MOLECULAR AND CELLULAR BIOLOGY, Mar. 1995, p. 1591–1601 Vol. 15, No. 3
0270-7306/95/$04.00+0

The Expression Pattern of the Murine *Hoxa-10* Gene and the Sequence Recognition of Its Homeodomain Reveal Specific Properties of *Abdominal B*-like Genes

GAIL V. BENSON, THE-HUNG EDWARD NGUYEN,† AND RICHARD L. MAAS*

Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, and Howard Hughes Medical Institute, Boston, Massachusetts 02115

Received 31 August 1994/Returned for modification 18 October 1994/Accepted 21 December 1994

Homeobox genes of the *Abdominal B* (*AbdB*) family constitute a distinct subset of vertebrate *Hox* genes. Analysis of the murine *Hoxa-10* gene, one member of this family, revealed several properties specific to this class. Two transcripts of *Hoxa-10*, a10-1 and a10-2, encode homeodomain proteins of 55 kDa (399 amino acids) and 16 kDa (96 amino acids), respectively. These proteins have identical homeodomains and C-terminal regions encoded by a common 3′ exon but differ significantly in the sizes of their N-terminal regions because of the usage of alternative 5′ exons. The 5′ exon of the a10-2 form is also present in transcripts of *Hoxa-9*, the next 3′ gene, indicating that splicing can occur between adjacent *AbdB Hox* genes within a cluster. Both *Hoxa-10* transcripts demonstrated identical patterns of expression in the posterior body and proximal limb bud, differentiating them from *AbdB* morphogenetic and regulatory transcripts and suggesting a role with other *AbdB Hox* genes in the patterning of these structures. Finally, a binding site selection identified the sequence AA(A/T)TTTTATTAC as the Hoxa-10 homeodomain consensus binding site, with a TTAT core sequence. Preferential recognition of a TTAT core therefore differentiates the AbdB class from Antennapedia (Antp) class gene products which bind a TAAT core. Thus, in vertebrates, structural similarities, coordinate transcriptional regulation, sites of expression, and binding site preferences all serve to distinguish *AbdB* from *Antp Hox* genes.

Homeotic mutations were originally defined by Bateson as those that transform one structure into the likeness of another (3). In many cases, the genes involved in these mutations have been found to represent developmentally regulated transcription factors unified by a common structural element, the homeobox (33, 43), a 183-bp sequence encoding a helix-turn-helix DNA-binding domain. In *Drosophila melanogaster*, various combinations of homeobox genes interact with each other and with target genes to determine the identity of segmental structures (28, 34). In certain developmental fields of vertebrates, including the limbs, rhombomeres, and somites, *Hox* genes also exhibit spatially restricted patterns of expression coincident with morphogenesis of these segmented structures (34). This finding has led to the concept of an analogous vertebrate *Hox* code, which posits that the particular combination of *Hox* genes expressed determines tissue identity (25). Recent gene targeting experiments with the mouse have demonstrated homeotic transformations consistent with this idea (34).

The genomic organization of homeobox genes has been strongly conserved throughout evolution. In mammals, there are 38 *Hox* genes arranged in four chromosomal clusters. Genes at the same relative position within different clusters exhibit the highest similarity to each other and also to genes at those positions within the *Drosophila* homeotic complex, suggesting that the clusters arose from duplication of a single ancestral cluster (13, 18). While both shared *cis*-acting elements and regulation by dynamic chromatin structure have been proposed (38, 44), the selection pressure that maintains *Hox* genes in a clustered arrangement is not known. However, it has been observed that several properties are dictated by physical position within the cluster. Moving from 5′ to 3′, *Hox* genes are expressed at progressively more anterior boundaries along the anterior-posterior axis and with an earlier onset of activation (34). In addition to structural and temporal colinearity, there may exist a functional colinearity: genes that are more 5′ can repress 3′ genes (31) and dominate when both are expressed (2, 34).

Abdominal B (*AbdB*) is the most 5′ gene within the *Drosophila* homeotic complex, while the homologous gene in mammalian clusters has been further duplicated to include either four or five *AbdB*-like genes at the 5′ end of the A, C, and D clusters. The *AbdB*-like genes share several properties which distinguish them as a specific subset of *Hox* genes. First, structurally, they share certain characteristic sequence motifs outside of the homeodomain and lack the YPWM peptide present in all other *Hox* genes (21). Within the homeodomain, they exhibit differences at some of the positions in the flexible N-terminal arm known to contact DNA (26, 35, 49). Second, the regulation of these genes appears to have some unique features. Retinoic acid treatment of animals or cultured cells either represses or fails to activate *AbdB*-like genes, in contrast to its effect on other *Hox* genes (24, 46). Additionally, in hematopoiesis, *AbdB*-like genes appear to be coordinately expressed as a group in myeloid-restricted lineages and coordinately inactivated in erythroleukemia cells, the converse of the case for other *Hox* genes (30). Finally, the expression patterns of *AbdB*-like genes suggest that they may control patterning of the posterior body and the limb bud; most extensively studied are the *Hoxd* cluster members, which are expressed in colinear domains along the posterior body axis, the limb bud, and the genitourinary tract (11, 12, 21).

Several fundamental questions remain concerning this subfamily. What is the expression pattern of the *AbdB*-like genes of other clusters? How is the regulation of *AbdB*-like genes achieved? How do the differences within the homeodomain

* Corresponding author. Phone: (617) 732-5119. Fax: (617) 738-5575.

† Present address: School of Medicine, University of California, San Francisco, CA 94143.

affect DNA binding? In order to address these questions, we have focused on *Hoxa-10*. Here we show that *Hoxa-10* is similar to *AbdB* in sequence, in the large number of alternative transcripts, and in a posterior restriction of expression. We also show that while Antp class homeodomains recognize a TAAT core sequence (20), Hoxa-10 exhibits a preference for a TTAT core binding site, and we suggest that this may be a unique feature of AbdB-like homeodomains. In addition, we show that hybrid transcripts of exons of *Hoxa-10* and *Hoxa-9* exist, providing another argument for the clustered arrangement of *Hox* genes in the genome.

MATERIALS AND METHODS

Amplification of *Hox* genes by PCR. To identify *Hox* genes expressed in the developing kidney, RNA was isolated by standard methods from embryonic day-15 mouse kidneys and reverse transcribed with random hexamer primer and RNase H^- reverse transcriptase (Superscript; Bethesda Research Laboratories). The reverse-transcribed cDNA was subjected to PCR with the primers 5′-GCTCTAGAGCAG(A/G/C)(C/T)(G/C)CTGGA(A/G)CTGGAGAAGGA(A/G)TT(C/T)-3′ (forward) and 5′-CGGGATCCTTCAT(C/G)C(G/T)(C/G)CG(G/A)TTCTGG/A)AACCA(G/A)AT-3′ (reverse) with denaturation at 94°C for 1 min, annealing at 42°C for 2 min, and extension at 72°C for 3 min for 35 cycles, followed by subcloning and plasmid sequencing. The primers used to amplify hybrid transcripts of *Hoxa-9* and *Hoxa-10* had the following sequences: 5′-GCTCAAGAAGTGATTGAGGCT-3′ (F) and 5′-CTGAAGCTTACAATACCTCCTCCA-3′ (R) (41). The sequences obtained were compared with those in the GenBank and EMBL databases and in the independent databases of homeobox sequences maintained by Thomas Bürglin (Massachusetts General Hospital, Boston) and Claudia Kappen (Yale University, New Haven, Conn.).

Screening of mouse cDNA libraries. The initial a10-3 cDNA clone was obtained by screening approximately 10^6 clones of an adult mouse kidney cDNA library in λgt10 with an end-labeled 79-bp oligonucleotide complementary to the *Hoxa-10* PCR product isolated by the process described above. a10-3 was labeled by nick translation and used to isolate other cDNAs from an embryonic day-11.5 BALB/c mouse λgt11 library and an adult CD-1 kidney library in λZap II. Fragments from these clones were then employed in subsequent screenings. In total, 35 independent clones were isolated and analyzed.

DNA sequencing. Genomic and cDNA fragments were subcloned into the plasmid vector pGEM3Z (Promega) or into Bluescript II SK^- (Stratagene). Double-stranded plasmid templates were sequenced by the dideoxy method. Compressions were resolved by adding 10% dimethyl sulfoxide to the annealing reaction and termination mixtures and running dimethyl sulfoxide reactions in parallel with conventional, dideoxyinosine, and deazaguanine reactions for both strands.

In vitro translation. DNA templates for the in vitro transcription of a9-2, a9-2′, and a10-2 were prepared by PCR. The forward primer contains the T7 promoter sequence and bp 463 to 482 of a10-2 (see Fig. 2B). The reverse primer contained either bp 1316 to 1336 (see Fig. 2A) for a10-2 or bp 464 to 487 (40) for a9-2 and a9-2′. Linearized a10-1 template or 5 μg of the above-described PCR templates was transcribed and capped under standard conditions. In vitro translation was performed by using nuclease-treated rabbit reticulocyte lysate (Promega) and [^{35}S]methionine (1,200 Ci/mmol). In some experiments, because comigration of the a10-2 form and hemoglobin resulted in band broadening, a wheat germ system (Promega) was used. Samples were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (10 or 15% polyacrylamide); this step was followed by fixation in isopropanol-water-acetic acid (25:65:10), treatment with scintillant (Amplify; Amersham), drying, and autoradiography.

RNase protection assays. The a10-3 clone (bp 112 to 1426) was linearized at a *Bgl*II site for preparation of a [^{32}P]UTP riboprobe with SP6 polymerase. The a10-2 clone (bp 141 to 626) was linearized with *Sal*I and transcribed with T3 polymerase. Analysis was performed as described previously (27). For each embryonic stage and adult tissue, 50 μg of total RNA was hybridized with 2×10^5 cpm of riboprobe for 12 h at 50°C and digested with 20 μg of RNase A per ml and 1.5 μg of RNase T_1 per ml. Protected fragments were electrophoresed in a 6% denaturing polyacrylamide gel and analyzed by autoradiography.

In situ hybridization. In situ hybridization was performed according to Sassoon and Rosenthal (42). Briefly, day-12.5 postcoitus embryos were fixed in 4% paraformaldehyde, dehydrated, and embedded in paraffin wax. Serial sections were deparaffinized, hydrated, subjected to proteinase K digestion, blocked with acetic anhydride, and then dehydrated. Probe was added at a concentration of 3.5×10^7 cpm/ml directly to the sections, and hybridization was carried out for 16 h at 50°C. Slides were then washed, treated with RNase A, and washed further. After a final dehydration they were dipped in Kodak NTB-2 emulsion and exposed for 1 week. The transcript-specific probes were transcribed from pGEM3Z plasmids containing either the first exon of a10-1 (bp 97 to 957) or the sequence between the splice donor sites of a10-3 and a10-2 (bp 313 to 431). The common probe contained the sequence from bp 1110 to 1426. These probes were prepared by linearizing with either *Sal*I or *Xho*I and transcribing RNA in vitro with T7 or SP6 polymerase in the presence of [^{35}S]UTP. A sense probe was also transcribed from the a10-2 sequence. The 854-nucleotide (nt) a10-1 probe and the 316-nt common probe were hydrolyzed to fragments of approximately 100 nt before use.

Synthesis of Hoxa-10 homeodomain-GST fusion protein. The homeobox of *Hoxa-10* was amplified by PCR and cloned into the *Sma*I site of a pGEX-2T vector, in frame with the glutathione *S*-transferase (GST) coding sequence. Following transformation of this plasmid into BL21(DE3) *Escherichia coli*, expression of the homeodomain-GST fusion protein was induced with 0.1 mM IPTG (isopropyl-β-D-thiogalactopyranoside). Cells were lysed in cold phosphate-buffered saline (PBS), and Triton X-100 was added to 1%. The lysate was incubated with glutathione agarose beads for 1 h at 25°C. Some of the protein was found in the pellet and could be eluted with 0.5 M NaCl–PBS and bound to beads. The beads were then washed three times in 50 bed volumes of cold PBS and eluted in 10 mM glutathione–50 mM Tris (pH 8.0). The protein was >95% pure as judged by Coomassie staining of SDS-PAGE gels. The activity of the protein was determined to be approximately 25% by titrating binding against a known amount of DNA. For electrophoretic mobility shift assay (EMSA), the homeodomain was cleaved from the GST carrier with 1% (wt/wt) thrombin in 2.5 mM $CaCl_2$–150 mM NaCl–50 mM Tris (pH 7.5).

