The EMBO Journal vol.11 no.5 pp.1825 - 1836, 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

Mai Har Sham[1], Paul Hunt, Stefan Nonchev, Nancy Papalopulu, Anthony Graham, Edoardo Boncinelli[2] and Robb Krumlauf[3]

Laboratory of Eukaryotic Molecular Genetics, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London, NW7 1AA, UK
[1]Present address: Dept of Biochemistry, University of Hong Kong, Li Shu Fan Building, 5 Sassoon Road, Hong Kong
[2]IIGB, CNR, Via Marconi 12, 80125 Naples, Italy
[3]Corresponding author

Communicated by F.Grosveld

In this study we have investigated the organization and regulation of the mouse *Hox-2.7* gene. There are several alternative transcripts some of which are conserved between mouse and humans. By Northern and *in situ* analysis we are able to identify at least three types of transcripts which are different in size and splicing pattern and have distinctly different boundaries of expression in the nervous system. One subset of the endogenous transcripts has a boundary of expression that corresponds to the adjacent *Hox-2.8* gene instead of *Hox-2.7*. In another type of transcript there is an alternative reading frame which predicts a protein that has homology to an enzyme ATPase and suggests that a non-homeobox containing gene may be located in the *Hox-2* cluster. A *Hox-2.7–lacZ* transgene is expressed in a similar pattern to the endogenous gene in that spatially-restricted domains of expression are seen in the branchial arches, neural tube, paraxial mesoderm (somites), cranial ganglia, neural crest and gut. However, the anterior boundaries of transgene expression only correspond to the subset of *Hox-2.7* transcripts which map to the *Hox-2.8* boundary. The proximity of a *Hox-2.7* promoter to regions which regulate the adjacent *Hox-2.6* gene and the expression of transgenic and endogenous transcripts in a *Hox-2.8* pattern, suggest that regulatory elements may be shared by neighbouring genes to establish the complete expression pattern.

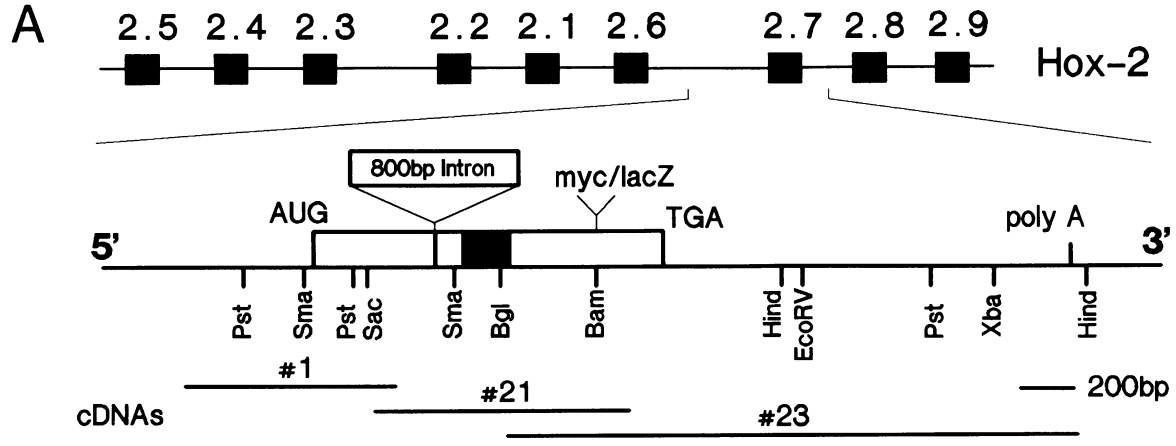
Keywords: homeobox genes/*Hox-2.7*/transgenic mice/transcriptional regulation/differential splicing

Introduction

The *Hox* family of vertebrate homeobox containing genes are candidate genes involved in regulating molecular patterning at the transcriptional level. Support for this idea comes both from their evolutionary conservation (Boncinelli *et al.*, 1988; Duboule and Dolle, 1989; Graham *et al.*, 1989; Kappen *et al.*, 1989) and spatial domains of expression during embryogenesis (reviewed in Holland and Hogan, 1988; Kessel and Gruss, 1990). Genes from the *Hox* complexes display highly ordered and overlapping domains of expression along embryonic axes and distinct coordinate patterns of expression have been observed in the limb (Dolle *et al.*, 1989; Izpisua-Belmonte *et al.*, 1991b; Nohno *et al.*, 1991; Yokouchi *et al.*, 1991), trunk (Gaunt *et al.*, 1988; Dressler and Gruss, 1989; Izpisua-Belmonte *et al.*, 1991a), prevertebrae (Kessel and Gruss, 1991), neural tube (Graham *et al.*, 1989; Wilkinson *et al.*, 1989) and branchial arches (Hunt *et al.*, 1991a, b). This has led to the suggestion that the *Hox* genes could provide a combinatorial set of signals for defining regional differences by analogy with their *Drosophila* homologues (Lewis, 1978) and that they may achieve this independently in different embryonic contexts (Hunt *et al.*, 1991a; Izpisua-Belmonte *et al.*, 1991b; Kessel and Gruss, 1991). Ectopic expression studies in transgenic mice showed that *Hox* genes could dominantly produce craniofacial abnormalities (Balling *et al.*, 1989) and vertebral transformations (Kessel *et al.*, 1990) through alterations in the patterns of expression. Recent experiments using homologous recombination and embryonic stem cells to generate mutations in *Hox* genes have provided direct evidence that these genes also play a normal role in patterning the head and trunk (Chisaka and Capecchi, 1991; Lufkin *et al.*, 1991).

Together the descriptive and experimental studies have lead to the suggestion that the functional domains of *Hox* genes are correlated with their coordinate patterns of expression. Therefore, one possible explanation for the conservation of *Hox* complexes is that it is important to maintain the mechanisms for spatial and temporal regulation of these complex patterns of expression. Northern analyses of *Hox* genes have demonstrated that they have complex transcription patterns (Krumlauf *et al.*, 1987; Graham *et al.*, 1988) and there is evidence for multiple promoters (Simeone *et al.*, 1987; Cho *et al.*, 1988), differential splicing and polyadenylation (Baron *et al.*, 1987; LaRosa and Gudas, 1988; Simeone *et al.*, 1988). In some cases these variations in transcript structure are conserved between species (Cho *et al.*, 1988; Savard *et al.*, 1988). The *Hox* genes are all oriented in the same direction with respect to transcription, and in the human *HOX 3* cluster several genes can be transcribed as a unit from a single 'master' promoter allowing multiple mRNAs to be generated by differential splicing (Simeone *et al.*, 1988). These complex and nested patterns suggest that transcriptional regulatory elements can have a long range influence in the clusters. In transgenic mice, analysis of *cis*-acting regulatory requirements for spatially restricted domains of *Hox* expression have shown that it is possible to obtain partial expression patterns outside of the clustered organization (Kress *et al.*, 1990; Zakany *et al.*, 1990; Schughart *et al.*, 1991), but that multiple elements are necessary to reconstruct the endogenous patterns in transgenic mice (Bieberich *et al.*, 1990; Puschel *et al.*, 1990, 1991; Whiting *et al.*, 1991). Some of these elements are located close to the coding region, however, others act as spatially-specific enhancers capable of directing expression from several *Hox-2* or heterologous promoters and could

