Characterization of a murine homeo box gene, Hox-2.6, related to the Drosophila Deformed gene

Anthony Graham, Nancy Papalopulu, Joanne Lorimer, John H. McVey, Edward G.D. Tuddenham, and Robb Krumlauf

Laboratory of Molecular Embryology, National Institute for Medical Research, London NW7 1AA, UK

The Hox-2 locus on chromosome 11 represents one of the major clusters of homeo-box-containing genes in the mouse. We have identified two new members (Hox-2.6 and Hox-2.7), which form part of this cluster of seven linked genes, and it appears that the Hox-2 locus is related by duplication and divergence to at least one other mouse homeo box cluster, Hox-1. The Hox-2.6 gene encodes a predicted protein of 250 amino acids, which displays extensive similarity in multiple regions to certain mouse, human, Xenopus, and zebra fish homeo domain proteins. The Drosophila Deformed (Dfd) gene also shares these same regions of similarity, and based on this sequence conservation, we suggest that Hox-2.6 forms part of a vertebrate ‘Dfd-like’ family. Hox-2.6 is expressed in fetal and adult tissues and is modulated during the differentiation of F9 teratocarcinoma stem cells. In situ hybridization analysis of mouse embryos shows that the Hox-2.6 is expressed in ectodermal derivatives: spinal cord, hindbrain, dorsal root ganglia, and the Xth cranial ganglia. In the central nervous system, expression is observed in the most posterior parts of the spinal cord, with the anterior limit residing in a region of the hindbrain and no expression in the mid- or forebrain. In mesodermal structures, Hox-2.6 is expressed in the kidney, the mesenchyme of the stomach and lung, and the longitudinal muscle layer of the gut. Expression has not been observed in derivatives of embryonic endoderm. The patterns of Hox-2.6 expression in both mesoderm and ectoderm are spatially restricted and may reflect a role for the gene in the response to or establishment of positional cues in the embryo.

[Key Words: Hox-2.6; homeo box; mouse embryo; Drosophila Deformed; in situ hybridization; teratocarcinoma cells]

Received June 3, 1988, revised version accepted August 25, 1988.

The homeo box was first discovered in Drosophila in genes that were implicated in controlling pattern formation by specifying segmental pattern and segment identity (for review, see Gehring 1987). This process involves the determination of positional values along the anteroposterior body axis. More recently, it has been shown that several Drosophila homeo-box-containing genes are expressed in pattern that suggest roles in positional specification, independent of the process of segmentation [Doyle et al. 1986; Hoey et al. 1986; Akam 1987; Gehring 1987].

The homeo box can encode a 61-amino-acid protein domain that has a putative helix-turn-helix motif. This motif is implicated in DNA-binding activity [Desplan et al. 1985; Fainsod et al. 1986; Cho et al. 1988], and it is possible that the biochemical function of the homeo domain is to direct DNA binding and potentially regulate transcription. Homeo-box-containing genes have been identified and cloned from the genomes of widely separated phylogenetic groups, including nonsegmented invertebrates and mammals [McGinnis 1985; Holland and Hogan 1986]. Homeo box genes are organized in clusters in the Drosophila and mammalian genomes. Murine homeo-box-containing genes represent a large gene family with at least 21 members, which are organized in several [at least four] gene clusters. The two major mouse homeo box clusters, Hox-1 and Hox-2, are located on chromosomes 6 and 11, respectively [Hart et al. 1985; Bucan et al. 1986]. To date, the Hox-1 cluster has at least seven members [Duboule et al. 1986; Rubin et al. 1987], and the Hox-2 cluster has five [Hart et al. 1987]; however, the size and organization of the clusters has not been completely established.

In general, very little overall similarity has been observed between the proteins of Drosophila homeo-box-containing genes and their vertebrate counterparts. The homeo box itself and a short hexapeptide region [Krumlauf et al. 1987] are the major related elements among many of these genes from different species. Recently, Regulski et al. [1987] presented data showing that the Drosophila Deformed (Dfd) gene exhibits similarity...
with two vertebrate genes not only in the homeo domain but also in other regions of the protein sequence. Information of this type suggests that vertebrate analogs of previously characterized Drosophila genes may exist. Therefore, it is important to determine and compare the coding information of more vertebrate homeo-box-containing genes to allow definition of closely related homeo box gene subfamilies, not based solely on the homeo box.

Whether conservation of structure reflects homologous functions for homeo box genes in the development of different species has yet to be shown. However, in view of the difficulty in defining mammalian genes responsible for regulating important developmental processes, they provide an opportunity to study molecules that are potentially involved in control during embryogenesis. The patterns of expression of mammalian homeo box genes during development are consistent with such a role (for review, see Feinberg et al. 1987). A number of groups have analyzed the expression of murine homeo box genes and found them to be temporally and spatially restricted in their patterns of expression (Gaunt et al. 1986; Dony and Gruss 1987). The murine homeo box genes appear to be expressed in similar but not identical overlapping domains, which often display differences in their anterior–posterior, dorsal–ventral, and medial–lateral levels of expression (Feinberg et al. 1987; Krumlauf et al. 1988; Gaunt 1988; Holland and Hogan 1988). The complex and overlapping patterns of transcription suggest that these genes may be involved in several different processes during embryogenesis, including the setup of positional cues.

The fact that vertebrate homeo box genes are found in similar types of clusters in all organisms so far analyzed is intriguing and may be a consequence of their modes of regulation or function. Thus, it is important to extend our understanding of the organization and expression of genes within these mouse clusters, to examine evolutionary or functional relationships between members of the homeo box gene family. This paper presents a further characterization of the Hox-2 locus, identifying two new genes (Hox-2.6 and Hox-2.7) and shows that there are at least seven linked members of this cluster. Analysis of the structure and expression of one of the new members, Hox-2.6, is detailed. The Hox-2.6 gene is expressed in a temporally and spatially restricted pattern during embryogenesis, and its predicted protein displays extensive identity to other homeo domain proteins. Based on multiple regions of similarity to the Drosophila Dfd gene, Hox-2.6 can be seen to be a member of a ‘Dfd-like’ family of vertebrate genes.

