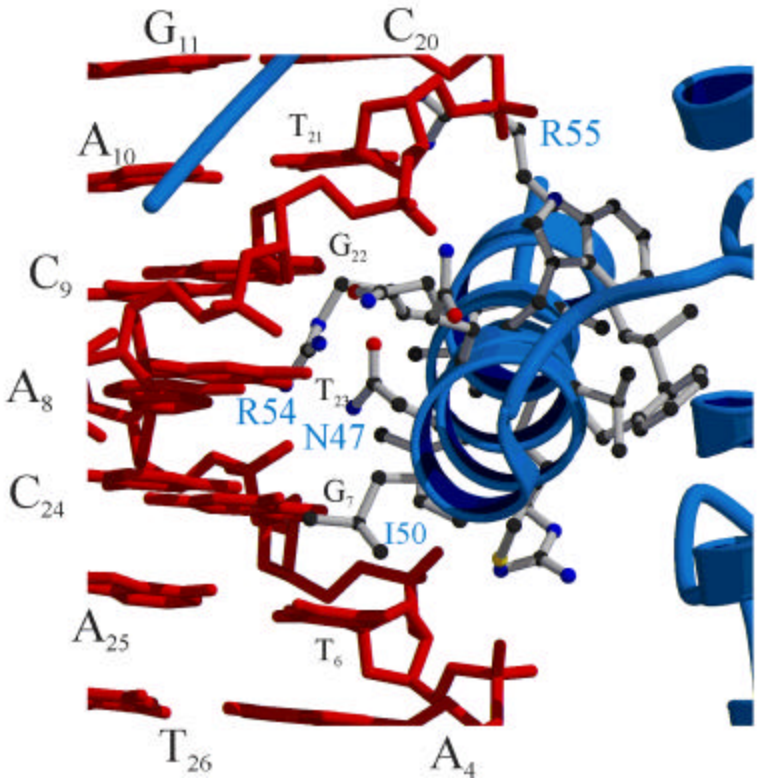
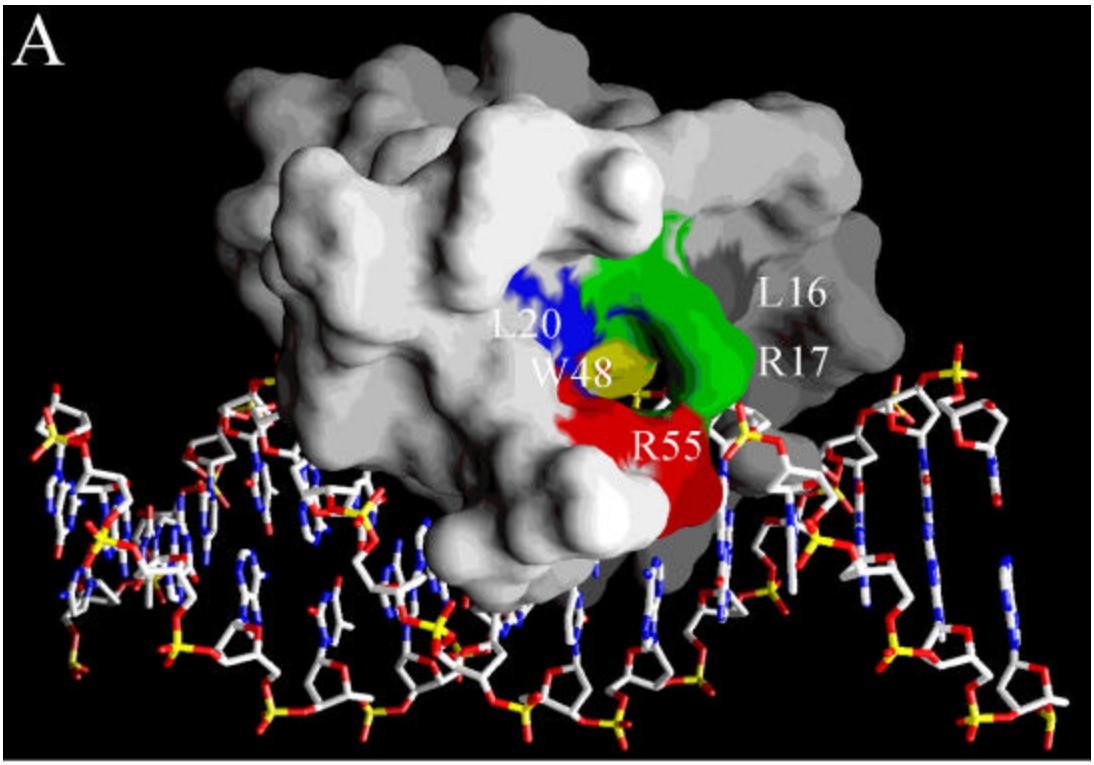