© Oxford University Press



B

```
      -470                -450                -430                -410                -390                -370
GGTTATTTCGGGGTGCCAGAGAGAGGGGCAATGGTCCACGAAGCAGTGGTGTTCTATTCCTAAGCGGTCCTCGAAGGAGTACAGTCTTGATTTTGTTTTAAAAAAAAATGGAAATGAAAA

      -350                -330                -310                -290                -270                -250
TGGATGTTTTCTTTTTCCCGTTAAAATTAAAGTTGAAGTGTGAAGCATCCAAGTCAGGTTAGATACTTGGTCTCAGCGTCTGCTGCAGAGATACTGGCGAGGGGAGAAGAAGAAAAAAAA

      -230                -210                -190                -170                -150▽               -130
AAAACTATTGATGTCAGTTCCCTTTTTAGTTCCTAAGATGGATTCGAGGCCCAGTCCTCTTCTCCCCTCTGTCTCTTCTCTCGCCCCGCAGGTCAGCCGCTTGGAACAGACCCGAGAGGA

      -110                -90                 -70                 -50                 -30                 -10
GGGGGGCAGAAAGGGGAGGGGGGTCCGGCGTGTCACGTGACCCCCAGGGGTGCCAATGTCCGGTCGTGAGGGTATCAGGCCCTTGCAAGTTGCCACCCACTGCCCGGGCCTCGCCCAGCG

       10                  30                  50                  70                  90                  110
ATGCAGAAAGCCACCTACTACGACAACACCGCAGCTGCGCTCTTCGGAGGCTACTCCTCGTACCCTGGCAGCAATGGTTTCGGCTACGACGGGCCTCCCCAGCCCCCCTTTCAGGCCGCC
M  Q  K  A  T  Y  Y  D  N  T  A  A  A  L  F  G  G  Y  S  S  Y  P  G  S  N  G  F  G  Y  D  G  P  P  Q  P  P  F  Q  A  A

      130                 150                 170                 190                 210                 230
ACGCACCTGGAGGGTGACTACCAGCGCTCAGCGTGTTCTCTGCAGTCCCTGGGCAACGCCGCCCCACATGCCAAGAGCAAGGAGCTCAATGGCAGCTGCATGAGGCCAGGCCTGGCCCCA
T  H  L  E  G  D  Y  Q  R  S  A  C  S  L  Q  S  L  G  N  A  A  P  H  A  K  S  K  E  L  N  G  S  C  M  R  P  G  L  A  P

      250                 270                 290                 310                 330                 350
GAGCCCCTGCCCGCACCTCCGGGCTCACCCCCACCCAGCGCCGCACCTACCAGTACCACTAGCAACAGCAATAACGGGGGTGGGCCCAGCAAAAGCTGCCCCCCCAAGTGCGGTTCCGGC
E  P  L  P  A  P  P  G  S  P  P  P  S  A  A  P  T  S  T  T  S  N  S  N  N  G  G  G  P  S  K  S  C  P  P  K  C  G  S  G

      370                 390                 410                 430                 splice              470
TCCAACTCCACCCTCACCAAACAGATATTCCCCTGGATGAAAGAGTCGAGGCAAACGTCCAAGCTGAAAAACAGCTCCCCCGGCACAGCAGAAGGTTGTGGTGGTGGCGGCGGCGGTGGT
S  N  S  T  L  T  K  Q  I  F  P  W  M  K  E  S  R  Q  T  S  K  L  K  N  S  S  P  G  T  A  E  G  C  G  G  G  G  G  G  G

      490                 510                 530                 550                 570                 590
GGCGGCGGCGGCGGCGGTGGCGGCGGCAGCAGCGGCGGGGGCGGCGGCGGCGGGGGAGGGGACAAGAGCCCCCCGGGGTCGGCCGCGTCCAAGCGGGCGCGCACGGCGTACACAAGC
G  G  G  G  G  G  G  G  G  S  S  G  G  G  G  G  G  G  G  G  G  D  K  S  P  P  G  S  A  A  S  K  R  A  R  T  A  Y  T  S

      610                 630                 650                 670                 690                 710
GCGCAGCTGGTGGAGCTGGAGAAGGAGTTCCACTTCAACCGTTATTTGTGCCGGCCGCGCCGGGTCGAGATGGCCAATCTGCTGAACCTCAGCGAGCGCCAGATCAAGATCTGGTTCCAG
A  Q  L  V  E  L  E  K  E  F  H  F  N  R  Y  L  C  R  P  R  R  V  E  M  A  N  L  L  N  L  S  E  R  Q  I  K  I  W  F  Q

      730                 750                 770                 790                 810                 830
AACCGTCGCATGAAGTACAAGAAAGACCAGAAGGCCAAGGGGCTGGCCTCGTCCTCTGGGGGTCCCTCTCCGGCCGGAAGCCCCCCGCAGCCCATGCAGTCCACGGCCGGCTTCATGAAC
N  R  R  M  K  Y  K  K  D  Q  K  A  K  G  L  A  S  S  S  G  G  P  S  P  A  G  S  P  P  Q  P  M  Q  S  T  A  G  F  M  N

      850                 870                 890                 910                 930                 950
GCCTTACACTCCATGACCCCCAGCTACGACAGCCCGTCCCCACCAGCTTTCGGCAAAGGCCACCAGAATGCCTACGCGCTGCCTTCCAACTACCAGCCCCCTCTTAAGGGTTGCGGCGCC
A  L  H  S  M  T  P  S  Y  D  S  P  S  P  P  A  F  G  K  G  H  Q  N  A  Y  A  L  P  S  N  Y  Q  P  P  L  K  G  C  G  A

      970                 990                 1010                1030                1050                1070
CCACAGAAGTACCCCCCGACCCCGGCGTCGGAGTATGAGCCCCACGTCCTCCAAGCTAACGGGGGCGCCTACGGGACGCCCACCATGCAGGGCAGTCCGGTGTATGTGGGCGGGGGTGGC
P  Q  K  Y  P  P  T  P  A  S  E  Y  E  P  H  V  L  Q  A  N  G  G  A  Y  G  T  P  T  M  Q  G  S  P  V  Y  V  G  G  G  G

      1090                1110                1130                1150                1170                1190
TACGCGGATCCGCTGCCGCCCCCTGCCGGCCCCTCCCTCTACGGCCTCAATCACCTTTCCCACCACCCCTCGGGGAACCTGGACTACAACGGGGCGGCCCCTATGGGCCCCAACCAGCAT
Y  A  D  P  L  P  P  P  A  G  P  S  L  Y  G  L  N  H  L  S  H  H  P  S  G  N  L  D  Y  N  G  A  A  P  M  G  P  N  Q  H

      1210                1230                1250                1270                1290                1310
CATGGACCCTGTGACCCTCACCCCACGTACACAGACCTCTCCTCTCACCACGCACCGCCTCAGGGTAGAATCCAAGAAGCGCCCAAACTGACACACCTGTGATGGGAGGGGCGGGCTGGG
H  G  P  C  D  P  H  P  T  Y  T  D  L  S  S  H  H  A  P  P  Q  G  R  I  Q  E  A  P  K  L  T  H  L  *

      1330                1350                1370                1390                1410
GAGGGAGGGGTGAGGTTAAGGGCCAGGGAGACGGCGGACTTGGGGGATGGGCGCGGTTTAGAGTCCTGAAAGAGGTGTGGGATGGGGTTG //  1.3kb

AAAAGGGCTGAACTCAGCTCCAAATGTCACGTATAGTCCCGAATTTGTAGAGATAAGAAGTTAATTCTCTCTTGGAACTTTCTTTGTTCTTCCGTAGTTATTTTTTTCCCCCTTAGTTAA

AAGAGTTGTCTGTCAAAACAATTCTTGAATAAACTTTCTGTTATTAATTTTAAAAAAAAAAAAA
```

Fig. 1. Structure and sequence of the *Hox-2.7* gene and its predicted protein. **(A)** Structure and restriction map of three overlapping cDNA clones for the *Hox-2.7* gene. The open rectangles above the restriction map indicate the open reading frame of the predicted protein and the position of the intron. The solid box indicates the position of the homeodomain in the middle of the open reading frame. The *Bam*HI site in the sequence near the carboxy terminal end of the protein was used in some constructs for inserting in frame the human c-*myc* epitope or the bacterial *lacZ* gene. **(B)** Nucleotide and predicted amino acid sequence generated from the overlapping *Hox-2.7* cDNA clones. The hexapeptide and homeodomain regions are surrounded by open boxes. The splice site is indicated by a filled triangle. The open triangle indicates the splice acceptor site where an alternative upstream exon is spliced onto this transcript (see Figure 5). The polyadenylation signal (underlined) is located 18 bp upstream of the poly(A)$^+$ recognition sequence at the end of cDNA 23.

exert long range effects (Tuggle *et al.*, 1990; Whiting *et al.*, 1991).

In a few cases the distribution of different proteins or transcripts from the same gene have been examined during embryogenesis (Oliver *et al.*, 1988; Savard *et al.*, 1988; Murphy and Hill, 1991). However, in general there is very little information on the structure and spatial distribution of the different transcripts from an individual gene and how they are related to gene regulation and function. Therefore, we have analysed in detail the transcription patterns of the *Hox-2.7* gene to characterize the types of transcripts, their spatial distribution and relationship to neighbouring genes in the *Hox-2* complex. We identified three types of transcripts which were differentially expressed in the neural tube and using a *Hox-2.7–lacZ* reporter gene in transgenic mice we generated a pattern of expression characteristic of a subset of these transcripts. Our findings suggest that *Hox-2.7* transcription is influenced by regulatory elements from the adjacent genes.

Results

Hox-2.7 cDNA sequence and predicted protein

As an initial step in characterizing the *Hox-2.7* gene, we isolated overlapping cDNA clones from an 8.5 days post coitus (dpc) mouse embryo library. The structure, restriction map and sequence of these clones are shown in Figure 1. There are many open reading frames (ORF) predicted from the sequence, but the longest ORF encodes a protein of 433 amino acids which is in-frame with the homeodomain. Our analysis shows that in the *Hox-2.7* gene, like other *Hox* genes, the ORF is disrupted by a single intron of ~800 bp located 122 bp upstream of the homeobox and has long 5′ and 3′ untranslated regions. A conserved hexapeptide sequence (Mavilio *et al.*, 1986; Krumlauf *et al.*, 1987) of Ile–Phe–Pro–Trp–Met–Lys can be identified upstream of the splice site, but the N-terminal amino acid sequence of the predicted *Hox-2.7* protein (Met–Gln–Lys) is very diverged from the general Hox consensus of Met–Ser–Ser. The predicted *Hox-2.7* protein is also unusual in having a long carboxy terminal domain flanking the homeodomain, which makes it the largest Hox protein to date. The amino terminal domain is rich in serine and proline, and is separated from the homeodomain by a run of 26 glycine residues interrupted by two serine residues. It has been proposed that a glycine-rich region in the *Drosophila* Ubx protein functions as a hinge region separating different domains of the protein and this motif could play a similar role in *Hox-2.7*.

Due to the presence of the multiple open reading frames, the unusually long carboxy terminal domain and the diverged N-terminal sequence of the predicted protein, we feel that it is necessary to prove in some independent way that the major predicted homeodomain ORF is actually used. We have inserted in the predicted frame an oligonucleotide encoding 12 amino acids (position 408–419) of the human c-*myc* protein, into a composite *Hox-2.7* cDNA which spans the entire sequence in Figure 1B. Following *in vitro* transcription and translation of marked and control constructs, a monoclonal antibody specific for the c-*myc* epitope (Evan *et al.*, 1985) was used to immunoprecipitate the translation products (Figure 2). None of the translation products from the control constructs are precipitated. However, the immunoprecipitation identifies a major protein of ~60 kDa which could only be encoded by the largest ORF. In addition, all of the smaller products from the *myc*-tagged construct are precipitated by the antibody. Multiple proteins were translated from all of the constructs despite the fact that a single RNA species of discrete size was used in each case. Since the *myc* tag is near the carboxy terminus and all of the translated products are precipitated by the antibody, we feel that the multiple bands represent internal initiations (which are in agreement with the observed product sizes) rather than degradation products. These results confirm that the predicted open reading frame is correct.

Northern blot analysis

To investigate the tissue distribution and sizes of *Hox-2.7* transcripts, we have analysed poly(A)$^+$ RNA extracted

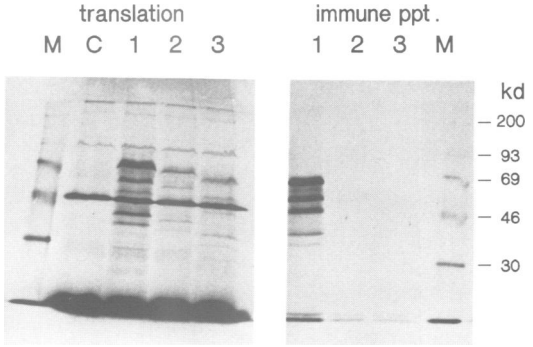


Fig. 2. *In vitro* translation and immunoprecipitation of Hox-2.7 proteins. Autoradiograph of SDS–PAGE [^{35}S]Methionine labelled products translated in a rabbit reticulocyte lysate system from three *in vitro* transcribed *Hox-2.7* RNAs: (1) *Hox-2.7*–sense *myc* fusion; (2) *Hox-2.7*–antisense *myc* fusion; (3) non-tagged *Hox-2.7* alone. The panel on the left (before) and right (after) represent immunoprecipitation with a monoclonal antibody against the *myc* epitope tag. Several polypeptides were translated from each of these transcripts, but only the *Hox-2.7*–sense *myc* products were precipitated by the anti-myc antibody. C, control translation with no added RNA; M, molecular weight markers with relative size in kilodaltons (kd).