Binding site selection assay. The following 70-bp oligonucleotide (N70) was synthesized for the selection: 5′-GTGACCGTCGAGACGGAATTCGCGGCCGCN$_{12}$CTCGAGGGATCCGTGCTCAGTCCCTAT-3′ (16, 48). The oligonucleotide contains a stretch of 12 random nucleotides flanked by *Eco*RI, *Not*I, *Xho*I, and *Bam*HI restriction sites used for subsequent cloning. The pool of oligonucleotides was made double stranded by annealing with a primer complementary to the fixed sequence (PB primer: CGATAGGGGACTGAGCACGGATCCCT) and extending with the Klenow fragment of DNA polymerase I in the presence of [α-^{32}P]dCTP, prior to gel purification on a 10% nondenaturing polyacrylamide gel. The binding reaction was done in 30 μl containing 20 mM Tris (pH 8.0), 50 mM KCl, 0.5 mM EDTA, 10% glycerol, 1mM dithiothreitol, 70 μg of bovine serum albumin (BSA) per ml, 250 μg of poly(dI-dC) per ml, 500 ng of Hoxa-10 homeodomain-GST fusion protein on glutathione agarose beads, and 100 ng of N70, and the mixture was shaken at room temperature for 1 hour. The fusion protein beads were then washed twice in ice-cold buffer, resuspended in 30 μl of dH_2O, and boiled for 3 min to release the bound sequences. The selected DNA was then amplified in a PCR mixture containing 0.2 mM (each) deoxynucleoside triphosphate (dNTP), 50 pmol (each) of PA (GTCAACGTCGAGACGGAATTCGCGG) and PB primers, 5 μl of selected DNA, and 1 U of *Taq* polymerase. PCR conditions were 20 cycles of 94°C for 1 min, 65°C for 30 s, and 72°C for 30 s. To ensure that all DNA was double stranded, a final extension cycle was performed in which 300 pmol of PA and PB primers was added to the reaction mixtures, followed by 94°C for 2 min, 65°C for 1 min, and 72°C for 10 min. The selection process, including the binding and PCR amplification, was repeated eight times using 5 μl of DNA from the previous PCR in lieu of N70 in subsequent cycles. After nine rounds of selection, the selected DNA was digested with *Eco*RI and *Bam*HI and subcloned into *Eco*RI-*Bam*HI-digested pGEM3Z. The sequences of 30 inserts were determined and aligned with the program Pileup (Genetics Computer Group, Madison, Wis.). The consensus sequence was tested by comparing the ability of the Hoxa-10 homeodomain to bind sequences from the binding site selection which varied at these positions. Sequences which approximated the consensus more closely bound with greater relative affinity (data not shown). EMSA analysis of oligonucleotide pools from each round suggested that most of the enrichment occurred during the first four rounds of selection. A selection performed with GST alone yielded only random sequences.

EMSAs. Probes for EMSA reactions were synthesized by annealing two overlapping 22-bp primers and extending with the Klenow fragment of DNA polymerase I in the presence of [α-^{32}P]dCTP. The primer sets for the three different probes were as follows: a-10 site (F), 5′-CGGATCCAAATTTTATTACGAA-3′ and a-10 site (R), 5′-GGAATTCGTAATAAAATTTGGA-3′; *AbdB* site (F), 5′-CGGATCCAAATTTTATGGCGAA-3′ and *AbdB* site (R), 5′-GGAATTCGCCATAAAATTTGGA-3′; and a-3 site (F), 5′-CGGATCCAAATTCAATTACGA-3′ and a-3 site (R), 5′-GGAATTCGTAATTGAATTTGGA-3′. Note that all positions are the same with the exception of 2 bp within the consensus sequences. The resulting 26-bp probes were gel purified on a 12% nondenaturing polyacrylamide gel. EMSA reactions were performed by incubating 10,000 cpm of probe with increasing concentrations of Hoxa-10 homeodomain (25, 90, 340, and 1,200 nM) in the presence of 10mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid) (pH 7.4)–50 mM KCl–1 mM β-mercaptoethanol–2.5 μg of BSA per ml–5 μg of poly(dI-dC)–20% glycerol. The reaction mixtures were incubated at 4°C for 30 min and then subjected to electrophoresis through an 8% nondenaturing polyacrylamide gel in 0.5× TBE (Tris-borate-EDTA) at 240 V and 4°C. Visualization was by autoradiography, and quantitation of the percentage of probe shifted was determined by PhosphorImager (Molecular Dynamics) analysis.

Nucleotide sequence accession numbers. The sequences of a10-1 and of a10-2 and a10-3 have been deposited in GenBank under accession numbers L08757 and L08758, respectively.

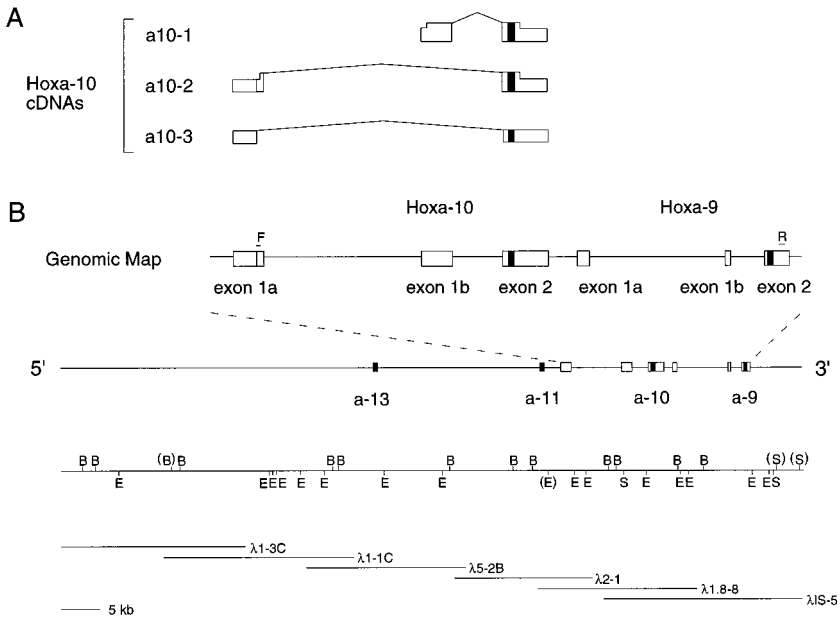


FIG. 1. (A) Scheme showing the relationship among three representative classes of cDNA clones for *Hoxa-10*, designated a10-1, a10-2, and a10-3. The a10-1 form encodes a 399-amino-acid protein with 325 residues N-terminal to the homeodomain; a10-2 encodes a 94-amino-acid protein with 20 amino acids N-terminal to the homeodomain. a10-3 is a noncoding form. (B) Genomic structure of the *AbdB* region of the *Hoxa* cluster, determined from overlapping phage clones. The exons depicted in panel A are shown in relation to other *AbdB* genes in the *Hoxa* cluster. The positions of the individual homeoboxes are shown in black. Restriction sites shown in parentheses are polymorphic. Abbreviations: E, *Eco*RI; B, *Bam*HI; S, *Sal*I.

RESULTS

Identification of three alternatively spliced *Hoxa-10* cDNAs. The *Hoxa-10* homeobox sequence was isolated during a reverse transcription-PCR screen of mouse embryonic day-15 kidney RNA by using degenerate primers corresponding to helices II and III of the homeobox. The homeobox was clearly in the *AbdB* subfamily and was most closely related to *Hox-d10* with 59 of 61 amino acids identical (13). Genetic and physical mapping studies confirmed the assignment of this sequence as *Hoxa-10*, and it was then used to isolate cDNA clones from adult kidney and embryonic day-11.5 libraries. Analysis of a total of 35 *Hoxa-10* cDNA clones revealed three classes of alternatively spliced transcripts (Fig. 1A). Exon mapping demonstrated that these transcripts are generated by splicing from either of two alternative 5′ exons to a common 3′ homeobox-containing exon (Fig. 1B). The a10-1 class of transcripts splices from a 5′ exon 2 kb upstream of the homeobox-containing exon, while the a10-2 and a10-3 transcript classes splice from donor sites 172 bp apart in a 5′ exon 8 kb upstream of the homeobox-containing exon.

Alternatively spliced transcripts encode distinct protein products. The choice of 5′ exon dictates the coding potential. The a10-1 class of transcripts has an open reading frame encoding a 399-amino-acid protein. The a10-2 class of transcripts has a short open reading frame of just 95 amino acids, but this open reading frame does contain an ATG within a favorable Kozak consensus in frame with the homeodomain. In contrast to a10-1 and a10-2, the a10-3 class of transcripts contains multiple termination codons upstream of the homeobox and lacks an in-frame ATG, indicating that it cannot encode a homeodomain-containing gene product.

The sequences and deduced translations of the a10-1 and a10-2 cDNAs are shown in Fig. 2A and B. The a10-1 and a10-2 proteins contain the same C-termini, encoded by the common 3′ exon. This includes 17 amino acids preceding the homeodomain, the homeodomain itself, and a short 14-amino-acid C-terminal tail. Thus, in addition to the AbdB-like homeodomain, both proteins contain a sequence immediately preceding the homeodomain conserved among all *AbdB*-like genes with an invariant tryptophan at position −6 from the homeodomain (21).

The a10-1 and a10-2 proteins differ in their N-termini, which are encoded by the alternate 5′ exons. Neither N-terminal region contains the YPWM peptide present in all Antp-like homeoproteins (21). The a10-1 protein has a large 307-amino-acid N-terminal region which contains two additional motifs conserved among *AbdB*-like *Hox* genes. An aspartic acid resides between amino acids 15 and 25, and a second conserved sequence surrounding a proline-tyrosine pair (glycine-tyrosine in the present case) is found in the central part of the protein (21). In addition, the N-terminal region of a10-1 is distinctive for two glycine-rich regions and a proline-glutamine-rich region (underlined in Fig. 2A). In contrast, the 5′ exon of the a10-2 transcript encodes only an additional 3 amino acids and thus lacks any AbdB conserved motifs or putative transactivating domains. In vitro translations of both transcripts verified the initiation codons and reading frames and produced proteins migrating at 55 and 15 kDa, respectively (Fig. 2D).

A comparison of *Hoxa-10* sequences with the paralogous genes *Hoxc-10* and *Hoxd-10* (13, 37) reveals 62 and 68% similarity, respectively, to a10-1. Several short regions of near identity occur in the N-terminus (residues 1 to 20, 148 to 163, and 198 to 214) and in the C-terminal tail. Transcripts corresponding to the a10-2 class have not yet been reported in these paralogs.