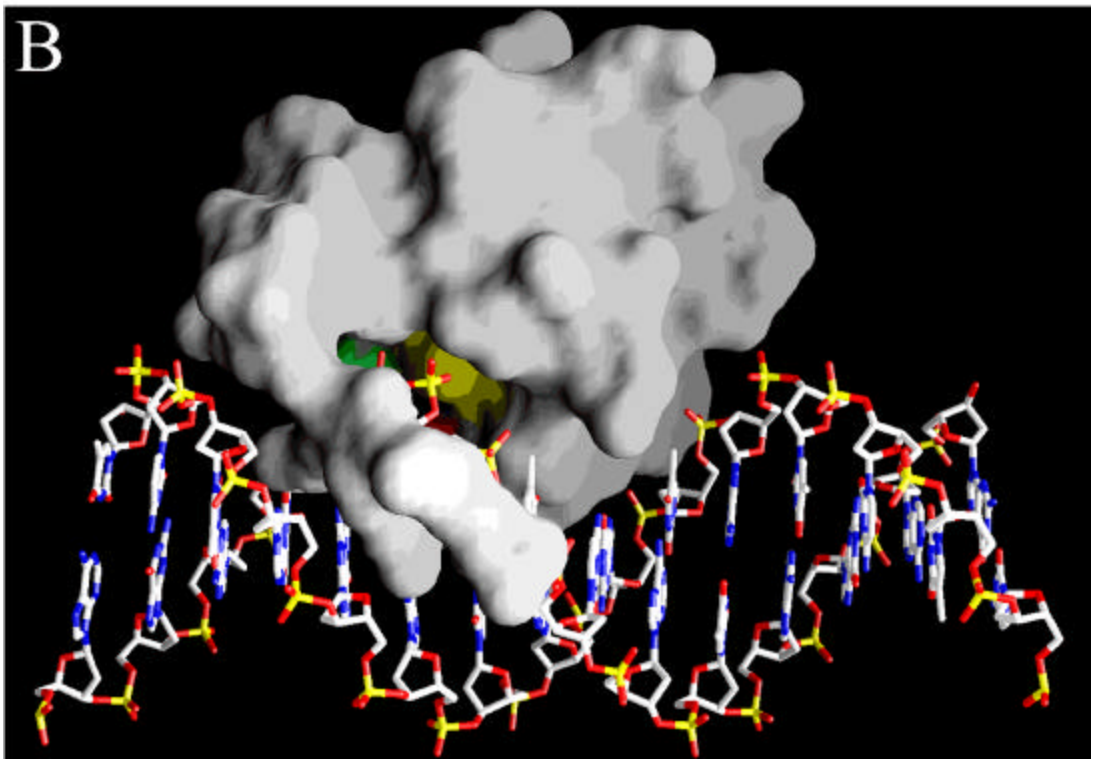
Figure 3.13 **Molecular surface representation of the putative hydrophobic pocket of AKR.** The solvent accessible surface of AKR complexed to DNA is shown. The respective amino acids that comprise the hydrophobic pocket are highlighted in white lettering. A, front and B, rear views of amino acids 2 to 61 of AKR complexed to its recognition element [320]. The sequence of the Opt-1 binding site used to generate the AKR/DNA model. This figure was prepared using the program GRASP [361].



Major Minor

5′-C ^{1}A ^{2}C ^{3}T ^{4}A ^{5}T ^{6}G ^{7}A ^{8}C ^{9}A ^{10}G ^{11}A ^{12}T ^{13}C ^{14}T ^{15}-3′

3′-G$_{30}$T$_{29}$G$_{28}$A$_{27}$T$_{26}$A$_{25}$C$_{24}$T$_{23}$G$_{22}$T$_{21}$C$_{20}$T$_{19}$A$_{18}$G$_{17}$A$_{16}$-5′



3.6 Discussion

3.6.1 *A revised model of AKR/DNA interactions*

Although the overall structures of homeodomain containing proteins vary significantly, they also exhibit similarities. For example, equivalent á-carbons for residues 9-58 of the homeodomains from *Drosophila* engrailed (en) and the yeast MATá2 can be superimposed with an average root mean-square deviation (r.m.s.d.) value of 1.0 Å [364]. Comparatively, superimposing the structures of several independently determined Antp homeodomains results in an r.m.s.d. value of 0.2 Å [365]. With this in mind, the data compiled by the Molecular Modelling Database (MMDB) indicates that if a model and template share approximately 50% identity, the deviation of the final model from the template structure is approximately 3Å. This reinforces the validity of the data generated through modelling studies such as those described here.

Previously, we generated a model of an AKR/DNA complex using coordinates derived from studies of the crystal structure of the MATá2/MATa1 dimer complexed to DNA [279]. However, due to the limitations of the server, only 53 amino acids of the homeodomain could be modelled. These problems were further complicated since the length of the DNA molecule to be used for docking with the protein was limited to 6 nucleotides. Consequently, we were not able to model the NH_2-terminal arm of AKR, nor could we make any predictions about residues that would make contacts to nucleotide positions outside of the hexanucleotide core. In the current study, advances in software design at the server have allowed us to model the homeodomain in its entirety using a 15-base-pair strand of DNA that encompasses the core binding element. In this report, we have refined the structure of our previous model as described in the *Materials and Methods* section. A space-filling model of the AKR homeodomain complexed to the Opt-1 site has been generated (Figure 3.10). This space-filling model illustrates clearly that the homeodomain of AKR binds the major groove of its recognition element in a manner identical to that which is highly conserved among

homeodomain-containing proteins. The NH_2-terminal arm behaves as a finger, completely inserting itself into the minor groove of the DNA where the Arg residues at positions 3 to 5 make the greatest number of contacts to nucleotides A10 to T13/A18 to T21 (Figure 3.11). Additional contacts are also observed to A8C9/G22T23 in the major groove of the hexanucleotide core binding element.