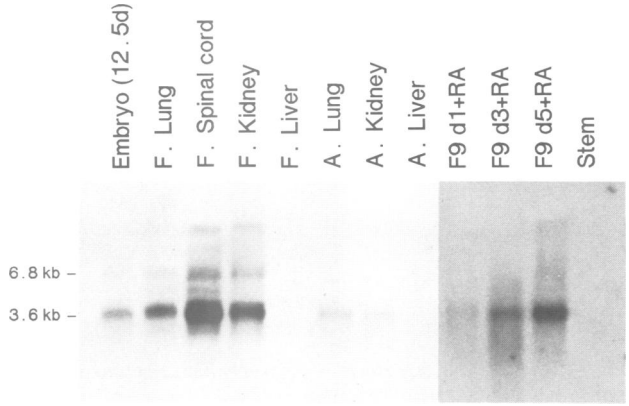


Fig. 3. Expression of *Hox-2.7* in F9 cells and fetal and adult tissues. Northern blot analysis of poly(A)$^+$ RNA (2 μg/lane) probed with the common *Hox-2.7* probe containing the homeobox region. The respective source and stage of the tissues (fetal = 14.5 dpc) and time-course of retinoic acid differentiation in F9 cell cultures is indicated above the lanes. In the RNA extracted from tissues, two abundant transcripts of 6.8 and 3.6 kb are detected in most samples and several minor or less abundant transcripts (including ones at 5.2, 10 and 15 kb) are also detected. In RA treated F9 cells, only the 3.6 kb transcript is detected.

from teratocarcinoma cell line and tissues at different stages by Northern hybridization. The probe, which contained the 3′ coding region flanking the homeodomain, hybridized to two major transcripts (3.6 and 6.8 kb) and several minor transcripts (Figure 3). In fetal tissues (14.5 dpc) transcripts were found in the lung, kidney and spinal cord, but not in liver, heart or brain [Figure 3; see also (Graham *et al.*, 1989)]. Expression persists in the same adult tissues, but at a much lower level. The ratio of the 6.8 and 3.6 kb transcripts is similar in all tissues, however, the relative distributions of the minor transcripts varies considerably between different tissues. This is readily seen by comparing expression of the 5.2 kb transcript in fetal kidney with that in the spinal cord. Some *Hox* genes can be induced in the retinoic acid dependent differentiation of mouse F9 teratocarcinoma cells and we examined *Hox-2.7* in this system. No expression is observed in untreated F9 stem cells, however, transcripts are induced by retinoic acid in parietal endoderm-like monolayer cultures (Figure 3). It is interesting that only the 3.6 kb transcript was preferentially induced and continued to accumulate during the differentiation time-course. This is also true of the human homologue as only the 3.6/3.4 kb transcript is expressed in EC cells (Stornaiuolo *et al.* 1990).

Comparison of human and mouse genomic and cDNA structures

Based on the fact that vertebrate *Hox* complexes are highly conserved, comparison between the human and mouse *Hox-2.7* sequences may give some insight into both structural organization and specific regions important in regulating gene expression. The coding region of the mouse *Hox-2.7* cDNA sequence is highly homologous to the human *Hox2G*, showing 91% homology at the nucleic acid level and 97% homology at the amino acid level [Figure 1 and (Acampora *et al.*, 1989)]. This homology also included the polyglycine region.

In contrast to the coding region, the 5′ untranslated

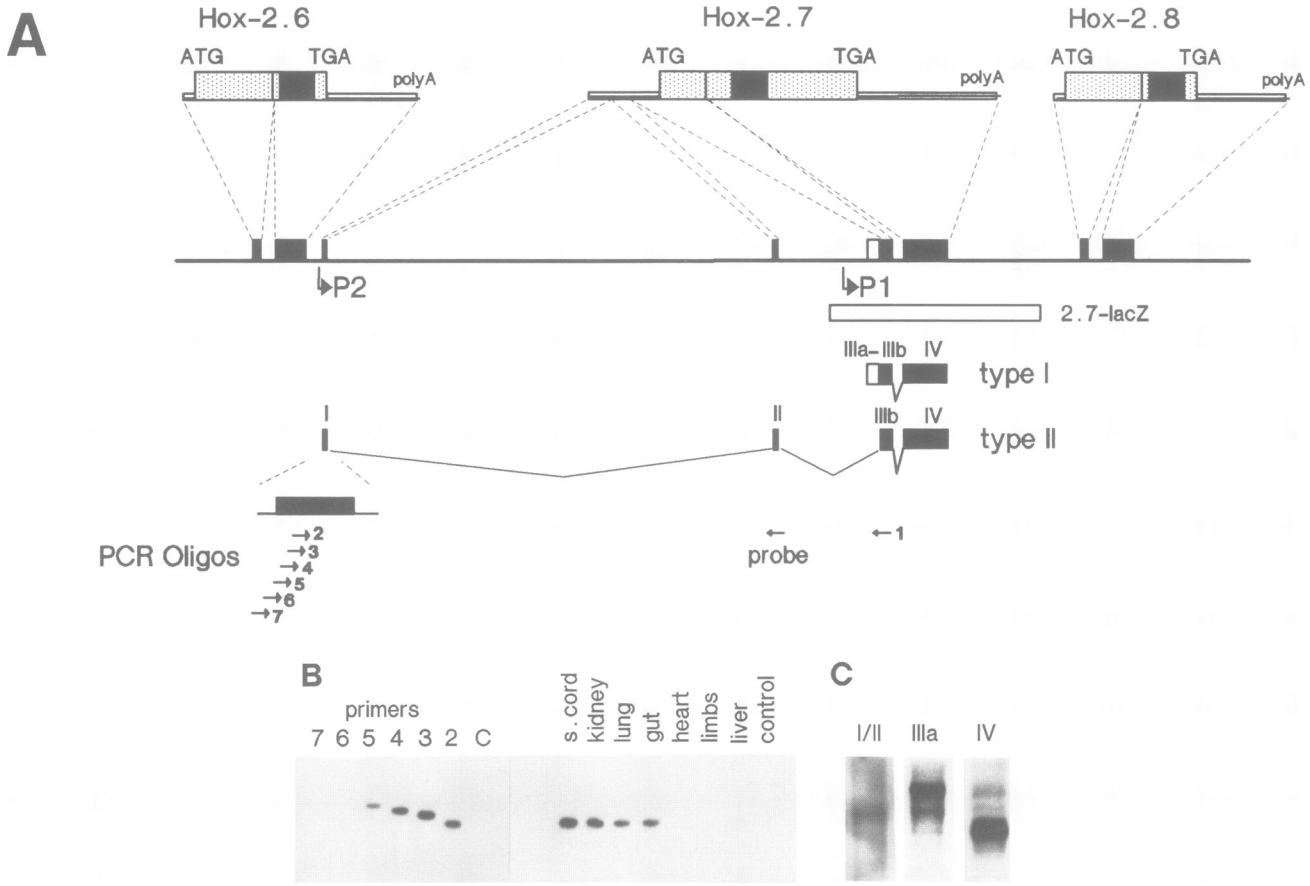


Fig. 4. Structure and expression of alternative transcripts from the *Hox-2.7* gene. (**A**) Organization of the mouse *Hox-2.6*, *Hox-2.7* and *Hox-2.8* genes, and the structure of two types of *Hox-2.7* transcripts. Type I transcripts are represented by the cDNA clones and organization shown in Figure 1 and has two exons (IIIa and b and IV) separated by an 800 bp intron. Type II transcripts have four exons, with exons I and II, located 25 kb and 8 kb upstream of the ATG respectively, spliced onto the middle of exon III (see open triangle in Figure 1). The structure of the type II transcript is identical to that of human *Hox2G* cDNA (Acampora *et al.*, 1989). The positions of two potential promoters P1 and P2 are indicated by arrows. Positions of the oligonucleotides used for PCR reactions in (**B**) are illustrated below the type II transcript. Primer 1 is derived from exon IIIb and is used in conjunction with primers 2, 3, 4, 5, 6 and 7 which are derived from exon I. An oligonucleotide derived from exon II is used as a probe for characterizing the PCR products. Open box illustrates the genomic fragment used to generate the *Hox-2.7–lacZ* transgenic construct. (B) Southern Blot of PCR products hybridized with an exon II oligonucleotide probe. To map the transcription start site, RT-PCR reactions were performed with total RNA extracted from 14.5 dpc embryonic spinal cord using primer 1 and primers 2–7. The tissue distribution of type II transcripts, were examined by similar PCR reactions from total RNA extracted from 14.5 dpc embryonic tissues using primers 1 and 2. (**C**) Northern blot of 13.5 dpc whole embryo poly(A)$^+$ RNA (2 μg/lane) hybridized to probes derived from exon I and II (type II transcripts), exon IIIa (type I transcripts) and exon IV (common probe spanning the homeobox) (see Materials and methods for details. Note that probes for type I and type II transcripts do not hybridize to the most abundant *Hox-2.7* 3.6 kb transcript, thereby defining a third class of transcript.

sequence of the mouse *Hox-2.7* cDNA differs considerably from that of a human *HOX-2G* cDNA derived from the N-tera2 cell line. The mouse cDNA sequence is collinear with the genomic sequence, whereas the human 5′ sequence is in fact encoded by two more exons spread over ~25kb upstream of the ATG codon (Acampora *et al.*, 1989). The organization of the two types of transcripts and their relationship to surrounding *Hox-2* genes is illustrated in Figure 4A. The type I transcript represented by the mouse cDNA contains two exons, III and IV. The type II transcript, based on the human cDNA, contains four exons where the upstream exons (I and II) are spliced onto exons III and IV. This splice occurs in the middle of exon III resulting in a region unique to the type I transcript (IIIa) and a region common to both types of transcripts (IIIb).

We synthesized two oligonucleotide probes for human exons I and II in order to determine whether these represented genuine differences between the species and whether both types of transcripts were present in the mouse. Both of these human oligo probes hybridize to genomic clones and we have mapped the positions of these potential upstream exons in the mouse *Hox-2.7* locus (Figure 4A). Exon II is positioned at 5 kb upstream of Exon IIIb and Exon I is located ~19 kb upstream of Exon II. This indicates that the sequences and their relative positions are conserved.