The 5′ exon of a10-2 can splice to exons of *Hoxa-9*. Unexpectedly, 136 bp of sequence within the 5′ exon of a10-2 demonstrated 76% identity with the 5′ end of a guinea pig cDNA clone of *Hoxa-9* (41). To determine whether this finding reflected strong conservation of 5′ exons or splicing between exons of two adjacent *Hox* genes, mouse adult kidney RNA was subjected to reverse transcription and PCR analysis by using a

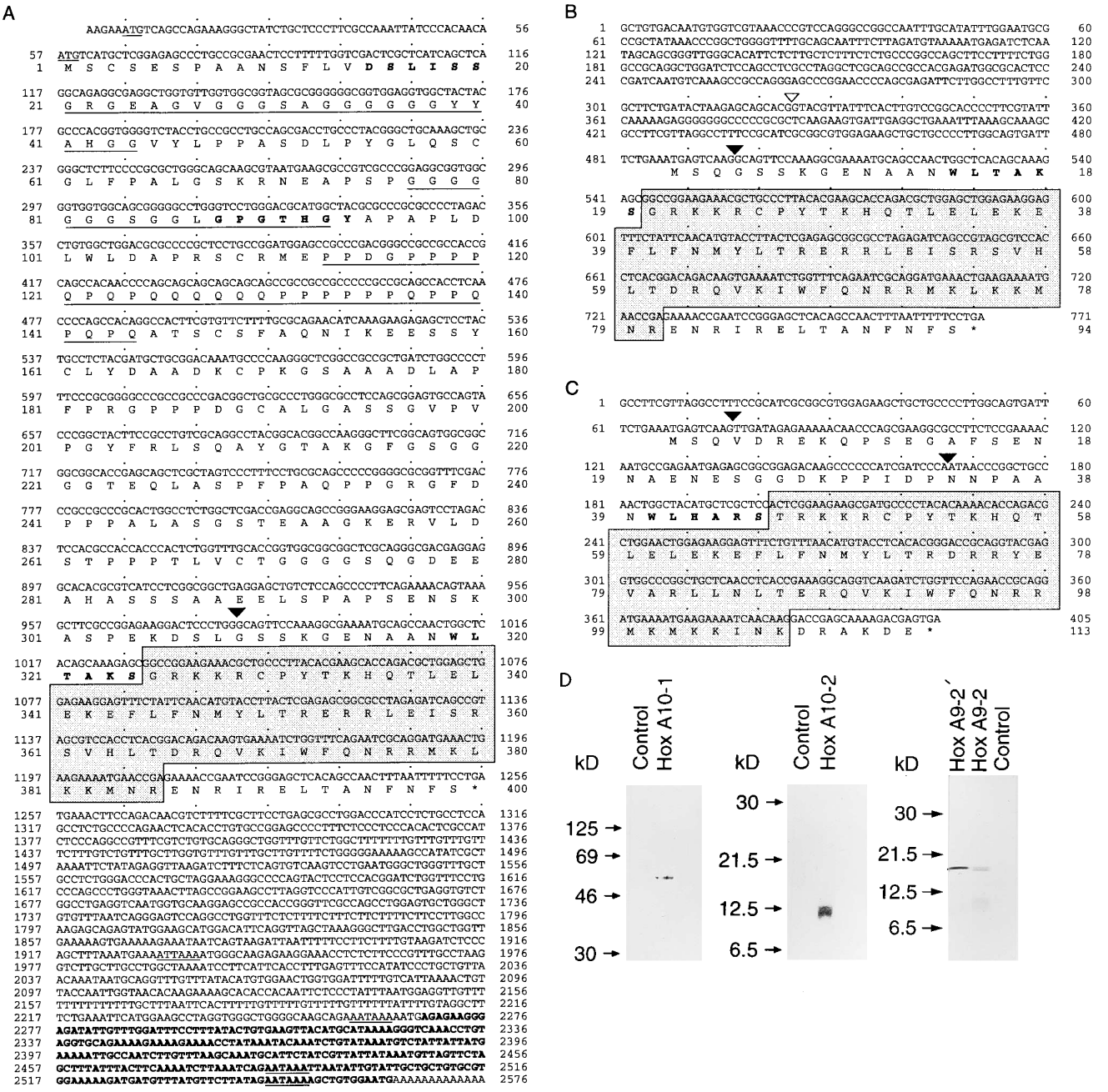


FIG. 2. Deduced protein sequences for the a10-1 (A) and a10-2 (B) forms of Hoxa-10. The homeodomain, common to both, is boxed and shaded. Sequence motifs highly conserved in other *AbdB Hox* genes are shown in bold, as is a portion of the 3′ untranslated region, which is highly conserved (>95% over 200 bp) between mouse and human (29). The splice junctions of a10-1 and a10-2 are indicated by solid triangles, while the a10-3 splice donor located in the a10-2 5′ untranslated region is indicated by an open triangle. Two glycine-rich and a proline-glutamine-rich domain in a10-1 are underlined, as is an additional in-frame 5′ ATG which appears not to be utilized as an initiation codon. Potential polyadenylation sites are also underlined. (C) Sequence of the PCR product a9-2′, demonstrating splicing of *Hoxa-10* exon 1a to the *Hoxa-9* gene. The intermediate exon, *Hoxa-9* exon 1b, lies between the two splice junctions (solid triangles) and is included in a9-2′ but is absent in a9-2. (D) In vitro translation demonstrating different size products encoded by a10-1, a10-2, a9-2, and a9-2′. Autoradiogram of SDS-PAGE gels showing synthesis of ^{35}S-labeled products and control reactions to which no RNA was added.

forward primer (F) derived from the 5′ exon of a10-2 (exon 1a) and a reverse primer (R) complementary to a sequence in the 3′ untranslated region of murine *Hoxa-9* (shown in Fig. 1A). Two products, 560 and 650 bp in length, were obtained. The sequences of the PCR products (Fig. 2C) revealed that the 5′ ends of both were identical to *Hoxa-10* exon 1a up to the a10-2 splice donor site. The 560-bp product, designated a9-2, was then joined, in frame, to a sequence identical to that reported for the homeobox-containing exon of murine *Hoxa-9*. The 650-bp product, a9-2′, contained 90 bp of intervening sequence between *Hoxa-10* exon 1a and the *Hoxa-9* homeobox exon, which was 94% identical to an intermediate *Hoxa-9* exon found in guinea pig. This analysis indicates that splicing can occur between the adjacent *AbdB*-like *Hox* genes, *Hoxa-10* and *Hoxa-9*. Identical reverse transcription-PCR products were obtained from day-10.5 embryonic limb bud RNA and day-11.5

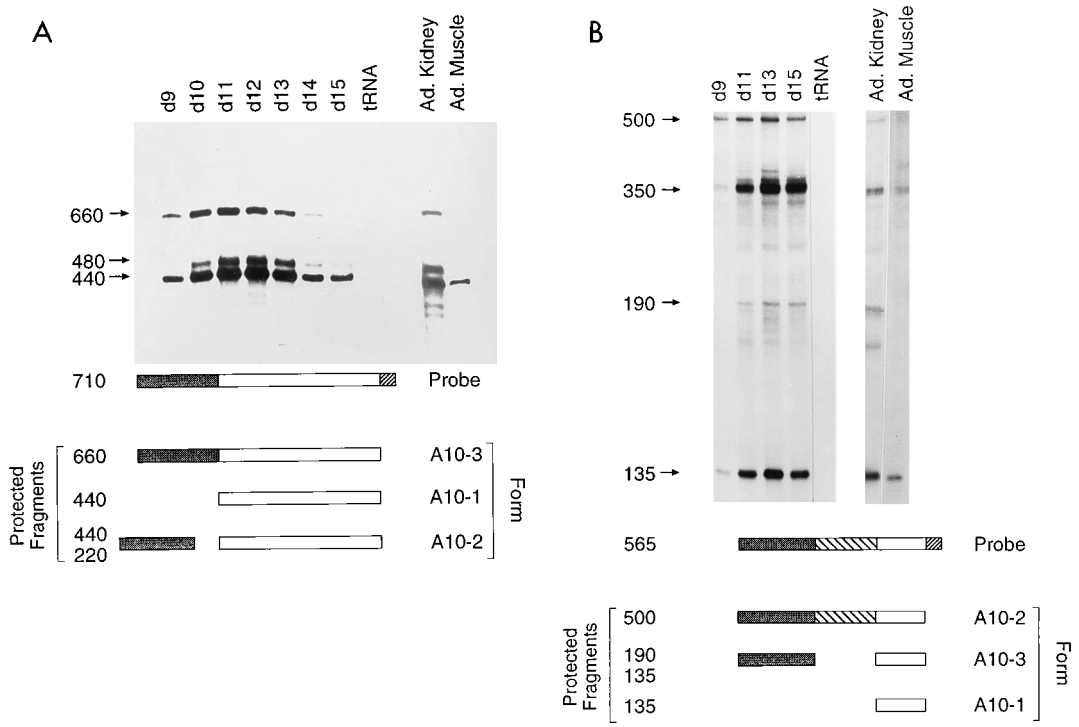


FIG. 3. Developmental RNase protection analysis demonstrating differential expression of a10-1, a10-2, and a10-3 transcripts. Fifty micrograms of total RNA from embryonic days 9 to 15, adult kidney and skeletal muscle, and tRNA were assayed with a riboprobe transcribed from either a10-3 (A) or a10-2 (B) cDNA templates. In the schematics, the exon 1a sequence up to the a10-3 splice donor is shaded, the exon 1a sequence between the a10-3 and a10-2 splice donors is shown by thick stripes, and the exon 2 sequence is shown in white. The vector sequence is shown by thin stripes. The numbers at left are sizes in nucleotides.

metanephric mesenchyme RNA. In vitro translation experiments, performed on both a9-2 and a9-2′, yielded protein products migrating at 16 and 17 kDa, respectively, demonstrating that like a10-2, these transcripts encode short homeodomain-containing forms (Fig. 2D).

Differential expression of the splice forms of *Hoxa-10*. Because a10-1 and a10-2 were potentially analogous to the morphogenetic (m) and regulatory (r) transcripts of *Drosophila AbdB*, which exhibit differential temporal and spatial expression (7), RNase protection analyses were performed to determine if there is differential expression of *Hoxa-10* transcripts. Equal amounts of total mouse embryonic RNA isolated from days 9 to 15 of gestation and adult kidney and skeletal muscle RNA were assayed for the presence of *Hoxa-10* transcripts. Two antisense riboprobes were used to discriminate between the three transcripts. The first was synthesized from an a10-3 cDNA and spans the a10-3 splice junction from exon 1a to exon 2 (Fig. 3A). This probe is protected by a10-1 transcripts only in the part derived from exon 2 (440 nt), by a10-2 transcripts as two fragments (220 and 440 nt), and by a10-3 transcripts along its full length (660 nt). The second riboprobe was synthesized from an a10-2 cDNA and spans the a10-2 splice junction from exon 1a to exon 2 (Fig. 3B). This probe is protected by a10-1 transcripts only in the part derived from exon 2 (135 nt), by a10-2 transcripts in its entirety (490 nt), and by a10-3 transcripts as two fragments (160 and 135 nt).

RNase protection assays done with these two probes demonstrated the presence of all three classes of *Hoxa-10* transcripts during development (Fig. 3). Increasing expression of all three *Hoxa-10* transcripts from day 9 to day 12 was observed, followed by decreasing expression to day 15, in agreement with a developmental Northern (RNA) analysis of embryonic poly(A)$^+$ mRNA isolated from these same stages (data not shown). Comparison of the different protected fragments during development revealed that a10-1 is the major transcript of the *Hoxa-10* gene from embryonic days 9 to 15. The a10-2 class of transcripts is at its highest relative abundance at day 9 and then decreases, while a10-3 demonstrates a low level of expression at day 9 and increases slightly thereafter. An additional protected fragment is seen with both probes (480 and 350 nt). Studies with additional riboprobes suggest that, although no cDNAs corresponding to these fragments have been isolated, the presence of these fragments is consistent with a splice acceptor upstream of the a10-3 splice donor.

Of the adult tissues analyzed, RNase protection assays revealed the presence of all three transcripts in the kidney, but only a10-1 was detected in skeletal muscle (Fig. 3). These results indicate that *Hoxa-10* transcripts are differentially expressed both temporally during development and in a tissue-specific fashion in the adult.

a10-1 and a10-2 transcripts have identical spatial distributions in posterior body and proximal limb bud at embryonic day 12.5. To determine whether different *Hoxa-10* transcripts demonstrated different spatial patterns of expression during embryogenesis, in situ hybridization of antisense riboprobes to serial sections of day-12.5 mouse embryos was performed. Transcript-specific riboprobes were transcribed from the first exons of a10-1 and a10-2, and a common *Hoxa-10* probe was transcribed from the shared homeobox-containing exon. A sense control probe was transcribed from the a10-2 template and did not show any hybridization.

