Specific DNA binding of the two chicken *Deformed* family homeodomain proteins, *Chox-1.4* and *Chox-a*

Hiroshi Sasaki+, Emiko Yokoyama and Atsushi Kuroiwa*,+
Department of Molecular Neurobiology, Tokyo Metropolitan Institute for Neurosciences, 2-6 Musashidai, Fuchu-city Tokyo 183, Japan

Received December 27, 1989; Revised and Accepted March 1, 1990 EMBL accession nos X17245, X17246

ABSTRACT

The cDNA clones encoding two chicken *Deformed* (*Dfd*) family homeobox containing genes *Chox-1.4* and *Chox-a* were isolated. Comparison of their amino acid sequences with another chicken *Dfd* family homeodomain protein and with those of mouse homologues revealed that strong homologies are located in the amino terminal regions and around the homeodomains. Although homologies in other regions were relatively low, some short conserved sequences were also identified. *E. coli*-made full length proteins were purified and used for the production of specific antibodies and for DNA binding studies. The binding profiles of these proteins to the 5'-leader and 5'-upstream sequences of *Chox-1.4* and *Chox-a* coding regions were analyzed by immunoprecipitation and DNase I footprint assays. These two *Chox* proteins bound to the same sites in the 5'-flanking sequences of their coding regions with various affinities and their binding affinities to each site were nearly the same. The consensus sequences of the high and low affinity binding sites were TAATGA(C/G) and CTAATTTT, respectively. A clustered binding site was identified in the 5'-upstream of the *Chox-a* gene, suggesting that this clustered binding site works as a cis-regulatory element for auto- and/or cross-regulation of *Chox-a* gene expression.

INTRODUCTION

The homeodomain was originally discovered in a group of *Drosophila* homeotic genes and a segmentation gene (1,2), and later in many developmentally regulated genes including spatial coordinate, segmentation and homeotic genes. All of these genes are known to play important roles in the normal morphogenesis of the *Drosophila* embryo (reviewed in 3,4,5).

Genetic studies in *Drosophila* have suggested that some of the homeobox containing genes encode trans-regulatory proteins (e.g. 6,7). The yeast mating-type proteins, MAT a1 and MATα2, are sequence specific DNA binding proteins that determine the cell differentiation of yeast. The homeodomain containing proteins were thought to be sequence specific DNA binding proteins since the helix-turn-helix motif that was found in the DNA binding domain of yeast mating type proteins was also found in the homeodomain. Recent studies have demonstrated that the bacterially produced homeodomain containing proteins, both in the case of fusion and full length products, recognize and bind to specific sequences of DNA *in vitro* (8−12). Even a protein composed solely of the homeodomain (13) and, moreover, a chemically synthesized *Antp* homeodomain peptide (14) can bind to specific DNA sequences. In addition, some *Drosophila* homeobox proteins were found by transient expression assay to behave as sequence specific transcriptional activators and/or repressors in cell culture (15−19). Further, some transcription factors, *e.g.* Oct-1, Oct-2 and Pit-1, which have an ability to recognize and bind to specific DNA sequences, were found to contain homeodomains (reviewed in 20). On the basis of this evidence, it has been proposed that the protein products of *Drosophila* homeobox genes act as transcriptional regulators (reviewed in 21).

By means of DNA sequence cross-homologies with various *Drosophila* homeoboxes, homeobox containing genes were also isolated in vertebrates (reviewed in 22). Their clustered gene organizations and developmental expression patterns resemble those of *Drosophila* homeobox genes (23,24). Recently, several experiments have suggested that the vertebrate homeobox genes are also involved in morphogenesis (25,27).

We chose the chicken embryo as a model system for studying vertebrate homeobox genes, since much experimented work has been done on embryology of the chicken. We cloned homeobox genes from chicken genomic and cDNA libraries, and studied their expression. Chicken *ho*meobox genes (*Chox*) are found in clusters similar to those found in mouse (*Hox*). Furthermore, the genes are expressed in temporally and spatially restricted patterns as are the homeobox genes of other animals (Kuroiwa and Yokoyama, in preparation). Therefore, we thought that the *Chox* genes might also have an important function in chicken development as transcriptional regulators.

We focused on two chicken *Deformed* (*Dfd*) family homeobox

* To whom correspondence should be addressed

+ Present address: Department of Cell Biology, Research Institute for Tuberculosis and Cancer, Tohoku University, 4-1 Seiryomachi, Aobaku, Sendai 980, Japan

genes from two different homeobox gene clusters. One is *Chox-1.4* (chicken homologue of *Hox-1.4*) and the other is *Chox-a* (chicken homologue of *Hox-5.1*). We isolated cDNA clones corresponding to both genes and determined their nucleic acid sequences. Their predicted amino acid sequences were compared with another chicken *Dfd* family homeodomain protein, *Chox-Z* protein, and with mouse *Dfd* family homeodomain proteins. The well conserved sequences were localized at amino terminal regions and around homeodomains. Although the sequence homologies in other regions were relatively low, some short conserved sequences were identified. In order to analyze the function of these homeodomain proteins, we prepared full length protein products of these homeobox genes using an *Escherichia coli* expression system and purified them to homogeneity, then compared their DNA binding activities *in vitro*. Both proteins bound to the 5'-leader and 5'-upstream sequences of *Chox-1.4* and *Chox-a* genes recognizing similar sequence motifs with various binding affinities.

MATERIALS AND METHODS

cDNA cloning and sequencing

A cDNA library from four day embryonic poly(A)$^+$ RNA was constructed using a standard method with minor modifications (28). Double stranded cDNA was ligated with an equimolar mixture of EcoRI-SmaI adaptors and phosphorylated SmaI linkers instead of phosphorylated EcoRI linker. The cDNA was directly ligated into the EcoRI site of lambda gt10. This cDNA library was screened with cloned chicken genomic DNA fragments containing *Chox-1.4* and *Chox-a* homeobox as probes. Both strands of the two longest cDNA clones (c302 for *Chox-1.4* and c71 for *Chox-a*) were sequenced by dideoxy-sequencing methods (29).

RNase protection assay

RNase protection assays were performed following a standard method (30). The SmaI-TaqI fragment of *Chox-1.4* genomic DNA and the BglII-BamHI fragment of *Chox-a* genomic DNA were subcloned into Bluescript (Stratagene). Antisense RNA probes were prepared following Stratagene's protocol. About a 1×10^6 cpm probe was hybridized with 1.5μg of chicken 4 day embryo poly (A)$^+$ RNA or yeast tRNA at 45°C overnight. Protected RNAs were analyzed on field gradient 6% acrylamide sequencing gels containing 8.3M urea with both DNA sequence markers and RNA markers.

Plasmids and protein production

A high copy number T7 expression vector, pUCT7, was constructed by transferring the BglII-AatII fragment of pET-3c (isolation number pAR3038, 31) which carries the T7 promoter and terminator, from the PvuII to the AatII site of a pUC18 vector. Each cDNA fragment was introduced into the vector so as to produce a full length protein from the first methionine without fusing them to a phage protein sequence. Protein production was done as described by Studier and Moffatt (32). The produced protein was purified from the supernatant of *E. coli* sonication lysate by 40% ammonium sulfate precipitation and S-Sepharose Fast Flow (Pharmacia) column chromatography. Column work was performed in buffer A (50mM HEPES-Na pH7.6, 1mM EDTA, 10mM 2-mercaptoethanol, 5% glycerol) and eluted with 0 to 1.2M linear gradient of NaCl. The purified protein was stored at −20°C or −80°C in buffer A containing 50% glycerol.

