Development 107, 343-359 (1989)
Printed in Great Britain © The Company of Biologists Limited 1989



# The mouse Hox-1.4 gene: primary structure, evidence for promoter activity and expression during development

BRIGITTE GALLIOT*, PASCAL DOLLÉ, MARC VIGNERON, MARK S. FEATHERSTONE**, AGNES BARON† and DENIS DUBOULE‡

*LGME du CNRS, Unité INSERM 184, Faculté de Médecine, 11 rue Humann, 67085 Strasbourg Cedex, France*

* Present address: ZMB4, Im Nevenheimer Feld 282, 6800 Heidelberg, FRG
** Present address: McGill Cancer Center, McGill University, 3655 Drummond Street, Montreal, Quebec, Canada, H3G 1Y6
† Present address: Ludwig Institute for Cancer Research, Courtauld Building, 91 Riding House Street, London W1P 8BT, UK
‡ Present address and address for correspondence: EMBL, Meyerhofstrasse 1, 6900 Heidelberg 1, FRG

## Summary

**This study reports the structure of the mouse homeobox-containing gene Hox-1.4 of the HOX-1 cluster, as well as its expression pattern during embryonic and fetal development. The overall structure of this gene includes two major exons, the second of which encodes the homeodomain. The putative Hox-1.4 protein displays similarities with products of homologous genes located at the same relative positions in other HOX clusters. A fragment extending 360 base pairs (bp) upstream of a transcriptional start site was shown to be able to promote transcription in transfected cells. This fragment is GC-rich and contains binding sites for the Sp1 transcription factor. *In situ* hybridization studies revealed the Hox-1.4 expression pattern during development. As already reported for several other murine Hox genes, Hox-1.4 is expressed in the fetal central nervous system (CNS), in structures derived from somitic mesodermal condensations (sclerotomes, prevertebrae) as well as in several mesodermal components of various organs and structures such as lungs, gut, stomach, intestine and meso- and metanephros. This expression pattern is in good agreement with recent proposals concerning the involvement of such genes in the establishment of the vertebrate body plan as well as the relationship between the positions of these genes within their clusters and the anteroposterior restriction of their expression domains.**

Key words: homeogene; homeoprotein; *in situ* hybridization; upstream elements.

## Introduction

The discovery of the homeobox in *Drosophila* (see Gehring, 1985, 1987 for reviews) has led to the characterization of many vertebrate genes harbouring this specific sequence. The 20 such homeogenes isolated so far in the mouse are found in four major clusters (Hart *et al.* 1985, 1987; Duboule *et al.* 1986; Colberg-Poley *et al.* 1985; Awgulewitsch *et al.* 1986; Featherstone *et al.* 1988), like the homeotic genes of *Drosophila* which are also clustered. Structural comparisons suggest that these clusters may have arisen in evolution by large-scale successive duplications since both the size of the intergenic spaces and the sequence of genes at equivalent positions are conserved. These genes are expressed during fetal development in the central and peripheral nervous systems, as well as in mesodermal derivatives (Gaunt, 1988; Holland and Hogan, 1988*a*). In these structures, various Hox genes display different but overlapping domains of expression along the anteroposterior (AP) body axis (Gaunt, 1988; Gaunt *et al.* 1988). It was recently proposed that, as in *Drosophila* (Lewis, 1978; Harding *et al.* 1985), the ordering of the Hox genes along their clusters might reflect the anteroposterior specificity of their expression. Thus, genes located in the 3′ part of the various complexes display anterior expression domains, whereas genes located near the 5′ end of the clusters are expressed more posteriorly (Gaunt *et al.* 1988; Duboule and Dollé, 1989; Graham *et al.* 1989). Among those genes that show an anterior expression boundary are those belonging to the Hox-1.4 subfamily; Hox-2.6 (Graham *et al.* 1988), Hox-5.1 (Featherstone *et al.* 1988) and Hox-1.4 (Duboule *et al.* 1986; Wolgemuth *et al.* 1986, 1987; Rubin, 1986). Previous work carried out with Hox-1.4 revealed that this gene is transcribed during fetal development and in spermatocytes.

Here we present the complete structure and sequence of the Hox-1.4 gene and show its structural relationship with the other members of this subfamily. We also

present data on the structure of its promoter as well as an extensive description of its expression pattern during normal fetal development.

## Materials and methods

### *cDNA libraries*

Poly (A)+ RNA was extracted according to Auffray and Rougeon (1980) from 150 10-day-old fetuses or a few adult testes. Oligo(dT)-primed libraries were prepared in the phage λgt10 (Le Bouc *et al.* 1986). About $2\times10^6$ clones were obtained before amplification in each case and, after amplification, over 1 million clones were screened with a *Hin*dIII 1.3 kb genomic fragment labelled by nick translation. Ten testis and two embryonic positive clones were isolated under high-stringency hybridization conditions (at 42°C in 50% formamide 6×SSC, 0.1% SDS, 5% dextran sulphate, 4×Denhardt's solution, washing at 60°C in 0.2×SSC, 0.1% SDS). The longest testis cDNA and the two fetal overlapping cDNAs (depicted in Fig. 1) were subcloned in pEMBL vectors (Dente *et al.* 1983).

### *DNA sequencing*

The sequences of genomic DNA and cDNAs clones were determined on both strands by the chemical sequencing procedures (Maxam and Gilbert, 1977) on either single- or double-stranded restriction fragments. Using universal or specific primers, four areas were verified by the chain termination dideoxy method (Sanger *et al.* 1977) adapted to double-stranded plasmid DNA (Chen and Seeburg, 1985).

### *Preparation of DNA probes and nuclease S1 analysis*

5 picomoles of various primers (MK100; MK101; Fig. 2) were end-labeled with T4 polynucleotide kinase in the presence of [$\gamma^{32}$-P]dATP (800 mCi mmol$^{-1}$) and further annealed to 5 μg of plasmid DNA previously denatured by NaOH treatment (Chen and Seeburg, 1985). The annealing was performed in a final volume of 25 μl, in the presence of 50 mM-NaCl, 10 mM-HCl pH 8, 10 mM-$MgCl_2$ at 65°C for 10 min and 42°C for 45 min. Then 5 μl of 5×Klenow buffer (50 mM-Tris–HCl pH 7.2, 10 mM-$MgSO_4$, 0.5 mM-DTT, 500 μM-mix dNTP, Klenow polymerase 3 i.u. μl$^{-1}$) were added and the reactions incubated at 25°C for 30 min. A 3 μl sample was removed before the polymerase was inactivated by extraction with phenol. The DNA was precipitated and resuspended in water. After restriction enzyme digestions were performed, the samples were denatured at 70°C for 10 min and run on a 6% polyacrylamide–50% urea gel. The nondigested aliquot was used as a marker.

Single-stranded probes were thus obtained of 160 nt and 180 nt long, extended from MK100 and, MK101, respectively. Hybridization of these probes in excess with 5 μg of poly (A)+ RNA or 50 μg of total RNA was carried out overnight in 20 μl of hybridization mix (50% formamide, 400 mM-NaCl, 40 mM-Pipes pH 6.2, 1 mM-EDTA) at 42°C. After S1 nuclease digestion (3 i.u. μg$^{-1}$ RNA; Appligene) at 25°C for 2 h, the protected fragments were analysed on a sequencing gel. For the experiment shown in Fig. 6, a double-stranded DNA probe was end labelled at the *Bam*HI site (B) within the second β-globin exon (Fig. 1), hybridized in 80% formamide at 50°C overnight with 50 μg of cytoplasmic RNA and further digested as above. The size of the protected fragments (209 bp) corresponds to the length between the *Bam*HI site and the beginning of the second β-globin exon. As control plasmid, the hybrid SV40 early promoter/rabbit β-globin plasmid pA0 (Zenke *et al.* 1986) was used.