Hybridization of a10-1- and a10-2-specific probes to sections through the limb bud at day 12.5 revealed that expression of both transcripts is restricted to a band of mesenchyme along the proximal-distal axis (Fig. 4A and B). The proximal and distal boundaries of a10-1 and a10-2 are identical and corre-

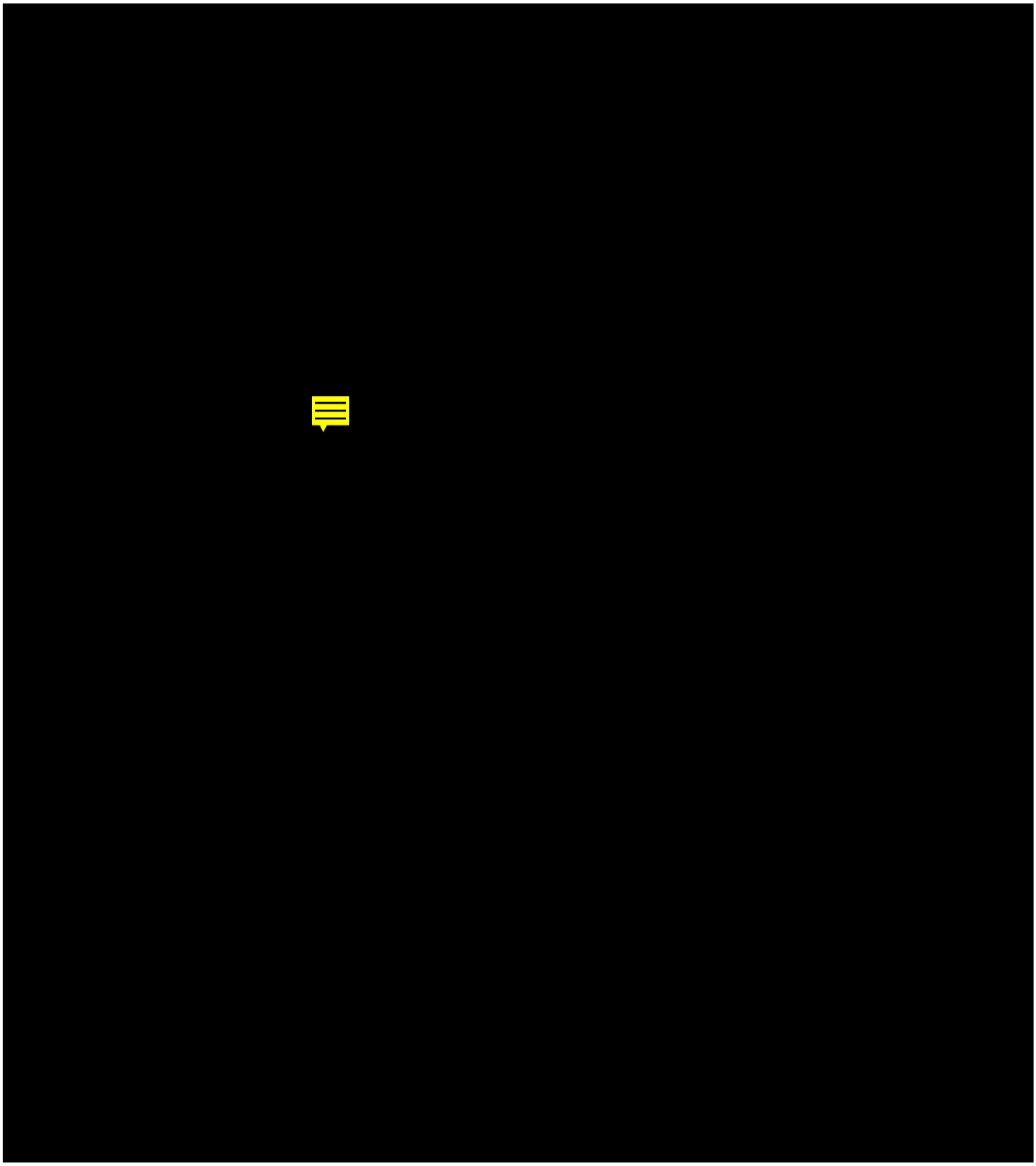


FIG. 4. *In situ* hybridization comparing the expression of *Hoxa-10* transcripts at day 12.5 of embryogenesis. Probes used were specific for transcript a10-1 (A and D) or transcript a10-2 (B and E) or were common to all transcripts of *Hoxa-10* (C and G). Shown are serial sagittal sections through the limb bud (A and B; bar, 100 μm) and through the body axis (C, D, and E; bar, 250 μm). Panels F and G show the kidney from panel C at higher magnification (bar, 50 μm), in bright field and dark field, respectively. Solid arrowheads indicate tubules that express *Hoxa-10*, while open arrowheads indicate more mature tubules which are not expressing *Hoxa-10*. A control sense probe did not demonstrate any hybridization. Abbreviations: a, anterior; g, gut; gt, genital tubercle; li, liver; L1, first lumbar prevertebra; m, metanephros; nt, neural tube; p, posterior; pd, paramesonephric duct; u, ureter; um, urogenital mesenchyme.

spond approximately to the zeugopod which will form the radius and ulna. Expression appears to be excluded from cartilaginous condensations within this region, however. A second set of hybridizing cells was also detected with both probes near the base of the condensing cartilage of the humerus. This second group of cells may correspond to migratory myoblasts from the somatopleural domain (Fig. 4C) (19).

The a10-1-specific, a10-2-specific, and common *Hoxa-10* probes also demonstrate similar patterns of hybridization to day-12.5 midsagittal sections. *Hoxa-10* expression is restricted along the body axis to the posterior region of the embryo (Fig. 4C to E). In the posterior somites strong expression was detected with a sharp anterior boundary at L1, while in the neural tube and dorsal root ganglia, the anterior limit of expression is slightly more rostral, consistent with the case for other *Hox* genes (21). A well-defined boundary of expression is also seen in derivatives of the intermediate mesoderm, where *Hoxa-10* is expressed in the mesenchyme of herniated gut (not shown), in the urogenital mesenchyme surrounding the ureter and mesonephric and paramesonephric ducts, and in the genital tubercle (Fig. 4D and E). The epithelium of the ureter and ducts does not express *Hoxa-10*, however. Likewise, in the metanephros, expression is seen in the stroma and at early stages of tubule development but is lost as tubule development progresses (Fig. 4F and G). By day 14 postcoitus, expression of *Hoxa-10* has declined in all sites except in the kidney and immediately surrounding the mesonephric and paramesonephric ducts, where expression remains very high (data not shown).

While the signal from the a10-2 probe was weaker, experiments directly comparing the levels of a10-1 and a10-2 expression showed that the expression patterns of both are identical in all of the tissues examined. In addition, the common probe, designed to detect all transcripts of *Hoxa-10*, did not show any additional sites of expression, implying that all *Hoxa-10* transcripts have an identical spatial pattern of expression at embryonic day 12.5.

The Hoxa-10 homeodomain binds a TTAT core recognition sequence. To investigate whether the differences in the AbdB-like homeodomain alter DNA recognition, the DNA-binding properties of the Hoxa-10 homeodomain were examined with a binding site selection assay (16, 48). The Hoxa-10 homeodomain was expressed as a GST fusion protein and incubated with a 70-bp double-stranded DNA fragment containing a random 12-bp core, the number of positions showing preferences in Ubx and Def selections (15). Poly(dI-dC) competitor was added in order to prevent nonspecific binding of DNA. Sequences that bound the homeodomain-GST fusion were amplified by PCR and subjected to further selection. Of 30 sequences examined after nine rounds of selection, a clear consensus of AA(A/T)TTTTATTAC was determined (Fig. 5). An apparent dissociation constant of 8×10^{-9} M was estimated by determining the amount of homeodomain-GST fusion protein required to shift 50% of a low concentration of Hoxa-10 consensus site probe in an EMSA (data not shown).

Although the opposite strand of the consensus sequence contains a TAAT, recognized as the core binding site of Antp-like homeodomains (20), TTTAT occurred more frequently: of the selected sequences, 40% contained a TAAT while 85% contained a TTTAT. In addition, the consensus determined closely approximated the consensus determined for the *Drosophila* AbdB homeodomain, TTTTATGGC. The AbdB core has been assigned as TTAT by alignment with the Ubx consensus (14), although in this sequence as in Hoxa10 the T at the −1 position is also strongly selected. To determine whether the Hoxa-10 homeodomain preferentially bound a TTAT core, EMSA was performed to compare the ability of the Hoxa-10

A

```
 1  .CAGCATTTAGTC....        16  .AAATTTTTATTG....
 2  .....TTTTATTAGGTT        17  .AAACTTTTATTA....
 3  TTTAATTTTATT.....        18  ....TTTTTATTAC...
 4  ..ATATTTTATGGC...        19  ..ATTTTTTATTAC...
 5  ...TATTTTATGACT..        20  ...CTTTTTATTAC...
 6  ..TATTTTTATGAC...        21  ..AATTTTTATTAC...
 7  .CAAATTTTATGA....        22  .CAATTTTTATTA....
 8  .AAAATTTTATGA....        23  ...AGTTTTATTACT..
 9  .AGAATTTTATGA....        24  ...AAGTTTATTACC..
10  .CGAATTTTATGA....        25  ..AGCTTTTATTAC...
11  .AAAATTTTACGA....        26  .GAGATTTTATTA....
12  .AAAGTTTTATTG....        27  .AGACTTTTATTG....
13  ..AAGTTTTATTGC...        28  .AGCGTTTTATGGC...
14  .CAATTTTTATTG....        29  ....TTTTTACACCTT.
15  .CAATTTTTATGG....        30  ...AAGGTCATTAAT..
```

B

Nucleotide	Frequency															
G	0	1	4	3	4	2	1	0	0	0	1	10	8	1	1	0	0
A	0	8	16	19	11	1	0	0	0	30	0	1	19	1	0	0	0
T	1	1	2	3	10	27	29	30	29	0	27	19	0	0	4	2	1
C	0	6	0	2	4	0	0	0	1	0	2	0	2	14	1	0	0
Total	1	16	22	27	29	30	30	30	30	30	30	30	29	16	6	2	1
Consensus	N	N	A	A	W	T	T	T	T	A	T	T	A	C	N	N	N
Position			-5	-4	-3	-2	-1	1	2	3	4	5	6	7			

FIG. 5. Determination of Hoxa-10 consensus binding-site from selected sequences. Sequences were aligned with the program Pileup (Genetics Computer Group).

homeodomain to bind each of three probes (Fig. 6A). The first probe contains the Hoxa-10 consensus sequence determined as described above. The second probe has a substitution at the 2 bp following the TTAT, destroying the ATTA and creating the consensus determined for the AbdB homeodomain (14). This

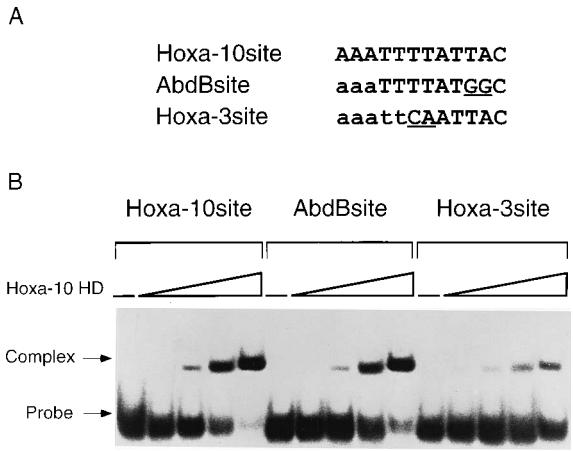


FIG. 6. Results of EMSA comparing binding of the *Hoxa-10* homeodomain to the Hoxa-10 consensus sequence, the AbdB consensus sequence, and the Hoxa-3 consensus sequence. (A) Sequence comparison of the probes used to assay the binding site specificity of Hoxa-10. The Hoxa-10 probe contains the consensus sequence determined as described above. The AbdB and Hoxa-3 probes each differ from the Hoxa-10 probe by 2 bp, so that they contain only a TTAT or ATTA core, respectively. Bases which are part of the consensus binding sites identified for the homeodomains of Hoxa-10, AbdB, and Hoxa-3 are indicated by uppercase letters. Bases outside of the consensus sequences have been assigned the identity of the base at the same position in the Hoxa-10 consensus and are indicated by lowercase letters. (B) Results of EMSA comparing the binding of the Hoxa-10 homeodomain to each of the three probes. Increasing concentrations of the Hoxa-10 homeodomain (25, 90, 330, and 1,200 nM) were incubated with radiolabeled 26-bp probes containing the consensus sequences shown. Binding reactions were carried out in the presence of 5 μg of poly(dI-dC) per ml and 2.5 mg of BSA per ml at 4°C for 30 min. Probe bound to the homeodomain (Complex) was separated from free probe by nondenaturing PAGE and visualized by autoradiography. A Hoxb-7 homeodomain-GST fusion protein control demonstrated a different order of relative binding preference for these probes, binding the Hoxa-10 and Hoxa-3 consensus site probes with greater relative activity than the AbdB consensus site probe.

probe thus contains a TTAT but not an ATTA. The third probe has a substitution at the first 2 bp of the TTAT, destroying the TTAT but preserving the ATTA. The substituted base pairs were chosen such that this probe contains the consensus determined for the Antp class Hoxa-3 homeodomain (6). Each probe differs by only the 2 bp shown.