Antibody production

Balb/c mice were immunized by injecting 50−80 μg of purified *Chox-1.4* or *Chox-a* protein in MPL+TDM emulsion (RIBI ImmunoChem Research, Inc). First booster immunizations were given three weeks after primary immunizations, and second booster immunizations one week later. The antisera were purified by DEAE-cellulose (Whatmann) column chromatography and protein A-Sepharose (Pharmacia) column chromatography. In order to remove the cross-reactivity between anti-*Chox-1.4* antibody and *Chox-a* and *E. coli* protein, anti-*Chox-1.4* protein IgG was loaded onto the *Chox-a* protein producing *E. coli* lysate column which had been made by binding *E. coli* lysate to the CNBr-Sepharose 4B (Pharmacia), and the pass-through fraction was used as the specific antibody. A similar manipulation was done to prepare anti-*Chox-a* protein antibody.

Immunoblot

Nuclei were prepared from 4 day old chicken embryos following the method of Gorski et al. (33). Electrophoresis was performed following the description by Laemmli (34). Fractionated protein was electrophoretically blotted onto nitrocellulose membrane. The immunodetection was made using dried milk as blocking agent, alkaline phosphatase conjugated anti-mouse IgG (Promega) as secondary antibody, and BCIP-NBT (BRL) as substrate, by standard procedures.

Immunoprecipitation of DNA fragments

Approximately one ng of the end-labeled DNA fragment mixture was incubated with 10ng of purified *Chox* protein in 20μl of binding buffer (10mM Tris-HCl pH7.8, 100mM KCl, 0.05% Triton X-100, 1mM EDTA, 1mM dithiothreitol, 5% glycerol) for 20min on ice. To this mixture, 4μl of anti-*Chox* protein IgG was added and the mixture was further incubated for 20 min on ice. To this mixture, 10 μl of IgGsorb (fixed *Staphylococcus aureus*; The Enzyme Center) solution was added, and the DNA-protein-IgG complexes were precipitated by centrifugation after incubation for another 10min on ice. The resulting pellets were washed with binding buffer and then with TBS-T (20mM Tris-HCl pH7.6, 137mM NaCl, 0.1% Tween 20) added together with various concentrations of NaCl. Washed pellets were resuspended in 100μl of TE containing 400μg/ml yeast RNA, 1% SDS, 10mM EDTA, and were extracted with a 1:1 phenol chloroform mixture. The DNA was ethanol precipitated from the aqueous phase and electrophoresed on 5% polyacrylamide gels.

DNase I footprint assay

5'end-labeled DNA was incubated with various amounts of purified protein in 20μl of binding buffer for 20min on ice. To this mixture, 20μl of ice-cold 10mM $CaCl_2$, 10mM $MgCl_2$ and 2−5μl of 5ng/μl DNase I (BRL) solution in 0.15M-NaCl was added and immediately incubated for one minute. The reaction was terminated by addition of 50μl of a stop solution (20mM EDTA, 1%SDS, 100μg/ml yeast RNA) and 100μl of 1:1 phenol-chloroform mixture. The DNA was ethanol precipitated from the aqueous phase and electrophoresed on 6% field-gradient sequencing gels. The same DNA fragment treated with a G+A chemical sequencing reaction (35) was run on the same gel as a marker.

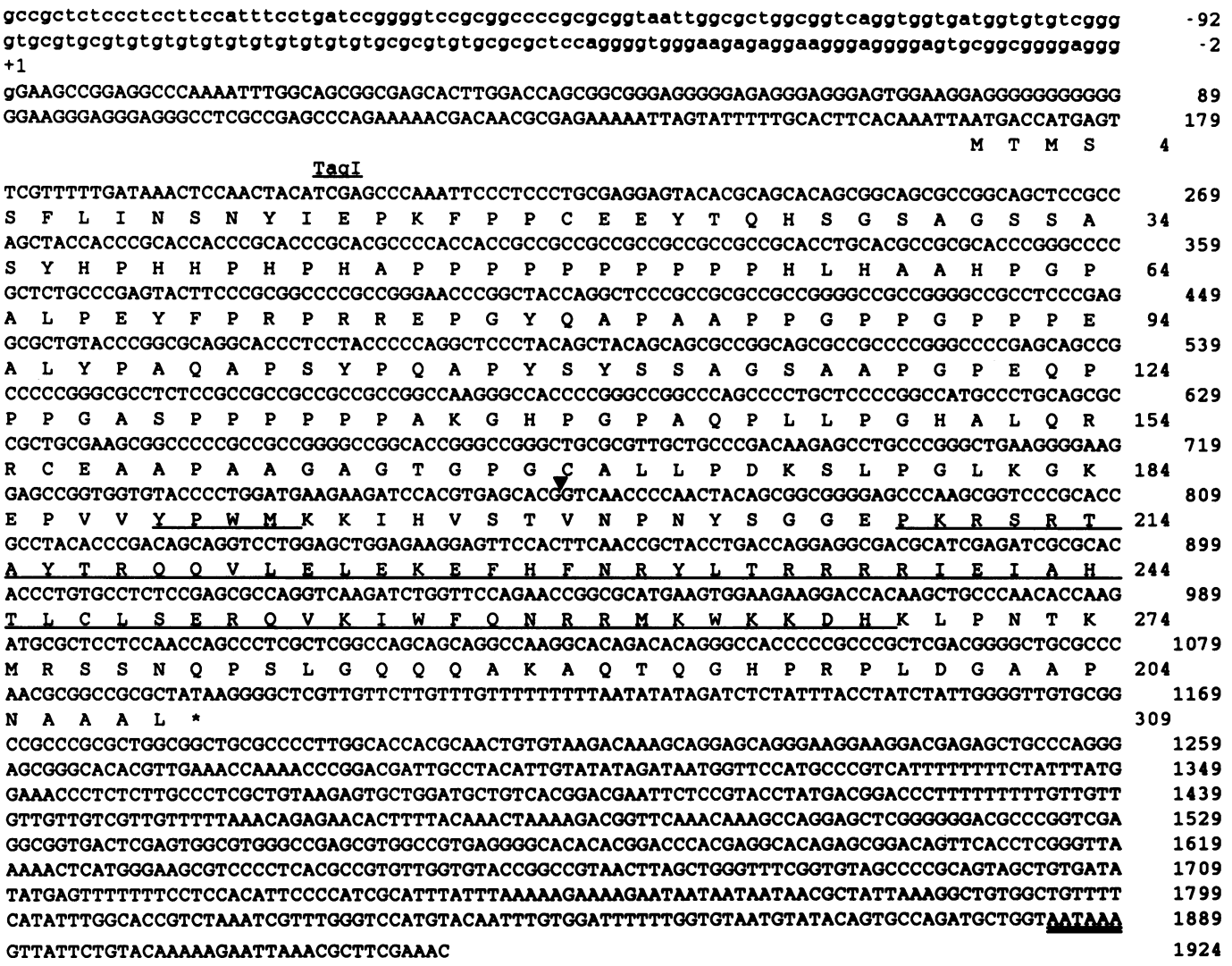

Figure 1. Nucleotide sequence of *Chox-1.4* cDNA and 5′ flanking genomic DNA and deduced amino acid sequence. The major transcriptional initiation site is indicated by +1 on the nucleotide sequence. The deduced amino acid sequence is shown under the nucleotide sequence by single letter codes. Homeobox and the YPWM box are underlined. The splicing junction point is indicated by the filled triangle. A putative poly (A) signal in the 3′-non-coding sequence is double underlined. The longest cDNA covers from +103 to +1924. The rest of this sequence is derived from a genomic clone.

RESULTS

Structure of Chox-1.4 and Chox-a cDNAs and predicted proteins

The *Chox-1* and the *Chox-5* clusters, which are the chicken homologues of the mouse *Hox-1* and the *Hox-5* clusters, contain at least ten and nine independent homeobox genes, respectively (Kuroiwa and Yokoyama, in preparation; Kuroiwa, unpublished data). We chose to use two genes, *Chox-1.4* (*Chox-1* cluster) and *Chox-a* (*Chox-5* cluster), for our binding studies and isolated cDNA clones from a chicken embryonic cDNA library.