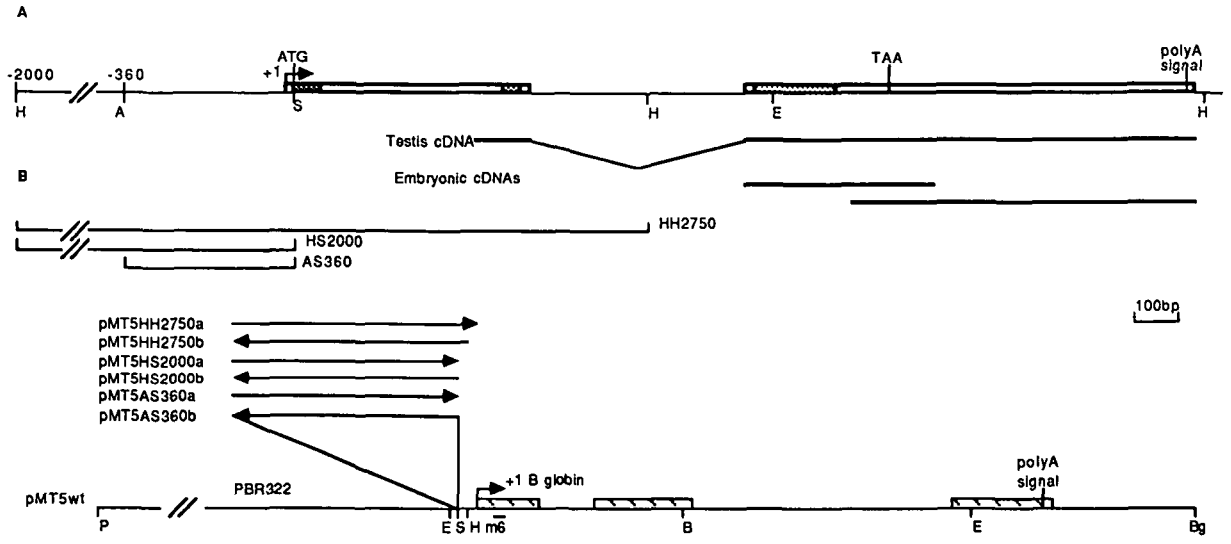


**Fig. 1.** (A) Genomic map of the murine Hox-1.4 gene. The two rectangles represent the exons. The arrow shows the orientation of transcription from the start site. The three hatched boxes indicate the three main regions of similarity with related homeogenes (see Fig. 3) from 5′ to 3′: the DNA portion coding for the first 22 aminoacids (aa), the hexapeptide, and the homeodomain. The longest testis cDNA clone and the two fetal overlapping cDNAs are outlined as thick lines. (B) Recombinant plasmids assayed for promoter activity. The following Hox-1.4 fragments were inserted in both orientations into the pMT5 vector in front of rabbit β-globin sequences: the 2750 bp *Hin*dIII–*Hin*dIII fragment (HH2750) encompassing 2000 bp upstream of the start site, the 2000 bp *Hin*dIII–*Sac*I fragment (HS2000) and the 360 bp *Apa*I–*Sac*I fragment (AS360) harboring only putative upstream sequences. These constructs are named pMT5HH2750a, pMT5HH2750b, pMT5HS2000a, pMT5HS2000b, pMT5AS360a, pMT5AS360b respectively. In each case 'a' refers to the 5′-3′ orientation and 'b' to the reverse orientation as indicated by arrows. The rectangles represent the β-globin exons. The 22mer oligonucleotide m6 is complementary to the first exonic sequence (+39 to +60). H, *Hin*dIII; E, *Eco*RI; S, *Sac*I; A, *Apa*I; Bg, *Bgl*II; B, *Bam*HI; P, *Pvu*II.

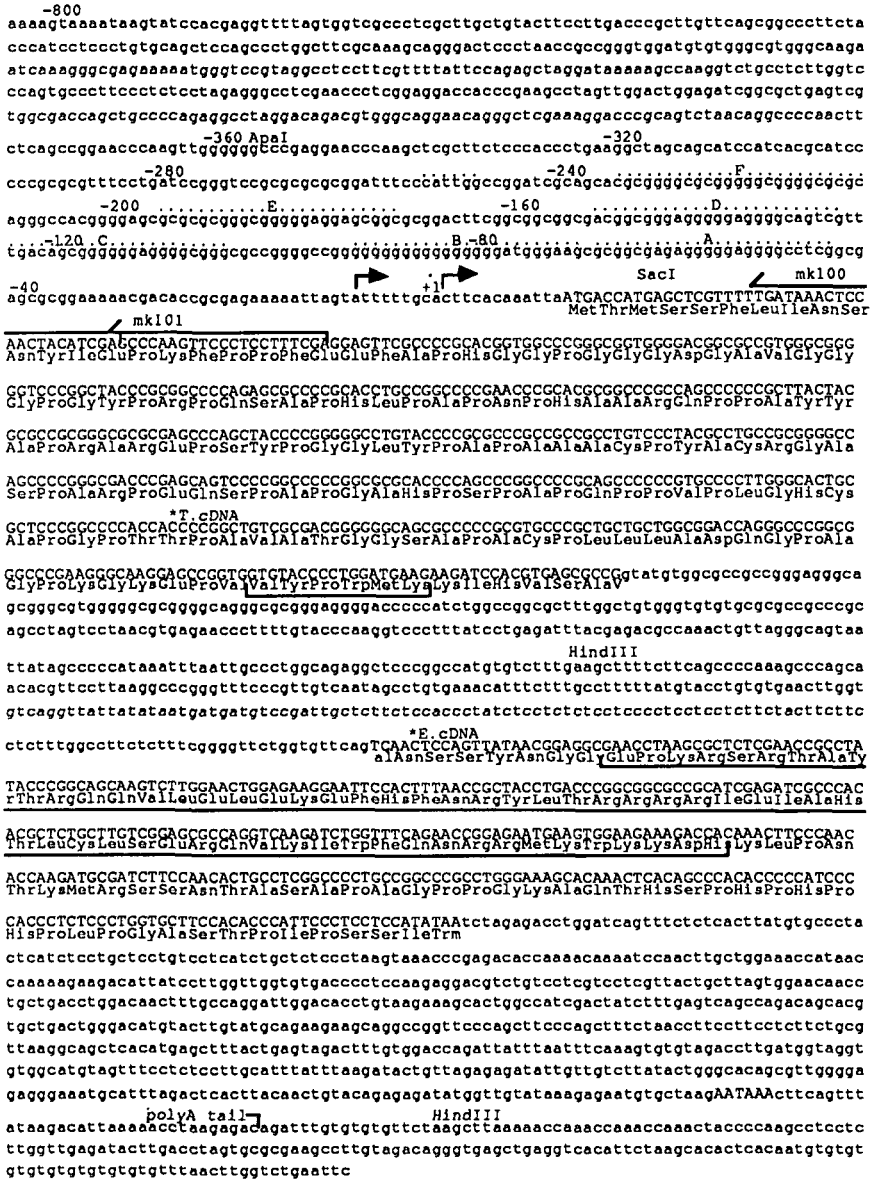

**Fig. 2.** DNA and predicted protein sequences of the murine Hox-1.4 gene. The sequences of the Hox-1.4 cDNA clones and genomic region have been determined. No divergence was observed between cDNA and genomic sequences. Sequences encoding the predicted protein are in capital letters and sequences corresponding to upstream, intronic and trailer regions are in small letters (except the polyadenylation signal AATAAA). The homeobox and hexapeptide sequences are underlined. The transcription start site is numbered +1. *T.cDNA and *E.cDNA refer to the 5′ end of the testis and embryonic cDNAs respectively, their 3′ end extending to the same poly A site. The MK100 and MK101 oligonucleotides were used for the determination of the start site (Fig. 5).

