Journal of Cellular Biochemistry 52:449–462 (1993)

Protein Product of the Somatic-Type Transcript of the *Hoxa-4 (Hox-1.4)* Gene Binds to Homeobox Consensus Binding Sites in Its Promoter and Intron

Ke Wu and Debra J. Wolgemuth

Departments of Biochemistry and Molecular Biophysics (K.W.), Genetics and Development (D.J.W.), and Obstetrics and Gynecology (D.J.W.), The Center for Reproductive Sciences (K.W., D.J.W.), Columbia University College of Physicians and Surgeons, New York, New York 10032

Abstract The murine *Hoxa-4* gene encodes a protein with a homeodomain closely related to those produced by the *Antennapedia*-like class of *Drosophila* genes. *Drosophila* homeodomain proteins can function as transcription factors, binding to several specific DNA sequences. One sequence that is frequently encountered contains a core ATTA motif within a larger consensus sequence, such as CAATTAA. The in vitro synthesized protein product of *Hoxa-4* was shown to bind to a subset of restriction fragments of the *Hoxa-4* gene itself as determined by gel retardation experiments. Direct examination of the sequences of the fragments bound by Hoxa-4 protein revealed the presence of four regions containing the core ATTA motif. Two regions contained sequences of the CAATTAA class and were located ~1 kb upstream from the putative somatic *Hoxa-4* promoter and within the intron. Two additional binding sites containing the consensus target sequence involved in autoregulation of *Drosophila Deformed* gene were identified: one immediately downstream of the putative embryonic transcription start site and one within the intron, respectively. Specific binding of the in vitro produced Hoxa-4 protein to oligonucleotides corresponding to these sequences was observed in gel retardation assays. The same results were obtained with Hoxa-4 protein produced in a Baculovirus expression system. Experiments using oligonucleotides containing base substitutions in positions 1, 3, 4, and 5 in the sequence CAATTAA showed severely reduced binding. The use of truncated mutant Hoxa-4 proteins in gel retardation assays and in transient co-transfection experiments revealed that the intact homeodomain was required for the binding. These results also suggested that the *Hoxa-4* gene has the potential to auto-regulate its expression by interacting with the homeodomain binding sites present in the promoter as well as in the intron.

Key words: DNA binding proteins, transcription factors, gene regulation

Genetic and molecular analyses have established that pattern formation in the *Drosophila* embryo is controlled by a cascade of activity of interacting genes. The presence of potential DNA-binding domains in the products of many of these genes suggests that they may function as transcription factors. One such domain is encoded by the homeobox, a 183 bp sequence found in several classes of development-regulating genes in *Drosophila* [McGinnis et al., 1984; Scott and Weiner, 1984; review in Gehring, 1987; Scott et al., 1989] and highly conserved in organisms as evolutionarily diverse as sea urchins and mammals. The presence of this domain in mammalian genes expressed in temporally and spatially restricted patterns during embryogenesis [review in Holland and Hogan, 1988; Kessel and Gruss, 1990] raises the possibility that these genes may regulate the expression of target genes during mammalian development.

The approximately forty mouse homeobox genes identified to date are organized in a linear array in clusters, known as *Hox-A, Hox-B, Hox-C,* and *Hox-D,* on mouse chromosomes 6, 11, 15, and 2, respectively [Duboule et al., 1990; Martin et al., 1987; Scott, 1992]. Comparison of the nucleotide sequence of the homeodomain and other regions of these mammalian genes with their *Drosophila* counterparts reveals that their position within this linear array has also been conserved throughout evolution [Duboule and Dolle, 1989; Gaunt et al., 1988; Graham et

Received February 25, 1993; accepted March 22, 1993.

Address reprint requests to Dr. Debra J. Wolgemuth, Associate Professor of Genetics and Development, Columbia University College of Physicians and Surgeons, 630 West 168th Street, New York, NY 10032.

al., 1989]. Furthermore, as in *Drosophila,* expression of the genes along the anterior-posterior axis of the mouse embryo follows a positional order which reflects their order within the cluster. The extent to which this organizational and sequence conservation reflects functional conservation remains to be demonstrated. However, McGinnis and colleagues [Malicki et al., 1990; McGinnis et al., 1990] have provided evidence that some specific biochemical and developmental functions have been conserved in vertebrate homologues of *Drosophila* genes. Mouse *Hoxb-6* (*Hox-2.2*) and human *HOXD4* (*Hox-4.2*) can substitute at least in part for the normal regulatory function of their *Drosophila* homologues *Antennapedia (Antp)* and *Deformed (Dfd),* respectively.

The 61 amino acid-long homeodomain consists of three alpha helical regions and a flexible N-terminal arm [Laughon and Scott, 1984; Qian et al., 1989; Shepherd et al., 1984]. The putative recognition helix, helix 3, is the most highly conserved region among different homeodomain proteins, suggesting that these proteins may recognize closely related DNA sequences. Examination of the sequences protected in DNase I footprinting and gel retardation assays has revealed that distinct consensus, high affinity binding sites can nonetheless be identified for specific homeodomains [Laughon, 1991]. The *Antp*-like class of homeodomains recognizes consensus sequences containing an ATTA core motif flanked by two or more nucleotides recognized only by specific homeodomains. For example, a consensus sequence that is frequently encountered in target sequences of the *fushi tarazu (ftz)* class of *Antp*-like genes recognizes the consensus sequence CAATTAA [Cho et al., 1988; Desplan et al., 1988; Hoey and Levine, 1988].

The mouse *Hoxa-4* homeobox-containing gene gives rise to two distinct transcripts expressed in restricted patterns during embryogenesis and in adult male germ cells [Duboule et al., 1986; Rubin et al., 1986; Wolgemuth et al., 1986, 1987]. We wished to determine how the unique tissue, temporal, and spatial specificity of *Hoxa-4* expression is established in the embryo and the adult and to identify downstream or target genes of *Hoxa-4.* Given the recently documented autoregulatory capacity of the *Drosophila* homologue of *Hoxa-4,* namely, *Dfd* [Bergson and McGinnis, 1990; Kuziora and McGinnis, 1988; Regulski et al., 1991], we began our studies on *Hoxa-4* regulation and target specificity by determining if the mouse gene has binding sites for its own protein product. In the present study, the ability of the somatic form of the *Hoxa-4* gene product to bind to sequences of the *Hoxa-4* gene was addressed using in vitro synthesized and Baculovirus-produced proteins and in vivo in cell lines. Specific binding was demonstrated to sequences in the promoter that appears to be used in somatic cells, including those in the developing embryo, and within the intron. Using a series of mutant and truncated forms of the Hoxa-4 protein, this binding was shown to be dependent upon the presence of the homeodomain. These observations suggest that *Hoxa-4* may have, among its functions, the potential to autoregulate its expression.

METHODS

Recombinant Constructs

All plasmid constructs were generated according to standard cloning procedures [Sambrook et al., 1989]. Plasmid p18CAT was constructed by inserting a blunt-ended 1.6 kb HindIII-BamHI CAT gene fragment from pSV_2CAT [Laimins et al., 1982] into the mung bean nuclease-treated EcoRI site of pUC18. 5′ deletion mutants of *Hoxa-4,* designated pHox-1.4-CAT plasmids, were constructed as follows (Fig. 1). Restriction fragments were 5′ blunt-ended with T7 DNA polymerase to yield fragments containing sequences flanking the putative embryonic promoter plus 80 base pairs of the first exon. These were directionally cloned into the SmaI-SacI sites of plasmid p18CAT. Thus, the *Hoxa-4* translation start (ATG) now present in all the promoter-CAT constructs was in frame with the translation frame of the downstream CAT gene.

To generate a full-length *Hoxa-4* cDNA, plasmid p(−52)Hox-1.4-CAT was cut with XbaI at the polylinker and treated with Bal31. The DNA was then cut with SacI and fragments of ~50–70 bp in size were selected by gel purification and cloned into the SmaI-SacI site of plasmid p18CAT. The resulting plasmid, p(+7)Hox-1.4-CAT, contains 38 base pairs of the 5′ untranslated sequences plus 14 base pairs of the coding sequence. A 1.5 kb SacI-HpaI (HpaI cuts within the CAT gene) fragment of plasmid p(+7)Hox-1.4-CAT was replaced by a *Hoxa-4* genomic SacI-FspI fragment which contains all the coding sequence (except the first 14 bp), the intron and the 3′ untranslated sequence, resulting in plasmid pHox-1.4 Fg. Plasmid pHox-1.4 Fc, containing a full-length *Hoxa-4* cDNA equivalent, was