The nucleotide sequences of the *Chox-1.4* and *Chox-a* cDNAs fused to genomic DNAs at their 5′ end and their deduced amino acid sequences are shown in Figures 1 and 2, respectively. In *Chox-1.4*, neither the TATA box-like sequence nor the CAAT box-like sequence is found in the upstream region of transcriptional initiation sites. In *Chox-a*, a TATA box and a putative SP1 binding site (CCCGCC) are found. An analysis of the corresponding genomic DNA indicates that these genes have two exons separated by an approximately 0.5kb and 0.34kb introns, respectively (data not shown).

The predicted *Chox-1.4* and *Chox-a* proteins are composed of 309 and 236 amino acid residues and have respective molecular weights of 33,222 and 26,970 daltons. The *Chox-1.4* and *Chox-a* proteins have *Dfd* type homeodomains and amino acid sequence homologies of their homeodomains to *Dfd* homeodomain are 88% and 90%, respectively. These *Chox* proteins also contain the conserved tetrapeptide sequences (YPWM box) in the upstream of the homeodomains that is a common feature of a number of other homeodomain proteins (36). The sequence homologies with mouse *Dfd* family homeodomain proteins will be elaborated in the DISCUSSION. The *Chox-1.4* protein contains an unusual abundance of proline residues (20%), most of which (74% of the total) are clustered in a proline rich domain. This proline rich domain is located amino acid position 38−148 where proline content is 41%.

Amino acid sequence homologies between the *Chox-1.4* and *Chox-a* proteins are located in two regions: one is the amino

```
ctcgagttttattcgtatcggagcccggcggccatcccgccccggcggatggggcggaggggtgggggggtagggagagggaggctctgct    -156
cagacaaaggctccccgcacctgggggcagggaggcggccgcacgcggggaatggccccaaaactccgggatcggccccggaggcgaagc     -66
                                                                    +1
ccgccgcggaggggtcgctcagctccgcgtatattatttagttaaattcgtggaaattctgaggcGCACACAAATCCGGTGATCGGCTCC      25
CCGCCTCCCCATTGGCCGGGCCGGTCACATGCACGCCTAACTTTATTCAGTTGACAGCAAGTAGGAGGGCTCTATGGAAGGAGAAAAAAA     115
                                                                                  BamHI
GACAACACGAGAAAAATTAGTATTTTCTACCTTCCGAAATTAATGGCCATGAGTTCGTATATGGTGAACTCTAAGTATGTGGATCCCAAA     205
                                                  M  A  M  S  S  Y  M  V  N  S  K  Y  V  D  P  K   16
TTTCCTCCTTGCGAGGAATATTTGCAGAGCAGCTACCTAGGCGAGCAGGGGGCCGAGTATTACGGGGCCTCGCAGGGCTCCGATTTCCAG     295
F  P  P  C  E  E  Y  L  Q  S  S  Y  L  G  E  Q  G  A  E  Y  Y  G  A  S  Q  G  S  D  F  Q    46
CACCAGGGGCTCTACCCACGGTCAAACTACAGCGAGCAGCCCTTTTGGCTCGCCGGTGCCCAGGGCTCCGCCGTGCAGCCGCGGGGTCAC     385
H  Q  G  L  Y  P  R  S  N  Y  S  E  Q  P  F  W  L  A  G  A  Q  G  S  A  V  Q  P  R  G  H    76
GGACAGGAGCAGTCCGCCCCTCCGAGCCACTTCCCCGGCCAAGCCGAGCATTGTCCTCCGCCTCCATGTCCAATCCCGAGCTGCGGCCAG     475
G  Q  E  Q  S  A  P  P  S  H  F  P  G  Q  A  E  H  C  P  P  P  P  C  P  I  P  S  C  G  Q    106
CAGCCAGCCCTCAAAGCCCCCCACGGGTCAGCAGTTAAGCAGCCAGCAGTAGTTTATCCCTGGATGAAGAAAGTCCATGTTAACTCGGTG     565
Q  P  A  L  K  A  P  H  G  S  A  V  K  Q  P  A  V  V  Y  P  W  M  K  K  V  H  V  N  S  V    136
AACCCCAATTACAGCGGAGGGGAACCTAAGCGCTCCCGAACAGCTTACACCAGGCAGCAAGTCCTAGAACTGGAAAAAGAATTTCATTTT     655
N  P  N  Y  S  G  G  E  P  K  R  S  R  T  A  Y  T  R  Q  Q  V  L  E  L  E  K  E  F  H  F    166
AACAGGTACCTGACGCGGCGCCGCCGCATCGAGATAGCGCACACTTTGTGTCTCTCTGAGCGCCAGATCAAGATCTGGTTTCAGAACAGA     745
N  R  Y  L  T  R  R  R  R  I  E  I  A  H  T  L  C  L  S  E  R  Q  I  K  I  W  F  Q  N  R    196
AGAATGAAGTGGAAAAAAGACCACAAACTGCCCAACACTAAGGGCAGGTCTTCCTCCTCCGCTTCTAACCCCCACTTACAGACTGTTCCC     835
R  M  K  W  K  K  D  H  K  L  P  N  T  K  G  R  S  S  S  S  A  S  N  P  H  L  Q  T  V  P    226
AAGGACCATCAGACTGACTTGACAACTTTATAGAAGAGGTGAAATTTTCCATTTCATCCTTCGATTCTGACCAGTACTGTACATAAGTGA     925
K  D  H  Q  T  D  L  T  T  L  *                                                              236
CACTATCCAAGCAGAACCTGCGTGTAGCAGCCAAGGACTCGAGAAACTTCCATCTCTCCGTAAGGTACTGTGGTACACAGGAGACTGAAT    1015
GACGCTACTCCGGACTTTATTTTATTTGCCTCTCGCTAGCACAGAAACAAAGGGGGGAAAAAAAAGCATCATACAGGCAGTAGGAATTGG    1105
GAAAGTCTTAAACGAGACCTTCTGTCCATAAAAAGTAATAAATCACTGAAAATGAAACTGTCCTGTGTTTCARKTTGGGATTTGAAAGTG    1195
AAGGTTTGATGCTTTATTCTGGGATTTATTATTATTATTTTGCTTTATTGTCTTTTTATCCCGCACAGCGAGTGTGATGCTTTTGGGGCA    1285
AGTCGAGGACAGGAAATTTATGGCACAGTCCTTTATTTGAACACAAAGCCTCGCGTTTGTGAACTTTTTTGTGCTGCTTTACCTTTATTG    1375
GTGTAGTGAAAATACTCACTCATTTCCATAACGTCTCAGAAGATGTGCATAGGAATTAAGTGAAAACGGGTTATTTTGTGTCCTGATGGT    1465
GTTAGCGTATTTATTTATTTGTTATGTAGAAGAGTG                                                          1501
```

Figure 2. Nucleotide sequence of *Chox-a* cDNA and 5′ flanking genomic DNA and deduced amino acid sequence. The symbols are the same as in Figure 1. Putative SP1 binding site and TATA box in the 5′-upstream are indicated by double underlining. The putative poly (A) signal is not found in the 3′ non-coding sequence, perhaps because the isolated cDNAs were not full length. The cDNA sequence starts from +189.

terminal region, and the other is the region around the YPWM box and homeodomain (Figure 8, large boxes). In the homeodomain, only one conservative amino acid difference (Val/Ile) was found. Since the homeodomain is a putative DNA binding domain, the sequence similarity within this domain suggests that both proteins bind to similar DNA sequences.

Determination of transcriptional initiation sites

Before beginning DNA binding studies of *Chox* proteins, their transcriptional initiation sites were determined by RNase protection assay (Figure 3).