### *Recombinant plasmids used for transfections*

Putative upstream Hox-1.4 sequences were subcloned in the pMT5 plasmid (gift from M. Zenke) which harbors the rabbit β-globin-coding sequence. The Hox-1.4 sequences were inserted in the multiple cloning site and consisted of the 2750 bp genomic *Hin*dIII–*Hin*dIII fragment in both orientations (HH2750a; HH2750b; Fig. 1), the 2000 bp *Hin*dIII–*Sac*I and the 360 bp *Apa*I–*Sac*I fragments in both orientations (HS2000a, HS2000b, AS360a, AS360b, Fig. 1). Large-scale plasmid preparations and purification of supercoiled DNA by two rounds of CsCl gradient centrifugation were as previously described (Zenke *et al.* 1986). The DNA sequence of the insertion in AS360a was verified before transfection.

### *Cell lines and DNA transfection*

Human HeLa and mouse F9 cells were grown in Dulbecco's modified Eagle's minimal essential medium (DMEM; GIBCO) containing 10% fetal calf serum, 500 i.u. of penicillin, 400 $\mu g\,ml^{-1}$ of gentamycin and 100 $\mu g\,ml^{-1}$ of streptomycin. Calcium-phosphate-mediated transfections of all cell lines were carried out using 10 $\mu g$ of the appropriate plasmid per 9 cm dish. Cells were seeded 6–10 h before transfection to give 50–60% confluence. The efficiency of transfection was controlled with the β-galactosidase pCH110 plasmid (Hall *et al.* 1983). In order to induce differentiation *in vitro*, F9 cells were treated the following day with retinoic acid (0.5 $\mu M$) for 2 days. Cytoplasmic RNA was then harvested by the Nonidet P40 lysis method of Groudine *et al.* 1981.

### *Primer extension mapping*

One picomole of the m6 oligonucleotide (5′-GCAC-CATTCTGTCTGTTTTGGG-3′), complementary to the β-globin sequence between portion +39 and +60, was labelled with [$\gamma$-$^{32}$P]dATP (800 Ci $mmol^{-1}$) and hybridized with 50 $\mu g$ of total RNA in 10 $\mu l$ of 250 mM-NaCl, 40 mM-Pipes pH 6.4, 1 mM-EDTA and 0.2% SDS, overnight at 42°C after an incubation at 60°C for 1 h. After alcohol precipitation, the samples were resuspended in 30 $\mu l$ of reverse transcriptase buffer containing 50 mM-Tris pH 8.3, 50 mM-KCl, 5 mM-$MgCl_2$, 5 mM-DTT, 500 $\mu M$-dNTP, 5 i.u. of AMV reverse transcriptase and incubated at 42°C for 45 min. Alcohol-precipitable material was collected and resuspended in 5 $\mu l$ of

95% formamide, denatured at 70°C for 10 min and run in a sequencing gel.

*Preparation of nuclear extracts*

DNAse I footprinting and gel retardation assays were performed according to Xiao *et al.* (1987) and Davidson *et al.* (1986). For these experiments, the following oligonucleotides were used: GC+:(5′-GCCCCTAACTCCGCCCAGTTC-3′) and its complementary strand, which corresponds to the GC box motif (Barrera-Saldana *et al.* 1985) and bind the transcription factor Sp1 (Briggs *et al.* 1986). GC−:(5′-GCCCCTAACTCAGTCCAGTTC-3′) and its complementary strand, which correspond to the same sequence with point mutations that prevent Sp1 binding. HC+: (5′-GCGGAGGGGAGGGA-CAGTC-3′) and its complementary strand, which are derived from the consensus sequence from Fig. 7B. HC−: (5′-GCTGGATGGTGAGTG-GAAGTC-3′) and the complementary strand, which represent the same consensus with several point mutations.

In situ *hybridization*

The *in situ* hybridization experiments were carried out on paraffin-embedded embryo and fetus sections according to Gaunt *et al.* 1986 with minor modifications (Duboule and Dollé, 1989). The probe is identical to that described in Gaunt *et al.* (1988). Exposure times were from 14 to 21 days.

## Results

*Structure of the Hox-1.4 gene*

The mouse Hox-1.4 gene maps to the HOX-1 cluster on chromosome 6 (Duboule *et al.* 1986; McGinnis *et al.* 1984; Bucan *et al.* 1986). Previous work (refs. above) showed Hox-1.4 to contain a homeobox related to the Antp group and to be expressed in embryos from day 9 to day 15 as two transcripts, a major one 1.4 kb pairs long and a minor one of 2.5 kb. In adult mouse testis, a shorter [1.25 kb] transcript was detected. A common acceptor splice site for testis and embryonic RNAs was mapped by $S_1$ nuclease analysis 27 bp upstream of the homeobox (Duboule *et al.* 1986). Using adult testis and day 10 fetal poly $(A)^+$ mRNA, we have constructed oligo(dT)-primed cDNA libraries in λgt10 (LeBouc *et al.* 1986). For each of these libraries, over 1 million clones have been screened and different Hox-1.4 cDNAs have been obtained and sequenced (Fig. 1). Both testicular and embryonic cDNAs extend 801 bp past the 3′ end of the homeobox to the poly A tail, 35 bp downstream of a classical polyadenylation signal. The deduced ORF, which translates the homeobox, ends 47 codons downstream of this domain and is followed by a 660 bp AT-rich untranslated region. Comparison of a partial testis cDNA (1126 bp long) with genomic sequences (Fig. 2) has revealed a 486 bp intron located just upstream of the homeobox, confirming the previously mapped acceptor splice site. The exon which is thus brought next to the homeobox encodes the hexapeptide Val-Tyr-Pro-Trp-Met-Lys found 5′ proximal to the homeodomain of many genes (Mavilio *et al.* 1986).