Characterization of alternative Hox-2.7 transcripts and open reading frames

To examine whether these potential upstream exons are actually transcribed in the mouse embryo, we performed

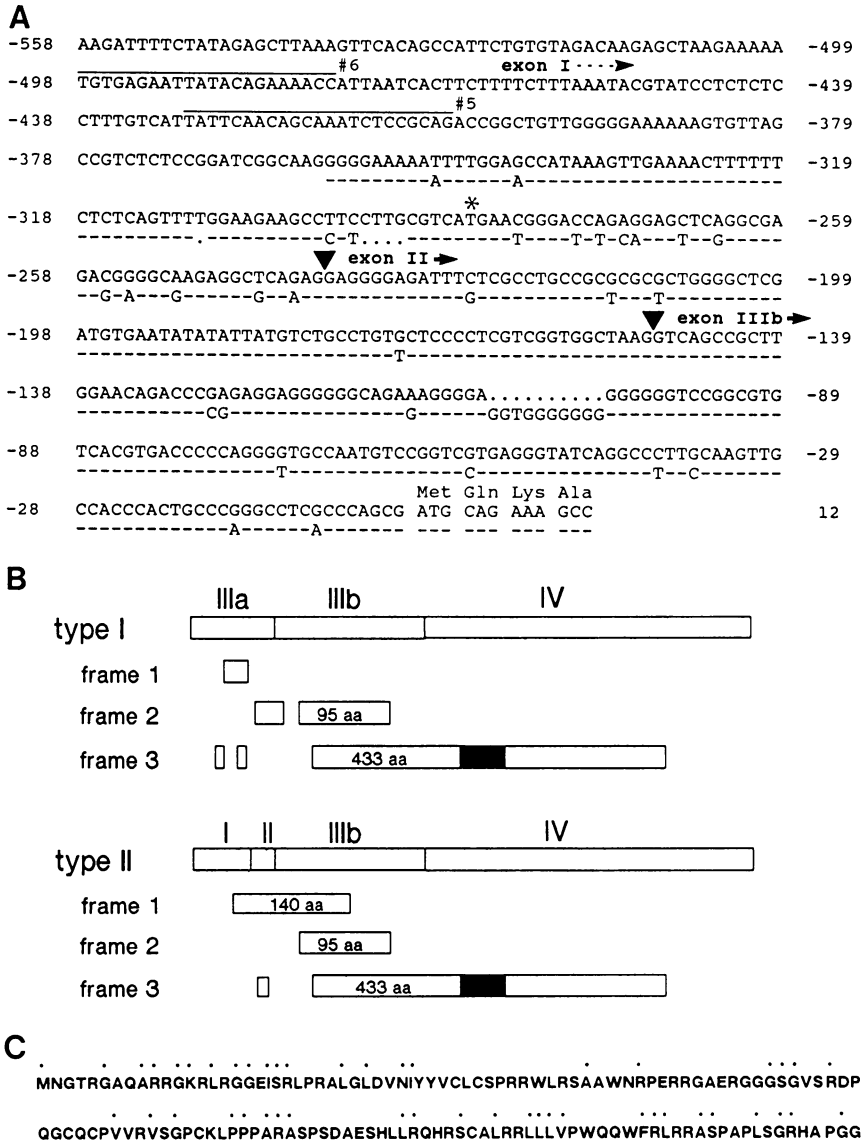


Fig. 5. Sequence and open reading frames of alternative *Hox-2.7* transcripts. (**A**) Nucleotide sequence of the 5′ untranslated region of mouse *Hox-2.7* type II transcripts and comparison with the human Hox2G cDNA sequence. Identical nucleotides are indicated by a dash and differences in the human sequence are stated. Absence of residue(s) is indicated by a dot(s). The splice sites are indicated by filled triangles. The oligonucleotide sequences of primers 5 and 6 used for PCR are marked with a line above. The asterisk marks the Met codon of the 140 amino acid open reading frame and the Met start of the 433 amino acid homeodomain protein is indicated above the sequence. (**B**) Schematic diagram showing the primary structure of the two types of transcripts and their multiple predicted open reading frames. Frame three contains the major 433 amino acid homeodomain (filled box) protein sequence. A 140 amino acid protein unique to the type II transcript is found in frame 1. (**C**) Sequence of the 140 amino acid polypeptide predicted from the upstream reading frames of type II transcripts and comparison with a family of human ATPase β-chain sequences (Ohita and Kagawa, 1986). Conserved amino acids with the human sequence are indicated by a dot above the lines.

PCR with RNA samples extracted from various embryonic tissues. As shown in Figure 4A and B (using PCR primer 1 in exon IIIb and primer 2 in exon I) the predicted DNA fragment of 280 bp is amplified from spinal cord RNA and an oligonucleotide probe from exon II hybridizes to the amplified product. This indicates that transcripts with sequence homologous to the human *Hox2G* exons I and II are expressed in mouse spinal cord and to verify this we sequenced the mouse PCR product. When the PCR sequence is compared with genomic sequence, it is clear that the fragment is indeed derived from the appropriate genomic regions and has two splice sites. Furthermore, the sequence generated from the PCR product shares 90% homology with the human exon I/II sequence and the splice sites are positioned in identical locations (Figure 5A). Therefore both types of transcripts are normally present in the mouse and the homology between the mouse and human *Hox-2.7* gene is not restricted to the coding region, but also extends to the 5′ untranslated regions.

These experiments do not define the 5′ end of the type II transcript and genomic sequence indicates that a region located upstream of the PCR sequence resembles the consensus splice acceptor site, $Py_{11}NCAG|G$. To map the 5′ end of exon I we have synthesized a series of oligonucleotide primers which extend further upstream and in conjunction with primer 1, PCR was performed with spinal cord RNA. Successively larger DNA fragments were amplified with primers 2−5 but no products were observed with primers 6 and 7 (Figure 4B). Therefore, the 5′ end of exon I lies in the sequence between primers 5 and 6, ~450 bp upstream of the translation start site. As there are no consensus splice sites in this region we believe that this represents the start of the type II transcripts.

The presence of two types of transcripts indicates that there may be two promoters regulating the expression of the *Hox-2.7* gene, a proximal promoter P1 which directs the transcription of the two exon transcript (type I) and a distal promoter P2 which directs the transcription of the four exon transcript (type II), as illustrated in Figure 4A. It is important to note that the differences between these transcripts are in the 5′ untranslated region and do not change the major predicted protein. However, in other reading frames there are several upstream ATG codons which mark alternative proteins in the different types of transcripts (summarized in Figure 5B). A predicted polypeptide of 95 amino acids, which has a N-terminal sequence (Met−Ser−Gly), similar to the *Hox* consensus(Met−Ser−Ser) is present in both forms of transcripts. In the type II transcript, there is a unique open reading frame which predicts a 140 amino acid polypeptide and the sequence shares 40−45% homology to the β-chain of the enzyme ATPase (Ohita and Kagawa, 1986), which is highly conserved from plants to mammals (Figure 5C). The entire ORF is not conserved in the sequence reported for the human cDNA due to a 10 bp insertion (9G and 1T) in a G-rich region, but the nucleic acid sequences are 92% identical.

Differential expression of alternative transcripts

Having demonstrated that at least two types of transcripts are present in the mouse, we wanted to see if they were differentially expressed. The characterization of the type II transcript was initially done with RNA from fetal spinal cord, therefore we examined its expression in other tissues by PCR (Figure 4B). Amplified products were also observed in kidney, lung and gut, indicating that the relative tissue distribution is similar between type I and type II transcripts.

To examine their size and spatial distribution, we generated a common probe based on the homeobox region (exon IV) and probes specific for exon I/II (type II) and exon IIIa (type I). First, Northern analysis shows that these probes hybridized to different subsets of mRNA (Figure 4C). The type II probe primarily hybridized to 5.2 kb transcripts and the type I probe to both 6.8 and 5.2 kb transcripts. Neither of these probes hybridized significantly to the 3.6 kb species which was the major transcript detected by the common probe and the only transcript induced by retinoic acid (Figure 3). This indicates that there must be other types of transcripts which do not contain exons I/II or IIIa but do contain the homeobox. In addition, the fact that the 5′ differences in the type I and II transcripts is small but the mRNA sizes fairly large suggests that there are other differences in splicing or polyadenylation.

In situ hybridization with the common probe is shown in Figure 6. We must stress that this common probe does not contain the *Hox-2.7* homeobox nor does it cross-hybridize with the *Hox-2.8* homeobox. It is termed common because it has 3′ sequences present in both transcripts. In the central nervous system at 9.5 dpc there is a sharp anterior limit of expression in the hindbrain which maps to the boundary between rhombomeres 4/5 [Figure 6e, see also (Wilkinson *et al.*, 1989; Hunt *et al.*, 1991b)]. In the neural tube at 12.5 dpc there is a clear dorsal restriction typical of *Hox-2* genes (Graham *et al.*, 1991) and the anterior limit persists; but in addition we detect a faint domain of expression which extends more anteriorly in the hindbrain (Figure 6a and b). Outside of the central nervous system *Hox-2.7* is expressed in a range of mesodermal derivatives and tissues which are originated from the neural crest (Graham *et al.*, 1989; Hunt *et al.*, 1991b). In mesodermal derivatives the common probe detects expression in lung, stomach, pancreas, metanephros and degenerating mesonephric tubules, and in prevertebrae with an anterior boundary at C1 (Figure 6b, c, d and h). In the branchial arches of the developing head, the probe hybridizes to the surface ectoderm and mesenchymal tissue of the third and posterior arches (Figure 6f). With respect to the neural crest, expression is observed in the mesenchyme of the thyroid gland and thymus, in the dorsal root ganglia and in the IX/X inferior cranial ganglion complex (Figure 6). Identical results are obtained with common probes derived from either the coding or the 3′ untranslated region of exon IV and these data confirm and extend our previous studies (Graham et al., 1989; Wilkinson *et al.*, 1989; Hunt *et al.*, 1991b).