Piper *et al.* (1999) suggest that the preference for a G at position 7 of the Pbx1/HoxB1 binding site may result from the lack of contacts by the Hox NH_2-terminal arm to base pair 7 [277]. The authors suggest that the presence of a leucine at residue 4, whose bulky hydrophobic side chain rules out minor groove contacts, causes the arm to bend away from the DNA. Our model suggests that residue 7 of the homeodomain NH_2-terminal arm is positioned such that substitution of the wild-type Asn residue with an Ile as in Meis1, or Val as in Prep1 may also protrude into the minor groove region. This would cause the NH_2-terminal arm to bend away from the DNA and result in the lower binding affinity observed for these proteins.

Our model predicts that the residues, which determine the binding specificity of the AKR homeodomain, reside at positions 50 and 54 in the helix 3 region of the AKR homeodomain. These residues determine the nucleotide identities at positions 6, 7 and 8 to 10, respectively, (5'-TGACA-3') of the hexanucleotide core binding site. Contacts established by the Asn47 to G22, T23 and C24 reinforce the contacts made by both Ile50 and Arg54. Electrostatic interactions made by Arg55 to the backbone at nucleotides T21 help to reinforce Arg54 contacts to this nucleotide. The additional contacts this residue makes to C20 help determine the identity of the complementary G at position 11 of the hexanucleotide core. This may explain the ability of AKR to recognize either a T or G at this position (Figure 3.12).

Our model also predicts the presence of a putative hydrophobic pocket in the AKR homeodomain consisting of Leu16, Arg17, Leu20, Trp48, Phe49 and Ala52. The crystal structures of Pbx1/HoxB1 and Ubx/Exd dimers indicate that the TALE loop contributes to the formation of a

hydrophobic pocket that is comprised of residues Phe20, Leu23a, Pro24, Tyr25, Arg53, Tyr56 and Lys57. Of these residues, only Tyr25 and Arg53 are conserved in AKR. In AKR, the residues at positions 20, 56 and 57 have been replaced by the hydrophobic amino acid Leu and Pro24 by Tyr (Figure 3.2). It is possible that this hydrophobic pocket mediates interactions between AKR and a putative cofactor protein in a manner analogous to the PBC/Hox interactions. Interactions of AKR with a cofactor are suggested by the high molecular weight of the ubiquitous complex present in nuclear extracts (Figure 3.5C). This complex displays a binding specificity similar to that of *in vitro* translated full-length AKR on both wt F' and a number of mutant binding sites (Figures 3.4 and 3.5). These results suggest that any interaction of AKR with a cofactor does not effectively modify the binding specificity determined for AKR in isolation. In addition, it is possible that variations in the residues comprising the hydrophobic pocket of TALE proteins could dictate the type of cofactor recruited. However, at present there is no evidence that AKR and its murine orthologue TGIF interact with other cofactors through the hydrophobic pocket. It has been shown that TGIF interacts with histone deacetylase 1 (HDAC1) in a DNA-independent manner and that this interaction is dependent upon residues outside the homeodomain in both the NH_2- and COOH-terminal regions of TGIF [366]. In addition, we have shown here that both the NH_2- and COOH-terminal regions of AKR are important in the downregulation of the estrogen-inducible activity of the proximal promoter of apoVLDLII by AKR (Figure 4.12, Chapter 4). Furthermore, the ability of Pbx or Exd to interact with their respective Hox partners resides both in the homeodomain and in a small COOH-terminal region of the protein. Thus, there is some evidence to suggest that residues outside the homeodomain can be important for protein-protein interactions in which these homeodomain proteins participate. Since we were only capable of modelling the homeodomain of AKR it was not possible to determine if additional residues outside of the homeodomain interact with the hydrophobic pocket.

3.6.2 *AKR binds its recognition element in a manner distinct from that of the Ile50-containing TALE proteins*