In the case of *Chox-1.4*, 446 nt SmaI-TaqI antisense RNA was used as a hybridization probe (Figures 3A and 8A). Two major protected bands were observed at 208 and 114 nt positions on the gel (Figure 3A lane 1). The major transcriptional initiation sites, mapped by the presence of the 208 nt protected band, is indicated as +1 in Figure 1. The shorter band observed at the 114nt position could possibly be due to an RNase protection artefact caused by the poly G tract at this position.

In the case of *Chox-a*, 550 bases BglII-BamHI antisense RNA was used for an RNase protection experiment (Figures 3B and 8A). A single major protected band appeared at the 201nt position, and two minor bands were also seen at the 243 and 160nt positions. The major transcriptional initiation site, mapped by the presence of the 201nt protected band, is indicated as +1 in Figure 2.

In each case, other probes elongating into the 5′ region gave the same results.

Chox-1.4 and Chox-a proteins produced in *E. coli* and chicken embryo

In order to study the DNA binding activity of the *Chox-1.4* and *Chox-a* proteins, we produced full length proteins using an *E. coli* expression system. Full length *Chox* proteins were visible with the gel electrophoresis of crude *E. coli* lysates, and were purified to electrophoretic homogeneity (Figure 4A). These purified proteins were used for the following *in vitro* DNA binding experiments and for preparation of specific polyclonal antibodies.

The intrinsic *Chox-1.4* and *Chox-a* proteins in the chicken embryonic nuclei were visualized by immunoblot analysis. The anti-*Chox-1.4* protein antibody detected a single band which showed the same mobility as the *E. coli* produced *Chox-1.4* protein on SDS electrophoresis gel; both protein products migrated as 40 kd proteins (Figure 4B). The difference between the expected and observed molecular weight may be explained by the great abundance of proline and basic amino acids already observed in the *Drosophila* homeobox gene product (i.e. 37). These results suggest that the majority of *Chox-1.4* protein in the chicken embryo is not affected by modifications such as truncation, phosphorylation or glycosylation that would change the mobility on the gel. Similar results were obtained by

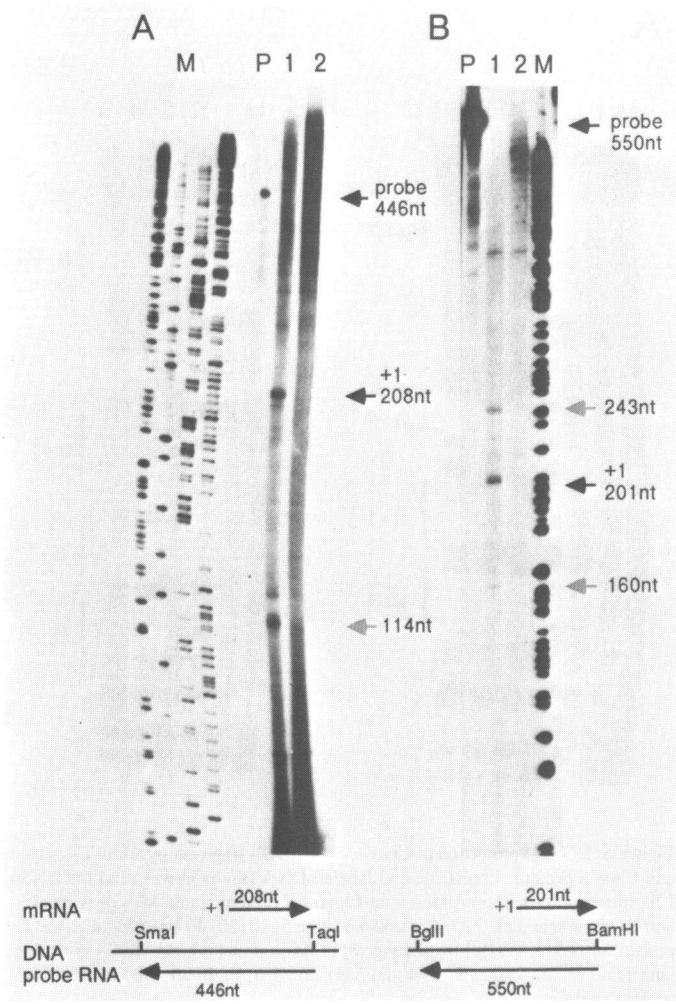

Figure 3. RNase protection assay of *Chox-1.4* and *Chox-a*. Panel A: RNase protection assay with *Chox-1.4* probe. Panel B: RNase protection assay with *Chox-a* probe. Lane M, marker DNA; lane P, probe RNA; lane 1, hybridized with chicken poly(A)$^+$ RNA; lane 2, hybridized with yeast tRNA. The probes used and detailed methods are described in MATERIALS AND METHODS.

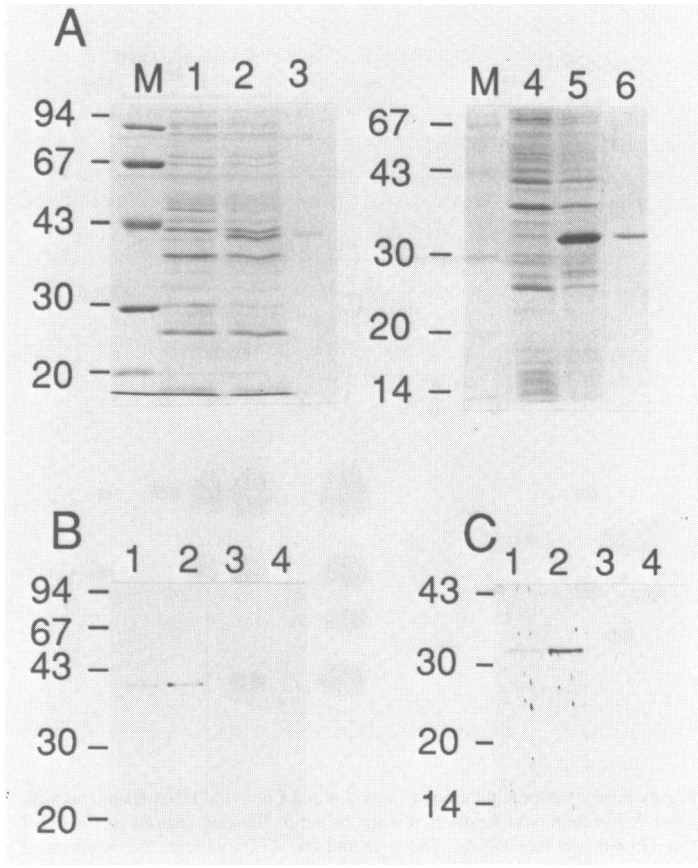

Figure 4. Purification of *E. coli* made *Chox* proteins and immunoblot detection of chicken embryonic *Chox* proteins. Panel A shows 10% (lanes 1–3) and 12.5% (lanes 4–6) SDS-polyacrylamide gels stained with Coomassie Brilliant Blue. Lane 1: 10μg of control *E. coli* lysate, lane 2: 10μg of *Chox-1.4* protein producing *E. coli* lysate, lane 3: 400ng of purified *Chox-1.4* protein, lane 4: 10μg of control *E. coli* lysate, lane 5: 10μg of *Chox-a* protein producing *E. coli* lysate, lane 6: 400ng of purified *Chox-a* protein, lanes M: molecular weight markers. Panels B and C show the results of immunoblot detection. The antibodies used were: lanes 1 and 2: anti-*Chox-1.4* (panel B) or anti *Chox-a* (panel C) IgG, lanes 3 and 4: non-immune IgG (panels B and C). Fractionated proteins are as follows: lanes 1 and 3: 35μg of nuclear proteins from four days old chicken embryos, lanes 2 and 4: 10ng of purified *Chox-1.4* (panel B) or *Chox-a* (panel C) protein from *E. coli*. Molecular masses of standard proteins are indicated with kilodaltons at the left side of each panel.

immunoblot analysis of *Chox-a* protein using anti-*Chox-a* antibody (Figure 4C). The very faint immunoreacting bands visible in lane 1 (about 15 and 18 kd) were nonspecific, since they were also detected when a non-immune mouse IgG fraction was used (lane 3).