The probes were radiolabeled in parallel and then incubated with increasing concentrations of the Hoxa-10 homeodomain in the presence of 5 μg of poly(dI-dC) per ml and 2.5 mg of BSA per ml. The Hoxa-10 homeodomain exhibited the highest relative binding activity for the Hoxa-10 and the AbdB consensus site probes and had a fourfold-less binding activity for the Hoxa-3 consensus site probe (Fig. 6B). At the maximal concentration tested, the a-10 homeodomain was able to shift 75% of the Hoxa-10 probe and 63% of the AbdB probe but only 19% of the Hoxa-3 probe. The same rank order of relative binding activities was obtained when performed under more stringent conditions in the presence of 25 and 50 μg of poly(dI-dC) competitor per ml (data not shown). Because the AbdB probe binds with nearly the same activity as the Hoxa-10 probe and lacks a TAAT, we conclude that the a-10 homeodomain prefers a TTAT core sequence.

DISCUSSION

Conservation of alternative splicing of *Hoxa-10*. Here, we report the identification of three alternatively spliced transcripts of *Hoxa-10*. The three transcripts differ in their first exon, producing significant differences in the encoded protein product. Indeed, only two transcripts, a10-1 and a10-2, are capable of producing a homeodomain-containing protein, while the third appears to be similar to noncoding isoforms reported for *Ultrabithorax* and *Hoxa-1* (32). Of the homeoprotein-encoding forms, a10-1 encodes a large 325-amino-acid N-terminal region with a proline-glutamine-rich domain, similar to those shown to function as transcriptional activators. a10-2 encodes a small N-terminal domain with only 20 amino acids prior to the homeodomain and lacks any obvious transcriptional activation domains.

Previously, two cDNAs, termed PL1 and PL2, were isolated from human myeloid leukemia libraries and mapped to the human *HOXA* cluster (29). Although there are several large frameshifts within the PL1 sequence, which we believe to be due to errors in the human sequence, and the first 140 bp of PL2 appears to be a chimeric artifact, our analysis indicates that these are the human homologs of a10-1 and a10-2. The presence of the a10-1 and a10-2 homologs in the human shows that there has been conservation of the alternatively spliced transcripts. Comparisons of the mouse and human sequences reveal that the homeobox itself is 95% identical at the nucleotide level and is 100% conserved in amino acid sequence. Interestingly, the highest homology between the mouse and human *Hoxa-10* genes occurs in the final 200-bp sequence of the 3′ untranslated region which is 97% identical (shown in bold in Fig. 2A). Therefore, this region may contain the sequence necessary for the regulation or stability of the transcript.

***Hoxa-10* demonstrates properties of *AbdB*-like genes.** The *Hoxa-10* gene displays several properties of *AbdB*-like *Hox* genes. In addition to an AbdB-like homeodomain, both proteins contain a conserved region which is distinctive for a tryptophan at position −6 preceding the homeodomain. a10-1 also contains two sequences common to the N-terminal region of AbdB-like proteins: an aspartic acid residue surrounded by a loose consensus, although in this case not containing a phenylalanine, is located within 15 to 20 amino acids of the start codon and a glycine-tyrosine pair within a consensus that fits that of the proline-tyrosine motif is located more centrally (21). Although no function has yet been ascribed to these motifs, they are conserved among *AbdB*-like *Hox* genes and the m and r forms of *Drosophila AbdB*.

As with *AbdB*-like genes of the *Hoxd* cluster and *Hoxc-10*, *Hoxa-10* is expressed in the developing limb bud (11, 37). As shown previously, it is part of a colinear pattern along the proximal-distal axis with *Hoxa-11* and *a-13* (19, 50). Expression described here at embryonic day 12.5 differs slightly from that reported, in that it shows a distinct distal as well as a proximal boundary. The difference may either be the result of using different probes or, alternatively, it may reflect a sharpening of boundaries at a slightly later stage, possibly through interaction with *Hoxa-11*. A second domain of expression, the somatopleural domain, observed here and for *Hoxa-10* and *Hoxa-11* previously, was postulated to represent muscle precursor cells as they migrate into the limb from the dermatomyotome (19). Interestingly, we have observed that adult skeletal muscle expresses *Hoxa-10*. These two findings support a possible role for *Hoxa-10* in muscle differentiation and maintenance.

The domains of expression of *Hoxa-10* along the body axis are similar to that seen for *AbdB*-like genes of the *Hoxd* cluster (12, 21). *Hoxa-10* expression is restricted to posterior domains in the somites, the neural tube, and the intermediate mesoderm in day-12.5 mouse embryos. In the derivatives of the intermediate mesoderm, *Hoxa-10* is expressed in the gut and urogenital tract, including the kidney, the mesenchyme surrounding the ureters and paramesonephric and mesonephric ducts, and the genital tubercle. These sites of expression overlap with those observed for *AbdB Hoxd* cluster members, raising the possibility of a relationship analogous to that seen in the developing limb where *AbdB* genes of these clusters define two different axes.

The discrete nature of *Hoxa-10* expression suggests that this gene may be involved in specifying the identities of structures within the urogenital system. In support of this hypothesis, retinoic acid treatment on day 9.5 of embryogenesis both repressed the expression of *Hoxa-10* and produced malformations in the neural tube, the gut, and the genital tubercle (24). In *D. melanogaster*, *AbdB* is expressed in the most posterior parasegments and in the genital imaginal disc (4). Mutations in regulatory sequences of *AbdB* can result in absence of the genitalia (22). Therefore, it is possible that *AbdB*-like genes are involved in the patterning of orthologous structures, as has been recently reported for Pax6 in *Drosophila* and mammalian eye development (39).

Splicing between *Hoxa-10* and *Hoxa-9*. One particularly surprising feature of the *Hoxa-10* gene is the splicing of exons to the next 3′ gene in the cluster, *Hoxa-9*. A database search identified a guinea pig cDNA (41) for a transcript initiating in the a10-2 5′ exon but which then splices to the homeobox-containing exon of *Hoxa-9*. The presence of similar transcripts in the mouse was confirmed by reverse transcription-PCR, but they were not detected by RNase protection analysis. Therefore, these may be transcripts with either low abundance or a limited distribution. Since the a9-2 and a10-2 transcripts have common exons, they would be predicted to be coordinately regulated, however. Such a mechanism of coordinate transcriptional regulation of adjacent *Hox* genes could act in hematopoiesis, where the *AbdB*-like genes appear to be expressed and repressed as a group (30).

It has been shown that three homeobox genes in the human *HOXC* cluster (C4, C5, and C6) also represent a single transcription unit (45). In the case of the *HOXC* cluster, the common 5′ exon is noncoding, whereas in the present case, the

common exon includes a functional initiation codon 10 bp upstream of the splice donor site. Consequently, the 5′ exon of a10-2 can participate in the synthesis of truncated forms of two *AbdB*-like *Hox* genes. Despite differences in the positions of their initiation codons, in both the A and C clusters, the transcripts generated by this splicing are similar: long proteins are encoded by exons proximal to the homeobox-containing exon, while homeodomain proteins with deleted or minimal N-termini are encoded by the distal common exon. Therefore, truncated homeoproteins may serve a key function in development. Furthermore, such splicing between homeobox genes may contribute to the pressure that has maintained homeobox genes in a clustered array throughout metazoan evolution.

a10-1 and a10-2 are not directly analogous to *AbdB* m and r. The structures of a10-1 and a10-2 have some similarities to the m and r functions of *Drosophila AbdB*. Transcripts of m and r are the result of alternative splicing; m transcripts, like a10-1, result from splicing of proximal exons, whereas r transcripts result from splicing of distal exons (51). The m protein has a large N-terminal domain with a glutamine-rich transcriptional activation domain, while the r protein is a truncated form lacking this region. The question therefore arises as to whether a10-1 and a10-2 are also similar to m and r in regulation and function: m specifies positional information in parasegments 8 to 13 while r represses m in parasegment 14 (4, 51). Examination of the expression of a10-1 and a10-2 by RNase protection analysis revealed that a10-2, like *AbdB* r (7), is expressed at highest levels earlier in development and that differential expression of a10-1 and a10-2 occurs in adult tissues. However, in situ hybridization at day 12.5 of embryogenesis demonstrated that a10-2 expression is not confined to more posterior domains. a10-1 and a10-2 appear to have an identical pattern of expression within the somites, neural tube, intermediate mesoderm, and limb bud. Therefore, we conclude that a10-1 and a10-2 are not directly analogous to m and r. Indeed, recent analysis of an *AbdB* homolog of *Schistocerca gregaria* suggests that diversification of m and r may not be a general feature of *AbdB*-like genes but may be specific to long-germ-band insects (23). Thus, the duplication of mammalian *AbdB* homologs within a cluster may substitute for the lack of transcripts with different spatial domains.

The identical patterns of expression of a10-1 and a10-2 suggest that they act within the same cells. The difference observed in the temporal expression patterns of the transcripts may be of functional importance in this context. A quantitative temporal colinearity has been postulated for *Hoxd* genes in the genital tubercle, with the last gene expressed at the highest abundance (11); the same phenomenon may be a feature of different transcripts of a single *Hox* gene. Recently, it has been shown that while proper differentiation of larval epidermis requires the continuous presence of homeotic gene products, in other cell types transient expression can have lasting effects (5). Therefore, early activation of a10-2 could modulate responses to later action by a10-1.

Because they contain the same homeodomain, a10-1 and a10-2 proteins, when coexpressed, might be expected to compete for the same binding sites and thereby modulate the expression levels of target genes. For instance, there is some evidence to suggest an antagonistic relationship between short and long forms of *XHoxc-6* and *HOXC4* (9, 32). Alternatively, a different capacity for interactions with other proteins may target a10-1 and a10-2 to different target sites in vivo; however, the region shown to mediate Matα's interactions with MCM1 in *Saccharomyces cerevisiae* (47) does correspond to the limited N-terminal sequence a10-1 shares with a10-2.

The Hoxa-10 homeodomain binds a TTAT core recognition sequence. Using a binding site selection assay, we identified the consensus binding site for the Hoxa-10 homeodomain as AA(A/T)TTTATTAC (the core is underlined). Despite the presence of a TAAT (ATTA in the consensus shown), which is the core binding site for Antp-like homeodomains, the TTAT sequence was more highly selected. It has been suggested that *Hox* genes recognize a single consensus site but with decreasing affinity according to position in the cluster (36); however, our results suggest that, at least in the 5′ region of the *Hox* cluster, the genes recognize a different consensus sequence. EMSAs comparing the binding of the Hoxa-10 homeodomain to its consensus sequence and to sequences modified to contain only the TTAT or TAAT revealed that the Hoxa-10 homeodomain binds to sequences containing a TTAT with high relative binding activity and to those containing a TAAT core with four-fold-reduced activity. Differences of this magnitude in in vitro binding assays have been shown to translate to significant differences in activation of target genes in vivo (10, 15).

The binding site determined for Hoxa-10 is remarkably similar to the consensus binding site recently reported for AbdB, TTTTATGGC (the core is underlined) (14), indicating a strong conservation of binding-site preferences through evolution despite 20 amino acid changes between the homeodomains of AbdB and Hoxa-10. Studies with chimeric homeoproteins have shown that the homeodomain is responsible for most of the functional specificity of the entire protein (20). Therefore, the conservation of the TTAT core preference may be an important mechanism for achieving functional specificity between *AbdB*-like genes and *Antp*-like *Hox* genes.

Molecular basis for TTAT core recognition. The second base pair of the core is contacted by the third amino acid in the N-terminal arm of the homeodomain (26, 49). However, amino acid substitutions at position 3 did not convert the Ubx homeodomain binding site, TAAT, to that of AbdB, TTAT, whereas a triple mutation at positions 3, 6, and 7 did (14). In Hoxa-10, two of these residues, K-3 and P-7, are identical to AbdB, whereas residue 6 is C in Hoxa-10 and K in AbdB. Therefore, it is likely that only residues 3 and 7 are involved in specification of the second base pair. The presence of a K-3 and a P-7 in all *AbdB*-like homeodomains suggests that they all recognize a TTAT core. This hypothesis is supported by the presence of a conserved sequence (AATTTATGA) which is found upstream of paralog 5 *Hox* genes and can be footprinted by HOXD9 and HOXD10 in vitro (1). These results may also have implications for core sequence recognition among other genes such as Hoxc-8, which has an S-3 and a T-7.