Results

Hox-2.6 gene structure and protein sequence

A number of overlapping genomic clones, which span the Hox-2 locus on chromosome 11, have been isolated from a mouse cosmid DNA library. Initially, a 220-bp fragment containing the Antp homeo box was used as a probe to isolate clones that span part of the locus. From these clones we were able to characterize the structure, sequence, and patterns of expression for one member of the locus, Hox-2.1 (Krumlauf et al. 1987). Subsequent genomic walking yielded more cosmid clones extending 5' and 3' of this region. Restriction mapping, sequencing, and hybridization analysis, using mouse and Drosophila homeo box probes, allowed construction of a map for the Hox-2 locus that positions seven homeo boxes over 80 kb [Fig. 1]. This map is in agreement with previous results assigning five genes, Hox-2.1 to Hox-2.5, to the locus (Hart et al. 1985, 1987b) and, in addition, identifies two new members of the locus, which we term Hox-2.6 and Hox-2.7. The Hox-2.6 and Hox-2.7 genes are, at present, the 3'-most members of the locus and lie ~13 kb and 31 kb downstream of Hox-2.1.

A 600-bp Hox-2.6 probe containing part of the homeo box and 3'-flanking sequences was generated from a subcloned 4.4-kb BamHI fragment [BgIII–BamHI, probe 1, Fig. 1]. This probe was used to isolate a number of Hox-2.6 cDNA clones from a λgt10 library, prepared from 8.5-day p.c. mouse embryo RNA (Fahrner et al. 1987). The two longest cDNA clones (2 and 7, Figs. 1 and 2) and genomic clones from the Hox-2.6 region were sequenced. Sequence analysis reveals that the cDNAs contain a long open reading frame encoded by two exons, which in the genomic DNA are separated by a 790-bp intron, and contain consensus splice donor-acceptor sites. Hox-2.6 is transcribed in the same direction as Hox-2.1, and it appears that all seven genes in the locus have the same 5'→3' orientation (Hart et al. 1987b). Figure 2 shows complete sequence of the two cDNAs and some additional 3' sequence from a genomic subclone. The size of the major Hox-2.6 transcript, estimated by Northern RNA analysis (see Fig. 4), is 2.4 kb, and the cDNA clones account for only 1300 bp. The 3' ends of these clones have no adjacent polyadenylation signal and appear to have arisen by specific internal priming from an A-rich region between position 998 and 1104 [Fig. 2]. There is a potential polyadenylation signal (AATAAA) at position 890 in the 3'-untranslated region; however, RNase protection experiments have indicated that this site is not used [data not shown]. Genomic probes 3' of the cDNA clones detect mRNA, suggesting that most of the sequence missing from our clones is derived from the 3'-untranslated end of the mRNA. Also contained within the 3'-untranslated region of Hox-2.6 is an AUUUA motif implicated in modulating the level of mRNA stability (Shaw and Kamen 1986).

Conceptual translation of the sequence in Figure 2 reveals that both cDNAs contain a long open reading frame. The first AUG, 36 bp after an in-frame stop codon, has been denoted the methionine-initiator codon. The predicted protein contains 250 amino acids, corresponding to a predicted molecular weight of 27,500, with an in-frame hexapeptide (Krumlauf et al. 1987) and homeo domain in the carboxy-terminal portion of the protein. Like Hox-2.1 and many other homeo-domain-containing proteins, Hox-2.6 is rich in prolines (18%), including a run of 14 consecutive prolines in the amino-terminal region.
Evolutionary relationships to the Hox-2.6 protein—a Dfd family

Sequence comparisons of the Hox-2.6 homeo domain with other homeo domain sequences are summarized in Table 1. The Hox-2.6 homeo domain shows 83% amino acid identity to that of Antp [McGinnis et al. 1984] and is therefore a member of the ANT-C/BX-C class of homeo domain. Considering murine homeo domains, the Hox-2.6 homeo domains show the closest similarity to those of Hox-1.4 and Hox-5.1, both with 93% amino acid identity [Duboule et al. 1986; Featherstone et al. 1988]. In addition, sequences immediately adjacent to the Hox-1.4 and Hox-5.1 homeo domains, plus the hexapeptides and amino-terminal regions, show extended region of identity with Hox-2.6 [Featherstone et al. 1988; B. Galliot, pers. comm.], these genes appear to form a subfamily of related genes.

Significant similarities to Hox-2.6 are also found in other vertebrate homeo-domain-containing proteins. These are the human HHO.c13 [Mavillo et al. 1986], the Xenopus Xho 1A [Harvey et al. 1986], and the zebra fish ZF 13 [Njolstad et al. 1988] genes. Both HHO.c13 and Xho 1A show 95% amino acid identity within the homeo domain, whereas ZF 13 shows 98% amino acid identity. All three proteins exhibit a high degree of identity in sequences immediately flanking the homeo domain, with those of ZF 13 being closest to Hox-2.6 [Njolstad et al. 1988].

An alignment comparison between Hox-2.6, Xho 1A, and HHO.c13 over the entire protein shows that sequences in addition to the homeo domain are highly conserved in these proteins (Fig. 3A). The three proteins are different in size (232–255 amino acids), but if positioned for maximal alignment, one can see substantial conservation in the amino terminal region. This is more extensive in the case of Xho 1A than in the case of HHO.c13, yet in both instances, the first 30–35 amino acids show a high degree of sequence similarity. Within these amino acids, many of the differences represent conservative changes, and there is a block of 12 amino acids that are identical in all three proteins. The other region of notable similarity in these proteins begins around the hexapeptide and extends through the homeo domain and the first five to six amino acids on its carboxy-terminal side. The remaining regions of these proteins are highly divergent.