Both Chox-1.4 and Chox-a proteins specifically bind to the 5′-flanking DNA fragments of Chox-1.4 and Chox-a genes *in vitro*

Since some homeodomain containing proteins such as *en* (8,9), *Ultrabithorax* (*Ubx*,12) and *Hox-1.3* (38) are known to bind specifically to DNA in their own 5′-upstream sequences, we also tested the specific DNA binding activity of the *Chox-1.4* protein to the 5′-upstream genomic DNA of the *Chox-1.4* coding region.

A 1.05kb EcoRI-TaqI DNA fragment containing the *Chox-1.4* translation start site, 5′-leader sequence and 5′-upstream sequence was digested with three different restriction enzymes (Sau3AI, SmaI and BstNI; Figure 5A). End labeled DNA fragment mixtures were used for an analysis of *Chox-1.4* protein binding by immunoprecipitation (39). The DNA-*Chox-1.4* protein complexes were precipitated with anti-*Chox-1.4* protein antibody. In each restriction enzyme digestion tested, the DNA fragments containing the transcription start site and translation initiation codon were specifically precipitated by the *Chox-1.4* protein and anti-*Chox-1.4* antibody (Figure 5A and B). This suggests that one or more *Chox-1.4* protein binding sites is located within the 255bp BstNI-TaqI DNA fragment.

Due to the sequence homology around the translation start sites and 5′-leader sequences between *Chox-1.4* and *Chox-a* genes (Figures 1, 2 and 7B), and to the resemblance of the spatial gene expression patterns in the spinal cord of chicken embryo (Kuroiwa and Yokoyama, in preparation), we postulated a cross regulatory mechanism of *Chox-1.4* and *Chox-a* gene expression through their protein products. Based on this hypothesis, we tested the specific DNA binding activity of *Chox-1.4* protein to the 5′-region of the *Chox-a* genomic DNA and *vice versa*.

As shown in Figure 5A, a 0.95kb EcoRI-BamHI DNA fragment containing the region upstream from the putative translation initiation codon of the *Chox-a* genomic DNA was digested with several restriction enzymes. The mixture of DNA

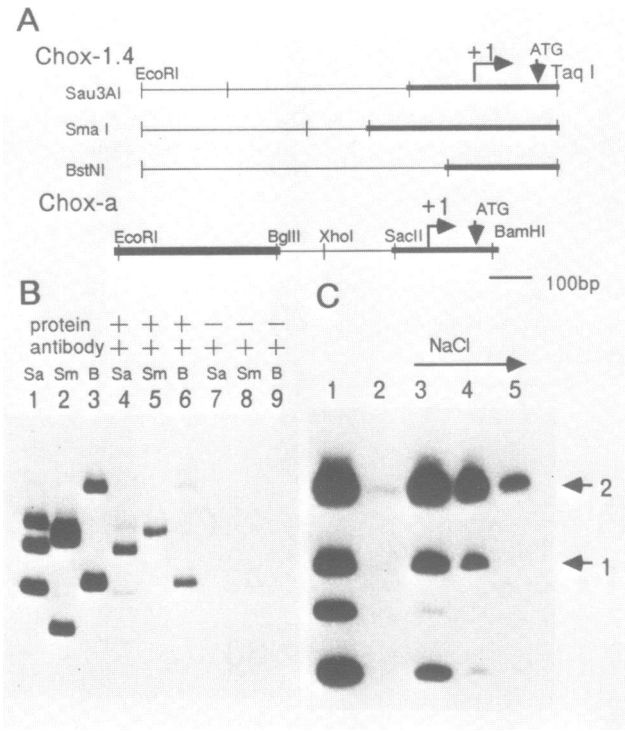

Figure 5. Immunoprecipitation of *Chox-1.4* and *Chox-a* 5′ DNA fragments with *Chox-1.4* protein. (A) Restriction maps of the 5′ flanking regions of *Chox-1.4* and *Chox-a* genomic DNAs. The +1s and the ATGs indicate the positions of major transcriptional initiation sites and putative translational initiation codons, respectively. The specifically precipitated fragments are shown by thick bars. (B) Autoradiographic profile of immunoprecipitated *Chox-1.4* DNA fragments. Lanes 1,2,3: untreated DNA, lanes 4,5,6: precipitated DNA fragments in complete system, lanes 7,8,9: precipitation profile without *Chox-1.4* protein. Lanes 1,4,7: Sau3AI digests, lanes 2,5,8: SmaI digests, lanes 3,6,9: BstNI digests. In all experiments precipitated complexes were washed with TBS-T supplemented with 100mM NaCl. (C) Autoradiographic profile of precipitated *Chox-a* DNA fragments. Lane 1: untreated DNA, lane 2: precipitation without *Chox-1.4* protein. The precipitated complexes for the complete system were washed with TBS-T (lane 3) or TBS-T supplemented with 100mM NaCl (lane 4) or with 200mM NaCl (lane 5) followed by elution of DNA fragments from the final pellets. Arrow 1 and arrow 2 indicate the SacII-BamHI fragment and EcoRI-BglII fragment, respectively.

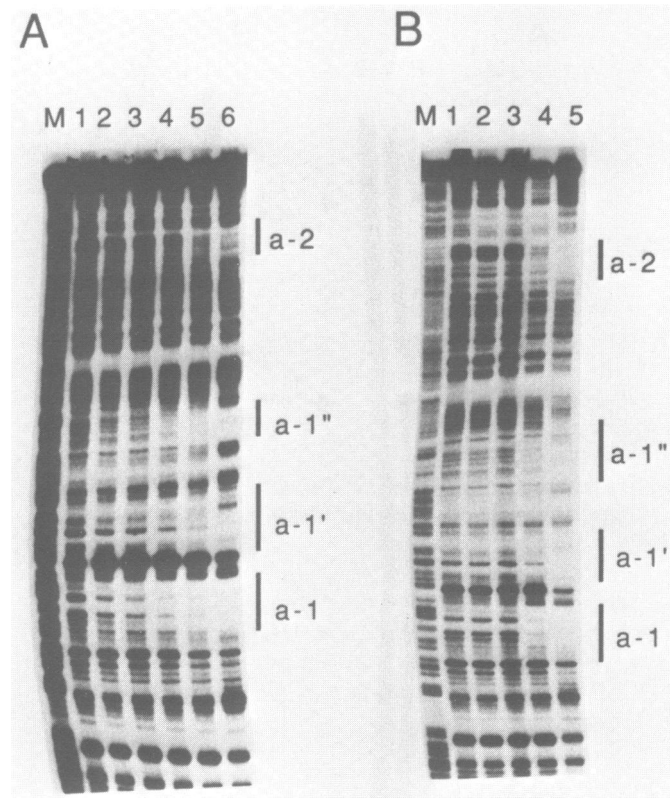

Figure 6. DNase I protection of *Chox-a* EcoRI-BglII fragment by *Chox-1.4* protein and *Chox-a* protein. EcoRI site end-labeled DNA fragment was used for DNase I footprinting. (A) Footprinting with *Chox-1.4* protein. Lane M: G+A marker, lane 1: no protein, lane 2, 3, 4, 5 and 6 contain 5, 10, 20, 40 and 80ng of *Chox-1.4* protein, respectively. (B) Footprinting with *Chox-a* protein. Lane M: G+A marker, lane 1: no protein, lanes 2, 3, 4 and 5 contain 1.6, 8, 40 and 200 ng of *Chox-a* protein, respectively.

fragments was analyzed for *Chox-1.4* protein binding by immunoprecipitation. Binding strength was measured by treating the DNA-protein complex with varying concentrations of NaCl in the washing buffer (Figure 5C). The DNA-protein complexes enduring challenge by higher NaCl concentrations are those that have the higher relative binding strength. As expected, *Chox-1.4* protein bound to the SacII-BamHI DNA fragment (arrow 1 in Figure 5C), where a sequence homology with *Chox-1.4* BstNI-TaqI fragment was observed (52 nucleotides of 60 are identical, Figure 7B). The binding strengths of *Chox-1.4* protein to *Chox-a* SacII-BamHI and *Chox-1.4* BstNI-TaqI fragments were nearly the same, since the *Chox-1.4* BstNI-TaqI fragment precipitated with *Chox-1.4* protein at 100mM-NaCl containing TBS-T (final NaCl concentration was 237mM because TBS-T itself contains 137mM-NaCl) but did not precipitate at 200mM-NaCl containing TBS-T (final NaCl concentration 337mM) (Figure 5B, C, data not shown). In addition, *Chox-1.4* protein also bound to the EcoRI-BglII fragment (arrow 2 in Figure 5C) with higher strength than to the SacII-BamHI fragment.