Alignment of the selected sequences also revealed strong preferences for certain nucleotides at positions flanking the core sequence. This finding is consistent with a 15-bp footprint (36) as well as the crystal and nuclear magnetic resonance structures determined for engrailed, Matα, and Antp homeodomains (26, 35, 49). The two nucleotides immediately 5′ of the core sequence (−1 and −2) are contacted by residue 7 in the N-terminal arm in MATα2 (49). This residue is proline in Hoxa-10 and AbdB (14), both of which demonstrate a preference for thymidines at these sites. Hoxa-10 also demonstrates preferences for positions −3, −4, and −5 from the TTAT, although these nucleotides do not appear to be contacted directly in the available crystal structures. While other binding site selection assays have not determined preferences at these positions (6, 15), our result may be a consequence of applying nine rounds of selection. At this level of stringency, subtle differences may become more apparent than with fewer rounds of selection.

The only difference in the binding site between Hoxa-10 and AbdB is in the 3′ flanking residues (14). These residues are

thought to be contacted by Q-50 and M-54 of helix III, both of which are conserved between AbdB and Hoxa-10. The preferred nucleotides at these positions also differ in the Dfd and ftz consensuses despite conservation of Q-50 and M-54 (15, 17). Therefore, these contacts may be fairly promiscuous, and the difference may be due to the influence of other residues. While AbdB and Hoxa-10 are nearly identical in helix III, the C-terminal tail, which is known to influence binding specificity (8), is highly divergent between Hoxa-10 and AbdB and was included in the AbdB homeoprotein used in the selections. Since the affinities of the Hoxa-10 homeodomain for these two sites in EMSA were only slightly different, it remains to be determined whether these nucleotide differences have in vivo significance. The phenotype seen when *Hoxb-9* was ectopically expressed in *D. melanogaster*, intermediate between *Antp* and *AbdB*, may suggest that this difference does have in vivo consequences (31). Therefore, as seen for *Antp*-like *Hox* genes (15), subtle differences in nucleotides flanking the TTAT core may constitute a means of providing fine specificity among *AbdB*-like *Hox* genes.

ACKNOWLEDGMENTS

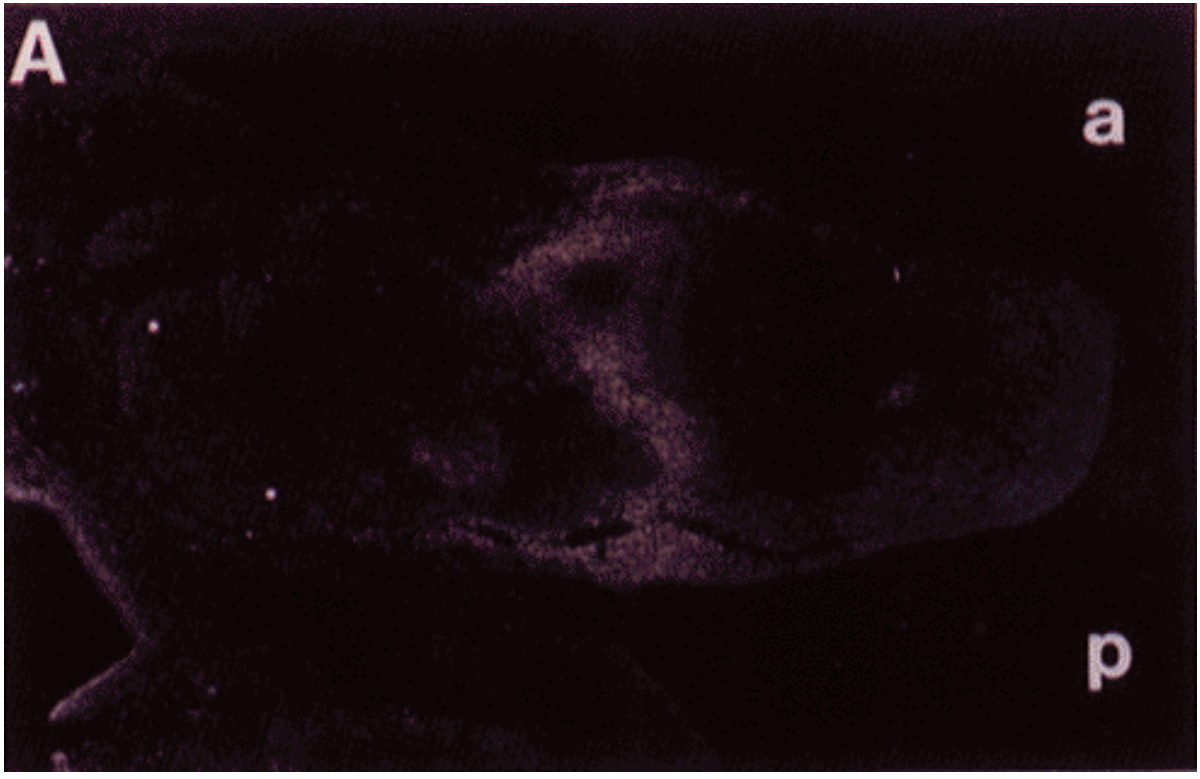
We gratefully acknowledge the generous assistance of Kerry Ressler and Susan Sullivan with the in situ hybridization technique.

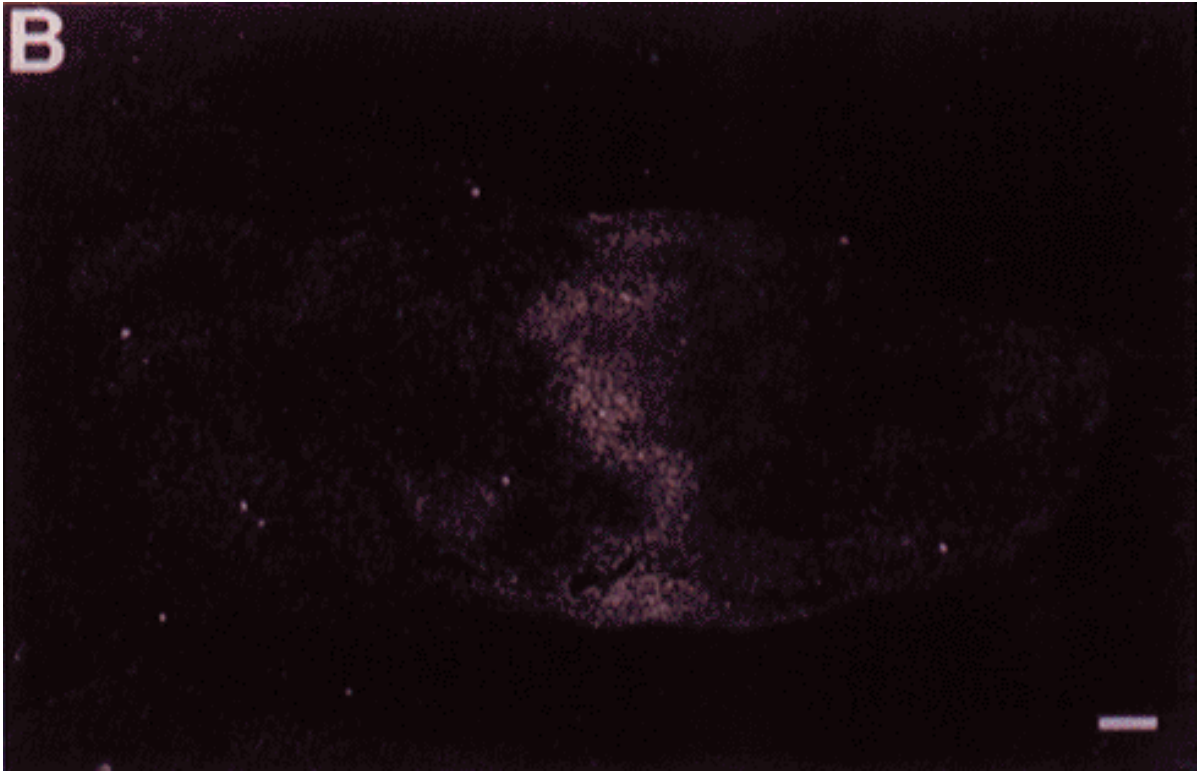
G.V.B. is supported by MSTP grant T32GM07753. T.-H.E.N. was an HHMI Summer Student during the course of these studies. R.L.M. is an Assistant Investigator at the Howard Hughes Medical Institute.