The genomic sequence upstream of the 5′ end of these partial cDNA clones reveals a long open reading frame which is continuous with the ORF of the homeobox and which displays at its putative *N*-terminus the triplet Met-Ser-Ser. This sequence presents striking similarities with the *N*-termini of most homeogenes studied so far (Mavilio *et al.* 1986; Duboule *et al.* 1989) and therefore likely localizes the Hox-1.4 translation start site (Fig. 2). Within the 800 bp upstream of this putative initiation codon, no canonical TATA box is detected, though there is a 50 bp AT-rich stretch immediately upstream which is preceded by a GC-rich region (over 80% GC over 250 bp). The Hox-1.4 putative translation product would thus be 285 amino acids long, with a relative molecular mass of about $30 \times 10^3$. This protein has a high content of proline (Pro≈18%) and alanine (Ala≈12%). The comparison of this putative Hox-1.4 protein with different homeogene products reveals striking similarities, particularly with those encoded by the mouse genes Hox-5.1 (Featherstone *et al.* 1988) and Hox-2.6 (Graham *et al.* 1988), the human genes C13 (Mavilio *et al.* 1986) and cp 19 (Simeone *et al.* 1988), the *Drosophila* homeotic gene *Dfd* (Regulski *et al.* 1987) and the *Xenopus* gene XHox1A (Harvey *et al.* 1986) (Fig. 3). These similarities are mainly localized in 3 regions including a 22 aa stretch at the *N*-terminus, the region of the hexapeptide and within the homeodomain itself (overlined in Fig. 3). In addition, DNA sequence comparisons reveal conservation immediately upstream of the start codons of these different genes (not shown). All these homeogenes thus belong to the Hox-1.4-like (*Dfd*-like) subfamily as predicted from their respective localizations along the different clusters (Duboule and Dollé, 1989; Boncinelli *et al.* 1988; Graham *et al.* 1989).

Downstream of the homeodomain, no striking similarities with other homeogene products are observed except with the human Hox-1.4 gene which presents a very comparable ORF (E. Boncinelli; pers. comm.). The 510 bp sequence between the Met-Ser-Ser triplet and the region encoding the conserved hexapeptide is extremely GC-rich and only partially covered by the testis cDNA clones (Fig. 1).

Two different oligonucleotides (Fig. 2, MK100, MK101) were used to map the 5′ end of the Hox-1.4 transcripts present in embryonic and F9 cell mRNAs, by S1 nuclease analysis. Two protected fragments, 54 and 62 nucleotides (nt) long (MK100) or 75 and 83 nt long (MK101) were obtained when fetal mRNAs were used whereas only the longer one was seen with differentiated F9 cells mRNA (Fig. 4). These two bands might represent different start sites. Alternatively, the larger one could be due to the particular sequence in this area where a stretch of five A residues precedes a stretch of five T residues (immediately upstream of position +1 in Fig. 2; also visible in the sequencing lane in Fig. 4). The resulting hairpin could prevent S1 digestion from occurring at the right place. We therefore map a possible start site 12 bp upstream from the translation start site shown in Fig. 2 (arrow), in a sequence (TTTTGCACTTC) that resembles the consensus eukaryotic cap site (Bucher and Trifonov, 1986). However, the absence of a consensus TATA box, and the observation of a transcript 45 nt longer by using

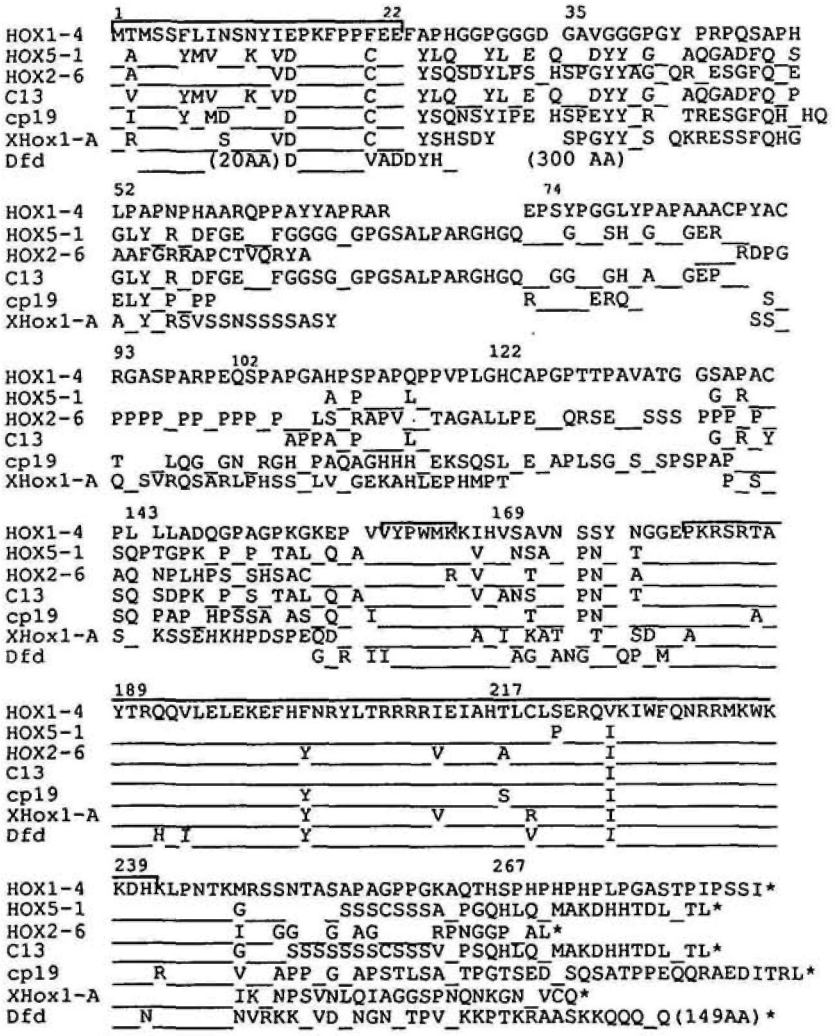


**Fig. 3.** Comparison of the Hox-1.4 predicted protein with other related homeogene products. The murine Hox-1.4 sequence, at the top, is used as the basis for comparison with other sequences from mouse (Hox-5.1, Hox-2.6), human (C13, cp19), *Xenopus* (XHox1A) and *Drosophila* (Dfd). For references, see the text. Hyphens represent identical amino acids whereas spaces are used to facilitate the best possible alignment. The three main regions of consensus are underlined: the *N*-terminal part, the hexapeptide and the homeodomain.

reverse transcriptase on the same mRNAs (data not shown), suggested the possibility of multiple start sites without excluding the presence of a short additional exon, located upstream. We therefore looked at the promoter activity of the region located immediately upstream of these putative start sites. 2 kb and 360 bp fragments extending upstream from the initiation codon were inserted in both orientations in front of rabbit $\beta$-globin coding sequences (plasmids pMT$_5$:HS2000a, HS2000b and AS360a, AS360b respectively, in Fig. 1). These plasmids were transfected into HeLa, Cos and F9 cells, and a primer complementary to the $\beta$-globin first exon (m6, Fig. 1) was used to detect the presence of $\beta$-globin mRNAs by the primer extension assay. A 106 nt long extension product was obtained after reverse transcription from the m6 primer annealed to RNA from cells transfected with constructs carrying the presumptive Hox-1.4 promoter in its correct 5′ to 3′ orientation in front of rabbit $\beta$-globin sequences (shown

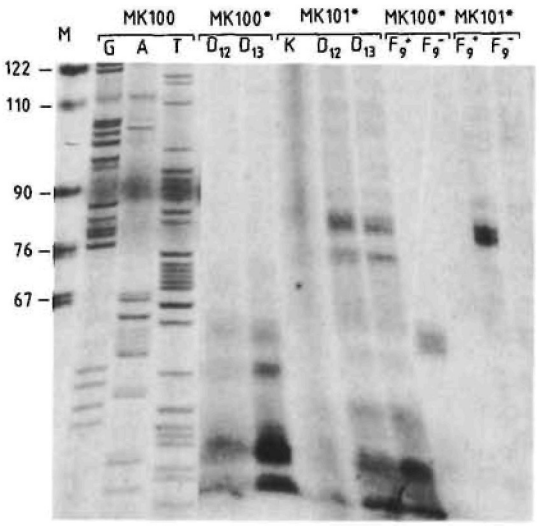


**Fig. 4.** S1 mapping of the murine Hox-1.4 transcription start site. S1 nuclease mapping was carried out using 2 single stranded probes extended from the end-labelled MK100 and MK101 primers (the MK101 probe is 20 nucleotides (nt) longer than the MK100). 50 $\mu$g of the various RNA samples were used in each assay. D12 and D13: total day-12 and day-13 fetal mRNA; K, total adult kidney mRNA; F9+, cytoplasmic RNA from untransfected, RA-treated F9 cells; F9−, cytoplasmic RNA from untransfected, untreated F9 cells. The G,A,T, panels show the sequence obtained with the dideoxy method using the MK100 oligonucleotide as primer.