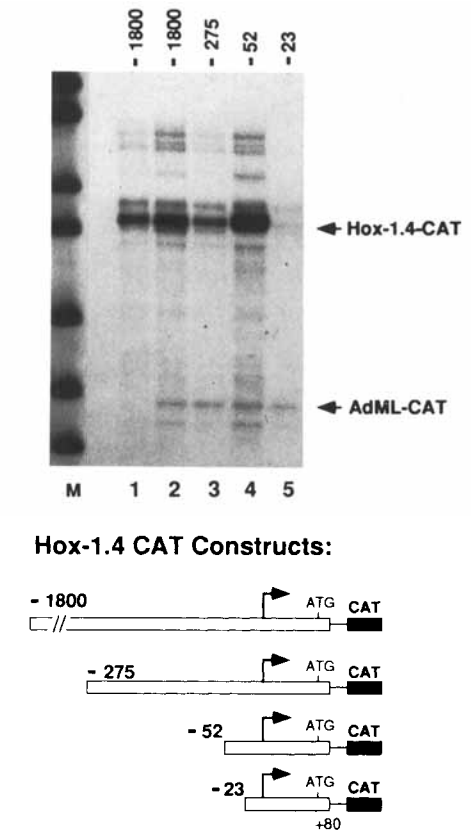

Fig. 1. Mapping the 5′ end of the *Hoxa-4* somatic-type transcript by in vitro transcription. 5′ deletion Hox-1 4-CAT plasmids (400 ng) were transcribed together with 100 ng internal control pAdML-CAT plasmid (except for lane 1 which received no pAdML-CAT) in the presence of undifferentiated F9 cell nuclear extracts. Transcripts were then analyzed by primer extension. Lane M contains HpaII-cut pBR322 end-labeled DNA markers of the following sizes (in nucleotides) 161, 148, 124, 111, 91, 77, and 68. Lanes 1–5 were loaded with the primer extended products of the Hox-1 4-CAT constructs indicated below. The arrows show the sizes (in nucleotides) of the indicated transcripts (Hoxa-4-CAT, 116, AdML-CAT, 75).

then generated from plasmid pHox-1.4 Fg by replacing a 685 bp genomic NruI-EcoRI fragment containing the *Hoxa-4* intron with the corresponding 200 bp of a cDNA clone (C.M. Viviano and D.J. Wolgemuth, unpublished observations).

All expression vectors were derived from a plasmid pRc/CMV (Invitrogen, San Diego, CA), which contains the CMV promoter and enhancer. Inserted sequences were expressed from the first ATG codon, and the T7 promoter was used for the in vitro production of sense transcripts.

The FTZ expression vector, plasmid pCMV-Ftz, was constructed by ligating a 2.6 kb HindIII-XbaI fragment from pF450 (provided by Dr. Gary Struhl) into HindIII-XbaI digested pRc/CMV vector. The Hoxa-4 expression vector pCMV Hox-1.4 was constructed by inserting the 1.1 kb blunt-ended XbaI fragment from pHox-1.4 Fc into the filled-in HindIII site of the pRc/CMV vector.

Expression vectors encoding mutant Hoxa-4 proteins were constructed as follows. For pCMV-Hox-1.4H, plasmid pHox-1.4 Fc was digested with SalI and HindII, filled-in with Klenow and self-ligated. A 0.5 kb HindIII-XbaI fragment of the resulting plasmid was cloned into pRc/CMV vector digested with HindIII and XbaI. The Hox-1.4H mutant construct deletes all amino acids encoded by the first exon and gives rise to a 114 amino acid-long carboxyl terminal protein with five foreign amino acid residues (Met-Pro-Ala-Gly-Arg) encoded by the pUC18 polylinker sequence at the N-terminal end. Plasmid pCMV-Hox-1.4B was constructed by ligating a 1.1 kb blunt-ended XbaI fragment from plasmid pHox-1.4 Fc, which was filled-in at the BglII site within the homeobox, into the filled-in HindIII site of the pRc/CMV vector. This mutation created a frame-shift, resulting in a 226 amino acid-long Hoxa-4 protein with 62 foreign amino acid residues at the C-terminal end. pCMV-Hox-1.4E was constructed by ligating a 1.1 kb blunt-ended XbaI fragment from plasmid pHox-1.4 Fc, which had a nonsense linker inserted at the EcoRI site within the homeobox, into the filled-in HindIII site of pRc/CMV vector. This mutation resulted in a prematurely terminated protein at amino acid residue 199. Reporter plasmid p(6x)-Hox-1.4-CAT was constructed by inserting multiple copies of the double stranded oligonucleotide 5′(tcgacAGGGCAATTAAATTTAg) 3′ and its complementary strand into the SalI site of the −109 tkCAT construct [McKnight, 1982]. Dideoxy sequencing [Zhang et al., 1988] was employed to determine both the structure and sequence of all plasmids used in these studies.

In Vitro Transcription

Preparation of nuclear extracts from undifferentiated F9 cells and the in vitro transcription reaction were essentially as described by Dignam et al. [1983]. The standard reaction consisted of 100 μg of protein of the nuclear extracts, 400 ng supercoiled Hox-1.4-CAT plasmid DNA plus 100 ng supercoiled pAdML-CAT DNA [Lufkin et al., 1989], each ribonucleotide triphosphate at 1.0 mM, 40 units RNasin in 50 μl total volume of 20 mM Tris (pH 7.9), 1 mM DTT, and

12 mM $MgCl_2$. After incubation at 30°C for 1 h, 150 μl stop buffer (100 mM NaCl, 25 mM EDTA, and 50 μg/ml tRNA) were added. RNA was purified by phenol-chloroform extraction, followed by ethanol precipitation.

The in vitro transcription products were analyzed by primer extension, essentially as described previously [Bodner and Karin, 1987], employing an end-labeled 20-nucleotide primer complementary to sequences lying immediately upstream of the initiation codon of the CAT gene. Primer extension products were analyzed on an 8% denaturing polyacrylamide gel.

In Vitro Synthesis of Hoxa-4 Protein and Its Mutant Derivatives

The CMV-Hox-1.4 constructs were linearized with appropriate restriction enzymes. T7 RNA polymerase was used for in vitro synthesis of *Hoxa-4* transcripts from these CMV-Hox-1.4 constructs. In vitro translation of Hoxa-4 was performed as described by Halazonitis et al. [1988]. The in vitro translated products were examined by SDS-PAGE to confirm their predicted relative molecular weights [Laemmli, 1970].

Production of Hoxa-4 Protein in Baculovirus Expression System

A BamHI-XbaI fragment which contains the full length *Hoxa-4* cDNA equivalent from plasmid pHoxa-4 Fc was cloned into BamHI-XbaI digested pVL1393 shuttle vector (Invitrogen, San Diego, CA). After calcium phosphate-mediated co-transfection of the *Autographa california* nuclear polyhedrosis virus genome and pVL1393-*Hoxa-4* DNA into *S. frugiperda* cells, integration of the hybrid polyhedrin-*Hoxa-4* gene into the viral genome was accomplished by double homologous recombination. The procedure used for selecting purification of the recombinant virus were carried out as described by Summers and Smith [1987]. Nuclear extracts were prepared according to the procedure described by Dignam et al. [1983] from subconfluent monolayers of infected cells.

DNA-Binding Assays

Oligodeoxynucleotides were prepared by the Protein Core Facility in The Center for Reproductive Sciences and were purified on an 8 M urea, 12% polyacrylamide gel. Complementary strands were annealed in 40 mM Tris (pH 7.5), 20 mM $MgCl_2$, and 50 mM NaCl, by heating at 65°C for 15 min and then cooling for 15 min each at 45, 37, and 25°C. Double-stranded probes were end-filled with ^{32}P-dCTP plus the other three cold dNTPs, employing T7 DNA polymerase. Free deoxynucleotides were removed by passing through a G-50 Sephadex column. Specific activities of the oligonucleotide products were found to be within 10% of one another, except as noted. For each reaction, 0.5 ng annealed oligonucleotides were used. DNA-protein binding analysis was performed as described previously [Pan et al., 1990]. Each 25 μl reaction contained 10 mM Tris (pH 8.0), 5% glycerol, 50 mM KCl, 1 mM EDTA, 1 mM dithiothreitol, 4 μg of poly(dI-dC) (Pharmacia LKB Biotechnology, Inc.), and 1.5 μl of translation mixture (or 2.5 μg of nuclear extract from infected cells, containing ~0.25 μg Hoxa-4 protein). Following incubation for 20 min at room temperature, 0.5 ng of labeled probe was added and incubation continued for an additional 10 min at room temperature. Reaction samples were applied to a pre-run 0.25XTBE, 10% polyacrylamide gel, and run at 30 V/cm. Dried gels were subjected to autoradiography at −80°C with intensifying screens. Bands corresponding to the bound DNA were excised, placed directly into Ecoscint A scintillation solution (National Diagnostics, Palmetto, FL) and counted.

Transient Co-Transfection Assays

Green monkey kidney (CV-1) cells were plated at a density of 1.5×10^6 per 60 mm plate in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum. One day later, cells were transfected with 2.5 μg/ml each of expression and reporter plasmids by the lipofectin-mediated method, essentially according to the manufacturer's protocol (Gibco BRL). DNA was diluted in 1.0 ml of Opti-MEM I reduced-serum medium before mixing with 1.0 ml of Opti-MEM I medium containing 30 μg of lipofectin™ reagent. Incubation was performed at room temperature for at least 15 min to allow the complex to form. Cells were rinsed twice with 3.0 ml of Opti-MEM I reduced-serum medium and the lipid/DNA complex was added to the dish. Cells were incubated for 5 h at 37°C and the medium was removed and replaced with 5 ml of DMEM containg 10% fetal calf serum. Cells were harvested 42 h after transfection and assayed for CAT activity as previously described [Gorman et al., 1982]. Each assay represents the

percent conversion of ^{14}C-labeled chloramphenicol by one half of the contents of a dish.