In situ analysis with probes for the type I and type II transcripts reveals that they have identical spatial patterns with the exception of the nervous system in 12.5 dpc mouse embryos. Expression is seen in lung, stomach, pancreas, developing kidney, thyroid, dorsal root ganglia, IX/X inferior cranial ganglia and in prevertebrae with a boundary at C1, similar to what we observed with the common probe (Figure 7B, C, E and F). In the CNS, the dorso-ventral distribution of both transcripts was normal. With respect to the A-P axis, the type II transcript is expressed at high levels but there is no clear boundary in the hindbrain, however, the type I transcript has a distinct anterior boundary of expression. To map the relative position of the anterior boundaries, we made a direct comparison of the hybridization patterns of the type I and common probes on

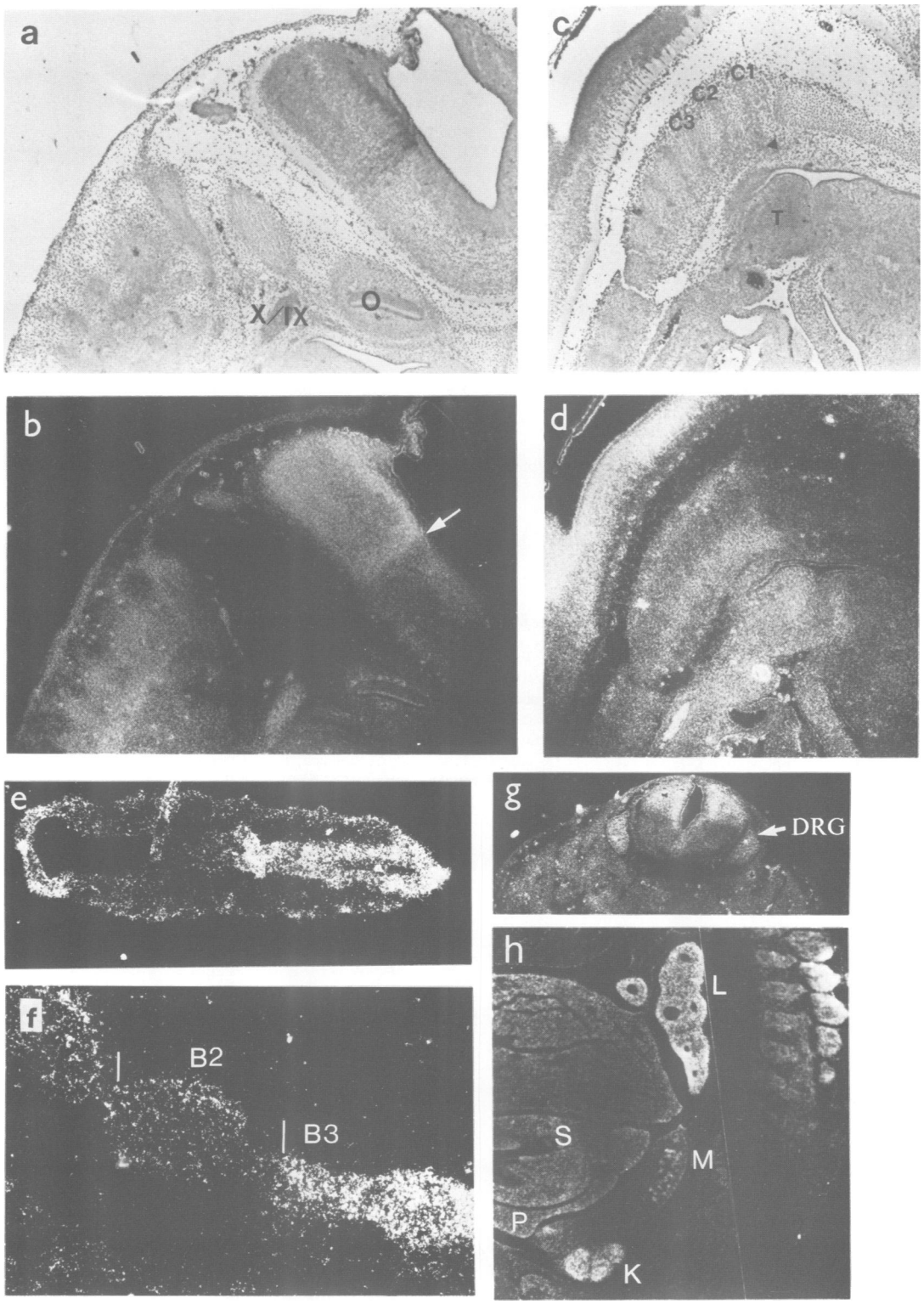


Fig. 6. *In situ* hybridization of mouse embryo sections with a *Hox-2.7* exon IV probe. This probe detects sequences common to several different types of transcripts. (**a**)−(**d**) and (**h**), sagittal sections of 12.5 dpc embryo; (**e**) and (**f**), coronal sections of 9.5 dpc embryo; (**g**), transverse section of 12.5 dpc embryo. The arrow in (b) denotes the major boundary of expression in the hindbrain, but there are some transcripts which extend into more anterior regions. Expression is detected upto the X/XI ganglion complex and the first cervical vertebrae (C1). B2 and B3, second and third brachial arches; C1, C2 and C3, first, second and third cervical prevertebrae; DRG, dorsal root ganglia; K, metanephric kidney; L, lung; M, mesonephric ducts; O, otocyst; P, pancreas; S, stomach; T, thyroid; X/IX, inferior glossopharyngeal and vagal ganglion complex.

near adjacent sagittal sections. The type I expression limit corresponds to that normally attributed to *Hox-2.7* at the junction of rhombomere 4/5 (Graham *et al.*, 1989; Wilkinson *et al.*, 1989) and also represents the major boundary observed with the common probe (see arrows Figure 7A and D). The common probe clearly hybridized to an additional domain which extended more anteriorly, up to the *Hox-2.8* boundary at rhombomere 2/3. This suggests

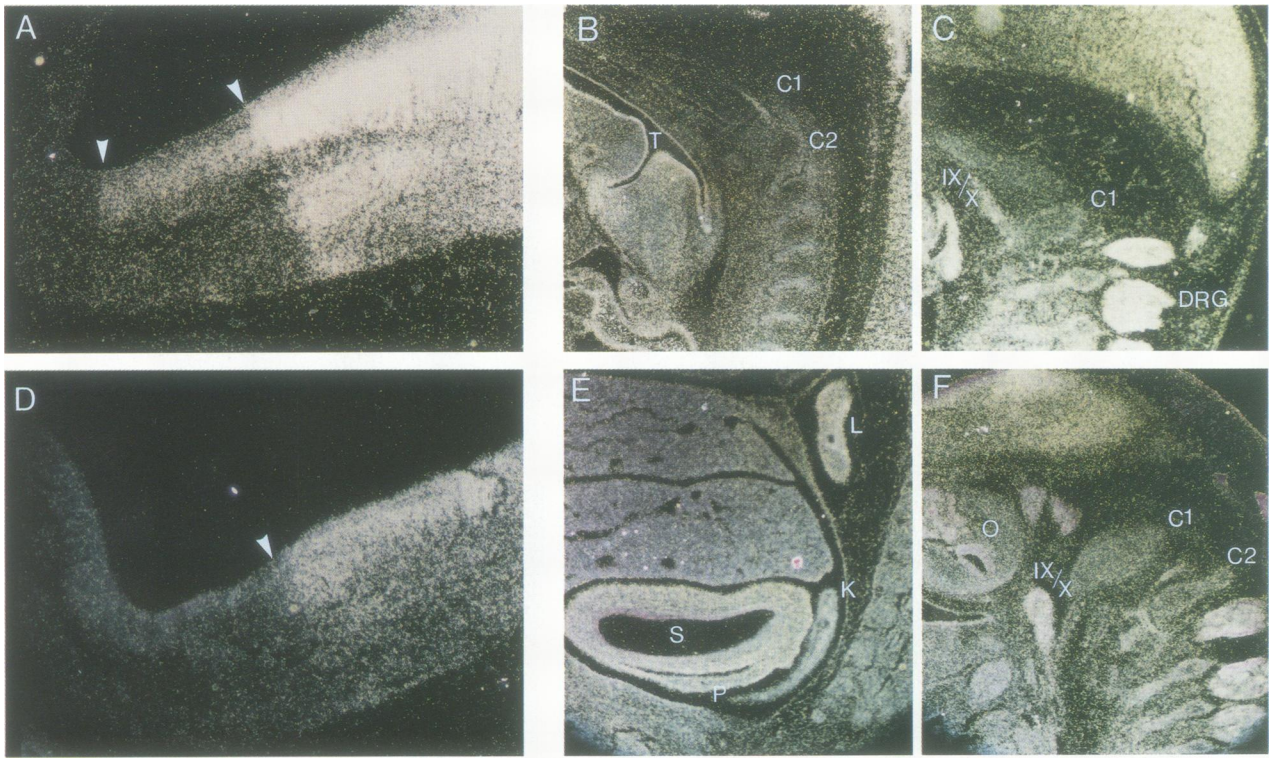


Fig. 7. Differential expression of alternative *Hox-2.7* transcripts. *In situ* hybridization of near adjacent 12.5 dpc mouse embryo sagittal sections with exon IV (common transcripts) (**A**), exon I/II (type II transcripts) (**B** and **C**) and exon IIIa (type I transcripts) (**D**, **E** and **F**) specific probes. It should be stressed that the common probe does not contain *Hox-2.7* homeobox sequences and does not cross-hybridize with the *Hox-2.8* homeobox. It is termed common because it contains 3′ untranslated sequences shared in both types of transcripts. Arrowheads in A and D indicate anterior boundaries of expression; note the absence of an additional domain of anterior expression in panel D with the type I specific probe. L, lung; K, Kidney; C1, C2 and C3, the first to third cervical vertebrae; O, otocyst; T, thyroid; S, stomach; P, pancreas; IX/X, inferior glossopharyngeal and vagal ganglion complex. In B, C and F boundaries of expression in the vertebrae map to C1.

that there is a third type of *Hox-2.7* transcript (type III) containing the homeobox, which is actually expressed in the more anterior domains.

Transgenic analysis

The complex *Hox-2.7* transcription pattern revealed by the northern and *in situ* analysis could be generated by differential splicing of RNAs from a common promoter, utilization of multiple promoters or a combination of both. To begin to distinguish between these possibilities we used transgenic mice to assay the expression of a *Hox-2.7* genomic construct, which has the bacterial *lacZ* gene inserted in-frame at the same *Bam*HI site previously used in the *myc* tagging experiments (Figure 1A). This fusion construct spans 9 kb of *Hox-2.7* genomic sequence, including 1 kb 5′ of the ATG initiation codon and 4 kb 3′ of the polyadenylation signal (Figure 4A). Therefore this construct does not contain exon I or II and excludes the effects of any distal *Hox-2.7* promoters. Transgenic mice were generated and in either transient (F_o) embryos or progeny from established lines, expression patterns as examined by whole mount staining and histological sectioning was identical.

The *Hox-2.7–lacZ* transgene is strongly expressed in seven out of nine transgenics lines and has an identical pattern in all cases. Staining patterns at two embryonic stages, 10.5 and 12.5 dpc, are shown in Figure 8. Whole mount embryos from these two stages show high levels of expression in the neural tube with sharp anterior boundaries. At 10.5 dpc expression extends from the most posterior regions to a rhombomere boundary in the hindbrain anterior to the otic vesicle (Figure 8A and B). Staining can also be observed in the second, third and more posterior branchial arches, the somites and the proximal region of the forelimb bud. In the later staged embryo the transgene is expressed in the apical ectodermal ridge of both the fore and hindlimb buds.