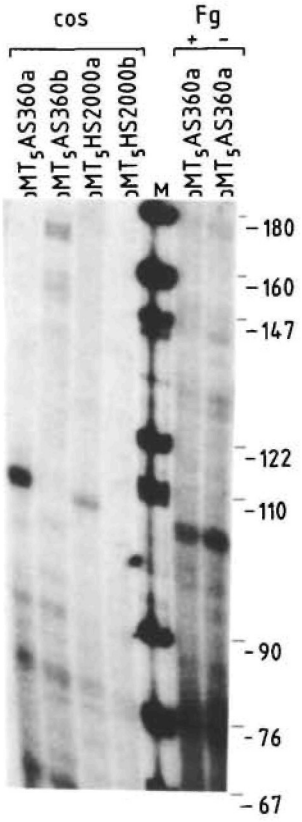


**Fig. 5.** Hox-1.4 promoter activity in transfected cells. Promoter activity was assayed by primer extension. Only the results obtained with either Cos or F9 cells treated (F9+) or not treated (F9−) with RA are shown. The various transfected constructs are indicated on the top and diagramed in Fig. 1. Band sizes (nt) are on the right. The two constructs with the 5′–3′ orientation (a) are able to promote the transcription of the rabbit β-globin gene as judged by the 106 nt long product extended from the m6 primer (see Fig. 1). Ten times less RNA was used in the case of Cos cells.

for Cos and F9 cells, Fig. 5). No specific extension products were detected when promoter fragments were cloned in the reverse orientation. The upstream end of the 106 nt extension product maps next to the 5′ end of the shorter fragment protected against S1 nuclease digestion when embryonic mRNAs are used (Fig. 4). Thus, both upstream Hox-1.4 fragments (2 kb and 360 bp) when inserted in the normal 5′-3′ orientation, were able to promote β-globin transcription in transfected cells. The same hybrid transcripts in approximately identical amounts were observed for each case. No differences in the amount of transcripts were detected when F9 cells, transfected with the 360 bp upstream region (pMT5AS360a), were treated with retinoic acid (RA Fig. 5). In contrast, when a 2750 bp fragment, including 2 kb of sequences upstream of the translation initiation site and extending 750 bp downstream, was used to promote transcription in F9 cells, the amount of β-globin transcripts increased with the addition of RA (Fig. 6). The first 750 bp of the coding sequence or the additional upstream region may therefore play a role in the response to RA. Thus, a 360 bp

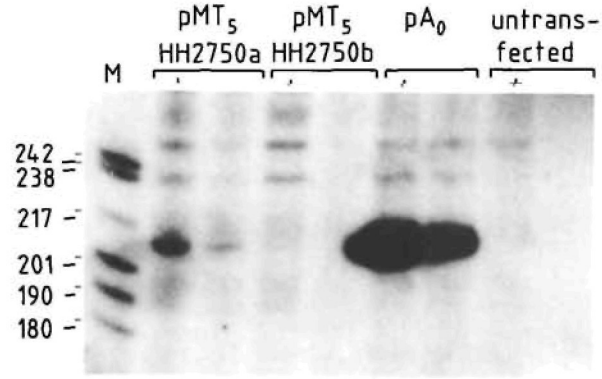


**Fig. 6.** The effect of RA on transfected F9 cells. The Hox-1.4/β-globin hybrid constructs used are diagramed in Fig. 1. Only the correct orientation (a) gave detectable amounts of β-globin transcripts, significantly higher after treatment of the cells with retinoic acid (+). The expected increase (3- to 5-fold) is observed with the control plasmid (pAO, see material and methods and Nomiyama *et al.* 1987).

sequence located immediately upstream of the Hox-1.4 coding sequence is sufficient to promote transcription in various transfected cells and likely harbours *cis*-acting promoter elements. This sequence contains several GC boxes as well as an inverted CCAAT box (position-250 in Fig. 2) also present in upstream sequences of most of the Hox-1.4-like homeogenes (B.G. unpublished results). In order to see whether these upstream elements may be functional in the context of the cell lines used for the transfection, an end-labelled DNA fragment corresponding to the upstream 360 bp was incubated with nuclear cell extracts obtained from various cell lines and DNAse I footprint experiments were carried out. Comparable results were obtained for all cell lines and only those obtained with Hela and F9 cells are shown (Fig. 7). Six major protections were detected, which correspond to six GC-rich sequences A to F in Fig. 7A. Three of these contained a genuine Sp1 consensus binding site (GC boxes, C,E,F; Fig. 7A, see Dynan and Tjian, 1983) whereas the others were related to it. In order to see if Sp1 is involved in these various protections, competition experiments were carried out in the presence of an Sp1-binding GC box oligonucleotide (GC+) and an oligonucleotide consensus for the 6 protections observed (HC+; see Fig. 7B). While the consensus oligonucleotide (HC+) could compete with all six binding sites and therefore deprotect them (Fig. 7A), the Sp1-binding oligonucleotide (GC+) competed clearly with only 2 of the protected sequences (E and F) and could partially deprotect three other sequences (A, C and D), consistent with the presence in E, F and C of genuine Sp1 consensus sequences. The inability of the GC+ oligonucleotide to deprotect sites B and, to some extent, A, C and D suggests the existence of another factor(s) which recognizes GC-rich motifs upstream of the Hox-1.4 transcription start site and which can bind the HC+ oligonucleotide. This conclusion is confirmed by the visualization of the various nucleoprotein complexes by the gel retardation assay. Hela cell extracts incubated with the HC+