As mentioned previously, the amino acid sequence and overall structure of the homeodomain is highly conserved even among evolutionarily distant organisms. Therefore, it is not surprising that these proteins bind to their regulatory elements in a very similar manner. Most homeodomains share a preference for a motif containing a 5'-TAAT-3' core and vary only in their preferences for bases immediately 3' to this core. However, we demonstrated previously that AKR binds a sequence, Opt-1, that contains the hexanucleotide core 5'-TGACAG-3'. This sequence matches the sequence of the MPRE, recognized by the Meis1/Prep1 proteins. We also showed that AKR establishes specific contacts with several G residues 3' of the hexanucleotide core of site F' in the apoVLDLII promoter (Figure 3.3 and reference [313]). These observations were corroborated by the results of binding site selection which indicated that, *in vitro*, AKR has the ability to specify nucleotides 3' of the Opt-1 hexanucleotide core element (5'-TGACAGATCT-3') [320]. The results of EMSA using the mutant binding sites FMC1 to FMC6 demonstrate that AKR maintains nucleotide preferences for bases at positions 5' of this core, which expands the total area recognized by the AKR homeodomain to a minimum of 13 nucleotides. Binding site selection studies performed with Meis/Pbx or Prep/Pbx complexes have failed to identify nucleotide preferences outside of the core regions selected by these proteins. The larger area recognized by AKR coupled with its demonstrated nucleotide preferences at specific positions outside of the hexanucleotide core may explain, in part, the higher affinity that this protein displays for the MPRE [320].

3.6.3 *Mutations in the NH_2-terminal arm and helix 3 of AKR result in subtle changes in the nucleotide preferences of the homeodomain*

The differences exhibited in the binding of AKR and the Ile50-containing TALE group of homeodomain proteins to their recognition elements may also reflect subtle changes caused by

amino acid differences in the two proteins, particularly through their NH_2-terminal arms. To determine the effects that the changes to the number of charged residues within the NH_2-terminal arm of AKR had upon the binding specificity of the protein, the Arg and Gly residues at positions 4 and 6 were mutated to Lys and Arg (R4K, G6R), respectively. These residues are found predominantly among members of the PBC class of homeodomain proteins, members of the PKnox and Meis classes, as well as members of the Bel group of homeodomain proteins (Figure 3.2) [248; 260; 332; 338]. With the exception of Bel1, whose recognition element is yet undetermined, all of these homeodomain proteins recognize elements containing a 4base pair core, 5'-TGAC/T-3'. However, the high affinity binding of these proteins requires interactions with a PBC co-factor. Previously, we demonstrated that both the R4K and G6R NH_2-terminal mutants exhibit increased binding to the F' site over that observed with the wild-type protein (Figure 2.4 and 2.6, Chapter 2). We suggested that these mutations augment the charged contacts with the DNA phosphate backbone, enabling the mutated homeodomains to tolerate a variety of nucleotide changes within the hexanucleotide element that abolished binding by the wild-type protein. To determine the effects of these mutations upon the binding site specificity of AKR, we repeated our PCR-based binding site selection using the mutated AKR proteins [320]. The results of these assays, shown in figure 3.8, indicate that both mutations increase the variability of the nucleotides present in the hexanucleotide core, specifically positions 1 to 3 and 5 for the R4K mutant (5-TGACAG-3') and positions 1 and 5 for the G6R mutant (5'-TGACAG-3'). This is consistent with the observed effects of these mutants on the DNA binding activities of the R4K and G6R mutants Figure 2.4, Chapter 2). The results of binding site selection indicate that the increased binding activity observed for the R4K and G6R mutants coincides with a general decrease in the stringency with which they are able to discriminate the nucleotides at positions throughout the hexanucleotide core.

Mutating Gly6 to Arg results in increased DNA binding. Gly is capable of adopting conformations that are sterically forbidden for other amino acids because of their bulkier side chains.

Consequently, the Gly at position 6 of the wild-type AKR homeodomain may confer a high degree of flexibility to the NH_2-terminal arm and enable AKR to reinforce its minor groove contacts resulting in a higher degree of binding specificity. In the G6R mutant, the flexibility is diminished and the basic charge of Arg generates a number of new contacts specifically with T21 and G22, which may explain the increased binding activity observed for this mutant. As with the G6R mutant, the R4K mutation increases DNA binding activity. Since both Lys and Arg are strongly basic and are positively charged at physiological pHs, it is difficult to reconcile the increased binding activity observed for the R4K mutant. Lys would not modify the number of H-bonds or charged interactions. However, the increased flexibility of the Lys residue may enable this mutant to accommodate a greater number of interactions than Arg, which would presumably stabilize interactions that are not favoured by the Arg in the wild-type protein. Structurally, Arg differs by a carbon atom from Lys. The larger Arg may be sterically hindered from maximizing all of its potential DNA interactions. With respect to the lack of observable DNA binding activity for the R4K/G6R double mutant, it is possible that the optimal representation of positively charged residues in the NH_2-terminal arm of AKR was exceeded such that the resulting electrostatic repulsion disrupted the minor groove contacts made by the protein. With this in mind, it is worth noting that there are only a small number of homeodomain proteins that incorporate >5 basic residues in their NH_2-terminal arms.

In contrast to AKR and TGIF, which have relatively high binding affinities for their recognition elements, the other Ile50-containing homeodomain proteins including Meis1 and Prep1 require a cofactor in order to bind tightly to the MPRE. We have demonstrated here that residues at positions 4 and 6 can significantly influence the binding affinity of AKR. Additionally, our model suggests that the residue at position 7 of the NH_2-terminal arm may also have an impact on the binding affinity of the homeodomain for its recognition element. The presence of hydrophobic residues at this position could disrupt the contacts made by the residues in the NH_2-terminal arm, destabilizing

the base specific contacts of the residues in helix 3 and resulting in the lack of detectable binding. The Meis/Prep proteins have Ile or Val residues respectively at position 7 compared to the Asn residue in the AKR homeodomain. Thus, the differences in the binding affinities of AKR and the other Ile50-containing TALE homeodomain proteins for their recognition elements may reflect amino acids present at key positions in the NH_2-terminal arm.