It was noted previously that HHO.c13 and Xho 1A shared related regions with the Drosophila homeo-domain-containing protein Dfd [Regulski et al. 1987]. We therefore compared Hox-2.6 with Dfd, and Figure 3B diagrams blocks of sequence similarity identified with an optimal alignment between the Hox-2.6 and Dfd proteins. The Dfd protein is much larger than Hox-2.6 (586 vs. 250 amino acids). This difference appears to be the consequence of several regions rich or repetitive in single amino acids, such as Gly, His, Gln, or Asn, which are not present in the mouse gene. The Hox-2.6 homeo domain has 89% amino acid identity with the Dfd homeo domain. The overall amino acid identity to Dfd is 40% (102/250 amino acids) and is more extensive
Figure 2. Nucleic acid and protein sequence of Hox-2.6 from cDNA and genomic clones. The longest open reading frame and its conceptual translation are shown, and numbering starts on the first AUG 36 bp downstream of an in-frame stop codon. The intron is indicated by the large arrowhead. The beginning and end of the two cDNA clones (2 and 7), shown in Fig. 1, are marked by small arrowheads and their respective numbers. The open boxes outline the hexapeptide and the homeo domain.

\[ \text{Hox-2.6 structure and expression} \]

The Hox-2 and Hox-1 clusters are related by duplication

To extend our analysis of the Hox-2 locus further, we have isolated the Hox-2.7 gene and sequenced the homeo domain [Fig. 4]. It shares only 68% amino acid identity [42/61] with the Antp homeo domain and, among other Drosophila homeo domains, is most closely related to S602Z1 [Doyle et al. 1986]. The predicted sequence in the homeo domain varies in only two positions with mhl9, also assigned to chromosome 11 [Lonai et al. 1987], and may represent the same gene. When compared with mouse homeo domains, Hox-2.7 displays the strongest identity with the Hox-1.5 domain and flanking regions [Fig. 4]. This represents another example [like Hox-2.6] of a gene in the Hox-1 cluster that is apparently more highly related to a gene in the Hox-1 cluster than to other Hox-2 members. A comparison of the sequence and organization of the seven genes in the Hox-2 cluster with the Hox-1 cluster sheds some light on the possible genealogy of these murine homeo box genes. The genes in both clusters are transcribed in the same 5'→3' orientation, and based on intergenic distances and sequence comparisons between members of the Hox-1 [Duboule et al. 1986] and Hox-2 clusters, it is possible to align some of the genes in these loci as shown in Figure 5.

Table 1. Homeo domain homologies to Hox-2.6

<table>
<thead>
<tr>
<th>Gene</th>
<th>Hox-2.6</th>
<th>Drosophila</th>
<th>Xenopus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antp</td>
<td>82%</td>
<td>[50/61]</td>
<td>McGinnis et al. 1984</td>
</tr>
<tr>
<td>Dfd</td>
<td>89%</td>
<td>[54/61]</td>
<td>Regulski et al. 1987</td>
</tr>
<tr>
<td>ZF 13</td>
<td>98%</td>
<td>[60/61]</td>
<td>P.R. Njostad et al. [pers. comm.]</td>
</tr>
<tr>
<td>Xbox 1A</td>
<td>95%</td>
<td>[58/61]</td>
<td>Harvey et al. 1986</td>
</tr>
<tr>
<td>HHO.c13</td>
<td>93%</td>
<td>[57/61]</td>
<td>Mavillo et al. 1986</td>
</tr>
<tr>
<td>Hox-1.4</td>
<td>93%</td>
<td>[57/61]</td>
<td>Duboule et al. 1986</td>
</tr>
<tr>
<td>Hox-5.1</td>
<td>93%</td>
<td>[57/61]</td>
<td>Featherstone et al. 1988</td>
</tr>
<tr>
<td>Hox-2.1</td>
<td>89%</td>
<td>[54/61]</td>
<td>Krumlauf et al. 1987</td>
</tr>
<tr>
<td>Hox-2.3</td>
<td>82%</td>
<td>[50/61]</td>
<td>Meijlink et al. 1988</td>
</tr>
<tr>
<td>Hox-2.7</td>
<td>75%</td>
<td>[46/61]</td>
<td>Fig. 4 [this paper]</td>
</tr>
</tbody>
</table>
In cases where only the homeo domain and flanking sequences have been determined, comparisons reveal that the pairs Hox-2.7–Hox-1.5 and Hox-2.2–Hox-1.2 (Colberg-Poley et al. 1985b; Hart et al. 1987a) display the highest levels of sequence identity with each other, sharing in common specific amino acid changes when compared with the Antp homeo domain. The Hox-2.6 gene shows the strongest similarity with the Hox-1.4 homeo domain and surrounding sequences (as described in the previous section), and several other regions of the two proteins are also related (D. Duboule, pers. comm.). Comparisons of other members of these clusters over the entire predicted protein sequences show that Hox-2.1–Hox-1.3 (Krumlauf et al. 1987; Odenwald et al. 1987; Fibi et al. 1988) and Hox-2.3–Hox-1.1 (Kessel et al. 1987; Meijink et al. 1987) are closely related to each other. The related region spans the hexapeptide region, the homeo domain plus surrounding sequence, and a