Sections of the same 12.5 dpc embryo are shown in Figure 8C-H. The transgene is expressed in most of the appropriate tissues, including the stomach, pancreas, thyroid, inferior cranial sensory and dorsal root ganglia, prevertebrae and neural tube, but not in the lung or kidney. The boundary of expression in the hindbrain maps near the pontine flexure and at the cellular level there is a clear distinction between expressing and non-expressing cells. Levels of expression are highest in the dorsal region of the neural tube, similar to the endogenous gene; however, a layer of cells on the ventral side of the neural tube have a more anterior domain of expression (Figure 8D, F and H). The rostral limit of expression corresponds to the most anterior domain observed by *in situ* analysis with the common probe (Figure 7A). In addition to the prevertebrae, part of the basioccipital bone was positively stained; and several inferior cranial ganglia, including the facial–acoustic ganglion complex VII/VIII and glossopharyngeal–vagal ganglion complex IX/X expressed the transgene (Figure 8G,H).

It is clear that the 1 kb of 5′ flanking DNA in the fusion construct contains a promoter region (P1) which helps to direct the expression of the transgene to a large subset of the endogenous *Hox-2.7* pattern. This occurs in the absence of upstream exons for the type II transcripts and suggests that multiple promoters are used and required to reconstruct

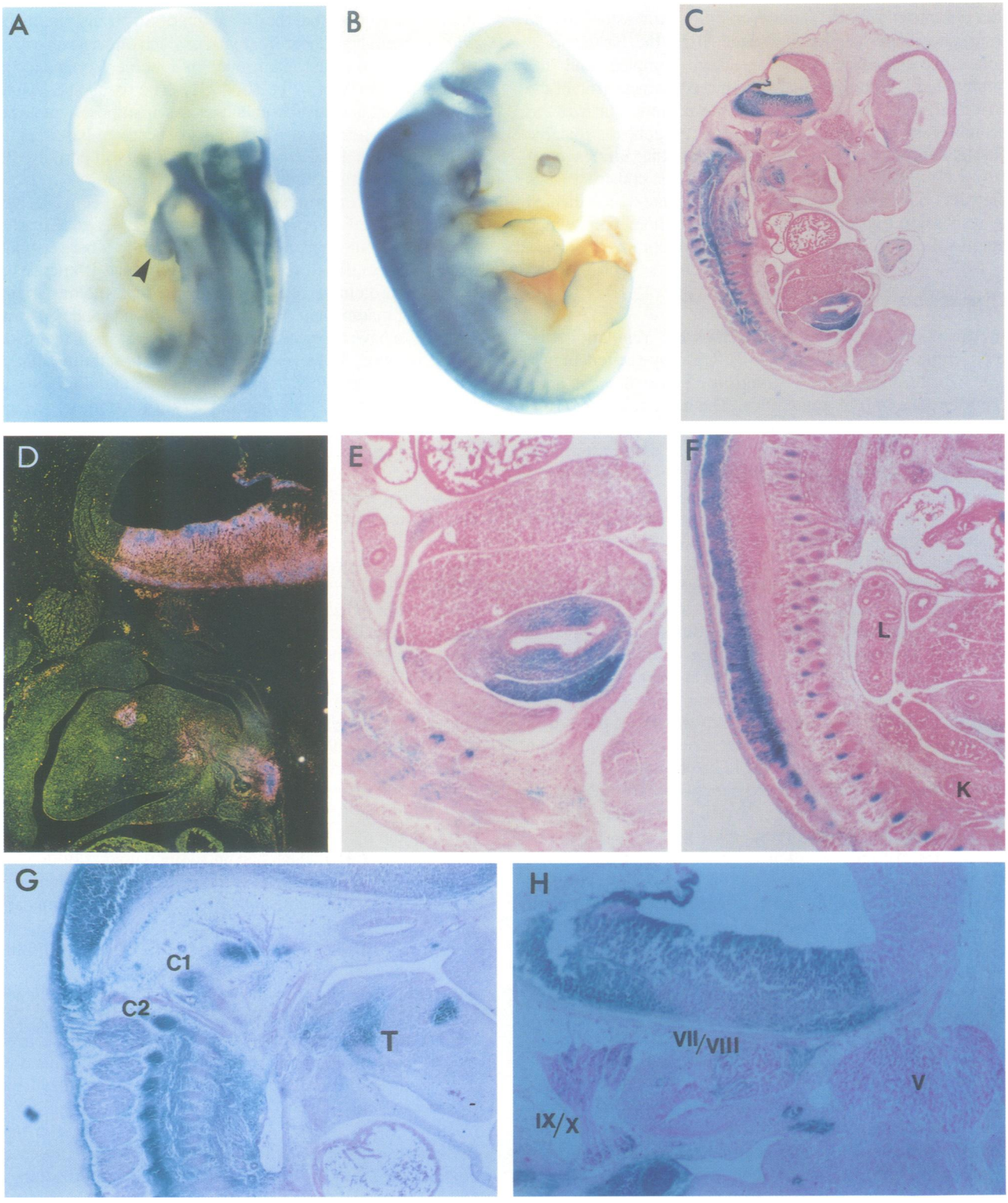


Fig. 8. Expression of a *Hox-2.7–lacZ* transgene. (**A**) A 10.5 dpc transgenic embryo stained in whole mount for β-galactosidase activity. The anterior limits of expression are in the hindbrain and the second branchial arch (marked by an arrowhead). (**B**) Whole mount stain of a 12.5 dpc embryo. (**C**)–(**H**) Paraffin embedded 6 μm sagittal sections of the same 12.5 dpc transgenic embryo shown in (B). The sections were counter stained with eosin (red). (D) Dark field image of a sagittal section where β-galactosidase staining appears pink. K, Kidney; L, Lung; T, Thyroid; C1/C2, first and second cervical vertebrae; V, trigeminal ganglion; VII/VIII, facial and acoustic ganglion complex; IX/X, inferior glossopharyngeal and vagal ganglion complex. Staining is observed in the dorsal root ganglia, pancreas, neural tube, vertebrae, thyroid and stomach. Note that expression of the transgene extends more anterior than the first cervical vertebrae into the basioccipital bone (panel G), is expressed in the VII/VIII ganglion complex (panel H) and extends to the pontine flexure (panels C and D). All of these sites are more anterior than the normal characteristic *Hox-2.7* boundaries and instead correlate with the normal boundaries of *Hox-2.8* expression.

the complete pattern. However, the expression of the transgene in the second branchial arch, the basioccipital bone, the facial–acoustic ganglion complex and the hindbrain, all represent anterior patterns characteristic of the *Hox-2.8* gene. This suggests that either there is a *Hox-2.8* promoter in the middle of the *Hox-2.7* gene or that the *Hox-2.7* promoter has become under the influence of *Hox-2.8* regulatory regions in the transgenic construct. This may also occur in the normal genomic context as indicated by a subset of transcripts with the *Hox-2.7* homeobox which map to the *Hox-2.8* boundary in the hindbrain.

Discussion

In this study we have approached the regulation and organization of the mouse *Hox-2.7* gene by sequencing, *in situ* hybridization and by using an *Escherichia coli lacZ* reporter gene in transgenic mice. The basic features of this gene and its predicted protein are similar to other *Hox* genes with the exception of a large domain on the carboxy terminal side of the homeodomain, which makes *Hox-2.7* the biggest *Hox* protein to date. Northern analysis reveals multiple transcripts which we are able to divide into at least three types on the basis of differences in size, splicing pattern and spatial distribution. These transcripts vary in their 5′ untranslated region and do not change the nature of the predicted protein; but they have distinctly different boundaries of expression in the nervous system. Our data suggests that multiple promoters are involved in generating the complex transcription pattern. We examined some of the *cis*-acting requirements for establishing the *Hox-2.7* pattern and found that a *Hox-2.7–lacZ* transgene is expressed in the same tissues as the endogenous gene, reconstructing an expression pattern typical of *Hox* homeobox genes. Spatially restricted domains of transgene expression are seen in the branchial arches, neural tube, paraxial mesoderm (somites), cranial ganglia, neural crest and gut. However, the boundaries of transgene expression do not represent those of the major *Hox-2.7* transcripts (Wilkinson *et al.*, 1989; Hunt *et al.*, 1991b), but correspond to those of a minor transcript which is expressed in a more anterior domain. These analyses indicate that promoters and regulatory regions of the *Hox-2* genes are interspersed within the cluster and that a single control region may affect the expression of more than one gene.

Distal transcripts are conserved and may encode an alternative protein

Most of the cDNAs for *Hox* genes which have been previously characterized, have a simple two exon structure and promoter initiation sites near the ATG of the predicted proteins. A major 3.6 kb *Hox-2.7* transcript has this structure which we denote as type I. In addition, on the basis of sequence homology to the human *HOX-2G* gene (Acampora *et al.*, 1989), we identified an alternative RNA (type II). This species is expressed at a low level and contains two additional 5′ exons transcribed from sequences a further 25 kb upstream of the ATG initiation codon, which are spliced into the 5′ untranslated region of the type I RNA. The position and sequence of the small upstream exons are not only conserved in mouse and human, but are also conserved in chick (Chaudhuri and Krumlauf, unpublished). This distal transcript is initiated near the 3′ end of the adjacent *Hox-2.6* gene, 1.5 kb downstream of its polyadenylation site. It is interesting to note that the first exon of this transcript maps within a neural enhancer (region A), which directs expression of the *Hox-2.6* gene to its proper rhombomere boundary in the hindbrain (Whiting *et al.*, 1991). The tissue and spatial specificity reside in this enhancer, as it imposes identical patterns on heterologous promoters. Therefore based on its proximity, this distal *Hox-2.7* promoter may in part also be regulated by this element.

Alternative open reading frames which could affect translational efficiency are found in all types of *Hox-2.7* transcripts, but they are generally small. There is one exception specific to the distal transcript and it encodes a predicted protein of 140 amino acids with homology to the β-chain of mammalian ATPase enzymes (Ohita and Kagawa, 1986). We have no evidence that this protein is actually made in the mouse, however, it opens the possibility that a transcription unit for a non-homeobox containing gene is located in the *Hox-2* cluster, as observed for transmembrane protein *amalgam* (Seeger *et al.*, 1988) in the *Drosophila* ANT-C complex.