The contributions of the NH_2-terminal arm residues to the DNA binding specificity of AKR are complemented by the direct contacts established by helix 3 residues. As mentioned previously, the binding specificity of the homeodomain is determined by the composition of amino acids at positions 47, 50, 51, 54 and 55. The variability of these residues generates a type of combinatorial code that defines the nucleotide requirements of the homeodomain within its recognition element [260; 268]. Therefore, we mutated these residues to those more commonly found in other homeodomain proteins in order to determine the effects upon the binding site specificity of AKR.

The residue at position 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]. When N47 of AKR was mutated to Ile (N47I mutant) the protein displayed a preference for a T or a C residue at position 4 of the hexanucleotide core whereas the wild-type protein always selected a C at this core position. In addition, the mutant protein specified a T at position -3 and +3 relative to the core sequence in over 50% of the sites selected, as does the wild-type protein (Figure 3.8 compare A and D). This mutant displays preferences for specific nucleotides outside the core compared to the wild-type protein. Although the Asn to Ile mutation enables this mutant protein to recognize a C or T at position 4 the higher K_d determined for this mutant binding to its optimal binding site suggests that the net effect of this mutation is to disrupt the tight binding of the protein. Although the Ile residue is capable of hydrophobic interactions with either C or T at position 4, our model indicates that this molecule is larger and less flexible than the wild-type Asn.

A Gly at position 50 is found predominantly among the members of the PBC class of TALE homeodomain proteins. These proteins recognize a binding element 5'-TGAT-3' that resembles the F' sequence (5'-TGACAT-3'). The I50G mutant did not exhibit any DNA binding activity. We cannot rule out the possibility that the lack of activity demonstrated by this mutant is a result of the incorrect folding of this bacterially produced protein. However, the crystal structures of a Pbx1/Hox or Ubx/Exd dimers on DNA show that Gly does not participate in contacts to its binding site core [277; 287]. Thus, the larger Ile residue in AKR may establish additional contacts and thus be important in the ability of AKR to bind DNA. Our model suggests that Ile50 establishes contacts to the first two bases (underlined) in the hexanucleotide core (5'-TGACAG-3') and thus plays a crucial role in determining the identities of the nucleotides at these two positions (Figure 3.11). A Gly residue at position 50 would essentially abrogate these contacts. Our inability to detect any DNA binding by this mutant supports this hypothesis (Figure 3.7).

Lys at position 50 of the homeodomain is found in the paired and six families of homeodomain proteins where it specifies a GG dinucleotide at the first two positions of the recognition element (5'GGCTTA-3') [338; 343; 363]. The I50K AKR mutant specified a binding site with A and T at positions -2 and +2. Significantly, the I50K mutant AKR exhibited a decreased specificity in the hexanucleotide binding site core. In sites selected by the wild-type protein positions 1 and 2 were invariably T and G, respectively whereas the I50K mutant selected T and G only 36% and 76% of the time, respectively. The modelling of the I50K mutant suggests that Lys50 is capable of establishing a greater number of contacts with the GG dinucleotides at positions 1and 2 whereas the wild-type Ile formed mainly hydrophobic interactions with the T and G residues at these positions in the F' site.

We proposed previously that the residue at position 54 makes a great number of base-specific contacts. There is also considerable variability in the identity of the residue at this position within groups of homeodomains. This provides a potential source of variation in sequence specificity

[338]. The Arg at position 54 was changed to Met, a residue that is found predominantly at this position in members of the Antp family of homeodomain proteins. 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 [365]. Previous results suggested that this mutation abrogated the ability of the protein to bind DNA [320]. Alternatively, the R54M mutation may have modified the binding specificity of the protein such that it no longer recognized F' or any of its derivative binding sites. To clarify the effects of the R54M mutation it was subjected to binding site selection. The results of binding site selection suggest that this mutant may be seriously impaired for binding DNA validating the indications of the model that this residue the major critical determinant of the binding specificity of AKR.

Lastly, the Ala at position 52 was changed into Lys. Similar to position 54, there is considerable variability with respect to the residue found at position 52 across families of homeodomain proteins. However, Lys is found at this position in the members of the PBC class of proteins as well as the *Caenorhabditis elegans* proteins Ceh-33, 34 and 40. The results of EMSA performed with this mutant demonstrated its ability to recognize only the optimal binding site of AKR. Binding site selection was performed with this mutant in order to determine whether these results were the consequence of a change in binding specificity or binding affinity. The mutant selected sequences with a 5'-TGACAG-3' core and specified for T at positions 4 and 15. Curiously, this mutant also selected for sequences with A at positions 5 to 14. The residue at 52 forms part of the putative hydrophobic pocket. Mutation of the wild-type Ala with Lys produces a charged interaction where the positive charges of Lys52 and Arg17 repel each other. This potential disruption of the hydrophobic pocket may have negative repercussions on the ability of the protein to bind DNA that would possibly manifest themselves in the comparatively lower K_d determined for this mutant.