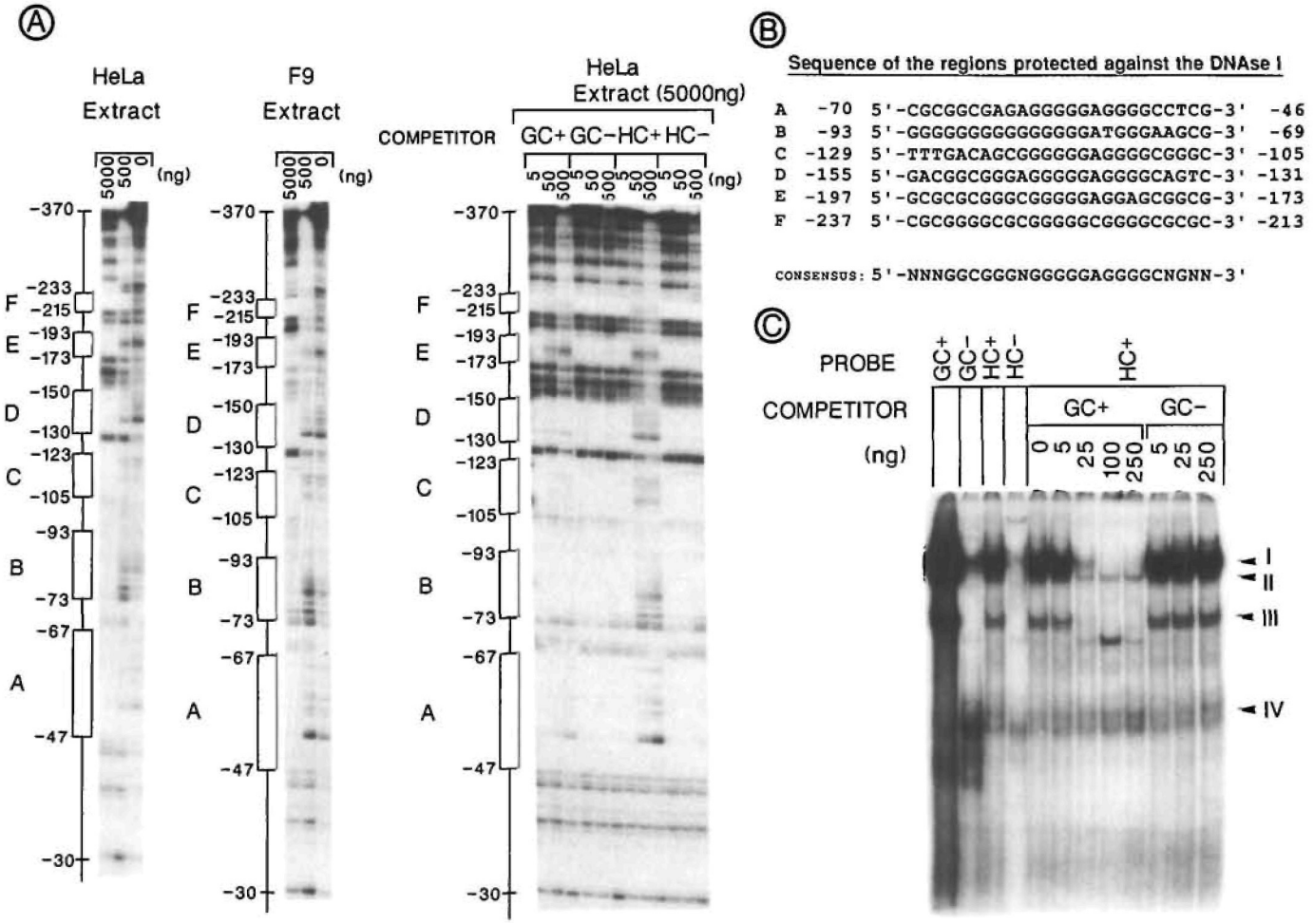


**Fig. 7.** DNA-binding activities upstream of the Hox-1.4 transcription start site. (A) DNAse I footprinting. The template used is the end-labelled pMT5A360a fragment (360 bp; Fig. 1). The left and middle panels show the protections obtained with Hela and F9 cell nuclear extracts at two different protein concentrations (top). The positions of the six major protected regions (A to F) are indicated on the left. The right panel shows a competition between a Hela cell nuclear extract (5000 ng) and various amounts of wild-type (GC+; HC+) and mutated (GC−; HC−) GC box and Hox-1.4 consensus oligonucleotides (see legend to 7B and in Materials and methods for sequences). While both GC− and HC− have no effect, HC+ can deprotect all six sequences whereas GC+ can deprotect only E, F, and, to a lesser extent, A, C and D. (B) Sequence comparison of protected regions. The letters correspond to those indicated under A. The consensus sequence is shown below. (C) Analysis of the nucleoprotein complexes formed after incubation with a Hela cell nuclear extract. The probes used are shown in the top and various amounts of cold competitor (GC+ or GC−) are indicated below. Four different complexes are identified and shown on the right side of the picture (I to IV).

oligonucleotide give four retarded bands by this technique. Competition with the Sp1-binding GC+ cold oligonucleotide disrupted the formation of some complexes (I and III; Fig. 7C) whereas complexes II and IV did not seem to be affected and therefore probably did not contain Sp1. The mutated Sp1 binding oligo (GC−) was unable to interfere with nucleoprotein complex formation. These results suggest that GC-rich elements of the Hox-1.4 promoter interact with Sp1 and other as yet unidentified transcription factors.

### *Expression of the Hox-1.4 gene*

The expression of Hox-1.4 during prenatal development was studied by *in situ* hybridization performed on serial sections of mouse embryos or fetuses from 7.5 to 15.5 days post coitum (p.c) using a probe located 3′ to the homeobox (see Materials and methods). As a control for each experiment, adjacent sections were hybridized with the sense probe and showed no significant hybridization. Neither was any signal seen upon hybridization of the antisense probe to sections that had been pretreated with RNAse A (Fig. 8C and D).

### *Expression of Hox-1.4 at early stages of development*

Hox-1.4 transcripts are detected at the earliest stage that we studied, i.e. 7.5 days p.c. At this stage, the mouse embryo is composed of three germ layers, and the neural groove and notochord are forming. Several sections through the posterior regions of the embryo show specific labelling, which seems to be restricted to the mesodermal germ layer (Fig. 9). The ectodermal layer does not seem to be labelled above background.

However, a definitive observation would require a more precise technique such as the use of antibodies. We therefore cannot exclude a weak but specific signal in restricted parts of the embryonic ectoderm. No hybridization is found in the allantois or any other extraembryonic tissues.

At day 9.5, more than 20 somites have formed, and the most anterior ones are differentiating. The neural