Differential distribution of Hox-2.7 transcripts

Our previous studies have shown that some *Hox-2* genes are expressed in restricted patterns which correlate with morphological structures in the branchial region of the head (Wilkinson *et al.*, 1989; for review see Hunt and Krumlauf, 1991; Hunt *et al.*, 1991b). On the basis of levels of expression detected by *in situ* hybridization, all the different *Hox-2.7* transcripts had anterior boundaries of expression which mapped to the first cervical vertebrae (C1), the IX/X ganglion complex and the third branchial arch, which are characteristic for the *Hox-2.7* gene. We cannot rule out that there are a small number of cells expressing specific transcripts in more anterior domains. However, differences in the neural tube were detected by *in situ* analysis and only the type I transcripts had the appropriate *Hox-2.7* boundary (r4/5) in the hindbrain. An exon IV probe detected another transcript (type III) which had an additional anterior domain of expression which extends to the *Hox-2.8* (r2/3) boundary. Although we were unable to define a clear anterior boundary for the type II transcripts, our analysis demonstrates that all classes of *Hox-2.7* RNA have different distributions in the neural tube.

An explanation for the subset of transcripts (type III) which map to the *Hox-2.8* boundary is that one of the *Hox-2.7* promoters in the normal chromosomal context is influenced by a region which establishes the primary *Hox-2.8* pattern. This is analogous to the relationship between the distal transcripts and the *Hox-2.6* neural enhancer, region A (Whiting *et al.*, 1991). An alternative explanation is that these transcripts could be derived from a *Hox-2.8* promoter embedded in the *Hox-2.7* gene. Further characterization of the initiation and termination sites will clarify whether this transcript is capable of encoding a *Hox-2.7* or *Hox-2.8* protein.

Expression of Hox-2.7 is influenced by neighbouring genes

The experiments in transgenic mice lend direct support for the idea that *Hox-2.7* expression can be regulated by elements from the adjacent *Hox-2.8* gene. The *Hox-2.7–lacZ* transgene which we examined did not contain the *Hox-2.8* gene, but was expressed in a pattern similar to that of *Hox-2.8*; as the anterior limits were mapped to the

basioccipital bone, the VII/VIII ganglion complex, the second branchial arch and r2/3 in the hindbrain. Transgene expression is analogous to that of the type III transcripts, which have a *Hox-2.8* boundary in the neural tube (described above), therefore we feel they reflect a normal subset of the *Hox-2.7* transcripts and not domains of ectopic expression. This demonstrates that the transgenic construct contains a regulatory region within or immediately surrounding the *Hox-2.7* gene, which can set the major *Hox-2.8* pattern.

Examination of transgene expression in branchial arch tissues revealed that at the anterior boundary only a small population of cells in the ganglia, bones and mesenchyme were positively stained, which would account for the lack of type III transcripts detected at these anterior limits when examined by *in situ* hybridization. The transgenic construct may also independently direct other aspects of the *Hox-2.7* patterns, but because the major *Hox-2.7* domains are a subset of *Hox-2.8*, we cannot confirm this possibility without extended deletion analysis.

In vertebrates, multiple promoters and differential splicing have been reported for other *Hox* genes (Simeone *et al.*, 1988) and variant transcripts from the same gene can display different tissue or spatial distributions (Murphy and Hill, 1991). Therefore it is not surprising that these mechanisms are also used for generating complex *Hox-2.7* transcription patterns. In addition, the proximity of a *Hox-2.7* promoter to regions which regulate the adjacent *Hox-2.6* gene and the expression of transgenic and endogenous *Hox-2.7* transcripts in a *Hox-2.8* pattern, suggest that regulatory elements may be shared by neighbouring genes to generate the complete expression pattern. If the variant *Hox* transcripts and their differential distributions are functionally required for normal development, then the organization of the multiple promoters and shared regulatory regions would need to be maintained for the appropriate spatial distributions. This provides one potential regulatory basis for conservation of *Hox* clusters in vertebrates.

Relationship to other paralogues

Hox-2.7 forms part of a paralogous group with the *Hox-1.5* and *Hox-4.1* genes, on the basis of sequence identity and positions in their respective clusters (Duboule and Dolle, 1989; Graham *et al.*, 1989). In the hindbrain and branchial arches these genes have the same boundaries of expression (Hunt *et al.*, 1991a), which correspond to the domains of the *Hox-2.7* type I transcripts detailed in this study. This suggests that regulatory regions may also be conserved between members of this group, however, it is not known whether the other paralogues utilize multiple promoters and have similar types of differentially spliced transcripts. Evidence that this paralogous group plays an important role in development has come from a targeted mutation in the *Hox-1.5* gene and mice homozygous for the mutation had phenotypic abnormalities concentrated in the head and thorax (Chisaka and Capecchi, 1991). The affected regions are generally correlated with the domains of *Hox-1.5* expression, but not all regions which express the gene are abnormal. This suggests that there may be functional redundancy or compensation by other members of the paralogous group and of the *Hox* network. Our *Hox-2.7–lacZ* experiments show that in the branchial region not all cells of the same structure are positively stained and support the idea that individual genes may be used to differentially pattern specific subsets of cells within a region.

One problem which arose in the *Hox-1.5* study was that structures more anterior than the normal *Hox-1.5* domain of expression, such as the second branchial arch, were also abnormal (Chisaka and Capecchi, 1991). However, based on the *Hox-2.7* transgenic and expression analysis in this study, there may be additional domains of *Hox-1.5* expression, more anterior than those previously established by *in situ* hybridization, which could account for these phenotypes. Further analysis of the regulation and function of the complex transcripts of *Hox-2.7* and the other members of this subfamily will therefore be important in understanding the molecular processes which pattern the vertebrate head.

Materials and methods

Isolation of cDNA clones and sequencing

Hox-2.7 clones were isolated from a cDNA library prepared from 8.5 dpc mouse embryos (Fahrner *et al.*, 1987) by screening at high stringency (0.1×SSC at 65°C) with a genomic *Sac*I–*Bam*HI fragment (Graham *et al.*, 1989) that contained the *Hox-2.7* homeobox as a probe. The inserts in the positive cDNA clones were removed and subcloned into Bluescript vector (Stratagene). Double-stranded DNA from the cDNA clones was sequenced by the dideoxy chain termination method using primers in the Bluescript vector or specific oligonucleotides and sequenase DNA polymerase (USB). All clones were sequenced on both strands and compressions were resolved using inosine derivatives run in parallel reactions. For fragments generated from polymerase chain reactions, specific oligonucleotides were used to sequence the double-stranded fragments directly.

Construction of fusion cDNA and clones tagged with myc

Full length *Hox-2.7* cDNA and genomic constructs were generated by combining overlapping regions from several clones. Fragments from the overlapping cDNA clones and a genomic cosmid clone pCos3.1 (Graham *et al.*, 1988) were joined together in Bluescript vector by several cloning steps to make a fusion cDNA. Briefly the *Sac*I fragment from cDNA1 was successively combined with the *Sac*I–*Bgl*II fragment from cDNA21, the *Bgl*II–*Hin*dIII fragment from pCos3.1. and the *Hin*dIII fragment from cDNA23 to make a complete continuous clone which preserved normal restriction sites and organization. A double-stranded oligonucleotide (5′-GATCCAGATCCTCCTCAGAAATCAGCTTTTGCTCCT-3′, 36 bp) with *Bam*HI sticky ends, encoding the critical 10 amino acids (Glu–Gln–Lys–Leu–Ile–Ser–Glu–Glu–Asp–Leu) of the human c-*myc* epitope recognized by the 9E10 antibody (Evan *et al.*, 1985) was inserted into the *Bam*HI site of the fused cDNA clone in both orientations.

Transcription and translation in vitro and immunoprecipitation

Fusion cDNA clones were linearized and transcribed according to Kreig and Melton (1988). The capped RNA transcripts were translated in the presence of [^{35}S]methionine using the rabbit reticulocyte system (Promega). The *in vitro* translated polypeptides were immunoprecipitated with an anti-myc antibody specific for the tagged epitope [MAb Myc1–9E10, kindly provided by Dr G.Evan (Evan *et al.*, 1985)], resuspended and separated by SDS–PAGE then analysed by autoradiography.

RNA isolation, Northern blotting and in situ hybridization

Total RNA for polymerase chain reaction was isolated from 14.5 day-old mouse tissues according to Kreig and Melton (1988). Poly(A)$^+$ mRNA from mouse embryonic tissues and F9 cells was isolated as described in (Krumlauf *et al.*, 1987; Krumlauf, 1991). Briefly, mouse tissues were harvested, rinsed in PBS, homogenized in 3 M LiCl and 6 M Urea and sonicated for 1 min on ice. The homogenate was stored overnight at 0–4°C and the RNA precipitate collected by centrifugation and washed once by resuspension in LiCl–Urea. The pellet was redissolved in 10 mM Tris–HCl (pH 7.6), 1 mM EDTA, 0.5% SDS then extracted with an equal volume of phenol–chloroform. The aqueous phase was collected, ethanol precipitated and redissolved in 10 mM Tris–HCl (pH 7.6), 1 mM EDTA, 0.5% SDS and poly(A)$^+$ RNA was selected by oligo(dT)–cellulose chromatography.

RNA samples were denatured at 60°C for 10 min in 70% formamide, 6% formaldehyde–1×MOPS and were separated in a 1.2% agarose–6.3% formaldehyde gel in 1×MOPS buffer (pH 7.0; 20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA). After electrophoresis the gel was sequentially soaked in 50 mM NaOH–0.1 M NaCl, 0.1 M Tris–HCl (pH 7.6), 2×SSC (each for 20 min), then blotted onto a Genescreen (Dupont) in 20×SSC overnight, coupled to the filter by UV crosslinking and then baked (Krumlauf,

1991). Filters were hybridized in 60% formamide, 1×Denhardt's, 20 mM NaPB pH 6.8, 100 μg/ml sheared salmon sperm DNA, 100 μg/ml yeast tRNA, 1% SDS and 10% dextran sulphate at 65°C for 12 h. The filters were washed in 2 ×SSC−0.1% SDS and 0.2×SSC−1% SDS at 70−80°C for 1−3 h. To eliminate non-specific hybridization in direct comparisons with probes specific for different exons of *Hox-2.7*, the filters were treated with 2 μg/ml RNase A (Sigma) in 2×SSC and washed in 2×SSC−0.2% SDS and 0.5% SSC−0.2% SDS both at 50°C for 30 min before autoradiography. All probes in both the Northern and *in situ* hybridizations were single stranded P^{32} labelled antisense RNA probes (riboprobes), synthesized from subcloned fragments in Bluescript (Promega) or region-specific PCR amplified products that contained a T7 polymerase promoter as described by Frohman and Martin (1989).