Since the amino acid sequences in the homeodomains of *Chox-a* protein and *Chox-1.4* protein were very similar, it seemed likely that both proteins would recognize similar sequences. Therefore, we repeated the immunoprecipitation analysis using purified *Chox-a* protein and anti-*Chox-a* antibody in place of *Chox-1.4* protein and anti-*Chox-1.4* antibody; the result was the same. The *Chox-a* protein bound to the *Chox-1.4* BstNI-TaqI fragment and to the *Chox-a* SacII-BamHI fragment with similar strength. The *Chox-a* protein also bound to the *Chox-a* EcoRI-BglII fragment with higher strength than to the SacII-BamHI fragment (data not shown).

Both Chox-1.4 protein and Chox-a proteins recognize the same nucleotide sequences

In order to determine the recognition sequences of the *Chox-1.4* and *Chox-a* proteins, we carried out a DNase I footprint analysis. At the same time, the relative binding affinity to each binding site was measured by changing *Chox* protein concentration. The sites protected at low protein concentration were defined as having strong affinities, and the sites requiring a high concentration of proteins for protection as having weak affinities. In the *Chox-a* EcoRI-BglII fragment, *Chox-1.4* protein bound to four different regions with varying affinities : to sites a-1 and a-1″ were strong, since these sites were protected at low protein concentration. The binding affinities to sites a-1′ and a-2 were weak since these sites required a higher concentration of proteins for protection (Figures 6A and 7). The protection patterns of both strands were similar

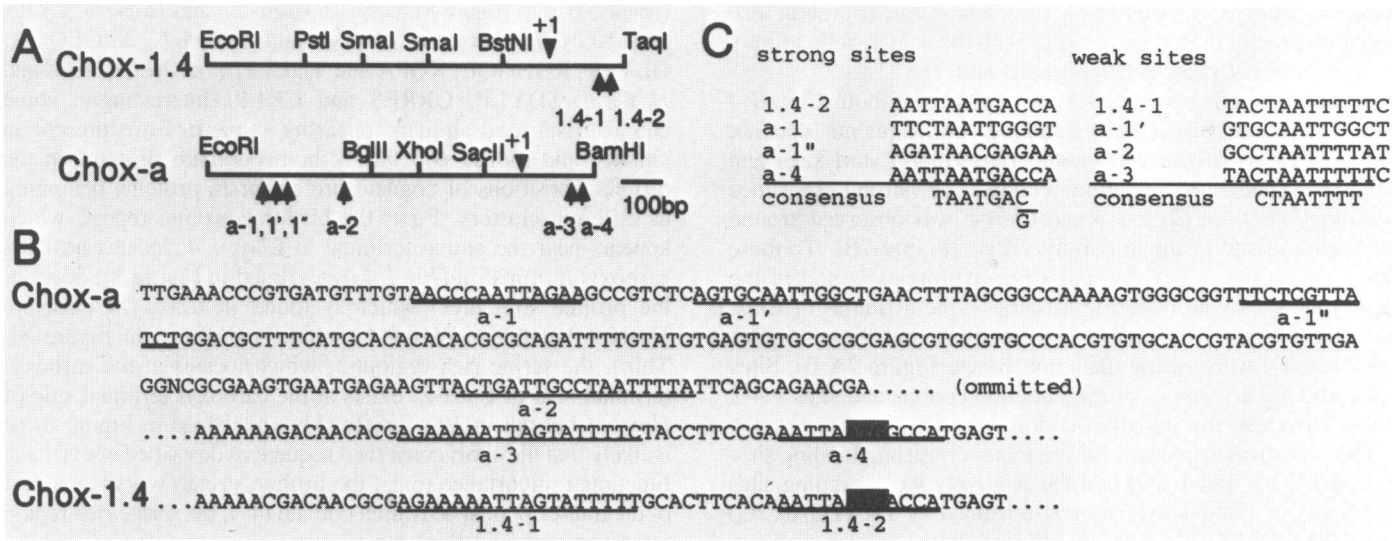

Figure 7. Summary of *Chox* protein binding sites. (A) Map of *Chox* protein binding sites. *Chox* protein binding sites and their names are indicated. *Chox-1.4* and *Chox-a* proteins bound to the same sites. +1s indicate the transcriptional initiation sites. (B) Nucleotide sequences around *Chox* protein binding sites. *Chox* protein binding sequences are indicated by underlines and their names. Strong binding sites are indicated with thick underlines, and weak sites with thinner underlines. Translational initiation codons are indicated by hatching. (C) Summary of *Chox* protein binding sequences. Binding sites were classified into strong and weak binding sites. The consensus sequence for each class is shown separately.

```
Chox-1.4   MTMSSFLINSNYIEPKFPPCEEYTQHSGSAGSSASYHP-HHPHPHAPPPPPPPPPPHLH-AAHPGPALPEYF-PRPRREP
Hox-1.4    ********************F**FAP*-*GP*GGDGAVG-GG*GYPR*QSA*HL*A*NP*-**RQP**--Y*A-**-A***
Chox-a     *A***YMV****VD********L*S*YLGEQG*E*----------------------Y-GASQ*SDFQHQG-LY**SNY
Hox-5.1    *A***YMV****VD********L*GGYLGEQG*D*----------------------YGGGAQ*ADFQPSG-LY**PDF
Chox-Z     *A**********VD********--------*HSD*L*-N*-------------------------S***Y-SSQ***S
Hox-2.6    *A**********VD********--------*QSD*L*SD*-------------------------S*G*YAGGQ***S

Chox-1.4   GYQAPAAPPGPPGPPPEALYP-AQAPSY----------PQAPYSYSSAGSAAPGPEQ-PPPGASPPPPPPAKGHPGP---
Hox-1.4    S*-------------*GG***-*P*AAC----------*Y*CRGA*P*-----R***S*A***H*S*------A*Q*---
Chox-a     SE*------------*FW*AG-**-GSA----------V*-PRGHGQEQ**--P*SHF*GQAEHC*-***---C*I*---
Hox-5.1    GE*------------*FGGGGPGP-GSA----------LP-ARGHGQEP*G--PGSHYGAP**R**A***---A*L*GAR
Chox-Z     TFQ-HE*MYQ*RSACS*Q***-SCQS*G----------H**-AVL*PR*HVH*PAGLQSHLSEPNH*CE*GTPS--*---
Hox-2.6    GFQ-*E*AFGRRA*CTVQR*A-*CRDPGPPPPPPPPPPPPP-PGL*PRAPVQ*TAG--ALLPEPGQRSEAVSSS*-*---

Chox-1.4   AQPLLPGHALQRRCEA-APAAGAGTGPGC-ALLPDKSLPGLKGKEPVVYPWMKKIHVSTV-NPNY-SGGEPKRSRTAYTR
Hox-1.4    PV**--**CAPGPTTP-*V*T*-*SA*A*PL**A*QGPA*P**********************A*-*SS*-N**********
Chox-a     SCGQQP--**-----K-**HGS*VKQ*-----------------A********V**NS*-****-**********
Hox-5.1    *CSQPT--GP-----KQP*PGT*LKQ*-----------------A********V**NSA-****-T**********
Chox-Z     ----P*SCSQ--------------------NSLNQ*PSNSSC**********V*****-****-**********
Hox-2.6    ----P*PC*Q--------------------NPLHP*PSHSAC**********R*V*****-****-A**********

Chox-1.4   QQVLELEKEFHFNRYLTRRRRIEIAHTLCLSERQVKIWFQNRRMKWKKDHKLPNTKMRSSNQ-------PS-LGQQQAKA
Hox-1.4    ********************************************************T-------A*APAGPPG**
Chox-a     ********************************I*******************G**SS-------A*NPHL*----
Hox-5.1    ********************************P***I*******************G**SSSSCSSSAAPGQHL*----
Chox-Z     **********Y*********V****S*******I*********************I**--N-------**---SSS*SL
Hox-2.6    **********Y*********V****A*******I*********************I**--G-------GT---AGA*--

Chox-1.4   QT---QGHPRPLDGA-APNAAAL
Hox-1.4    **HSPHP**H**P**ST*IPSSI
Chox-a     --------TV*K*-H-QTDLTT*
Hox-5.1    --------*MAK*-H-HTDLTT*
Chox-Z     *I---PPAASQSRSS-G*-*SS*
Hox-2.6    GG---PP----GRPN-GG-PP**
```