Figure 3. Evolutionary relationships between Hox-2.6 and homeo domain proteins of other species. (A) Comparison of a related subclass of vertebrate homeo domain proteins, including mouse (Hox-2.6), human (HHO.c14, Mavillo et al. 1986), and Xenopus (Xhox 1A, Harvey et al. 1986). The boxed region represents the homeo domain, and dashes indicate amino acid identity. (B) Diagrammatic representation of the relationship between Hox-2.6 and Drosophila Dfd. The shaded areas represent amino acid identity, and dashed lines demark related areas in the two proteins. (C) Generalized structure showing characteristic features and conserved regions based on comparison of several vertebrate homeo domain proteins. [AA] amino acid.
large region (30–50 amino acids) at the amino-terminal end of the proteins [Fibi et al. 1988; Papalopulu et al. 1988; J. Garbern et al., unpubl.]. Taken together, the data demonstrate that at least five genes in Hox-2 are related to five genes in Hox-1 via duplication and divergence of an ancestral cluster (Fig. 5). No mouse counterparts have been found as yet for the Hox-1.6 gene in the Hox-2 cluster, or for the Hox-2.4–Hox-2.5 genes in the Hox-1 cluster. However, it seems likely that further characterization of the overlapping regions of these clusters will reveal more related members and define the limits of the duplication.

**Developmental expression of Hox-2.6**

To begin to define the pattern of Hox-2.6 expression in the embryo and adult, we analyzed poly(A)+ RNA extracted from teratocarcinoma cell lines and tissues at various stages by Northern hybridization. In view of the extensive sequence identity between Hox-2.6 and other subgroup members, such as Hox-1.4 and Hox-5.1, it was important to select probes that were specific for the Hox-2.6 gene. The carboxy-terminal region of these related proteins and the 3′-untranslated regions are highly diverged; therefore, the probe used was a single-stranded antisense RNA from this region of Hox-2.6 (Fig. 1, probe 1). Although it does contain the last 15 amino acids of the homeo domain, using high-stringency washing conditions with RNase A treatment [see Methods; Krumlauf et al. 1987], this probe does not detect transcripts or cloned DNA from the related genes (Hox-1.4 and Hox-5.1) in control experiments and is specific for Hox-2.6.

In tissues expressing Hox-2.6, the probe detects at least six transcripts ranging from 2.4 to 10 kb, including two major species at 2.4 and 4 kb [Fig. 6A]. The complex pattern of transcription reflects multiple transcripts from the Hox-2.6 gene and is analogous to results that we obtained with another member of the locus, Hox-2.1 [Krumlauf et al. 1987]. The direct comparison of these adjacent genes in the Hox-2 cluster [Fig. 6] shows that they both have multiple transcripts that differ in size and intensity from each other. The bands do not appear to be unprocessed precursors, because intron probes do not hybridize to the multiple transcripts. It is possible that they represent RNAs capable of encoding multiple proteins from the same gene, as has been observed for a *Xenopus* homeo box gene [Cho et al. 1988].

The major tissues that express Hox-2.6 are the lung, spinal cord, and kidney, and a selected time course of expression in these tissues is shown in Figure 6A. The expression in lung is observed in all fetal, neonatal, and adult stages, and we have also detected transcripts by Northern analysis as early as 14.5 days p.c. The levels or complexity of expression in the lung do not appear to change with time, but the 5-kb transcript is reduced in abundance when compared with other tissues. In the spinal cord, Hox-2.6 is expressed at the same levels in 14.5-day and 16.5-day embryos and is also present in the 18.5-day embryo, but we have not examined expression in the adult spinal cord. In contrast to the constant relative levels of Hox-2.6 mRNA in the lung and spinal cord during development, expression in the kidney undergoes a 10-fold drop between neonatal (1 week) and adult (6 weeks) stages. This same pattern of kidney expression is observed with the Hox-2.1 gene [Krumlauf et al. 1987]. Hox-2.1 is expressed in the adult kidney [Jackson et al. 1985] [although not observable on the exposure shown in Fig. 6B] at greatly reduced levels when compared with fetal and neonatal stages. Hox-2.6 is also expressed in adult testes and fetal liver, stomach, and gut but not in adult heart, liver, or spleen [data not shown]. In general, it seems that Hox-2.6 is expressed at a higher level in most tissues and developmental stages examined relative to Hox-2.1, with the exception of spinal cord. Both Hox-2.6 and Hox-2.1 show a reduction at the level of high molecular transcripts in the lung, suggesting tissue-specific differences in transcription or processing of these complex RNA species.

Based on the isolation of cDNA clones from an 8.5-day p.c. library, Hox-2.6 is expressed in 8.5-day p.c. embryos. It is difficult to obtain sufficient material to examine
earlier stages by Northern blots, but there have been a number of reports describing patterns of expression of murine homeo box genes in the F9 embryonal carcinoma cell line (Colberg-Poley et al. 1985a, b; Breier et al. 1986; Baron et al. 1987; Kessel et al. 1987; Rubin et al. 1987). The F9 cells are able to differentiate in vitro in response to chemical inducers to form derivatives of the extraembryonic endoderm, and we have used them to determine whether Hox-2.6 might be active in early stages of embryonic differentiation. Expression of Hox-2.6 was analyzed, by Northern blots, in stem cells and at 1, 3, and 5 days after retinoic-acid-induced differentiation (Fig. 6A). Hox-2.6 transcripts are present in undifferentiated F9 stem cells at low levels. Transcript levels are increased after 1 day of differentiation to near maximal amounts and maintained at this high level throughout the remainder of the differentiation time course. This is in direct contrast to Hox-2.1, which is undetectable in F9 stem cells and accumulates to high levels after 1 day of induction (Fig. 6B). This accumulation is transitory, and levels continually fall during further differentiation stages to very low levels by day 5. Treatment of F9 stem cells with the protein synthesis inhibitor cycloheximide also induced Hox-2.6 and Hox-2.1 RNA levels (Fig. 6). In the case of Hox-2.1, this induction is not mediated by transcription but reflects mRNA instability involving the AUUUA motif (Shaw and Kamen 1986) in the 3'-untranslated region of the Hox-2.1 mRNA (R. Krumlauf et al., in prep.). Similar sequences reside in the 3' end of the Hox-2.6 mRNA (Fig. 2), and cycloheximide induction suggests that Hox-2.6 transcripts are also unstable in F9 stem cells. Thus, both Hox-2.1 and Hox-2.6 show an increase in expression following cycloheximide treatment and initial differentiation of F9 stem cells. However, the kinetics or temporal pattern of their response differs and depends on the de-