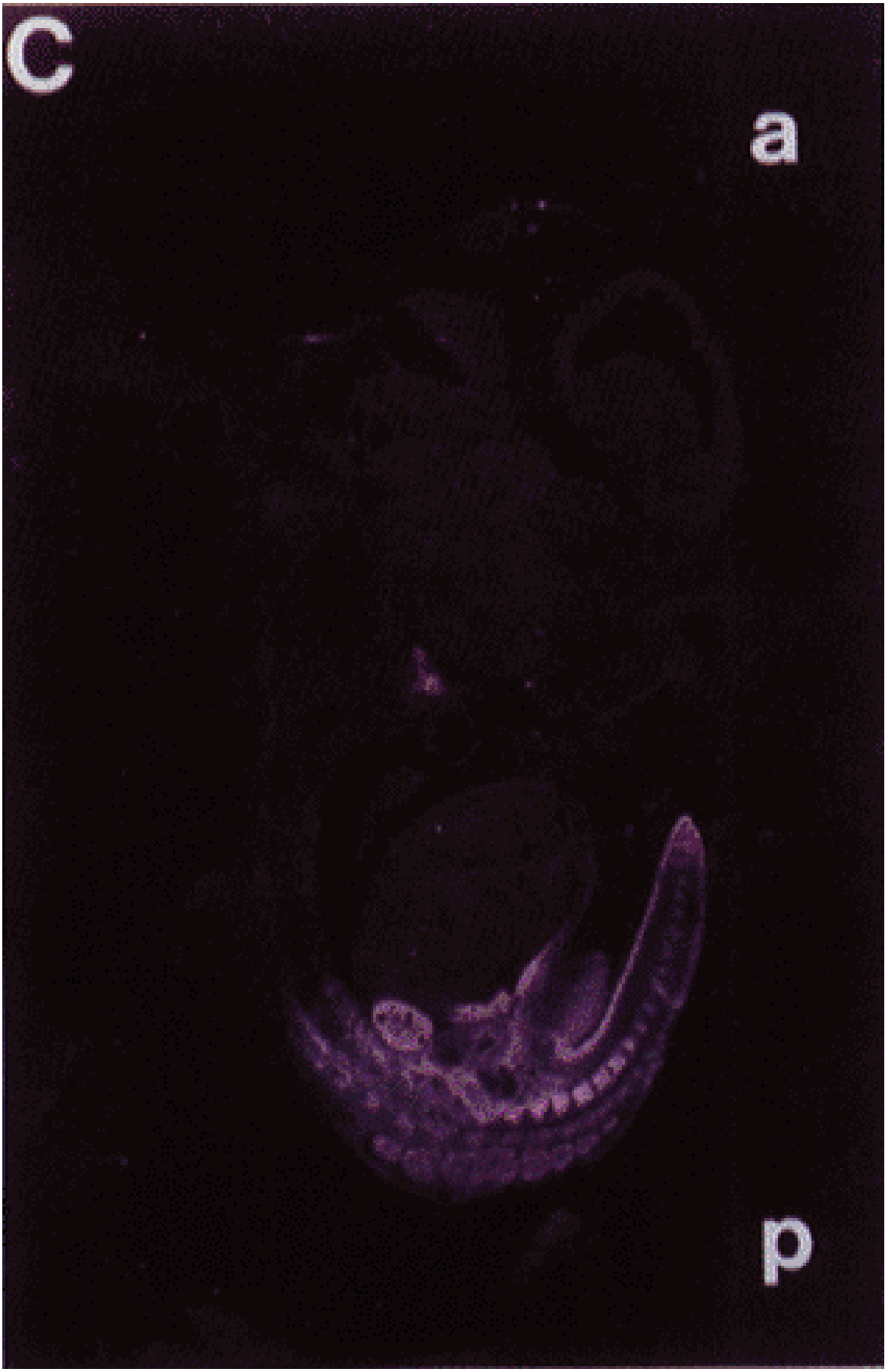
REFERENCES

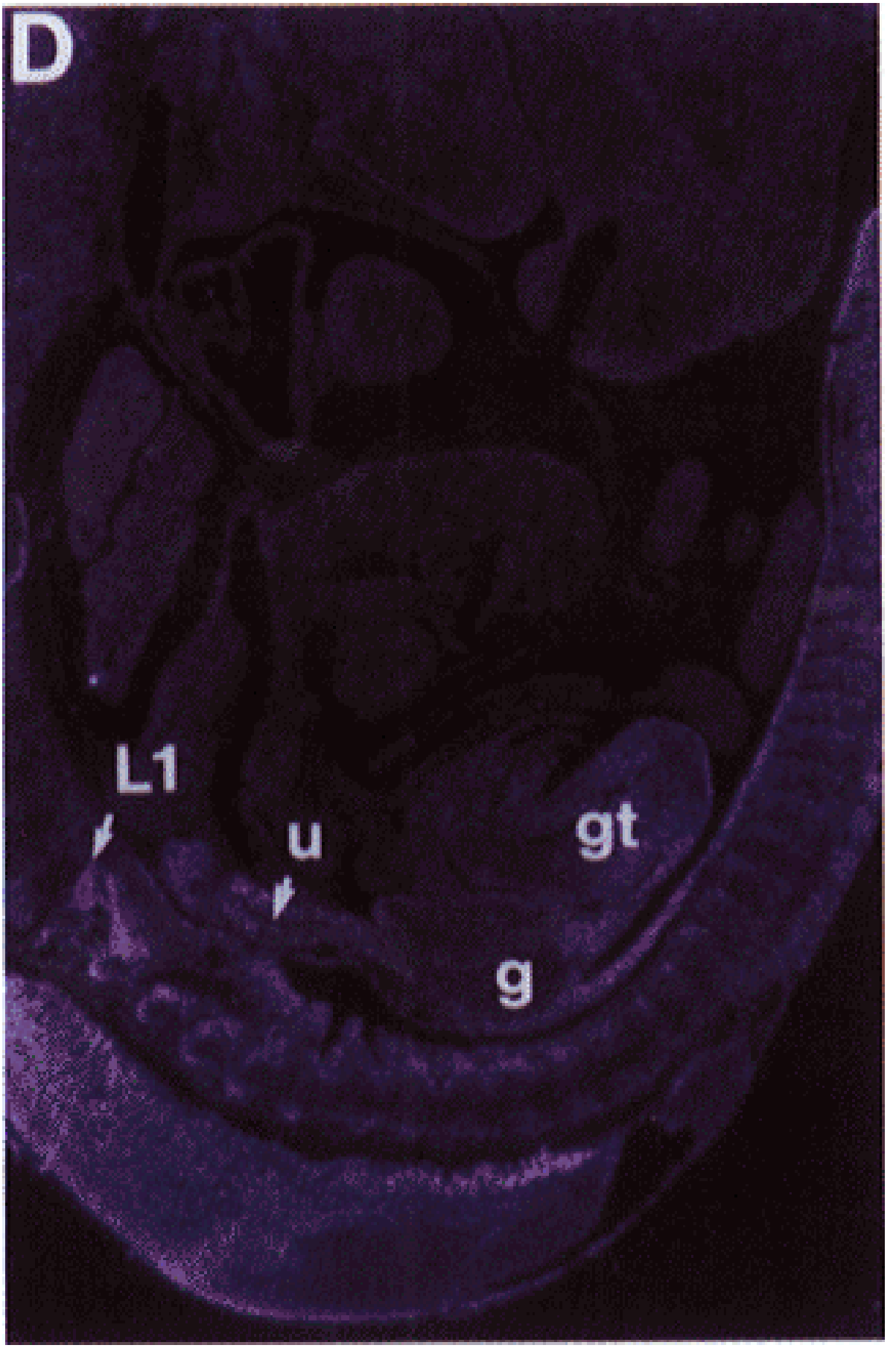
1. **Arcioni, L., A. Simeone, S. Guazzi, V. Zappavigna, E. Boncinelli, and F. Mavilio.** 1992. The upstream region of the human homeobox gene HOX3D is a target for regulation by retinoic acid and HOX homeoproteins. EMBO J. **11:**265–277.
2. **Bachiller, D., A. Macías, D. Duboule, and G. Morata.** 1994. Conservation of a functional hierarchy between mammalian and insect *Hox*/*HOM* genes. EMBO J. **13:**1930–1941.
3. **Bateson, W.** 1894. Materials for the study of variation. MacMillan & Co., London.
4. **Casanova, J., E. Sanchez-Herrero, and G. Morata.** 1986. Identification and characterization of a parasegment specific regulatory element of the *abdominal-B* gene of *Drosophila*. Cell **47:**627–636.
5. **Castelli-Gair, J., S. Greig, G. Micklem, and M. Akam.** 1994. Dissecting the temporal requirements for homeotic gene function. Development **120:**1983–1995.
6. **Catron, K. M., N. Iler, and C. Abate.** 1993. Nucleotides flanking a conserved TAAT core dictate the DNA binding specificity of three murine homeodomain proteins. Mol. Cell. Biol. **13:**2354–2365.
7. **Celnicker, S., D. J. Keelan, and E. B. Lewis.** 1989. The molecular genetics of the bithorax complex of *Drosophila*: characterization of the products of the *Abdominal B* domain. Genes Dev. **3:**1424–1436.
8. **Chan, S.-K., and R. S. Mann.** 1993. The segment identity functions of *Ultrabithorax* are contained within its homeo domain and carboxy-terminal sequences. Genes Dev. **7:**796–811.
9. **Cho, K. W. Y., J. Goetz, C. V. E. Wright, A. Fritz, J. Hardwicke, and E. M. DeRobertis.** 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.
10. **Dessain, S., C. T. Gross, M. A. Kuziora, and W. McGinnis.** 1992. Antp-type homeodomains have distinct DNA binding specificities that correlate with their different regulatory functions in embryos. EMBO J. **11:**991–1002.
11. **Dollé, P., J. C. Izpisúa-Belmonte, H. Falkenstein, A. Renucci, and D. Duboule.** 1989. Coordinate expression of the murine Hox-5 complex homeobox-containing genes during limb pattern formation. Nature (London) **342:**767–772.
12. **Dollé, P., J. C. Izpisúa-Belmonte, C. Tickle, and D. Duboule.** 1991. HOX4 genes and the morphogenesis of mammalian genitalia. Genes Dev. **5:**1767–1776.
13. **Duboule, D., and P. Dollé.** 1989. The structural and functional organization of the murine Hox gene family resembles that of *Drosophila* homeotic genes. EMBO J. **8:**1497–1505.
14. **Ekker, S. C., D. G. Jackson, D. P. Kessler, B. I. Sun, K. E. Young, and P. A. Beachy.** 1994. The degree of variation in DNA sequence recognition among four *Drosophila* homeotic proteins. EMBO J. **13:**3551–3560.
15. **Ekker, S. C., D. P. von Kessler, and P. A. Beachy.** 1992. Differential DNA sequence recognition is a determinant of specificity in homeotic gene action. EMBO J. **11:**4059–4072.
16. **Epstein, J., J. Cai, T. Glaser, L. Jepeal, and R. L. Maas.** 1994. Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. J. Biol. Chem. **269:**8355–8361.
17. **Florence, B., R. Handrow, and A. Laughon.** 1991. DNA binding specificity of the fushi tarazu homeodomain. Mol. Cell. Biol. **11:**3613–3623.
18. **Graham, A., N. Papalopulu, and R. Krumlauf.** 1989. The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. Cell **57:**367–378.
19. **Haack, H., and P. Gruss.** 1993. The establishment of murine *Hox-1* expression domains during patterning of the limb. Dev. Biol. **157:**410–422.
20. **Hayashi, S., and M. P. Scott.** 1990. What determines the specificity of action of *Drosophila* homeodomain proteins? Cell **63:**883–894.
21. **Izpisúa-Belmonte, J. C., H. Falkenstein, P. Dollé, A. Renucci, and D. Duboule.** 1991. Murine genes related to the *Drosophila AbdB* homeotic genes are sequentially expressed during development of the posterior part of the body. EMBO J. **10:**2279–2289.
22. **Karch, F., B. Weiffenbach, M. Pfeifer, W. Bender, I. Duncan, S. Celnicker, M. Crosby, and E. B. Lewis.** 1985. The *Abdominal* region of the *bithorax* complex. Cell **43:**81–96.
23. **Kelsh, R., I. Dawson, and M. Akam.** 1993. An analysis of *Abdominal-B* expression in the locust *Schistocerca gregaria*. Development **117:**293–305.
24. **Kessel, M.** 1992. Respecification of vertebral identities by retinoic acid. Development **115:**487–501.
25. **Kessel, M., and P. Gruss.** 1991. Homeotic transformations of murine vertebrae and concomitant alteration of *Hox* codes induced by retinoic acid. Cell **67:**89–104.
26. **Kissinger, C. R., B. Liu, E. Martin-Blanco, T. B. Kornberg, and C. O. Pabo.** 1990. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 resolution: a framework for understanding homeodomain-DNA interactions. Cell **63:**579–590.
27. **Krieg, P. A., and D. A. Melton.** 1987. In vitro RNA synthesis with SP6 RNA polymerase. Methods Enzymol. **155:**313–324.
28. **Lewis, E. B.** 1978. A gene complex controlling segmentation in *Drosophila*. Nature (London) **276:**565–570.
29. **Lowney, P., J. Corral, K. Detmer, M. M. LeBeau, L. Deaven, H. J. Lawrence, and C. Largman.** 1991. A human *Hox* 1 homeobox gene exhibits myeloid-specific expression of alternative transcripts in human hematopoietic cells. Nucleic Acids Res. **19:**3443–3449.
30. **Magli, M., P. Barba, A. Celetti, G. De Vita, C. Cillio, and E. Boncinelli.** 1991. Coordinate regulation of *HOX* genes in human hematopoietic cells. Proc. Natl. Acad. Sci. USA **88:**6348–6352.
31. **Malicki, J., L. Bogarad, M. Martin, F. Ruddle, and W. McGinnis.** 1993. Functional analysis of the mouse homeobox gene *HoxB9* in *Drosophila* development. Mech. Dev. **42:**139–150.
32. **Mavilio, F.** 1993. Regulation of vertebrate homeobox-containing genes by morphogens. Eur. J. Biochem. **212:**273–288.
33. **McGinnis, W., R. L. Garber, J. Wirz, A. Kuroiwa, and W. J. Gehring.** 1984. A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. Cell **37:**403–408.
34. **McGinnis, W., and R. Krumlauf.** 1992. Homeobox genes and axial patterning. Cell **68:**283–302.
35. **Otting, G., Y. Qian, W. Muller, M. Affolter, W. Gehring, and K. Wüthrich.** 1988. Secondary structure determination for the Antennapedia homeodomain by nuclear magnetic resonance and evidence for a helix-turn-helix motif. EMBO J. **7:**4305–4309.
36. **Pellerin, I., C. Schnabel, K. M. Catron, and C. Abate.** 1994. Hox proteins have different affinities for a consensus DNA site that correlate with the positions of their genes on the *hox* cluster. Mol. Cell. Biol. **14:**4532–4545.
37. **Peterson, R. L., D. F. Jacobs, and A. Awgulewitsch.** 1992. *Hox-3.6*: isolation and characterization of a new murine homeobox gene located in the 5′ region of the *Hox-3* cluster. Mech. Dev. **37:**151–166.
38. **Pfeiffer, M., F. Karch, and W. Bender.** 1987. The *bithorax* complex: control of segmental identity. Genes Dev. **1:**891–898.
39. **Quiring, R., U. Walldorf, U. Kloter, and W. Gehring.** 1994. Homology of the *eyeless* genes of *Drosophila* to the *Small eye* gene in mice and *Aniridia* in humans. Science **265:**785–789.
40. **Rubin, M. R., W. King, L. E. Toth, I. S. Sawczuk, M. S. Levine, P. D'Eustachio, and M. C. Nguyen-Huu.** 1987. Murine *Hox-1.7* homeo-box gene: cloning, chromosomal location, and expression. Mol. Cell. Biol. **7:**3836–3841.
41. **Rubin, M., and M. C. Nguyen-Huu.** 1989. Alternatively spliced *Hox-1.7* transcripts encode different protein products. DNA Sequence **1:**15–124.
42. **Sassoon, D., and N. Rosenthal.** 1993. Detection of messenger RNA by *in situ* hybridization. Methods Enzymol. **225:**384–404.
43. **Scott, M. P., and A. J. Weiner.** 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.
44. **Sham, M. H., P. Hunt, S. Nonchev, N. Papalopulu, A. Graham, E. Boncinelli, and R. Krumlauf.** 1992. Analysis of the murine *Hox-2.7* gene: conserved alternative transcripts with differential distributions in the nervous system