RESULTS

Mapping the 5′ End of the Somatic-Type Transcript of the *Hoxa-4* Gene

Previous promoter fusion studies by Galliot et al. [1989] showed that a DNA fragment extending 360 bp upstream from the putative translation start site of the somatic *Hoxa-4* transcript was sufficient to promote transcription in the transfected cell lines. S1 nuclease analysis using total RNAs extracted from embryos and from untransfected or transfected cell lines mapped two transcription start sites, 21 bp and 12 bp upstream from the translation start site [Galliot et al., 1989]. However, a transcript that was 45 nt longer was observed using primer extension analysis on the same RNAs. These discrepancies might have been due to the presence of a short additional upstream exon.

To precisely define the 5′ end of the somatic form of the *Hoxa-4* transcript, we used an in vitro transcription approach to map the start site. A series of Hox-1.4-CAT fusion constructs were generated (Fig. 1) and transcribed in vitro using nuclear extracts of undifferentiated F9 cells which express low levels of *Hoxa-4* RNA (data not shown). An internal control construct containing the adenovirus major late promoter (pAdML-CAT) was included in each reaction. A representative experiment is shown in Figure 1. All the Hox-1.4-CAT constructs examined yielded a major transcript of 116 nucleotides in length. After subtracting the nucleotides corresponding to the CAT sequences, the in vitro transcription experiments map the transcription start site of *Hoxa-4* to 66 bp upstream of the putative translation initiation site. This result was in agreement with the previous primer extension results of Galliot et al. [1989] and thus maps the correct 5′ end of the somatic transcript to 66 nt upstream of the translation start site. The previous localization of the 5′ ends at +46 and +55 by S1 mapping [Galliot et al., 1989] might have resulted from the presence of stretches of A/T residues located at +36 and +46, which could potentially form a hairpin loop in the RNA and prevent S1 digestion from occurring at the correct place.

Sequential 5′ deletion of sequences between positions −1800 to −52 in Hox-1.4-CAT had little effect on the in vitro transcription directed by the *Hoxa-4* promoter. However, the sequences between position −52 and −23 were required for optimal in vitro transcription (Fig. 1, compare lanes 4 and 5). Previous analysis by DNase footprinting of nuclear extracts from F9 cells to the *Hoxa-4* promoter region yielded six binding sites [Galliot et al., 1989]. Four of the six footprint sites are located between positions −275 and −52. One site is between −52 and −23. Another site is located further downstream. The results presented in Figure 1 showed that deletion of the first four footprint sites did not affect transcription in vitro of the *Hoxa-4* promoter. However, when the fifth footprint site (between −52 and −23) was also deleted, leaving a promoter containing only one footprint site, detectable transcription was lost.

Hoxa-4 Protein Binds to Its Promoter and Within Its Intron

As mentioned above, a goal of our studies is to identify potential targets of the protein product of the *Hoxa-4* gene. The autoregulatory function of the *Hoxa-4 Drosophila* homologue *Dfd* suggested examining the *Hoxa-4* gene for evidence for autologous binding sites. We began our analysis using DNA gel retardation assays on sequences ~1 kb upstream of the above defined putative somatic promoter and within the gene itself. The source of protein was in vitro synthesized full-length Hoxa-4 protein. A full-length construct corresponding to the somatic transcript of the *Hoxa-4* gene was generated by fusing a partial cDNA with genomic *Hoxa-4* sequences (see Materials and Methods). The full-length Hoxa-4 protein encoded by this composite construct is 285 amino acids in length, with the homeodomain spanning residues 179–239. The full-length protein made in rabbit reticulocyte lysate was used in gel retardation assays with 12 different DNA fragments, generated by restriction enzyme digestion, ranging from 65 to 345 bp in length, that together span the 1.2 kb of 5′ sequence under investigation as well as the entire 5′ exon and the intron (Fig. 2A). The numbered lanes in Figure 2B correspond to the numbered restriction fragments in Figure 2A. Four of the *Hoxa-4* fragments, 2, 7, 11, and 12, were found to bind to the Hoxa-4 protein (Fig. 2B) and not to control lysate which was not primed with *Hoxa-4* mRNA (data not shown). Two fragments (11 and 12) which exhibited strong Hoxa-4 binding are located within the intron. Another fragment (fragment 7) is located in the putative somatic proximal pro-

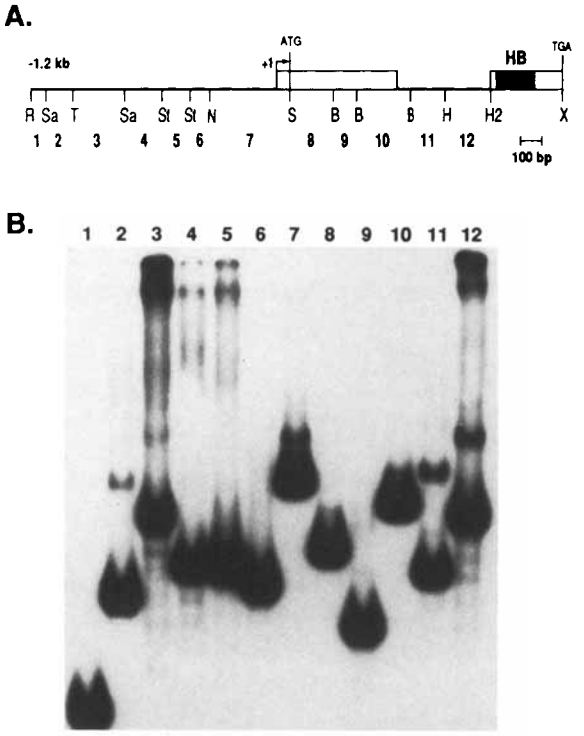

Fig. 2. The Hoxa-4 protein binds to sequences 5′ to its somatic promoter and in its intron. **A**: partial restriction map and summary of the *Hoxa-4* gene structure based on our own data and studies of Galliot et al. [1989]. The numbers below the restriction map indicate DNA fragments used for gel retardation assays. For example, fragment 1 is a 65 bp EcoRV-Sau3AI fragment, while fragment 2 is a 129 bp Sau3AI-BstUI fragment. R, EcoRV; S, SacI; N, NheI B, BssHII; H, HindIII; H_2, HincII; Sa, Sau3AI, St, StuI; T, BstUI; X, XbaI. BssHII, BstUI, and Sau3AI have multiple sites in this region and only the sites employed are shown. **B**: Gel retardation assays. A full length Hoxa-4 protein (corresponding to the *Hoxa-4* somatic-type transcript) made in rabbit reticulocyte lysate was used for gel retardation assays. The in vitro translated Hoxa-4 protein was incubated with each of the indicated DNA fragments after end-labeling with ^{32}P-dNTP. The protein-DNA reaction mixtures were electrophoresed on a 4% nondenaturing polyacrylamide gel, dried and autoradiographed. The lane numbers correspond to the fragments indicated in the restriction map shown in A.

moter, while a fourth fragment (fragment 2), which exhibited slightly weaker binding, is located ~1 kb upstream of the promoter. The small shift observed with fragment 7 is likely due to the fact that the homeodomain binding site is very close to the end of the fragment. No specific binding was observed for the other fragments examined.

Hoxa-4 Protein Binds in a Nucleotide-Specific Manner to Consensus Sequences

The sequence of the portion of the *Hoxa-4* gene used in the gel retardation experiments was determined and the four Hoxa-4 protein-binding fragments were examined for the presence of consensus sequences which have been shown to be target sites for *Drosophila* homeodomain proteins [Regulski et al., 1991; Scott et al., 1989]. Fragments 2 and 11 both contained a CAATTAA-type consensus sequence (Fig. 3), which has been shown to be a target site for binding of several *Drosophila* homeodomain proteins [Desplan et al., 1988; Hayashi and Scott, 1990; Laughon, 1991]. Fragments 7 and 12 contained a TCATTA motif (Fig. 3), which is the consensus target sequence involved in autoregulation of the *Drosophila Dfd* gene [Regulski et al., 1991], the *Drosophila* homologue of *Hoxa-4*.

To investigate the possible binding of Hoxa-4 protein to these consensus sequences, Hoxa-4 protein made in rabbit reticulocyte lysate was used for gel retardation assays with oligonucleotides corresponding to these sequences (listed in Table I). As shown in Figure 4, specific binding of in vitro translated Hoxa-4 protein to each of the four potential homeodomain binding sites (HDBS) found in the *Hoxa-4* gene was observed (Fig. 4, lanes 2, 5, 8, and 11). The control lysate which was not primed with exogenous RNA (Fig. 4, lanes 1, 4, 7, and 10) did not exhibit specific binding, although nonspecific, high molecular weight complexes appeared in both the control lysate as well as in the presence of Hoxa-4 protein. Specific binding of Hoxa-4 protein was completely eliminated in the presence of competing unlabeled oligonucleotides (Fig. 4, compare lanes 2 versus 3, 5 versus 6, 8 versus 9, and 11 versus 12). It should be noted that the concentration of Hoxa-4 protein in the translation lysate was low, representing $<0.05\%$ of the total protein as estimated by Coomassie staining. Nevertheless, Hoxa-4 protein was specifically bound to these consensus sequences, indicating a high affinity of Hoxa-4 protein for these particular sequences.