**Figure 6.** Northern analysis of Hox-2.6 and Hox-2.1 in F9 teratocarcinoma cells and mouse tissues from different developmental stages. The tissue of origin and stage of poly(A)+ RNA sample are indicated above each lane. In all tissue samples, 2 μg of RNA per lane was loaded, and the F9 cell samples contained 2.5 μg per lane. (A) Same filter as B reprobed with Hox-2.6; (B) filter probed with Hox-2.1. (RA) Cells treated with retinoic acid (5 × 10⁻⁸); (CHX) cells treated with cycloheximide (75 μg/ml) for 6 hr.
degree of differentiation of the F9 cells. In summary, the Northern analysis shows that Hox-2.6 and Hox-2.1 display similar overlapping patterns that only differ in the timing or level of expression in cell lines and a number of mouse tissues.

Hox-2.6—in situ hybridization

In situ hybridization has been used to extend and complement the analysis of expression by Northern blots that described the general patterns of Hox-2.6 expression. Antisense probes (probes 1 and 2, Fig. 1) and a sense probe were hybridized to sections of 12.5- and 14.5-day mouse embryos. We chose these stages, as they represent the stages of maximal expression of Hox-2.6. The sense probe acts as a control for nonspecific probe ‘stickiness’ to the sections. This probe did not highlight any features of the 12.5-day mouse embryo except red blood cells.

Of the three germ layers of the embryo, Hox-2.6 is expressed in derivatives of mesoderm and ectoderm but not in derivatives of endoderm. More specifically, Hox-2.6 is expressed in the following ectodermal derivatives: spinal cord, hindbrain, and some ganglia of the peripheral nervous system (Fig. 7). Hox-2.6 is not detected in the forebrain or midbrain; hence, expression within the central nervous system is restricted to more posterior regions. Transcripts are detected all along the spinal cord, from posterior portions (Fig. 7A) into the hindbrain (Fig. 7B,D). The anterior limit of expression maps to the hindbrain and does not seem to correlate with any obvious morphological change. This is clearly observed at 14.5 days, where one can see expression of Hox-2.6 extending along the spinal cord past the cervical flexure into the hindbrain (Fig. 7B). There appears to be no posterior limit of expression, and Hox-2.6 can be detected in the most posterior portions of the CNS. In contrast to Hox-2.1 and other homeo box genes that have been analyzed by in situ hybridization, Hox-2.6 is expressed uniformly in transverse sections of the spinal cord [Awgulewitsch et al. 1986; Toth et al. 1987; Holland and Hogan 1988] (Fig. 7C). It shows no dorsoventral or mediolateral spatial restriction at 12.5 days. Hox-2.6 is expressed in the peripheral nervous system in dorsal root ganglia and at very high levels in a ganglion that is probably the nodose ganglion of the Xth cranial nerve [Graham et al. 1988]. Figure 7D shows expression of the gene in the hindbrain, in a dorsal root ganglion, and in the nodose ganglion. Expression of Hox-2.6 has not been observed in other cranial ganglia. The CNS is derived from the neural tube, and the peripheral nervous system has a neural crest origin. Thus, Hox-2.6 is expressed in both neural tube and neural crest derivatives.

Expression of Hox-2.6 in mesodermal structures is also spatially restricted. Hox-2.6 is expressed in the mesenchyme of the stomach and the lung, in the kidney, and in the longitudinal muscle layer of the gut. Figure 8 shows clearly the restriction of the expression of this gene to mesodermal tissues and not to endodermal tissue. This is obvious in the lung at 12.5 and 14.5 days, where the inner endodermal epithelial layer of cells is not expressing and the outer mesenchymal layer of cells is expressing Hox-2.6 (Fig. 8A,B). This mesodermal–endodermal restriction can also be seen in the stomach (Fig. 8C), whereas we can observe Hox-2.6 transcripts in the muscle layer of the gut in Figure 8D. The restriction of expression to the longitudinal muscle and not the circular muscle is more obvious at 14.5 days (Fig. 8D). Expression of Hox-2.6 has not been observed in other mesodermal organs such as the heart, the limbs, or those derived from somites. Those tissues that do express Hox-2.6 are derived from intermediate and lateral plate mesoderm.

This overall pattern of expression is very similar to the one observed for Hox-2.1 [Holland and Hogan 1988]. One exception is the extent of expression that each gene shows along the rostral caudal axis. Most homeo box genes examined by in situ hybridization are expressed in the CNS; however the domains are overlapping but not identical and often display differences in the anterior limits of expression [Gaunt et al. 1988]. Figure 7 shows near adjacent sections of the hindbrain region of a 12.5-day mouse embryo that have been probed for either Hox-2.6 or Hox-2.1 (the Hox-2.1 probe is probe 1 of Krumlauf et al. 1987). One can clearly see that the expression of Hox-2.6 extends more anteriorly than that of Hox-2.1. The limits of expression of either gene do not seem to correspond to any clear morphological or biological cues at this embryonic stage. Thus, Hox-2.6 shows similar but subtly different patterns of expression to those of Hox-2.1 and, in common with all other homeo box genes, is spatially restricted in its patterns of expression during development. In view of the structural similarity between Hox-2.6 and other members of the Dfd-like subgroup, Hox-1.4 and Hox-5.1, it will be very interesting to determine whether their patterns of expression are also highly related.