Figure 8. Sequence comparison of chicken and mouse *Dfd* family homeodomain proteins. *Chox-a, -Z, Hox-1.4, -5.1* and *-2.6* are shown as a maximum matching to *Chox-1.4*. The amino acids identical to *Chox-1.4* are indicated by stars. Gaps are indicated by bars. Highly conserved regions are enclosed by large boxes. Conserved sequences between chicken proteins and their mouse homologues are enclosed in small boxes. The histidine-proline repeats are underlined, proline stretches are double underlined, and serine rich regions are thickly underlined.

(data not shown). To this DNA fragment, *Chox-a* protein also bound to exactly the same four sites with the nearly same affinity as the *Chox-1.4* protein (Figures 6B and 7).

A similar binding pattern was observed using both *Chox-1.4* BstNI-TaqI and *Chox-a* SacII-BamHI DNA fragments (data not shown). These fragments contain transcription start sites and putative translation initiation codons. A strong sequence homology between *Chox-1.4* and *Chox-a* was observed around the region just 5′ of the initiation codons (Figure 7B). To these DNA fragments, *Chox-1.4* and *Chox-a* proteins bound to the same four sites with differing affinities : the affinities of *Chox* proteins to sites 1.4-1 and a-3 were weak and those to sites 1.4-2 and a-4 were strong (data not shown, Figure 7A,B). Sites 1.4-1 and a-3 are just 5′ of the initiation codons and sites 1.4-2 and a-4 overlap the initiation codons.

The consensus sequences for the relatively strong binding sites (a-1, a-1″, a-4 and 1.4-2) and the relatively weak binding sites (a-1′, a-2, a-3 and 1.4-1) were determined as TAATGA(C/G) and CTAATTTT, respectively (Figure 7C).

DISCUSSION

Amino acid sequences of chicken Dfd family homeodomain proteins

We have determined all the amino acid coding sequences of the cDNA clones corresponding to the products of three different chicken *Dfd* family homeobox genes (*Chox-1.4* and *Chox-a*, this paper, and *Chox-Z* (40)). Comparing the amino acid sequences of these three homeodomain proteins, we found two highly conserved regions among them, as also described in mouse *Dfd* family homeodomain proteins (e.g. 41). One region is located around the amino terminal and the other is located around the YPWM box and homeodomain (both regions are enclosed with large boxes in Figure 8). These regions of chicken *Dfd* family homeodomain proteins have sequence homologies with *Drosophila Dfd* protein (42) and other *Dfd* family homeodomain proteins such as mouse *Hox-1.4* (41), *−5.1* (43), *−2.6* (44), human c13 (36), cp19 (45) and *Xenopus Xhox-1A* (46) (Figure 8, data not shown). The presence of strong conservation of the species suggests that these domains have important functions common to *Dfd* family proteins (e.g. DNA recognition, nuclear localization, protein-protein interaction, protein stability, etc.).

Comparison of these three chicken homeodomain proteins with mouse *Dfd* family homeodomain proteins suggests that the *Chox-1.4, -a* and *-Z* are chicken homologues of mouse *Hox-1.4, -5.1,* and *-2.6,* respectively. The reason is that the amino acid sequence homologies in the conservative regions described above are highest in this combination and the homologies in whole proteins are also strongest (*Chox-1.4* and *Hox-1.4*, 57%; *Chox-a* and *Hox-5.1*, 74%; *Chox-Z* and *Hox-2.6*, 66%) (Figure 8). These sequence homologies are somewhat lower than the homology between *Hox-5.1* and c13 (human homologue of *Hox-5.1*), which is 91%. These rather weak homologies between chicken genes and mouse genes are perhaps based on the distance of their species (mouse and human belong to the same mammalian class while chicken belongs to a different class). It may be that only functionally important sequences are conserved and functionally different and/or relatively unimportant sequences are different between chicken and mouse.

In less homologous regions, however, we found some short conserved sequences. Those whose amino acid sequences are identical in more than four continuous residues are as follows (small boxes in Figure 8): between *Chox-1.4* and *Hox-1.4*, LYPA and KGKEP; between *Chox-a* and *Hox-5.1*, TLGEQGA, GLYPR, RGHGQE, KQPA and TDLTTL; between *Chox-Z* and *Hox-2.6*, SDYLP, QRRES and CKEP. Interestingly, some characteristic and similar appearing sequences are present in chicken and mouse *Dfd* family homeodomain proteins in the different positions of cognate protein or in proteins belonging to different clusters. First, the histidine-proline repeat, which appears near the amino terminal in *Chox-1.4*, locates near the carboxyl terminal in *Hox-1.4* (underlined in Figure 8). Second, the proline stretch(es) which is found in *Chox-1.4* exists in *Hox-2.6* but not in *Hox-1.4* (double underlined in Figure 8). Third, the serine rich region(s), which locates at the carboxyl terminal side of *Chox-Z*, exists at the carboxyl terminal side of *Hox-5.1* but not in *Hox-2.6* (thickly underlined in Figure 8). It is likely that the short conserved sequences described above have functional importance (e.g., the proline stretch works as a part of the transcriptional activation domain (47), the serine rich region affects phosphorylation, etc.).

The close similarity in amino acid sequences of the homeodomains suggests that *Dfd* family homeodomain proteins recognize and bind to the same DNA sequences. Here, we have shown that *Chox-1.4* and *Chox-a* proteins bind to the same sequences displaying nearly the same mode of binding affinity at each site. In transcriptional regulation by a DNA binding protein factor, protein−protein interaction through a specific functional domain to the basic transcriptional machinery might take place to mediate activation or suppression of transcription. It is likely that these *Dfd* family *Chox* proteins, which have similar DNA binding features, have different effects on the basic transcriptional machinery through their different amino acid sequences.

DNA binding studies

In this paper, we prepared of purified and full length *Chox-1.4* and *Chox-a* proteins for DNA binding studies. We used purified proteins to remove the possibility of contamination which might interfere with or alter DNA binding specificities of these proteins. We used full-length proteins because we could know the effects of regions outside of the homeodomains, and to remove the possibility that fused and/or partial proteins have different DNA binding characters.