The specificity of the Hoxa-4 protein-DNA binding was further investigated using Hoxa-4 protein produced from the Baculovirus expression system. As seen in Figure 5, incubation with nuclear extracts from cells producing Hoxa-4 protein resulted in the formation of complexes, which are distinct from the free oligonucleotides (Fig. 5, lanes 2, 5, 8, and 11). The formation of Hoxa-4-associated complexes was blocked in competition experiments containing unlabeled oligonucleotides (Fig. 5, lanes 2 versus 3, lanes 5 versus 6, lanes 8 versus 9, and 11 versus 12). The binding of Hoxa-4 protein appeared to be strongest to HDBS2, intermediate

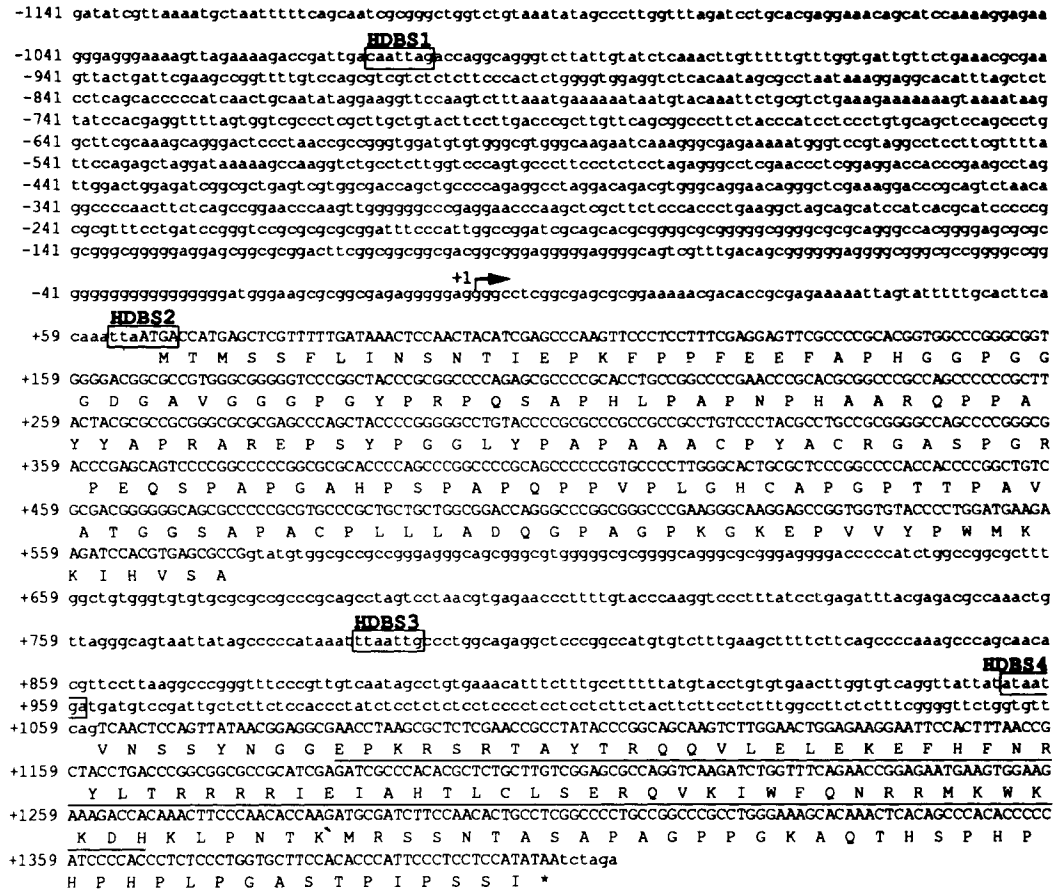

Fig. 3. Nucleotide and amino acid sequences of the murine *Hoxa-4* gene. Nucleotides are numbered on the left. The coding sequences are in capital letters and sequences corresponding to upstream, intronic and trailer regions are in small letters. The 5′ end of transcription was determined by in vitro transcription and is indicated by an arrow. The homeobox is underlined. The putative homeodomain binding sites (HDBS) are boxed.

to HDBS3 and HDBS4, and least efficient to HDBS1. This apparent difference in binding efficiency was also observed in the experiments using the in vitro translated Hoxa-4 protein (see Fig. 4). Additional semi-quantitative data supporting this observation are presented in the mutational analyses in Figure 6 and in Table I.

Interaction of Hoxa-4 Homeodomain With Its Putative Target Sequence

To understand the level of specificity of the Hoxa-4 protein interaction with its target sequences, a series of mutant target sequence oligonucleotides were synthesized, and their ability to interact with Hoxa-4 protein in gel retardation experiments was evaluated. Table I, part A, shows the sequence of the wild type and mutant versions of the CAATTAA-class consensus binding sequences. As observed in Figure 6A, the sequences CTATTAAA and CAATTAAT were efficiently recognized by the Baculovirus-produced protein (Fig. 6A, lanes 3 and 9). This suggested that bases at positions 2 and 8 were not important for protein-DNA binding, at least in vitro. However, base substitutions at positions 6 and 7 reduced the binding efficiencies approximately 2-fold. Base changes of C to G at position 1 or of A to T or T to A at positions 3, 4, and 5 more severely reduced binding (Fig. 6A, lanes 2, 4, 5, and 6), suggesting that the specific bases at these positions were indeed important for binding. With respect to the binding of the variants of the HDBS4 sequence (Table I, part B and Fig. 6B, lane 2), a single base alteration in the left side ATTA core motif greatly reduced the binding efficiency. On the other hand, base changes in the right side ATTA as well as in between these two palindromically arranged ATTA motifs had little effect on the binding (Fig. 6B, lanes 3 and 4).

TABLE I. Lists of Oligo Sequences Used for Gel Retardation Assays*

											Relative Binding $\times 10^3$ cpm
A.			1	2	3	4	5	6	7	8	
HDBS3	caggg	C	A	A	T	T	A	A	A	tttag	12.2
3G1	caggg	G	A	A	T	T	A	A	A	tttag	3.6
3T2	caggg	C	T	A	T	T	A	A	A	tttag	10.4
3T3	caggg	C	A	T	T	T	A	A	A	tttag	3.5
3A4	caggg	C	A	A	A	T	A	A	A	tttag	1.2
3A5	caggg	C	A	A	T	A	A	A	A	tttag	3.2
3T6	caggg	C	A	A	T	T	T	A	A	tttag	5.0
HDBS1	ttga	C	A	A	T	T	A	G	A	ccagg	6.2
3T8	caggg	C	A	A	T	T	A	A	T	tttag	12.5
B.											
HDBS4	aca	T	C	A	T	T	A	T	A	A	T acc 11.2
4C	aca	T	C	A	T	C	A	T	A	A	T aac 2.5
4CG	aca	T	C	A	T	T	A	C	G	A	T aac 7.8
4G	aca	T	C	A	T	T	A	T	A	G	A T aac 7.5
C.			1	2	3	4	5	6	7	8	
HDBS3	caggg	C	A	A	T	T	A	A	A	tttag	12.3
3A1	caggg	A	A	A	T	T	A	A	A	tttag	4.0
3G1	caggg	G	A	A	T	T	A	A	A	tttag	3.6
3T1	caggg	T	A	A	T	T	A	A	A	tttag	16.1
3C2	caggg	C	C	A	T	T	A	A	A	tttag	7.5
3G2	caggg	C	G	A	T	T	A	A	A	tttag	5.2
3T2	caggg	C	T	A	T	T	A	A	A	tttag	10.5
HDBS2	catgg	T	C	A	T	T	A	A	T	ttgtg	16.0
3TG	caggg	T	G	A	T	T	A	A	A	tttag	7.1
3TT	caggg	T	T	A	T	T	A	A	A	tttag	4.5
HDBS4	aca	T	C	A	T	T	A	T	A	taataac	11.5

*The altered bases are underlined. In addition to the sequences shown, oligonucleotides contained SalI site overhangs for end-filled labeling (see Methods). The relative binding represents an arbitrary assessment based on counts of radioactivity in the bound DNA. Relative specific activities of the oligos were within 10% of one another in these assessments.

To reveal the intrinsic relationship between these two types of consensus sequences (CAATTA and TCATTA), we carried out systematic base substitution studies at position 1 and 2. The results are presented in Table I, part C and Figure 6C. Position 1 showed a very strong preference for cytosine or thymine, while guanine and adenine at position 1 reduced binding approximately 3-fold. In the presence of cytosine at position 1, either adenine or thymine was adequate at position 2, while cytosine and guanine reduced binding approximately 2-fold. In the presence of thymine at position 1, adenine or thymine is preferred at position 2, with guanine and cytosine reducing binding ~2- and 3-fold, respectively.