Discussion

In this paper we present data relating to the characterization of a new murine homeo box gene, Hox-2.6. This gene and Hox-2.7, another new gene we have identified, form part of a group of seven tightly linked homeo box genes in the Hox-2 cluster located on mouse chromosome 11 [Fig. 1]. To date, nearly 25 murine homeo box-containing genes have been identified. Therefore, the Hox-2 locus represents part of a large gene family related by at least one common element, the homeo box. Sequence analysis reveals a structure for Hox-2.6 that is similar to that of many other murine and vertebrate homeo box genes. Typically, Hox-2.6 is comprised of two exons separated by an intron of about 1 kb [790 bp]. The protein-coding information [250 amino acids] is located in the 5’ portion of the mRNA, which has a long 3’-untranslated region. A conserved heptapeptide is located close to the homeo box but on a separate exon, with the splice site located 15–40 bp upstream of the homeo box. The predicted proteins are very rich in prolines and serines, and the distribution of these prolines
Figure 7. Expression of Hox-2.6 in ectodermal derivatives, as shown by in situ hybridization to embryo sections. (A) Hox-2.6 expression in posterior spinal cord of a 12.5-day p.c. embryo. (B) Expression of Hox-2.6 at the junction of the hindbrain and spinal cord of a 14.5-day p.c. embryo. (C) Transverse section of thoracic spinal cord of a 12.5-day p.c. embryo probed for Hox-2.6. (D) Expression of Hox-2.6 in hindbrain and constituents of the peripheral nervous system of a 12.5-day p.c. embryo. Bright-field and dark-ground pictures are shown. (RBC) Red blood cells; (SC) spinal cord; (HB) hindbrain; (DRG) dorsal root ganglion; (NG) nodose ganglion; (PV) pre-vertebrae.

throughout the non-homeo-domain portion of the protein should generate a nonhelical structure. Some of these features are summarized in Figure 3C, which presents a generalized structure for these proteins.

Hox-2.6 was found to be related to other mouse homeo-domain proteins (Hox-1.4 and Hox-5.1), based on sequence identity in several domains of the protein. This pattern of identity between different mouse genes appears to be the result of duplication and divergence of an ancestral homeo box cluster. Sequence comparisons of the Hox-1 and Hox-2 clusters reveal that at least five members of each cluster are highly related to members in the opposite cluster, suggesting a duplication event [Fig. 4]. The relationship of other homeo box clusters to Hox-1 and Hox-2 is not well established. However, based on the identity between Hox-2.6 and Hox-5.1, some other genes in the Hox-3 and Hox-5 clusters may also be related by duplication to Hox-1 and Hox-2. Consequently, some mouse genes may form a subgroup with as many as four highly related members.
Figure 8. Expression of Hox-2.6 in mesodermal derivatives, as shown by in situ hybridization, to embryo sections. (A) Hox-2.6 transcripts in the lung mesenchyme in transverse sections of 12.5-day p.c. embryo. (B) Lung of a 14.5-day p.c. embryo probed for Hox-2.6. (C) Hox-2.6 expression in the stomach mesenchyme of a 12.5-day p.c. embryo. (D) Expression of Hox-2.6 in the longitudinal muscle of a 14.5-day p.c. embryo. Bright-field and dark-ground pictures are shown. (E) Endoderm; (M) mesenchyme; (PV) pre-vertebrae; (LM) longitudinal muscle; (CM) circular muscle.

In a larger context, one can note relationships between a number of homeo box genes from a range of different organisms. Hox-2.6 not only shows similarity to two mouse genes (Hox-1.4 and Hox-5.1) but also to a human gene, HHO.cl3 (Mavillo et al. 1986), a Xenopus gene, Xhox 1A (Harvey et al. 1986), a zebra fish gene, ZF 13 (Njolstad et al. 1988), and the Drosophila gene, Dfd (Regulski et al. 1987). High levels of sequence identity are present in regions other than the homeo domain, including the hexapeptide and amino-terminal sequences.

We suggest that these genes represent a Dfd-like family, based on their multiple regions of similarity. However, in view of the fact that there appear to be at least three murine genes that all show similar degrees of identity to Dfd, it is clear that there is not a precise murine equivalent to the Dfd gene.

It may generally be possible to classify several of the nonlinked homeo box genes from different species into subgroups, based on conserved sequences in their amino-terminal region (referred to as the conserved
amino-terminal subfamily domain in Fig. 3C). This conservation in several species supports the idea that these domains may be functionally significant. *Hox-2.1* forms another such subgroup with the zebrafish *ZF 21* gene [P.R. Njolstad et al., pers. comm.] and the Xenopus X1H box 4 gene [Fritz and De Robertis 1988]. Some of these subfamilies may also be related to other *Drosophila* homeo domain proteins in a manner analogous to *Hox-2.6* and *Dfd*.

The *Hox-2.6* gene, as analyzed by in situ hybridization and Northern blots, is expressed in fetal and adult tissues. Major sites of expression are the lung, kidney, central and peripheral nervous systems, stomach, gut, and testis. Within visceral organs, *Hox-2.6* is expressed in derivatives of mesoderm but not endoderm. This is shown in Figure 8, which illustrates that expression is restricted to the mesoderm of the stomach, gut, and lung, and is not seen in the endodermal epithelium. Within ectodermal derivatives, *Hox-2.6* is expressed in the CNS, in the spinal cord and hindbrain, and in the PNS, in the dorsal root ganglia, and in the nodose ganglia, as shown in Figure 7.