Using soluble, full-length and purified proteins, we showed that both *Chox-1.4* and *Chox-a* proteins bound to *Chox-1.4* and *Chox-a* 5′ flanking regions of these genes with varying affinities. The binding recognition sequences and the binding affinities of the two proteins were the same. This suggests that the amino acid sequences characteristic of each protein do not affect their DNA binding activities analyzed by the method employed here. Consensus sequences of TAATGA(C/G) for high binding affinity sites and CTAATTTT for low binding affinity sites were derived. Since both sequences contain TAAT, we suggest that TAAT is a core sequence and/or is essential for the binding of the homeodomain and that surrounding sequences determine the difference in affinity. We have recently identified sequences that possess higher affinity for the *Chox-1.4* protein than the sequences described here; these also contain the TAATGA motif but differ in the flanking sequences (our unpublished observation).

We identified clustered *Chox-1.4* and *Chox-a* protein binding sites in the 5′-upstream sequence of the *Chox-a* gene (Figure 7). This suggests two possibilities with respect to regulation of *Chox-a* gene expression. First, in the case of the *Drosophila* homeobox

genes *Antp* (18) and *Ubx* (19), an autoregulatory mechanism through the protein products has been demonstrated. It is possible that the clustered binding sites found in the 5'-flanking sequence of the *Chox-a* gene function as a *cis*-element of an autoregulation system. Second, since the *Chox-1.4* protein can also bind to this sequence and *Chox-1.4* and *Chox-a* genes are found to be expressed in the same regions of an embryo with only minor differences (Kuroiwa and Yokoyama in preparation), *Chox-1.4* protein also has the potential to bind this sequence in *Chox-a* producing cells. It is easy to imagine that the two proteins act synergistically to regulate *Chox-a* gene expression, as postulated for *Drosophila* homeobox genes (17). An analysis of the function of this *cis* element in transcription is under way.

ACKNOWLEDGMENTS

pAR3038 was the generous gift of Drs. A.H.Rosenberg, J.J.Dunn and F.W.Studier. We are grateful to Dr. Leslie Pick for criticizing the manuscript. This work was supported in part by Grants-in-Aid from the Ministry of Education, Science and Culture of Japan.

REFERENCES

1. McGinnis,W., Levine,M.S., Hafen,E., Kuroiwa,A and Gehring,W.J. (1984) Nature, 308, 428–433.
2. Scott,M.P. and Weiner,A.J. (1984) Proc. Natl. Acad. Sci. USA, 81, 4115–4119.
3. Gehring,W.J. and Hiromi,Y. (1986) Annu Rev. Genet. 20, 147–173.
4. Akam,M. (1987) Development, 101, 1–22.
5. Scott,M.P. and Carrol,S.B., (1987) Cell 51, 689–968.
6. Hafen,E., Levine,M. and Gehring,W.J. (1984) Nature, 307, 287–289.
7. Kuziora,M.A. and McGinnis,W. (1988) Cell, 55, 477–485.
8. Desplan,C., Theis,J. and O'Farrell,P.H. (1985) Nature, 318, 630–635.
9. Desplan,C., Theis,J. and O'Farrell,P.H. (1988) Cell, 54, 1081–1090.
10. Fainsod,A., Bogarad,L.D., Ruusala,T., Lubin,M., Crothers,D.M. and Ruddle,F.H. (1986) Proc. Natl. Acad. Sci. USA, 83, 9532–9536.
11. Hoey,T. and Levine,M. (1988) Nature, 332, 858–861.
12. Beachy,P.A., Krasnow,M.A., Gavis,E.R. and Hogness,S.D. (1988) Cell, 55,1069–1081.
13. Müller,M., Affolter,M., Leupin,W., Ottig,G., Wüthrich,K and Gehring,W.J. (1988) EMBO J. 7, 4299–4304.
14. Mihara,H. and Kaizer,E.T. (1988) Science, 242, 925–927.
15. Jaynes,J.B. and O'Farrell,P.H. (1988) Nature, 336, 744–749.
16. Driever,W. and Nüsslein-Volhard,C. (1989) Nature, 337, 138–143.
17. Han,K., Levine,M.S. and Manley,J.L. (1989) Cell, 56, 573–583.
18. Winslow,G.M., Hayashi,S., Krasnow,M. Hogness,D.S. and Scott.M.P. (1989) Cell, 57, 1017–1030.
19. Krasnow,M.A., Saffman,E.E., Kornfeld,K. and Hogness,D.S. (1989) Cell, 57, 1031–1043.
20. Herr,W. and Strum,R.A. (1988) Genes Dev., 2, 1513–1516.
21. Levine,M. and Hoey,T. (1988) Cell, 55, 537–540.
22. Holland,P.W.H. and Hogan,B.L.M. (1988) Genes Dev., 2, 773–782.
23. Duboule,D. and Dolle,P. (1989) EMBO J., 8, 1497–1505.
24. Graham,A., Papalopulu,N. and Kramlauf,R. (1989) Cell, 57, 367–378.
25. Cho,K.W.Y., Goetz,J., Wright,C.V.E., Fritz,A., Harhwicke,J. and DeRobertis,E.M. (1988) EMBO J. 7, 2139–2149.
26. Harvey,R.P. and Melton,D.A. (1988) Cell, 53, 687–697.
27. Wolgemuth,D.J., Behringer,R.R., Mostoller,M.P., Brinster,R.L. and Palmiter,R.D. (1989) Nature, 337, 464–467, 1229–1235.
28. Gubler,U. and Hoffman,B.J. (1983) Gene, 25, 263–269.
29. Sanger,F. (1981) Science, 214, 1205–1210.
30. Gilman,M. (1987) Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D. Seidman,J.G. Smith,J.A. and Strul,K. (ed).,In Current Protocols in Molecular Biology. Greene Publishing Associates and Wiley-Interscience, unit 4.7.
31. Rosenberg,A.H., Lade,B.N., Chui,D., Lin,S., Dunn,J.J. and Studier,F.W.(1987) Gene, 56, 125–135.
32. Studier,F.W. and Moffatt.B.A. (1986) J. Mol. Biol., 189, 113–130.
33. Gorski,K., Carneiro,M. and Schibler,U. (1986) Cell, 47, 767–776.
34. Laemmli,U.L. (1970) Nature, 227, 680–685.
35. Maxam,A.M. and Gilbert,W. (1977) Proc. Natl. Acad. Sci. USA, 74, 560–564.
36. Mavilio,F., Simeone,A., Giampaolo,A., Faiello,A., Zappavigna,V., Acampora,D., Poiana,G., Russo,G., Peschle,C. and Boncinelli,E.(1986) Nature, 324, 664–668.
37. Krause,H.M., Klementz,R. and Gehring,W.J. (1988) Genes Dev. 2, 1021–1036
38. Odenwald,W.F., Garbern,J., Arnheiter,H., Tournier-Lasserve,E. and Lazzarini,R.A. (1989) Genes Dev., 3, 158–172.
39. McKay,R.G.D. (1981) J.Mol.Biol., 145, 471–488.
40. Sasaki,H. and Kuroiwa,A (1990) Nucleic Acids Res., 18, 184.
41. Galliot,B., Dollé,P., Vigneron,M., Featherstone,M. Baron,A. and Doboule,D. (1989) Development, 107, 343–389.
42. Regulski,M., McGinnis,N., Chadwick,R. and McGinnis,W. (1987) EMBO J. 6, 767–777.
43. Featherstone,M.S., Baron,A., Gaunt,S.J., Mattei,M.-G., Duboule,D.(1988) Proc. Natl. Acad. Sci. USA, 85, 4760–4764.
44. Graham,A., Papalopulu,N., Lorimer,J., McVey,J.H., Tuddenham,E.G.D. and Kramlauf,R. (1988) Genes Dev. 2, 1424–1438.
45. Simone,A., Pannese,M., Acampora,D., D'Esposito,M. and Boncinelli,E. (1988) Nucleic Acids Res. 16, 5379–5390.
46. Harvey,R.P., Tabin,C.J. and Melton,D.A. (1986) EMBO J. 5, 1237–1244.
47. Mermod,N., O'Neill,E.A, Kelly,T.J. and Tjian,R. (1989) Cell 58, 741–753.