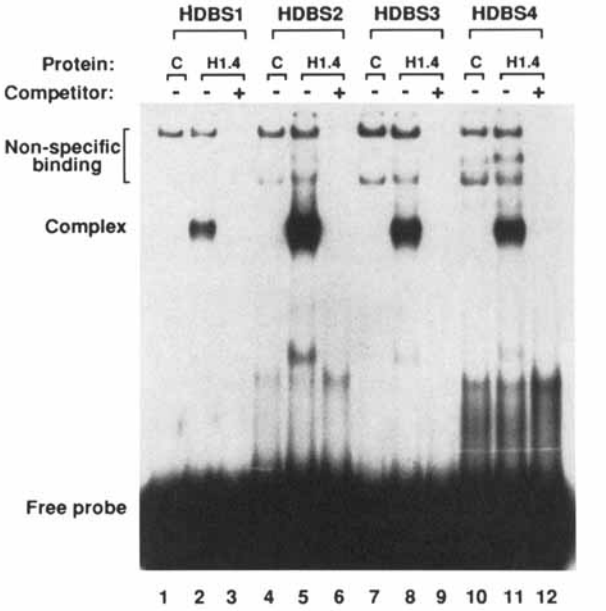

Fig. 4. Binding of HDBS with in vitro translated Hoxa-4 protein. Gel retardation assays were carried out with in vitro translated protein and 22-base pair oligodeoxynucleotide probes corresponding to HDBS1, HDBS2, HDBS3, and HDBS4 (summarized in Table I). The lanes indicated (+) contained a 500-fold molar excess of unlabeled oligodeoxynucleotide probes in the binding reaction. The reactions in lanes 1, 4, 7, and 10 were incubated with rabbit reticulocyte lysate without the addition of exogenous RNA and served as control.

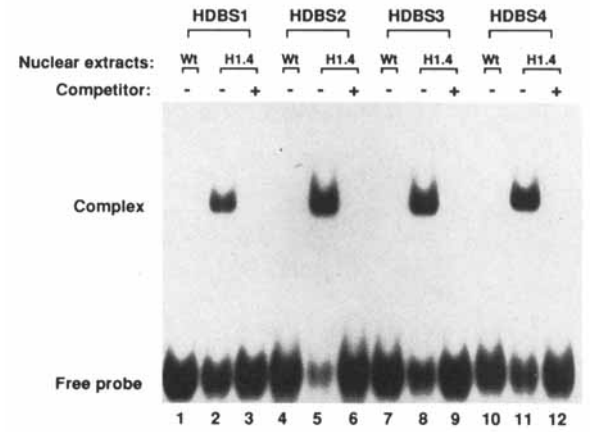

Fig. 5. Binding of HDBS with Hoxa-4 protein produced from baculovirus expression system. Analyses were performed as in source of Figure 4, except that the Hoxa-4 protein was derived from nuclear extracts prepared from both *Hoxa-4* recombinant baculovirus infected insect cells and wild-type virus infected insect cells.

Intact Homeodomain Is Essential for the Specific In Vitro Binding Observed

To identify the portion of the Hoxa-4 protein important for DNA binding, mutant forms of the protein were generated (Fig. 7A,B) and tested for their ability to form complexes with two

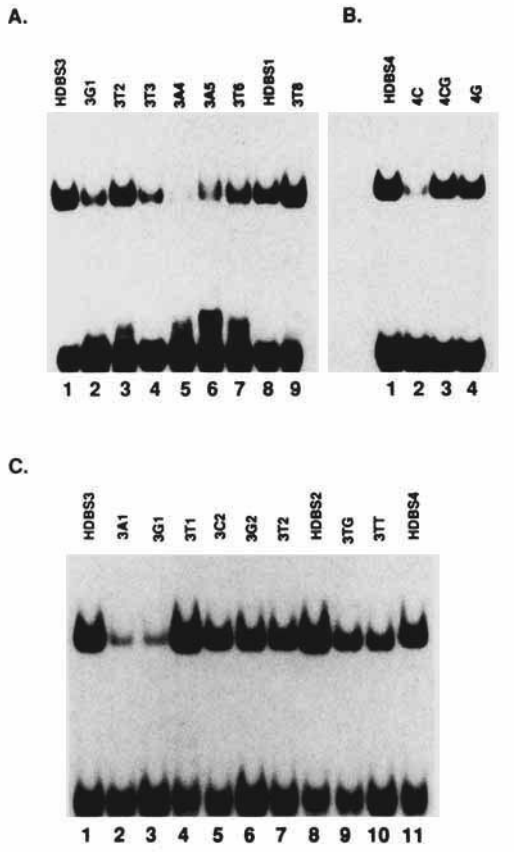

Fig. 6. Recognition of mutant oligodeoxynucleotides corresponding to the HDBS by baculovirus produced Hoxa-4 protein. Analyses were performed as described for Figure 4, except that mutant oligodeoxynucleotide probes (summarized in Table I) were employed. The source of Hoxa-4 protein was nuclear extracts of the Hoxa-4 recombinant baculovirus infected insect cells. Relative binding was assessed by radioactivity in excised bands.

putative target sequences in gel retardation assays (Fig. 7C). The four proteins described in Figure 7A were synthesized in rabbit reticulocyte lysate in the presence of ^{35}S-methionine and analyzed by SDS-PAGE (Fig. 7B). Each of the Hoxa-4 proteins was translated in vitro with the same relative efficiency and the sizes of the products were approximately consistent with those deduced from the predicted amino acid sequences present in the constructs. Both the full length protein (Hox-1.4) and the N-terminal truncated form (Hox-1.4H, which contained the intact homeodomain) showed comparable DNA-binding activities to two putative target sequences (HDBS1 and HDBS4) (Fig. 7C). Hox-1.4E is a prematurely terminated version which lacks amino acid residues 200 to 285. Hox-1.4B contains a frame-shift mutation at amino acid 226 in the homeodomain and substitutes 62 new amino acids at its carboxy terminal in place of amino acid residues 227 to 285. Both these mutant proteins were deficient in DNA-binding activity. These results suggested that the C-terminal portion, which included the homeodomain of the Hoxa-4 protein, mediated DNA binding, and that an intact homeodomain was essential.

Binding of Hoxa-4 Protein In Vivo

To further extend the in vitro binding results, the mutant Hoxa-4 proteins were examined for their ability to bind putative target sequences in tissue culture cells. Transient co-transfection experiments were performed with a reporter plasmid which contained six homeodomain binding sites of the HDBS3 class upstream of a reporter gene (−109 TK-CAT; see Materials and Methods). Expression plasmids capable of driving the production of each of the four Hoxa-4 proteins (normal and mutant versions) were co-transfected with the reporter construct in CV-1 cells. All the Hoxa-4 proteins expressed in CV-1 cells exhibited little or no effect on the induction of expression of the reporter construct (Fig. 8A). A similar lack of activation was observed in transient co-transfections in other cell lines, including F9, COS-1, and HeLa cells (data not shown).

To determine whether lack of transcriptional activation was due to a lack of in vivo binding, we took advantage of the fact that the *Drosophila* FTZ protein has been shown to bind efficiently to the same consensus sequences as that found in HDBS3 [Desplan et al., 1988]. In vitro binding studies showed that FTZ protein bound to the HDBS3 site as efficiently as did Hoxa-4 protein (data not shown). Furthermore, the FTZ protein has been shown to function as a transcriptional activator [Jaynes and O'Farrell, 1988] and was confirmed to function as an activator in our system as well (Fig. 8A). Competition experiments were then performed in which FTZ and Hoxa-4-producing constructs were co-transfected with the reporter construct to test if Hoxa-4 could compete for the binding site recognized by FTZ and therefore affect the activation of the reporter construct by the FTZ protein. Indeed, activation of the reporter gene by FTZ was dramatically reduced by the presence of the intact Hoxa-4 protein and the N-terminal mutant which contains the intact homeodomain (Fig. 8B). This effect of the Hoxa-4 protein was entirely dependent upon the presence of an intact homeodomain, since mutant forms of the Hoxa-4 protein lacking the intact homeodomain