Increasing the number of basic residues found within the NH_2-terminal arm appears to stabilize interactions not permitted by the wild-type protein. This permitted the R4K and G6R mutants to recognize a variety of nucleotides at positions 1 to 5 of the hexanucleotide core, 5'-TGACAG-3'.

This results in a general loss of nucleotide specificity at a given position along the core. In contrast, changes to amino acid residues within helix 3 modified the preferences of the mutant proteins at specific positions within the hexanucleotide core. These changes verified the contacts predicted for these amino acids by our molecular model to specific nucleotides within the core. Alignment of the residues within the NH_2-terminal arms and helix 3 regions of 148 individual homeodomain proteins indicated that the increasing number of basic residues occurring in the NH_2-terminal arm of a particular homeodomain protein coincided with a decrease in the degree of variability of the amino acids that are found at positions 47, 50, 52 and 54 of helix 3. As shown in figure 3.14, as the number of basic residues found in the NH_2-terminal arm decreases, the variability of the amino acids at these positions in the helix 3 also decreases such that homeodomain proteins with 1 or less basic residues show little variance in the identities of the residues found at 47, 50, 52 and 54. Notably, the residue at position 55 was found to be predominantly a Lys or Arg and our model has indicated that along with Arg3 to 5 this residue contacts the 3' G of the 5'-TGACAG-3' core. AKR contains a unique combination of amino acid residues, which make up helix 3, specifically at positions 47, 50, 52 and 54. This uniqueness manifests itself in the high stringency and specificity with which AKR contacts the nucleotides at positions 1 to 5 of the hexanucleotide core. It is possible that increasing the stability of the protein's interactions with DNA inevitably allows these residues to establish contacts normally too weak to be observed.

Figure 3.14 **The number of basic residues in the NH_2-terminal arm appears to be directly proportional to the variability of the amino acid residues that comprise helix 3 of the homeodomain.** The amino acid sequences of over 148 homeodomain proteins were aligned and grouped based upon the number of basic residues in their NH_2-terminal arm. Within a group of proteins the amino acid residues at positions 47, 50, 52, 54 and 55 were tabulated and are plotted as the number of different amino acid residues found at a given position as a function of the number of basic residues in the NH_2-terminal arm. The alignment results of the amino acid sequences from homeodomain proteins with A, five basic amino acid residues (33 individual sequences); B, four basic amino acid residues (36 individual sequences; C, three basic amino acid residues (25 individual sequences); D, two amino acid residues (24 individual sequences); E, one amino acid residue (20 individual sequences).

APPENDIX A List of Tables for Chapter 3

Table 3.1 Oligonucleotides used to introduce point mutations into $GSTAKR_{1\text{-}178}$.

Mutant	Primer No.	Sequence
	RGD981	5'-GGCTGGCAAGCCACGTTTGG-3'
	5319[a]	3'-GCGAGTCGTTCTGACACAGG-5'
	5117[a]	3'-CCGTTTGAGTACCGGATTCTTAAG-5'
R4K	RGD982	5'-GGAAACGTGGCAAACCTACCC-3'
	RGD983	3'-GCCTCCTTTGCAACCGTTGGATGGG-5'
G6R	RGD1020	5'-CGTAGGAACCTACCCAAAGAG-3'
	RGD1019	3'-CCTCTGCATCCTTGGATGGGTTTCTC-5'
R4K/G6R	RGD1020	5'-CGTAGGAACCTACCCAAAGAG-3'
	RGD1019	3'-CCTTTGCATCCTTGGATGGGTTTCTC-5'
N47I	RGD1084	5'-GGTCTGCATCTGGTTTATCAAC-3'
	RGD1085	3'-GATGTCCAGACGTAGACCAAAT-5'
I50G	RGD1016	5'-TTGGCAACGCACGCCGCAGG-3'
	RGD1015	3'-CGTTGACCAAACCGTTGCGTGCGGCGTCC-5'
I50K	RGD1087	5'-CTGGTTTAAGAACGCACGCCGC-3'
	RGD1088	3'-CGTTGACCAAATTCTTGCGTGCG-5'
A52K	RGD1025	5'-TATCAACAAACGCCGCAGGC-3'
	RGD1024	3'-CGTTGACCAAACCGTTGTTTGCGGCGTCCG-5'
R54M	RGD1571	5'-CGCATGAGGCTCTTACCTGATATGC-3'
	RGD1570	3'-GTTGCGTGCGTACTCCGAGAATGG-5'

Table 3.2 Oligonucleotide binding sites used in these studies.