The pattern of *Hox-2.6* expression that we observe in the 12.5-day embryo is complex and does not correspond to any obvious biological or morphological signal. It seems probable that *Hox-2.6* is responding to anteroposterior positional cues but is not doing so in any simple or exclusive fashion. It is clear that two structures derived from the same germ layer and sharing the same anteroposterior position do not necessarily both express this gene. This is illustrated well in the case of the lung mesenchyme, which does express this gene, and the limbs, which do not. Both structures share a similar anteroposterior position but exhibit differential expression of this gene. Thus, if position is of importance in determining the patterns of *Hox-2.6* expression, it is likely to have been modified by other cues to produce the observed pattern at 12.5 days. This is comparable to the patterns of expression that are observed in later stage *Drosophila* embryos, where the more simplistic earlier patterns of expression have been modified during development [Martinez-Arias et al. 1987]. Hence, a simple early pattern of expression in mouse embryos may be masked later in embryogenesis.

There are similar regions of expression shared between vertebrate and *Drosophila* homeo box genes in comparable stages of development. It has been shown that a number of *Drosophila* homeo box transcripts accumulate to their highest levels in the embryonic CNS [for review, see Doe and Scott 1988]. Expression in the CNS is observed with a large number of homeo box genes from a wide range of species. Expression of homeo box genes in the CNS has been described for humans [Simeone et al. 1986], for Xenopus [Carrasco and Malacinski 1987; Condie and Harland 1987], and the zebrafish [Njolstad et al. 1988]. With regard to the mouse, we have shown that *Hox-2.6* is expressed in the CNS, and this has also been demonstrated with *Hox-1.2* [Toth et al. 1987], *Hox-1.3* [Dony and Gruss 1987], *Hox-1.4* [Toth et al. 1987, Wolgemuth et al. 1987], *Hox-1.5* [Fainsod et al. 1987; Gaunt 1987], *Hox-1.6* [Baron et al. 1987], *Hox-2.5* [Feinberg et al. 1987], *Hox-2.3* [Deschamps et al. 1987], *Hox-2.1* [Holland and Hogan 1988], and En-1 and En-2 [Joyner and Martin 1987]. These genes are expressed in an overlapping pattern along the anteroposterior axis of the CNS, as illustrated for *Hox-2.1* and *Hox-2.6* in Figure 9, but they also exhibit subtle differences in their mediolateral and dorsoventral planes of expression. *Hox-2.6* is expressed uniformly in cross sections of the spinal cord [Fig. 5C], yet *Hox-2.1* is expressed in a dorsally restricted manner [Holland and Hogan 1988].

Homeo box genes are clearly implicated in determining the identity of cells in the *Drosophila* blastoderm, and recent studies would also suggest that they may act in an analogous manner in the *Drosophila* nervous system [Doe et al. 1988]. For example, it has been reported that *ftz* in *Drosophila* and *mec-3* in *Caenorhabditis elegans* [Way and Chalfie 1988] are necessary for specification of neuronal identity. Thus, homeo box genes in vertebrates, as well as in invertebrates, may possibly act in controlling neuronal fate. The specification of neuronal fate is certainly a very complex system, which may require a family of proteins that can act in concert to define cell fate/position. The large size of the family of murine homeo box genes and the fact that they all seem to be expressed in the CNS make them good candidates for performing such a role.

The three *Dfd*-related murine genes, *Hox-2.6*, *Hox-1.4*, and *Hox-5.1*, may act potentially in an antagonistic or synergistic manner. It will be of great interest to compare their patterns of expression and regulation in more detail. The one feature that may be shared is the extent of expression of these genes in the ectoderm along the rostrocaudal axis at comparable stages of development. *Dfd* is expressed near the rostral end of the *Drosophila* embryo [Martinez-Arias et al. 1987], and we have shown that *Hox-2.6* is expressed along the spinal cord into the hindbrain of the murine embryo. *Hox-1.4* is also expressed in the murine hindbrain [Toth et al. 1987], as is *Hox-5.1* [Featherstone et al. 1988]; and Northern analysis shows *HHO.c13* transcripts in the brain but not in the forebrain [Mavillo et al. 1986]. We have also presented evidence illustrating that *Hox-2.6* is expressed more anteriorly than the adjacent *Hox-2.1* gene [Fig. 9], and it has been reported that *Hox-2.5* is not expressed in the hindbrain but is restricted to the spinal cord [Feinberg et al. 1987]. Therefore, it would appear that there is a correlation between the anterior limit of expression of a gene in the ectoderm and the relative position within a cluster, as has also been observed in *Drosophila* [Akam 1987]. This correlation may also extend to the subfamily to which that given gene is a member.

**Methods**

**Isolation of clones and sequencing**

Cosmid clones were isolated from a mouse [129] genomic library, prepared with the vector pcos2EMBL (a gift from Anna-Marie Frischauf), by screening with subcloned probes from the *Hox-2.1* gene [Krumlauf et al. 1987]. The clones were mapped
and flanking regions subcloned to be used for further screening of overlapping cosmids. Homeo box regions in the cosmids were identified by hybridization with a Drosophila Antp [provided by Walter Gehring] and mouse (Hox-2.1) homeo box probe at reduced stringency according to Holland and Hogan [1986]. Homeo box fragments were then subcloned and used to screen a mouse Agt10 cDNA library [Fahrn et al. 1987] prepared from 8.5-day embryos for Hox-2.6 clones. All cosmid and cDNA subcloning was performed using the vector Bluescript pKSM13+ [Stratagene]. Double-stranded DNA from the cDNA and genomic clones was sequenced by the dideoxy chain-termination method, using the array of primers in the Bluescript vector and the Sequenase polymerase [US Biochemical Corporation] according to manufacturer’s directions. All sequencing was performed in both directions using both the standard and dITP reactions to avoid the compressions in GC-rich regions.