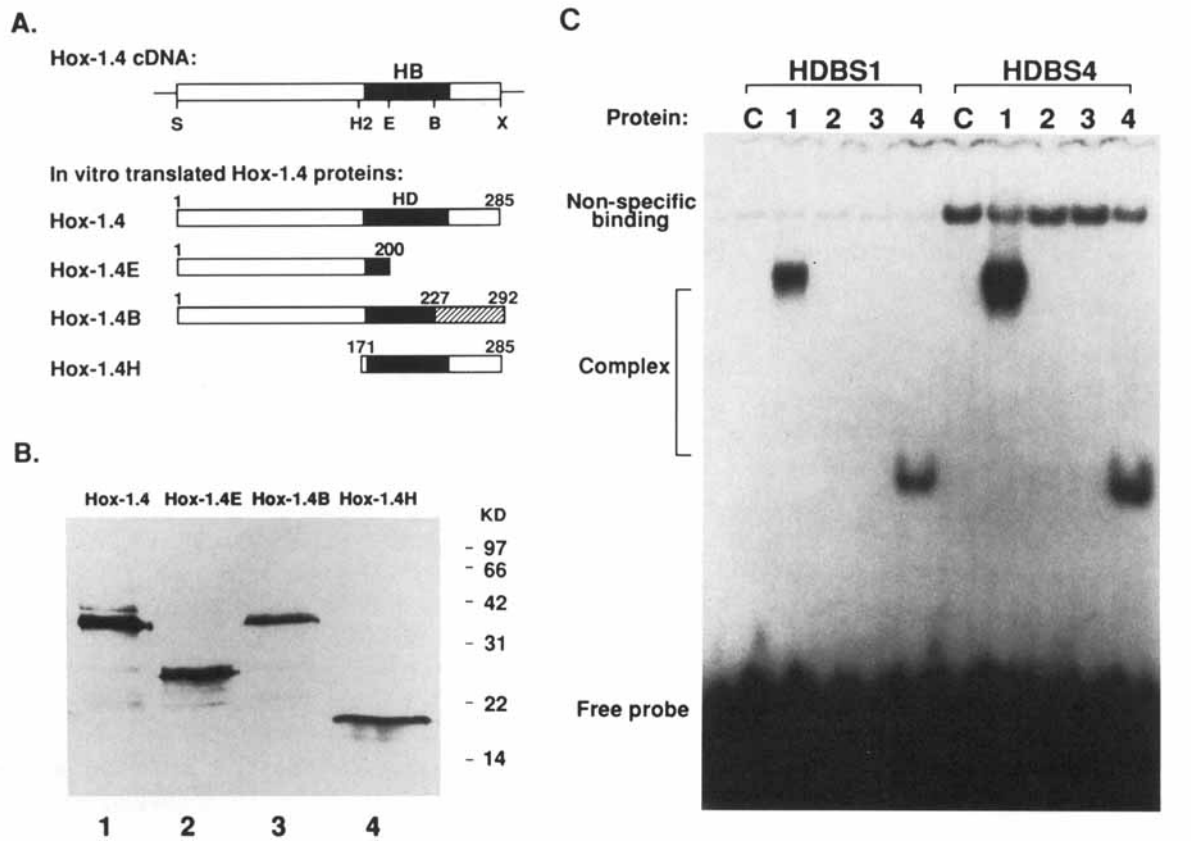

Fig. 7. Mapping the DNA binding domain of Hoxa-4 protein. **A**: In vitro translated Hoxa-4 protein and its mutant derivatives. Hoxa-4 protein and its mutant derivatives were in vitro translated from in vitro synthesized mRNAs using rabbit reticulocyte lysate. Hox-1.4 represents the full-length Hoxa-4 protein translated in vitro. Hox-1.4E represents a truncated protein produced by inserting a nonsense mutation at the EcoRI site in the homeobox. Hox-1.4B represents a translation frame shift mutant of Hoxa-4 protein at the BglII site in the helix 3 in the homeobox. Hox-1.4H is a truncated form which contains a protein domain encoded by the 3′ exon. S, SacI; H2, HincII; E, EcoRI; B, BglII; X, XbaI. **B**: SDS gel analysis of in vitro synthesized Hoxa-4 proteins. ^{35}S methionine-labeled Hoxa-4 proteins synthesized in vitro with rabbit reticulocyte lysate were separated on a 15% SDS polyacrylamide gel, dried, and exposed to X-ray film. The sizes of in vitro synthesized Hoxa-4 proteins are consistent with the sizes deduced from the amino acid sequences, except for the truncated protein Hox-1.4H, which is rich in basic amino acids and runs slower on the SDS gel. **C**: Binding of HDBS with in vitro synthesized Hox-1.4E protein and its mutant derivatives. Analyses were performed as described for Figure 4, except that in vitro synthesized Hoxa-4 protein and its mutant derivatives were employed. Lanes marked C (control) received only rabbit reticulocyte lysate. Lane 1, Hox-1.4 protein; lane 2, Hox-1.4B protein; lane 3, Hox-1.4E protein; lane 4, Hox-1.4H protein.

failed to suppress FTZ-mediated activation of the reporter gene (Fig. 8B).

DISCUSSION

In this study, we have shown that the protein product of the somatic form of the *Hoxa-4* gene is a sequence-specific DNA binding protein that recognizes sequences containing the core ATTA motif characteristic of the consensus binding sequences of *Antp*-like class homeodomain proteins [Hayashi and Scott, 1990; Odenwald et al., 1989]. Two of the sequences identified are in the promoter region of *Hoxa-4*, while the other two are located in the *Hoxa-4* intron. The integrity of the Hoxa-4 homeodomain was shown to be essential for the binding to the putative target sequences, in both in vitro binding assays and tissue culture cells.

Co-transfection experiments revealed that the Hoxa-4 protein was unable to transactivate transcription of a reporter plasmid containing a multimerized *Hoxa-4* consensus binding site. This failure to activate transcription was not due to a lack of in vivo binding since the Hoxa-4 protein was able to compete for the same binding site and repress FTZ-activated transcription in the same system. This suppression of FTZ-mediated activation appeared to be dependent on the binding of the Hoxa-4 product to the Hoxa-4 consensus binding site. A Hoxa-4 N-terminal mutant

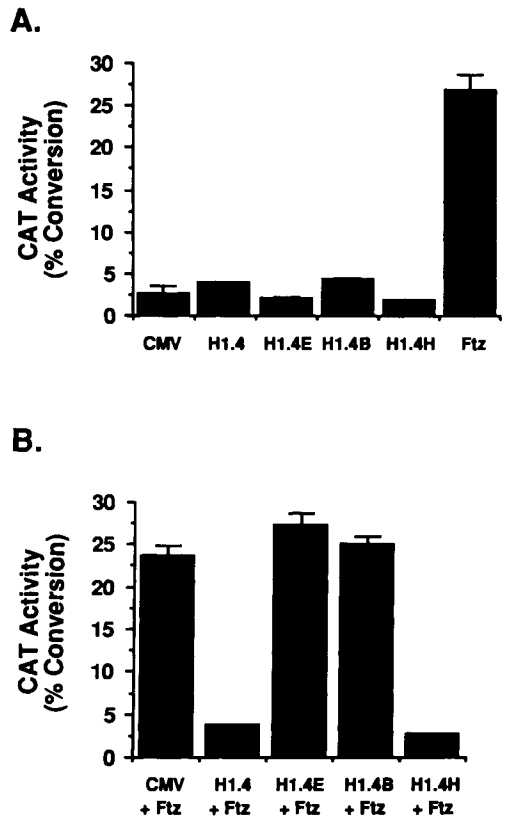

Fig. 8. Hoxa-4 homeodomain binding can compete to suppress activation by FTZ homeoprotein. A reporter construct bearing six copies of *Hoxa-4* HDBS3 inserted in front of tkCAT was co-transfected with CMV-Ftz and either CMV-Hox-1.4 or constructs yielding mutant versions of the Hoxa-4 protein (**A**). The influence of the indicated homeoprotein on the ability of the FTZ-producing plasmid to induce CAT activity in the reporter plasmid was assessed by co-transfection (**B**). Equivalent results were obtained in at least two separate experiments. The bars indicate the average percent CAT conversion in duplicate transfections within the same experiment.

protein with an intact homeodomain showed the same efficiency of suppression as compared with the full-length protein, whereas mutant forms of the Hoxa-4 protein lacking the intact homeodomain failed to suppress FTZ-mediated activation of the reporter gene. These results also suggested that it was unlikely that expression of Hoxa-4 protein might have activated a downstream *Hox* gene which then repressed FTZ activation function in the co-transfection assays. The failure of Hoxa-4 to transactivate may be due to a lack of cofactors which act combinatorially. For example, it has been suggested that other nuclear factors may be critical for homeobox protein function, such as in the autoregulation of the *even-skipped* homeobox gene [Jiang et al., 1991]. Moreover, the product of the chicken homolog of *Hoxa-4* can function as a transcriptional activator in a cell type specific manner [Sasaki et al., 1992]. If Hoxa-4 acts in concert with other factors bound to neighboring sites to function, the multimerized artificial Hoxa-4 consensus binding site used in this study may not contain a proper binding sites for both the Hoxa-4 protein and potentially interacting cofactors. Alternatively, such interacting cofactors may be absent in the cells employed.

Mutational analysis has suggested that the FTZ and UBX homeodomains recognize base sequence at each position of a consensus sequence, NNATTA [Ekker et al., 1991; Florence et al., 1991; Percival-Smith et al., 1990]. Flanking sequences were weakly discriminated by the FTZ homeodomain. Bases at position 4 and 5 are contacted in the major groove of the DNA molecule by helix 3 of the homeodomain in the three-dimensional structures of the ANTP and EN homeodomain complexes [Kissinger et al., 1990; Otting et al., 1990]. The bases at positions 1 and 2 have been shown to interact with amino acid residue 50 in a homeodomain and to be the determinant factor for specificities of FTZ and BICOID homeodomain binding to DNA [Hanes and Brent, 1989, 1991]. Our mutational analyses provided further support for the direct role of the bases at positions 1, 4, and 5 in protein-DNA interaction.