Primer No.	Sequence
RGD783	5'-CCGGGAATTCATATGGATCC(N_{24})AGATCTCGAGAAGCTTCGGC-3'
RGD784	5'-CCGGGAATTCATATGGATCC-3'
RGD785	3'-TCTAGACTCTTCGAAGCCG-5'
7914[a]	5'-GGCCTCTATGATAGGGTTGCCTGAAAATGTAGG-3'
7915[a]	5'-GGCCTTAAGGACAGGGTTGCCTGAAAATGTAGG-3'
7920[a]	5'-GGCCTCTATGACAGAAATGCCTGAAAATGTAGG-3'
7917[a]	5'-GGCCTTAATGACAGATCTGCCTGAAAATGTAGG-3'
7918[a]	5'-GGCCTTAATGACAGGGTTGCCTGAAAATGTAGG-3'
RGD984	5'-GTTTATGAAAGGGGCCTCTATGACAGATCTGCCTGAAAATGTAGG-3'
1511	5'-GGCCTCTATGACATGGTTGCCTGAAAATGTAGG-3'
7919[a,b]	5'-CCTACATTTTCAGGCA-3'

Table 3.3 Oligonucleotide binding sites used in these studies.

Primer No./Label	Sequence
758/FMC1U	5'-CCTCTTAATCATGGTTGCCT -3'
759/FMC1L	3'-AATTAGTACCAACGGACTTT-5'
760/FMC2U	5'-CCTAGATGACATGGTTGCCT-3'
761/FMC2L	3'-CTACTGTACCAACGGACTTT-5'
762/FMC3U	5'-CCTTAATGACATGGTTGCCT-3'
763/FMC3L[c]	3'-TACTGTACCAACGGACTTT-5'
768/FMC4U[c]	5'-CCTTTATGACATGGTTGCCT-3' 3'-TACTGTACCAACGGACTTT-5'
769/FMC5U[c]	5'-CCTCAATGACATGGTTGCCT-3' 3'-TACTGTACCAACGGACTTT-5'
770/FMC6U[c]	5'-CCTGCATGACATGGTTGCCT-3' 3'-TACTGTACCAACGGACTTT-5'

[a] The oligonucleotides presented in Tables 3.1 to 3.3 that were synthesized by the Cortec DNA Service Laboratories Inc. are listed according to their numbered designation. All of the remaining oligonucleotides were synthesized using a Beckman Oligo 1000 DNA synthesizer. [a,b] The 7919 oligonucleotide was end-labelled with ^{32}P and used to synthesize the complementary strand of the primers 984, 1511 and 7914 to 7918. [c] The 763 oligonucleotide was end-labelled with ^{32}P and used to reverse complement the 762, 768, 769 and 770 oligonucleotides. The acronym FMC is used to represent F' Mutant Core, the U and L symbolize the upper and lower strands, respectively.

Table 3.4 The Staden code of nucleotide ambiguity.

Staden Code	Meaning
A	A
C	C
G	G
M	A or C
R	A or G
W	A or T
Y	C or T
K	G or T
H	A or C or T
D	A or G or T

The Staden letter code of ambiguity used to describe the results of the target site selection performed with the various GST-AKR$_{1\text{-}178}$ mutants. This table was adapted from one available at http://www.angis.su.oz.au.

Table 3.5 The equilibrium binding constants (K_d) of the various mutant proteins binding to their selected sites.

Mutant	Selected Site	K_d	r^2
AKR	F'	3.1×10^{-10} M	0.99
AKR	Opt-1	5.2×10^{-11} M	0.95
AKR	G	5.1×10^{-10} M	0.85
R4K	OMBS1	1.2×10^{-10} M	0.98
G6R	OMBS2	2.8×10^{-10} M	0.90
N47I	OMBS3	2.8×10^{-9}M	0.48
I50K	OMBS4	9.4×10^{-10} M	0.91
A52K	OMBS5	3.9×10^{-9} M	0.44

The K_d values where determined from a triplicate trial performed at constant amounts of protein and increasing concentrations of binding sites. The data were subjected to linear regression analyses using Prism 2.1 (GraphPad Software Inc.). The value r^2 quantifies goodness of fit. It is a fraction between 0.0 and 1.0, and has no units. Higher values indicate that the model fits the data better. When r^2=0.0, the best-fit curve fits the data no better than a horizontal line going through the mean of all Y values. When r^2=1.0, all points lie exactly on the curve with no scatter. OMBS (Optimal mutant binding site) designates the oligonucleotide binding sites synthesized using the information gathered from the site selection experiments.

REFERENCES FOR CHAPTER 3

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.

366. Wotton D, Lo RS, Lee S, Massague J: A Smad transcriptional corepressor. *Cell* 1999, **97**:29-39.

367. Elbrecht A, Lazier CB, Protter AA, Williams DL: Independent developmental programs for two estrogen-regulated genes. *Science* 1984, **225**:639-641.

368. Schippers IJ, Kloppenburg M, Van Waardenburg R, AB G: *Cis*-acting elements reinforcing the activity of the estrogen-response element in the very-low-density apolipoprotein II gene promoter. *Eur.J.Biochem.* 1994, **221**:43-51.

369. Bullock WO: *BioTechniques* 1987, **5**:376-378.

370. Sapan CV, Lundblad RL, Price NC: Colorimetric protein assay techniques. *Biotechnol.Appl.Biochem.* 1999, **29**:99-108.