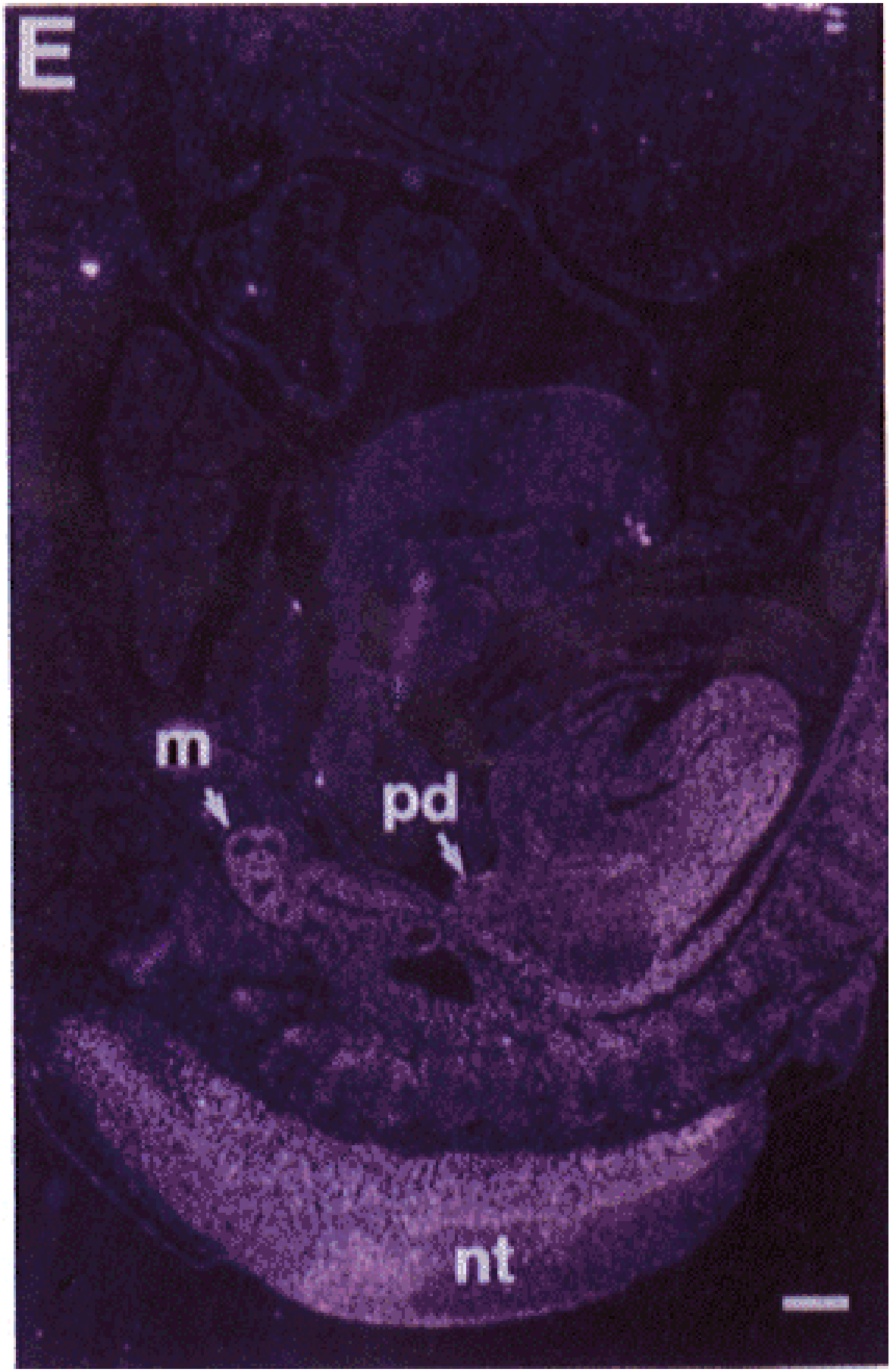
and the potential for shared regulatory regions. EMBO J. **11:**1825–1836.
45. **Simeone, A., M. Pannese, D. Acampora, M. D'Esposito, and E. Boncinelli.** 1988. At least three human homeoboxes on chromosome 12 belong to the same transcription unit. Nucleic Acids Res. **16:**5379–5390.
46. **Stornaiuolo, A., D. Acampora, M. Pannese, M. D'Esposito, F. Morelli, E. Migliaccio, M. Rambaldi, A. Faiella, V. Nigro, A. Simeone, and E. Boncinelli.** 1990. Human *HOX* genes are differentially activated by retinoic acid in embryonal carcinoma cells according to the position within the 4 loci. Cell Differ. Dev. **31:**119–127.
47. **Vershon, A. K., and A. D. Johnson.** 1993. A short, disordered protein region mediates interactions between the homeodomain of the yeast alpha 2 protein and the MCM1 protein. Cell **72:**105–112.
48. **Wilson, D., G. Sheng, T. Lecuit, N. Dostatni, and C. Desplan.** 1993. Cooperative dimerization of paired class homeo domains on DNA. Genes Dev. **7:**2120–2134.
49. **Wohlberger, C., A. Vershon, B. Liu, A. Johnson, and C. Pabo.** 1991. Crystal structure of the MATa2 homeodomain operator complex suggests a general model for homeodomain DNA interactions. Cell **67:**517–528.
50. **Yokouchi, Y., H. Sasaki, and A. Kuroiwa.** 1991. Homeobox gene expression correlated with the bifurcation process of limb cartilage development. Nature (London) **353:**443–446.
51. **Zavortink, M., and S. Sakonju.** 1989. The morphogenetic and regulatory functions of the *Drosophila Abdominal B* gene are encoded in overlapping RNAs transcribed from different promoters. Genes Dev. **3:**1969–1981.

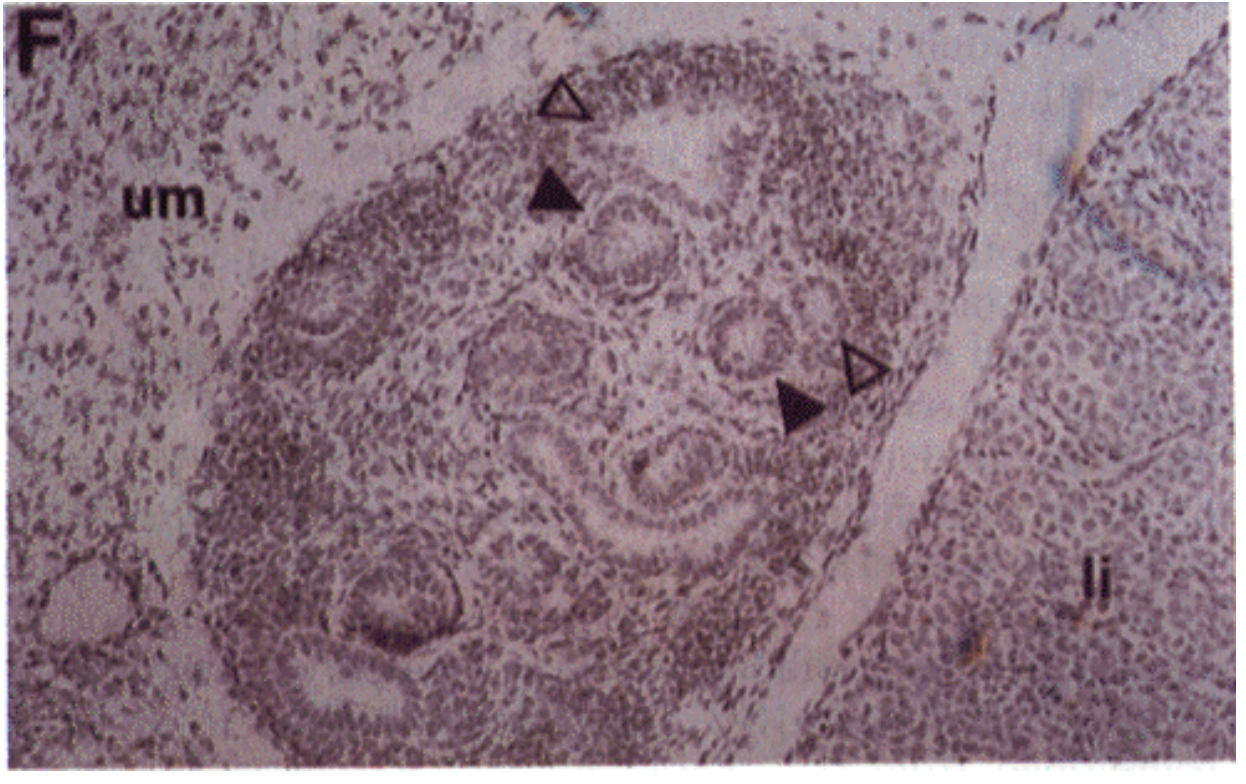

A
a
p

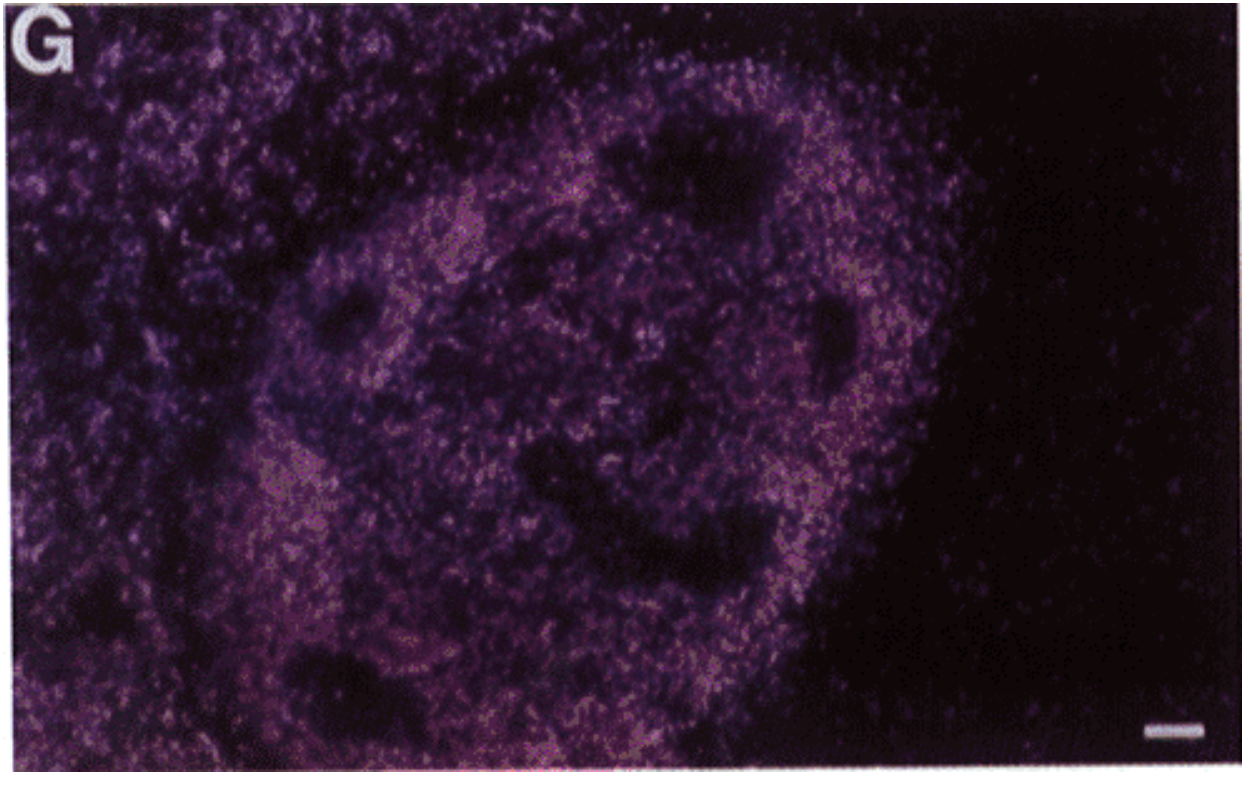

B


c
a
p


D
L1
u
gt
g


E
m
pd
nt


F
um
li


G