Our results with the second set of mutant consensus binding sequences (which are variants of putative target sequence HDBS4) were interesting with respect to the arrangement of the core ATTA sequence. The putative target sequence HDBS4 has two ATTA boxes arranged in a palindromic orientation. A single base substitution in the left side ATTA box significantly reduced or abolished binding (Fig. 6B, lane 4C), while single base substitutions in the right side ATTA box only slightly affected binding efficiency (as did the base changes between the two ATTA boxes). This indicated that the Hoxa-4 protein interacted in a selective, sequence-specific manner with the TCATTA sequence and did not bind indiscriminantly to the short sequence ATTA. Thus, sequences surrounding the ATTA core play a role in Hoxa-4-DNA binding. Finally, base substitution analysis suggested that either cytosine or thymine was optimal at position 1 (as outlined in the putative consensus sequence CAATTAA), while adenine or guanine reduced binding ~ 3-fold. Base substitution with a thymine at position 2 had little effect on bind-

ing, while base changes with a cytosine or guanine reduced binding 1–2-fold. Our data were consistent with the results from studies on FTZ homeodomain-DNA interactions [Florence et al., 1991].

In studies of the ANTP and EN homeodomain-DNA complexes, the side chain of Gln-50 (which is also conserved in the Hoxa-4 protein) was shown to make hydrophobic contact with the nucleotide at position 1 [Kissinger et al., 1990; Otting et al., 1990]. A thymine at position 1 would change the position of the amide and gamma CH2 moieties of Gln-50, since the C-7 methyl group of thymine protrudes further into the major groove. Our data suggest that such reposition with thymine at position 1 permitted the presence of an adenine or a cytosine at position 2. In the presence of a thymine at position 1, either guanine or thymine at position 2 resulted in reduced affinity of binding. We predict that the consensus core sequences for Hoxa-4 binding are CA/TATTA and TA/CATTA.

Both autoregulation and cross-regulation have emerged as apparent modes by which *Drosophila Antp*-class homeodomain proteins regulate transcription of target genes. For example, UBX protein is positively autoregulated in the visceral mesoderm of *Drosophila* embryos [Bienz and Tremml, 1988], while it negatively regulates *Antp* in the central nervous system [Carroll et al., 1986; Hafen et al., 1984]. The sequences sufficient for UBX-mediated autoregulation in the visceral mesoderm appear to involve both a sequence upstream of the promoter and a homeodomain binding sequence downstream of the *Ubx* transcription start site [Müller et al., 1989]. UBX protein is a dual functional transcriptional factor and can repress an *Antp* P1 promoter in co-transfection experiments [Krasnow et al., 1989]. Similar homeodomain binding sites downstream of the transcription start sites of the *Antp* gene [Beachy et al., 1988] were required for *Ubx*-mediated repression. The strongest binding site (HDBS2) for the Hoxa-4 protein that we have observed in this study is also located immediately downstream of the transcription site. Moreover, this sequence matches the TCATTA sites present in the *Drosophila Dfd* gene promoter, where they are required for the maintenance of *Dfd* expression in vivo through an autoregulatory pathway [Regulski et al., 1991]. Furthermore, the chicken Hoxa-4 protein has been shown to bind a site which is identical to the HDBS2 identified in this study [Sasaki et al., 1990]. Interestingly, the DNA sequences surrounding this site are identical in the mouse and chicken genes. This high level of cross-species sequence conservation may indicate a functional conservation as well.

Recent studies on the regulatory interactions in cultured cells between genes in the human *HOX3D* and *HOX4C* loci revealed that similar regulatory networks likely exist in vertebrates as well [Arcioni et al., 1992; Zappavigna et al., 1992]. The *HOX3D* promoter was shown to be transactivated by the protein products of the upstream *HOX3C* gene and the paralogues of the genes further upstream in the *HOX4* cluster. Similarly, a 90 bp DNA fragment located downstream of the *HOX4C* major transcription start site was shown to mediate transactivation by HOX4C and the HOX4D proteins. This region exhibited a high level of cross-species sequence conservation. The transactivation was further shown to require the presence of the ATTA-related binding sites. Activation of a reporter gene by *Hoxd-4* through a 217 bp *Hoxd-4* upstream sequence in transient transtection was dependent upon two ATTA-related binding sites which are present in this 217 bp fragment [Pöpperl and Featherstone, 1992]. Thus, the homeodomain binding sites we identified in the present study could also have potential autoregulatory and cross-regulatory capacities.

The presence of the putative target sites within the intron may be significant in light of the observation that the appropriate expression of the *Hoxa-7* and *Hoxb-4* genes apparently required their introns [Püschel et al., 1991; Whiting et al., 1991]. Expression of *Hoxb-4* in the majority of the neural and mesodermal domains required regions which included the intron and flanking sequences. Similarly, an element which contained the intron and parts of the two flanking exons was required to specify *Hoxa-7* expression to prevertebrae. Further, experiments in which the *Hoxa-4* gene was expressed in the appropriate tissues in transgenic mice involved constructs containing the intron [Wolgemuth et al., 1989; Behringer et al., 1993]. The observation that the Hoxa-4 protein is capable of binding (in vitro and most probably in tissue culture cells) to the putative target sequences raises the possibility that Hoxa-4 may bind to these sites in vivo and thus have an autoregulatory function. It will be possible to establish whether autoregulation plays a role in *Hoxa-4* expression by studying the expression in mice in which the endogenous *Hoxa-4* gene is mutated by homologous recombination and by defining the *Hoxa-4*

binding sites required for proper regulation. Such experiments will shed light on the degree to which regulatory mechanisms of homologous homeobox genes between vertebrates and invertebrates are conserved.

ACKNOWLEDGMENTS

We thank Dr. Walter Gehring and Dr. Markus Affolter for comments and suggestions during the course of these experiments, Dr. Axel Wadewitz for preparing Baculovirus nuclear extracts, Dr. Gary Struhl for providing plasmid pF450, Dr. Martin Winer and Dr. Stuart Ravnik for comments on the manuscript, and Jacqueline Zengotita for help in preparing the manuscript. This work was supported in part by ACS grant CD-437 and grants from the NIH, HD18122 and 28374.

REFERENCES

Arcioni L, Simeone A, Guazzi S, Zappavigna V, Boncinelli E, Mavilio F (1992) The upstream region of the human homeobox gene HOX3D is a target for regulation by retinoic acid and HOX homeoproteins EMBO J 11 265–277

Beachy PA, Krasnow MA, Gavis ER, Hogness DS (1988) An Ultrabithorax protein binds sequences near its own and Antennepedia P1 promoters Cell 55 1069–1081

Behringer RR, Crotty DA, Tennyson VM, Brinster RL, Palmiter RD, Wolgemuth DJ (1993) Sequences 5 of the homeobox of the *Hox*-1 4 gene direct tissue-specific expression of *lac Z* during mouse development Development 117 823–833

Bergson C, McGinnis W (1990) The autoregulatory enhancer element of the *Drosophila* homeotic gene deformed EMBO J 9 4287–4297

Bienz M, Tremml G (1988) Domain of Ultrabithorax expression in *Drosophila* visceral mesoderm from autoregulation and exclusion Nature 333 576–578

Bodner M, Karin M (1987) A pituitary-specific trans-acting factor can stimulate transcription from the growth hormone promoter in extracts of nonexpressing cells Cell 50 267–275

Carroll SB, Layman RA, McCutcheon MA, Riley PD, Scott MP (1986) The localization and regulation of Antennepedia protein expression in *Drosophila* embryos Cell 47 113–122

Cho KWY, Goetz J, Wright CVE, Fritz AF, Hardwicke J, De Robertis EM (1988) Differential utilization of the same reading frame in a *Xenopus* homeobox gene encodes two related proteins sharing the same DNA-binding specificity EMBO J 7 2139–2149

Desplan CJ, Theis Jk, O'Farrell PH (1988) The sequence specificity of homeodomain-DNA interaction Cell 54 1081–1090

Dignam JD, Lebowitz RM, Roeder RG (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei Nucleic Acids Res 11 1475–1478

Duboule D, Baron A, Mahl P, Gallhot B (1986) A new homobox is present in overlapping cosmid clones which define the mouse Hox-1 locus EMBO J 5 1973– 1980

Duboule D, Dolle P (1989) The structural and functional organization of the murine Hox gene family resembles that of *Drosophila* homeotic genes EMBO J 8 1497–1505

Duboule D, Boncinelli E, DeRobertis E, Featherstone M, Lonai P, Oliver G, Ruddle FH (1990) An update of mouse and human Hox genes nomenclature Genomics 7 458–459

Ekker SC, Young KE, Kessler DP, Beachy PA (1991) Optimal DNA sequence recognition by the Ultrabithorax homeodomain of *Drosophila* EMBO J 10 1179–1186

Florence B, Handrow R, Laughon A (1991) DNA-binding specificity of the fushi tarazu homeodomain Mol Cell Biol 11 3613–3623