Cell culture

F9 stem cells were maintained in Dulbecco’s modified eagle’s medium [DMEM], supplemented with 10% fetal calf serum on gelatinized plastic tissue culture dishes. The cells were induced to differentiate by adding retinoic acid [5 X 10^-8 M], cAMP [10^-4 M], and IMBX [10^-4 M] to cultures, and the cells were maintained in this differentiation media for 1–5 days, with the media being changed every 2 days. In the case where stem cells were treated with the protein synthesis inhibitor cycloheximide, the inhibitor was added to the cultures to a final concentration of 75 μg/ml for a total of 6 hr before harvesting. Under these conditions in control experiments, 95% of protein synthesis is blocked and 60% of the cells can recover to reinitiate protein synthesis when the inhibitor is removed. Cells were harvested for RNA at various time points by washing twice with PBS and lysing them on the plate in LiCl urea.

RNA isolation and Northern hybridization

Poly(A)^+ mRNA was isolated from F9 cells or mouse tissues essentially as described in Krumlauf et al. [1987]. Briefly, the samples were homogenized in 3 M LiCl–6 M urea on ice for 2
min and stored overnight at 4°C to precipitate the RNA. The RNA was harvested by centrifugation, redissolved in 10 mM Tris-HCl (pH 7.6), 1 mM EDTA, and 0.5% SDS and extracted several times with phenol : chloroform. Following ethanol precipitation, the poly(A)+ mRNA was isolated by oligo(dT) cellulose chromatography. The RNA samples were electrophoresed in denaturing formaldehyde-agarose gels, transferred to GeneScreen (DuPont) in 20 x SSC, and coupled to the membrane by UV cross-linking and baking, as described in Krumlauf et al. [1987]. The filters were hybridized using 3P-labeled single-stranded antisense RNA probes produced by in vitro transcription of subcloned Hox-2.1 and Hox-2.6 cDNA clones with T7 or T3 RNA polymerase. Probes were selected from regions of the genes that had a minimum identity with other subgroups members to minimize any cross hybridization. Hybridization conditions were 60% formamide, 5 x SSC, 0.1% bovine serum albumin, 0.1% Ficoll 400, 0.1% polyvinylpyrrolidone, 20 mM sodium phosphate (pH 6.8), 1% SDS, 7% dextran sulfate, 100 μg/ml trNA, and 10 μg/ml poly(A) at 65°C for 16 hr. The filters were washed in 1 x SSC, 0.5% SDS at 80°C for 2 hr, and in the cases where high-stringency conditions were required to ensure that the signal was not derived from related genes, the filter was treated further with RNase A. The filters were incubated in 2 x SSC, 0.1% SDS at 52°C for 1 hr. Probe was removed from filters (not treated with RNase A) for rehybridization by washing in 70% formamide at 80°C for 30 min and reexposing the filter to ensure probe removal.

In situ hybridization

The protocol used was basically that of Wilkinson et al. [1987], with some modifications, and is essentially as follows. Mouse embryos were fixed in 4% paraformaldehyde in phosphate-buffered saline (PBS) overnight at 4°C, and then embedded in paraffin wax. Six-micrometer sections were cut and dried onto gelatin-subbed slides. The sections were dewaxed in xylene, treated with 0.2 M HCl, refixed in paraformaldehyde, subjected to proteinase K, and then treated with acetic anhydride. After rehybridization, probes were redissolved at a final activity of 1.2 x 10^6 dpm/kb/μl in hybridization buffer (50% formamide, 0.3 M NaCl, 20 mM Tris, 5 mM EDTA, 10 mM NaPO₄, 0.1% dextran, 1 x Denhardt’s, rRNA). Hybridization was overnight at 50°C. The sections were washed in 65°C in 50% formamide, 2 x SSC, 100 mM dithiothreitol (DTT) for 40 min, followed by incubation with RNase A at 20 μg/ml in 0.5 M NaCl, 10 mM Tris, 5 mM EDTA for 30 min. Sections were washed again in 50% formamide, 2 x SSC, and 100 mM DTT for 30 min, followed by two 15-min washes in 2 x SSC and 0.1 x SSC. Sections were dehydrated through alcohol solutions containing 0.3 M NH₄OAc. Slides were dipped in a mix of Ifford K5 nuclear emulsion and glycerol/water [6 ml emulsion in 8.82 ml H₂O/0.18 ml glycerol] and kept at 4°C until developed. Exposures were between 10 and 12 days. Sections were stained in 0.02% toluidine blue for 1 min and then mounted in Permount.

Acknowledgments

We thank Dr. Bernhard Hermann for help in screening the cosmid libraries and isolating pCos 5.4 and 3.1, and Drs. Denis Duboule, Mark Featherstone, Brigitte Galliot, Pal Njolstad, Jim Garbern, Ward Odendwald, and Edoardo Boncinelli for sharing unpublished sequence information and ideas about the organization and evolution of homeo box genes. We also thank Drs. David Wilkinson, Peter Holland, Frank Ruddle, Mike Snow, Andy McMahon, Dennis Summerbell, and Mai Har Sham for valuable discussions and comments on the manuscript. N.P. was supported by the Greek State Scholarship Foundation and A.G. by a Medical Research Council studentship.

References


Dony, C. and P. Gruss. 1987. Specific expression of the Hox-1.3 homeo box gene in murine embryonic structures originating from or induced by the mesoderm. EMBO J. 6: 2965–2975.


Characterization of a murine homeo box gene, Hox-2.6, related to the Drosophila Deformed gene.

A Graham, N Papalopulu, J Lorimer, et al.

*Genes Dev.* 1988, 2:
Access the most recent version at doi:10.1101/gad.2.11.1424

| References | This article cites 58 articles, 21 of which can be accessed free at: http://genesdev.cshlp.org/content/2/11/1424.full.html#ref-list-1 |
| License |
| Email Alerting Service | Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here. |