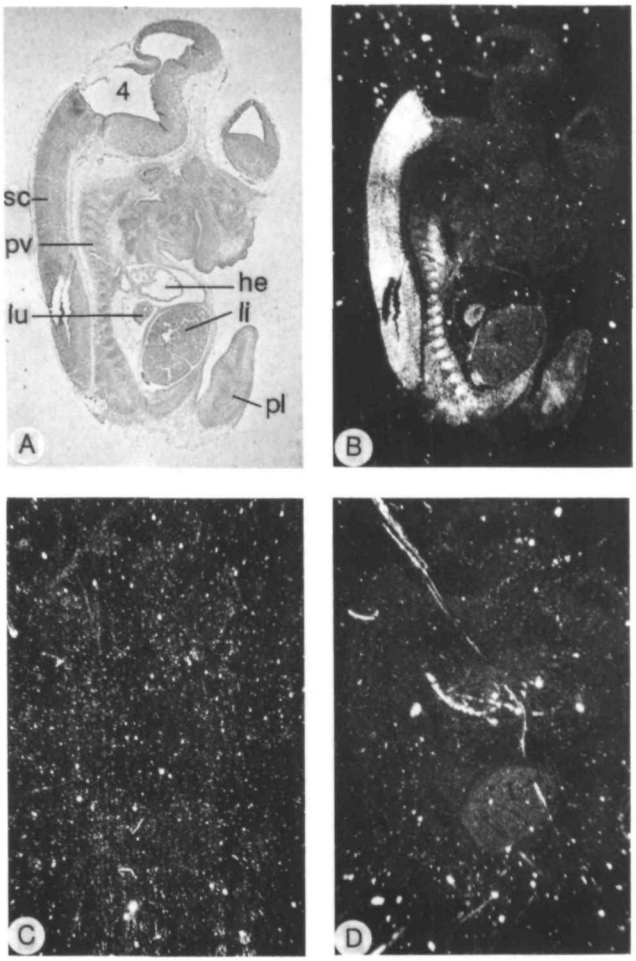


**Fig. 8.** Expression of the Hox-1.4 gene as seen by *in situ* hybridization using sagittal sections through a 13.5-day-old fetus and specificity controls. The section was hybridized to the Hox-1.4 antisense probe and is shown under bright-field (A) or dark-field (B) illumination. Adjacent sections hybridized with the Hox-1.4 sense probe (C) or first treated with RNAse A and hybridized to an antisense probe (D) are shown as controls. 4, fourth ventricle; sc, spinal cord; pv, prevertebra; lu, lungs; he, heart; li, liver; pl, posterior limb. A,B,C,D: magnification is 16×; exposure time was 20 days.

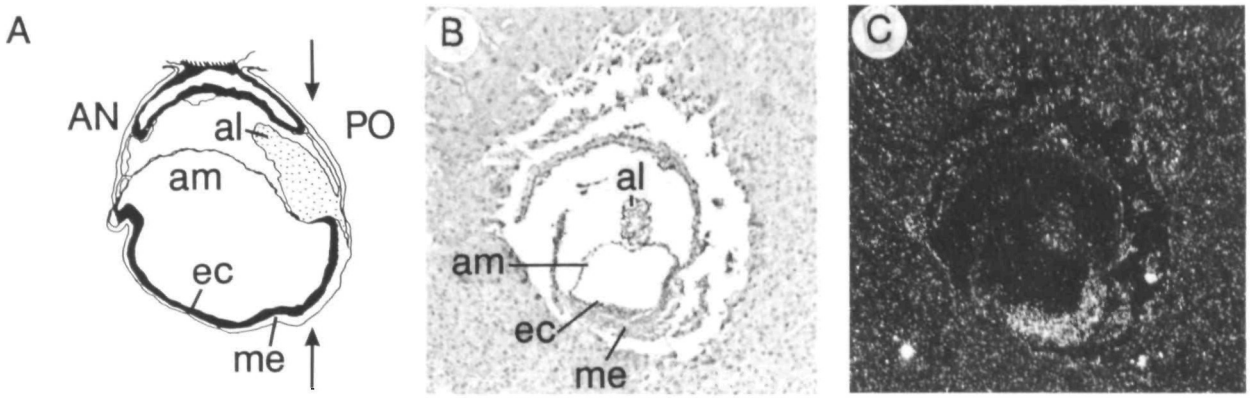


**Fig. 9.** Expression of Hox-1.4 in the 7.5-day-old mouse embryo. (A) Scheme of a medial longitudinal section through a 7.5-day-old embryo. The arrows indicate the plane of section shown in B and C. (B–C): Transverse section through the posterior part of a 7.5-day-old embryo hybridized with the Hox-1.4 antisense probe. AN, anterior; PO, posterior; al, allantois; am, amnion; ec, ectoderm; me, mesoderm.

tube has closed except in the regions of the anterior and posterior neuropores. At this stage, the expression of Hox-1.4 is higher and clearly restricted to the neural tube (CNS), the lateral plate mesoderm (Fig. 10 B), the mesodermal layer of the primitive intestine (Fig. 10 D, stomach primodium), and the para-axial mesoderm, which differentiates into segmented nephrotomes (Fig. 10 F). There is a sharp boundary of hybridizing cells between the wall of the stomach, where the hybridization is strong, and the anlagen of the liver, which is negative (Fig. 10 D, arrow). A homogeneous signal is also detected in the limb buds.

Hybridization is thus found in mesodermal and neuroectodermal tissues, but is restricted along the AP axis so that more anterior regions like the cephalic and cardiac tissues do not express Hox-1.4 (Fig. 10 D).

*Expression of Hox-1.4 in the nervous system*

In all developmental stages studied, a strong signal is observed in the CNS. The signal is most intense in the