371. Seiler-Tuyns A, Merillat AM, Haefliger DN, Wahli W: The human estrogen receptor can regulate exogenous but not endogenous vitellogenin gene promoters in a *Xenopus* cell line. *Nucleic Acids Res.* 1988, **16**:8291-8305.

372. Holth LT, Sun JM, Coutts AS, Murphy LC, Davie JR: Estrogen receptor diminishes DNA-binding activities of chicken GATA-1 and CACCC-binding proteins. *DNA Cell Biol.* 1997, **16**:1477-1482.

373. Liu Y, Teng CT: Estrogen response module of the mouse lactoferrin gene contains overlapping chicken ovalbumin upstream promoter transcription factor and estrogen receptor-binding elements. *Mol.Endocrinol.* 1992, **6**:355-364.

374. Maxam, AM and Gilbert, W Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 1980, **65**:499-559.

375. Sherf B, Navarro SL, Hannah RR, and Wood KV Dual Luciferase™ Reporter Assay: An advanced Co-Reporter Technology integrating Firefly and *Renilla* Luciferase Assays. *Promega Notes* 1996, **57**:2-12.

376. Grant CE, Kurz EU, Cole SP, Deeley RG: Analysis of the intron-exon organization of the human multidrug- resistance protein gene (MRP) and alternative splicing of its mRNA. *Genomics* 1997, **45**:368-378.

377. Gray SG, Ekstrom TJ: Effects of cell density and trichostatin A on the expression of HDAC1 and p57Kip2 in Hep 3B cells. *Biochem.Biophys.Res.Commun.* 1998, **245**:423-427.

378. Austin RJ, Biggin MD: A Domain of the *even-skipped* Protein Represses Transcription by Preventing TFIID Binding to a Promoter: Repression by Cooperative Blocking. *Mol.Cell.Biol.* 1995, **15**:4683-4693.

379. Tenharmsel A, Austin RJ, Savenelli N, Biggin MD: Cooperative binding at a distance by *even-skipped* protein correlates with repression and suggests a mechanism of silencing. *Mol.Cell.Biol.* 1993, **13**:2742-2752.

380. Razin A: CpG methylation, chromatin structure and gene silencing-a three-way connection. *EMBO J.* 1998, **17**:4905-4908.

381. Yoshida M, Horinouchi S, Beppu T: Trichostatin A and trapoxin: novel chemical probes for the role of histone acetylation in chromatin structure and function. *BioEssays* 1995, **17**:423-430.

382. Cole SP, Bhardwaj G, Gerlach JH, Mackie JE, Grant CE, Almquist KC, Stewart AJ, Kurz EU, Duncan AM, Deeley RG: Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. *Science* 1992, **258**:1650-1654.

383. Chen H, Lin RJ, Xie W, Wilpitz D, Evans RM: Regulation of hormone-induced histone hyperacetylation and gene activation via acetylation of an acetylase. *Cell* 1999, **98**:675-686.

384. Biggin MD, McGinnis W: Regulation of segmentation and segmental identity by Drosophila homeoproteins: the role of DNA binding in functional activity and specificity. *Development* 1997, **124**:4425-4433.

385. Chan SK, Jaffe L, Capovilla M, Botas J, Mann RS: The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. *Cell* 1994, **78**:603-615.

386. Van Dijk MA, Murre C: Extradenticle raises the DNA binding specificity of homeotic selector gene products. *Cell* 1994, **78**:617-624.

387. Peltenburg LTC, Murre C: Specific residues in the PBX homeodomain differentially modulate the DNA-binding activity of Hox and engrailed proteins. *Development* 1997, **124**:1089-1098.

388. Li S, Moy L, Pittman N, Shue G, Aufiero B, Neufeld EJ, Leleiko NS, Walsh MJ: Transcriptional repression of the cystic fibrosis transmembrane conductance regulator gene, mediated by CCAAT displacement protein/cut homolog, is associated with histone deacetylation. *J.Biol.Chem.* 1999, **274**:7803-7815.

389. Choi CY, Kim YH, Kwon HJ, Kim Y: The homeodomain protein NK-3 recruits groucho and a histone deacetylase complex to repress transcription. *J.Biol.Chem.* 1999, **274**:33194-33197.

390. Bach I, Rodriguez-Esteban C, Carriere C, Bhushan A, Krones A, Rose DW, Glass CK, Andersen B, Izpisua BJ, Rosenfeld MG: RLIM inhibits functional activity of LIM homeodomain transcription factors via recruitment of the histone deacetylase complex. *Nat.Genet.* 1999, **22**:394-399.

391. Choi CY, Lee YM, Kim YH, Park T, Jeon BH, Schulz RA, Kim Y: The homeodomain transcription factor NK-4 acts as either a transcriptional activator or repressor and interacts with the p300 coactivator and the groucho corepressor. *J.Biol.Chem.* 1999, **274**:31543-31552.

392. Crane-Robinson C, Wolffe AP: Immunological analysis of chromatin: FIS and CHIPS. *Trends.Genet.* 1998, **14**:477-480.

VITA