Gallhot B, Dolle P, Vigeron M, Featherstone MS, Baron A, Duboule D (1989) The mouse Hoxa-4 gene Primary structure, evidence for promoter activity and expression during development Development 107 343–359

Gaunt SJ, Sharpe PT, Duboule D (1988) Spatially restricted domains of homeobox gene transcripts in mouse embryos Relation to a segmented body plan Development 104 71–82

Gehring WJ (1987) Homeobox genes in the study of development Science 236 245–1252

Gorman CM, Moffat LF, Howard BH (1982) Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells Mol Cell Biol 2 1044–1051

Graham A, Papalopulu N, Krumlauf R (1989) The murine and Drosophila homeobox gene complexes have common features of organization and expression Cell 57 367–378

Hafen E, Levine M, Gehring WJ (1984) Regulation of Antennapedia transcript distribution by the bithorax complex in Drosophila Nature 307 287–289

Halazonitis TD, Georgopoulos K, Greenberg ME, Leder P (1988) c-Jun dimerizes with itself and with c-Fos, forming complexes of different DNA binding affinities Cell 55 917–924

Hanes SD, Brent R (1989) DNA specifity of the bicoid activator protein is determined by homeodomain recognition helix residue 9 Cell 57 1275–1283

Hanes SD, Brent R (1991) A genetic model for interaction of the homeodomain recognition helix with DNA Science 251 426–430

Hayashi S, Scott MP (1990) What determines the specificity of action of Drosophila segmentation gene fushi terazu Cell 63 883–844

Hoey T, Levine M (1988) Divergent homeobox proteins recognize similar DNA sequences in Drosophila Nature 332 858–861

Holland PWH, Hogan BLM (1988) Expression of homeobox genes during mouse development A review Genes Dev 2 773–782

Jaynes JB, O'Farrell PH (1988) Activation and repression of transcription by homeodomain-containing proteins that bind a common site Nature 336 744–749

Jiang J, Hoey T, Levine M (1991) Autoregulation of a segmentation gene in *Drosophila* Combinatorial interaction of the even-skipped homeobox protein with a distal enhancer element Genes Dev 5 265–277

Kessel M, Gruss P (1990) Murine developmental control genes Science 249 374–379

Kissinger CR, Liu B, Martinn-Blanco E, Kornberg TB, Pabo CO (1990) Crystal structure of an engrailed homeodomain DNA complex at 2 8 A resolution A frame work for understanding homeodomain-DNA interactions Cell 63 579–590

Krasnow MA, Saffman EE, Kornfeld K, Hogness DS (1989) Transcriptional activation and repression by Ultrabithorax proteins in cultured *Drosophila* cells Cell 57 1031–1043

Kuziora MA, McGinnis W (1988) Autoregulation of a *Drosophila* homeotic selector gene Cell 55 477–485

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4 Nature 227 680–685

Laimins LL, Khoury G, Gorman C, Howard B, Gruss P (1982) Host-specific activation of transcription by tandem repeats from simian virus 40 and moloney murine sarcoma virus Proc Natl Acad Sci USA 79 6453–6457

Laughon A (1991) DNA binding specificity of homeodomains Biochemistry 30 11367–11367

Laughon A, Scott MP (1984) Sequence of a *Drosophila* segmentation gene Protein structure homology with DNA-binding proteins Nature 310 25–31

Lufkin T, Jackson AE, Pan WT, Bancroft C (1989) Proximal rat prolactin promoter sequences direct optimal, pituitary cell-specific transcription Mol Endocrinol 3 559–566

Malicki J, Schughart K, McGinnis W (1990) Mouse Hox-2 2 specifies thoracic segmental identity in *Drosophila* embryos and larvae Cell 63 961–967

Martin GR, Boncinelli E, Duboule D, Gruss P, Jackson I, Krumlauf R, Lonai P, McGinnis W, Ruddle F, Wolgemuth D (1987) Nomenclature for homeobox-containing genes Nature 325 21–22

McGinnis N, Kuziora MA, McGinnis W (1990) Human Hox-4 2 and *Drosophila* Deformed encode similar regulatory specificities in *Drosophila* embryos and larvae Cell 63 969–975

McGinnis W, Levine ML, Hafen E, Kuroiwa A, Gehring WJ (1984) A conserved DNA sequence in homeotic genes of the *Drosophila melanogaster* Antennapedia and Bithorax complexes Nature 308 428–433

McKnight SL (1982) Functional relationships between transcriptional control signals of the thymidine kinase gene of herpes simplex virus Cell 31 355–365

Muller J, Thuringer F, Biggin M, Zust B, Bienz M (1989) Coordinate action of a proximal homeoprotein binding site and distal sequence confers the Ultrabithorax expression pattern in the visceral mesoderm EMBO J 8 4143–4151

Odenwald WF, Garbern J, Arnheiter H, Tournier-Lasserve E, Lazzarini RA (1989) The Hox-1 3 homeobox protein is a sequence specific DNA-binding phosphoprotein Genes Dev 3 158–172

Otting G, Qian YQ, Billeter M, Muller M, Affolter M, Gehring W, Wuthrich K (1990) Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution EMBO J 9 3085–3092

Pan WT, Liu Q, Bancroft C (1990) Identification of a growth hormone gene promoter repressor element and its cognate double- and single-stranded DNA-binding proteins J Biol Chem 265 7022–7028

Percival-Smith A, Muller M, Affolter M, Gehring WJ (1990) The interaction with DNA of wild-type and mutant fushi tarazu homeodomains EMBO J 9 3967–3974

Popperl H, Featherstone MS (1992) An autoregulatory element of the murine Hox-4 2 gene EMBO J 11 3673–3680

Puschel AW, Balling R, Gruss P (1991) Separate elements cause lineage restriction and specify boundaries of Hox-1 1 expression Development 112 279–287

Qian YQ, Billeter M, Otting G, Muller M, Gehring WJ, Wuthrich K (1989) The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution comparison with prokaryotic repressors Cell 59 573–580

Regulski M, Dessain S, McGinnis N, McGinnis W (1991) High-affinity binding sites for the Deformed protein are required for the function of an autoregulatory enhancer of the Deformed gene Genes Dev 5 278–286

Rubin MR, Toth LE, Patel MD, D'Eustachio P, Nguyen-Huu MC (1986) A mouse homeo box gene is expressed in spermatocytes and embryos Science 233 663–667

Sambrook J, Fritsch EF, Maniatis T (1989) "Molecular Cloning A Laboratory Manual," 2nd Ed Cold Spring Harbor, NY Cold Spring Harbor Press

Sasaki H, Yokoyama E, Kuroiwa A (1990) Specific DNA binding of the two chicken Deformed family homeodomain proteins, CHoxa-4 and Chox-a Nucleic Acids Res 8 1739–1747

Sasaki H, Yamamoto M, Kuroiwa A (1992) Cell type dependent transcription regulation by chick homeodomain proteins Mechanisms Dev 37 25–36

Scott MP (1992) Vertebrate homeobox gene nomenclature Cell 71 551–553

Scott MP, Tamkun J, Hartzell GW (1989) The structure and function of the homeodomain Biochim Biophys Acta Rev Cancer 989 25–48

Scott MP, Weiner AJ (1984) Structural relationships among genes that control development Sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila Proc Natl Acad Sci USA 81 4115–4119

Shepherd JC, McGinnis W, Carrasco AE, De Robertis EM, Gehring WJ (1984) Fly and frog homeodomains show homologies with yeast mating type regulatory proteins Nature 310 70–71

Summers TD, Smith GE (1987) "A Manual of Methods for Baculovirus Vector and Insect Cell Culture Procedures" Tex Agric Exp Stn Bull , No 1555

Whiting J, Marshall H, Cook M, Krumlauf R, Rigby PWJ, Stott D, Allemann RK (1991) Multiple spatially-specific enhancer are required to reconstruct the pattern of Hoxb-4 gene expression Genes Dev 5 2048–2059

Wolgemuth DJ, Engelmyer E, Duggal RN, Gizang-Ginsberg E, Mutter GL, Ponzetto C, Viviano CM, Zakeri ZF (1986) Isolation of a mouse cDNA coding for a developmentally regulated, testis-specific transcript containing homeo box homology EMBO J 5 1229–1235

Wolgemuth DJ, Viviano CM, Gizang-Ginsberg E, Frohman M, Joyner A, Martin GR (1987) Differential expression of the homeo box-containing gene Hox-1 4 during mouse male germ cell differentiation and embryonic development Proc Natl Acad Sci 84 5813–5817

Wolgemuth DJ, Behringer R, Mostoller M, Brinster RL, Palmiter RD (1989) Transgenic mice overexpressing the mouse homeobox-containing gene Hoxa-4 exhibit abnormal gut development Nature 337 464–467

Zappavigna V, Renucci A, Izpisua-Belmonte J-C, Urier G, Peschle C, Duboule D (1992) Comparison of mouse and human Hox-4 complexes defines conserved sequences involved in the regulation of Hox-4 4 EMBO J 10 4177–4187

Zhang H, Scholl R, Browse J, Somerville C (1988) Double stranded DNA sequencing as a choice for DNA sequencing Nucleic Acid Res 16 1220