The probes used were: common probe, *Bgl*II−*Bam*HI fragment of the 3′ coding region which contains part of the homeobox, or a *Bam*HI−*Hin*dIII fragment spanning the 3′ coding and 3′ untranslated regions (see Figure 1); type I probe, 300 bp PCR fragment derived from exon IIIa specific for type I transcript; type II probe, PCR fragment derived from exon I and II specific for the type II transcript. The method for generating RNA probes from PCR fragments was as described in Frohman and Martin (1989). The *in situ* hybridization protocol was performed exactly as described in Wilkinson and Green (1990).

Oligonucleotide primers for polymerase chain reaction

The sequences of the primers used for PCR in characterizing the type II transcript were: primer 1, TGATACCCTCACGACCGGACATTGGCA; primer 2, TTGCGTCATGAACGGGACCAGAGGAG; primer 3, GGAAAAATTTTGGAGCCATAAAGTTG; primer 4, AGTGTTAGCCGTCTCTCCGGATCG; primer 5, TATTCAACAGCAAATCTCCGCAG; primer 6, AAATGTGAGAATTATACAGAAAACC; and primer 7, AGTTCACAGCCATTCTGTGTAGAC. The sequence of the exon II oligonucleotide probe was AGAGCGAGCGGCAGGCGACAAATCTC.

Generation of DNA construct and transgenic mice analysis

The *Kpn*I−*Xba*I fragment from pCos3.1 was subcloned into pPolyIII (Lathe *et al.*, 1987). The *lacZ* gene was then inserted in-frame into the *Bam*HI site. The 5′ *Kpn*I fragment and 3′ *Xba*I−*Eco*RI fragment were subsequently cloned into this clone to make a construct which was essentially a *Kpn*I−*Eco*RI genomic clone with *lacZ* inserted at the *Bam*HI site. For microinjection the fragment was excised with *Xho*I, purified from agarose gel with Geneclean (Bio 101) and then passed through spin-X column (Costar).

(CBA×C57BL10)F_1 mice were used throughout these experiments as embryo donors, stud males, pseudopregnant females, vasectomized males and mature females for breeding. Transgenic mice were produced as described by Hogan *et al.* (1986) and β-galactosidase (LacZ) staining according to Whiting *et al.* (1991). Briefly, embryos to be stained for LacZ were fixed in 1% formaldehyde, 0.2% glutaraldehyde, 2 mM $MgCl_2$, 5 mM EGTA, 0.02% NP40 in PBS at 4°C for 30−90 min depending on size. They were then washed in three changes of PBS plus 0.02% NP40 at room temperature for 30 min each and stained in the dark in 1 mg/ml X-gal, 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$, 2 mM $MgCl_2$, 0.01% sodium deoxycholate, 0.02% NP40 in PBS at room temperature. Embryos to be sectioned were fixed further in 4% paraformaldehyde overnight at 4°C, dehydrated and embedded in paraffin wax (Wilkinson and Green, 1990). 6 μm sections were cut, the sections dewaxed and counterstained with eosin.

Acknowledgements

We thank Dr Gerrard Evan for providing the anti-myc 9E10 antibody, Dr Chitrita Chaudhuri for sharing the information on chick homologies, Dr Denis Duboule for sharing discussions and sharing unpublished data, and other members of the laboratory for discussions. A.G. and P.H. were supported by MRC studentships, N.P. by the Greek State Scholarship Foundation and S.N. by a grant from the Wellcome Trust.

References

Acampora,D., D'Esposito,M., Faiella,A., Pannese,M., Migliaccio,E., Morelli,F., Stornaiuolo,A., Nigro,V., Simeone,A. and Boncinelli,E. (1989) *Nucleic Acids Res.*, **17**, 10385−10402.

Balling,R., Mutter,G., Gruss,P. and Kessel,M. (1989) *Cell*, **58**, 337−347.

Baron,A., Featherstone,M.S., Hill,R.E., Hall,A., Galliot,B. and Duboule,D. (1987) *EMBO J.*, **6**, 2977−2986.

Bieberich,C., Utset,M., Awgulewitsch,A. and Ruddle,F. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8462−8466.

Boncinelli,E., Somma,R., Acampora,D., Pannese,M., D'Esposito,M., Faiella,A. and Simeone,A. (1988) *Hum. Reprod.*, **3**, 880−886.

Chisaka,O. and Capecchi,M. (1991) *Nature*, **350**, 473−479.

Cho,K., Goetz,J., Wright,C., Fritz,A., Hardwicke,J. and De Robertis,E. (1988) *EMBO J.*, **7**, 2139−2149.

Dolle,P., Izpisua-Belmonte,J.C., Falkenstein,H., Renucci,A. and Duboule,D. (1989) *Nature*, **342**, 767−772.

Dressler,G.R. and Gruss,P. (1989) *Differentiation*, **41**, 193−201.

Duboule,D. and Dolle,P. (1989) *EMBO J.*, **8**, 1497−1505.

Evan,G., Lewis,G., Ramsay,G. and Bishop,J. (1985) *Mol. Cell. Biol.*, **5**, 3610−3616.

Fahrner,K., Hogan,B. and Flavell,R. (1987) *EMBO J.*, **6**, 1265−1271.

Frohman,M. and Martin,G. (1989) *Technique*, **1**, 165−170.

Gaunt,S.J., Sharpe,P.T. and Duboule,D. (1988) *Development*, **104**, Suppl., 169−179.

Graham,A., Papalopulu,N., Lorimer,J., McVey,J., Tuddenham,E. and Krumlauf,R. (1988) *Genes Dev.*, **2**, 1424−1438.

Graham,A., Papalopulu,N. and Krumlauf,R. (1989) *Cell*, **57**, 367−378.

Graham,A., Maden,M. and Krumlauf,R. (1991) *Development*, **112**, 255−264.

Hogan,B., Costantini,F. and Lacy,E. (1986) *Manipulating the Mouse Embryo*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Holland,P. and Hogan,B. (1988) *Genes Dev.*, **2**, 773−782.

Hunt,P. and Krumlauf,R. (1991) *Cell*, **66**, 1075−1078.

Hunt,P., Gulisano,M., Cook,M., Sham,M., Faiella,A., Wilkinson,D., Boncinelli,E. and Krumlauf,R. (1991a) *Nature*, **353**, 861−864.

Hunt,P., Wilkinson,D. and Krumlauf,R. (1991b) *Development*, **112**, 43−51.

Izpisua-Belmonte,J., Falkenstein,H., Dolle,P., Renucci,A. and Duboule,D. (1991a) *EMBO J.*, **10**, 2279−2289.

Izpisua-Belmonte,J.-C., Tickle,C., Dolle,P., Wolpert,L. and Duboule,D. (1991b) *Nature*, **350**, 585−589.

Kappen,C., Schugart,K. and Ruddle,F. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5459−5463.

Kessel,M. and Gruss,P. (1990) *Science*, **249**, 374−379.

Kessel,M. and Gruss,P. (1991) *Cell*, **67**, 89−104.

Kessel,M., Balling,R. and Gruss,P. (1990) *Cell*, **61**, 301−308.

Kreig,P. and Melton,D. (1988) In Berger,S. and Kimmel,A. (eds) *In vitro RNA synthesis with SP6 RNA polymerase*, Vol. 155. Academic Press, New York pp. 397−415.

Kress,C., Vogels,R., De Graaff,W., Bonnerot,C., Meijlink,F.N.J.-F. and Deschamps,J. (1990) *Development*, **109**, 775−786.

Krumlauf,R. (1991) In Murray,E. (ed.) *Northern blot analysis of gene expression*, Vol. 7. The humana press Inc, Clifton, NJ, pp. 307−323.

Krumlauf,R., Holland,P., McVey,J. and Hogan,B. (1987) *Development*, **99**, 603−617.

LaRosa,G.J. and Gudas,L.J. (1988) *Mol. Cell. Biol.*, **8**, 3906−3917.

Lathe,R., Vilotte,J. and Clark,A. (1987) *Gene*, **57**, 193−201.

Lewis,E. (1978) *Nature*, **276**, 565−570.

Lufkin,T., Dierich,A., LeMeur,M., Mark,M. and Chambon,P. (1991) *Cell*, **66**, 1105−1119.

Mavilio,F. *et al.* (1986) *Nature*, **324**, 664−667.

Murphy,P. and Hill,R. (1991) *Development*, **111**, 61−74.

Nohno,T., Noji,S., Koyama,E., Ohyama,K., Myokai,F., Kuroiwa,A., Saito,T. and Tanaguchi,S. (1991) *Cell*, **64**, 1197−1205.

Ohita,S. and Kagawa,Y. (1986) *J. Biochem.*, **99**, 135−142.

Oliver,G., Wright,C.V., Hardwicke,J. and De-Robertis,E.M. (1988) *EMBO J.*, **7**, 3199−3209.

Puschel,A., Balling,R. and Gruss,P. (1990) *Development*, **108**, 435−442.

Puschel,A., Balling,R. and Gruss,P. (1991) *Development*, **112**, 279−288.

Savard,P., Gates,P.B. and Brockes,J.P. (1988) *EMBO J.*, **7**, 4275−4282.

Schughart,K., Bieberich,C., Eid,R. and Ruddle,F. (1991) *Development*, **112**, 807−812.

Seeger,M., Haffley,L. and Kaufman,T. (1988) *Cell*, **55**, 589−600.

Simeone,A. *et al.* (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4914−4918.

Simeone,A., Pannese,M., Acampora,D., D'Esposito,M. and Boncinelli,E. (1988) *Nucleic Acids Res.*, **16**, 5379−5390.

Stornaiuolo,A. *et al.* (1990) *Cell Differ. Dev.*, *31*, 119−127.

Tuggle,C.K., Zakany,J., Cianetti,L., Peschle,C. and Nguyen-Huu,M.C. (1990) *Genes Dev.*, **4**, 180−189.

Whiting,J., Marshall,H., Cook,M., Krumlauf,R., Rigby,P., Stott,D. and Allemann,R. (1991) *Genes Dev.*, **5**, 2048−2059.

Wilkinson,D. and Green,J. (1990) In Rickwood,D. and Cockcroft,D.L. (eds) *In Situ Hybridization and Three-Dimensional Reconstruction of Serial Sections: A Practical Approach*. IRL Press, Oxford, pp. 155−171.

Wilkinson,D., Bhatt,S., Cook,M., Boncinelli,E. and Krumlauf,R. (1989) *Nature*, **341**, 405−409.

Yokouchi,Y., Sasaki,H. and Kuroiwa,A. (1991) *Nature*, **353**, 443−445.

Zakany,J., Tuggle,C.K. and Nguyen-Huu,C.M. (1990) *J. Physiol.*, **84**, 21−26.

Received on October 30, 1991; revised on February 12, 1992