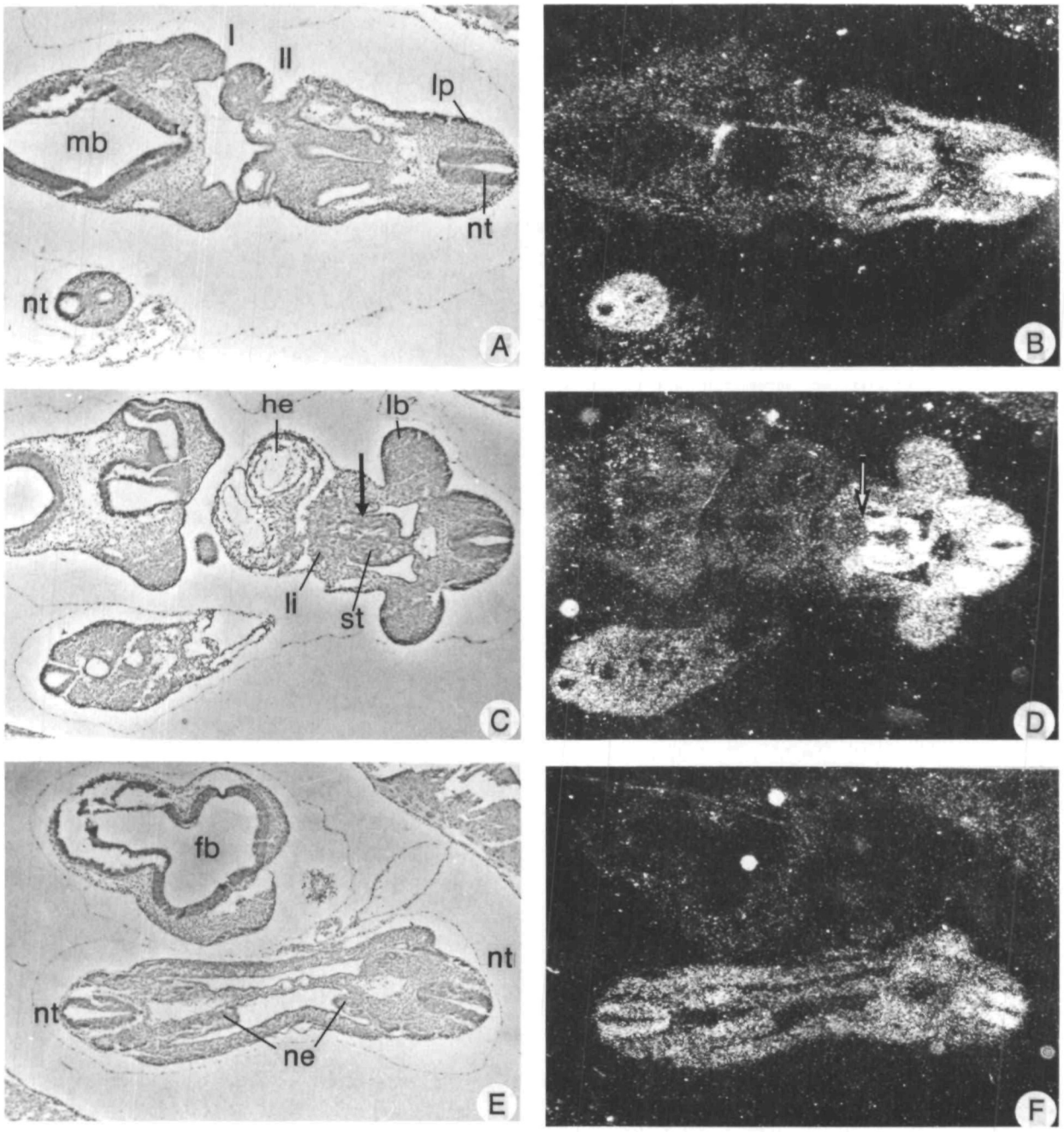


**Fig. 10.** Expression of Hox-1.4 in the 9.5-day-old mouse embryo. (A–F) Three serial transverse sections of a 9.5-day p.c. embryo hybridized with the Hox-1.4 antisense probe are shown. The sections are at the levels of the first branchial bars and midbrain (A, B), the heart (C, D) and the primitive kidney (E, F). I-II, first and second branchial bars; nt, neural tube; mb, midbrain; lp, lateral plate mesoderm; he, heart; lb, forelimb bud; st, stomach (primordium); li, liver; fb, forebrain; ne, nephrotome. The arrow indicates the limit between the stomach and the liver primordium.

hindbrain and upper spinal cord and decreases posteriorly, while remaining above background levels up to the most posterior regions of the spinal cord (Fig. 11). At 12.5 days p.c., the anterior boundary of expression lies within the myelencephalon, immediately posterior to the otic vesicle (Fig. 11 B to D). The boundary extends more anteriorly in the ventral floor of the fourth ventricle than in the roof plate. At 15.5 days p.c., the ventral boundary lies in the anterior part of the medulla oblongata, next to the limit between the medulla and the pons (data not shown). There is therefore a sharp boundary between Hox-1.4 expressing and nonexpressing neuroblasts in a region of high cellular mobility. The posterior boundary seems to become more anterior as development proceeds but this may be due to the overall decrease in the amount of transcripts, increasing the contrast between highly expressing areas (myelencephalon) and poorly expressing areas (posterior part of the neural tube). The hybridization is found only in areas of grey matter (which contain the cell bodies), whereas the presumptive areas for white matter (which have a fibrillar morphology) are negative.

At 12.5 days p.c., the ependymal cavity is surrounded by a layer of neuroepithelial cells of high mitotic activity, whereas the mantle layer is composed of neuroblasts, precursors of neurons and glial cells. Transverse sections of fetuses at this stage show that labelling in the spinal cord is not uniform: the neuroepithelial layer is not labelled (or very weakly), whereas a strong signal is observed in the mantle layer, especially in the ventral horns and dorsal regions, at the level of the cervical spinal cord (Fig. 11 E). From the thoracic spinal cord downwards, only the dorsal portion of the spinal cord is intensely labelled (Fig. 11 F and G). In addition, all the detected spinal ganglia are labelled,

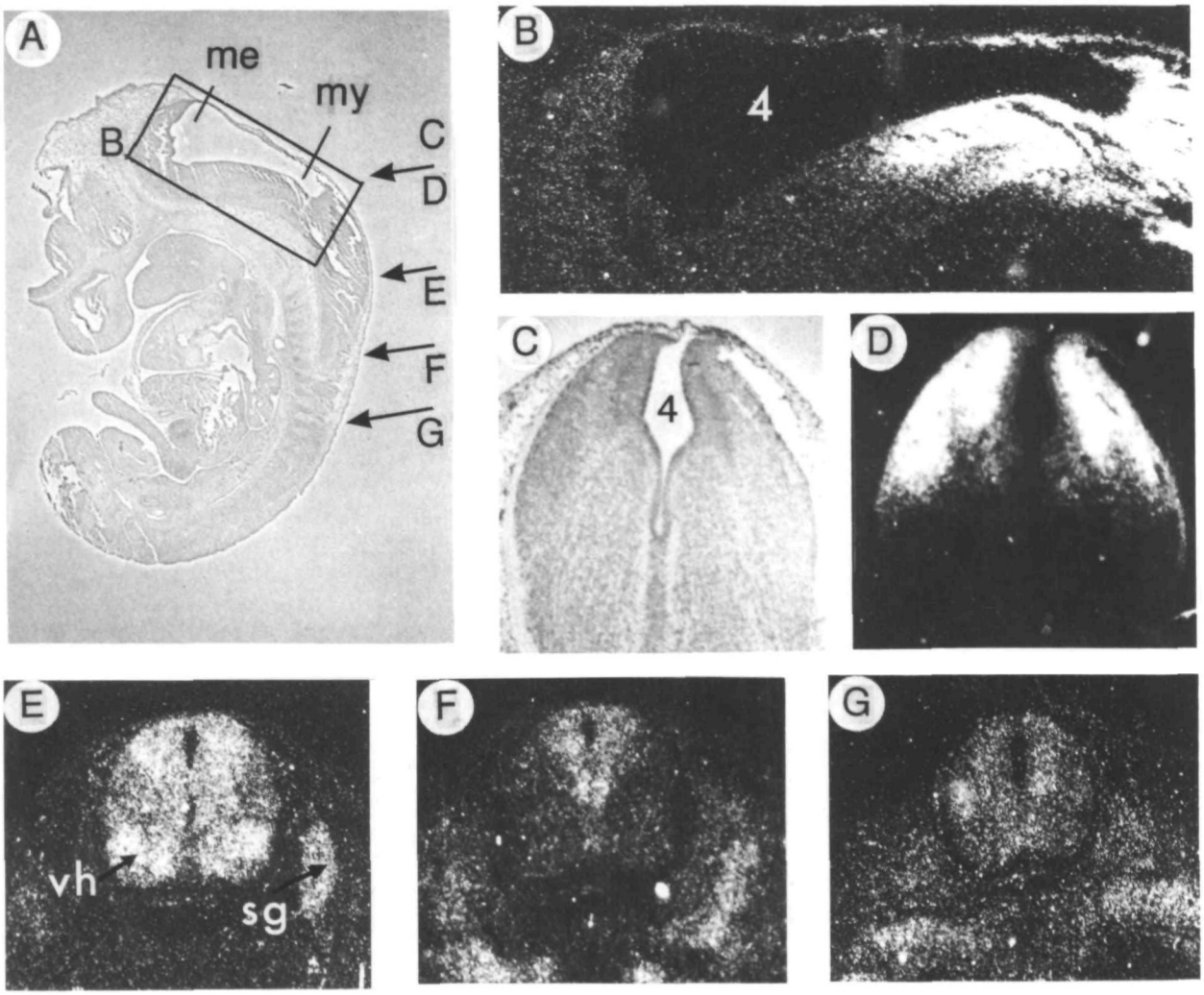


**Fig. 11.** Expression of Hox-1.4 in the nervous system of a 12.5-day-old mouse fetus. (A) Sagittal section (the plane of section is slightly skewed) of a 12.5- day-old fetus. The arrows indicate the planes of the transverse sections shown in panels C to G. (B) Enlargement of the hindbrain region (boxed in A) showing the anterior boundary of Hox-1.4 expression within the myelencephalon, also shown on a transverse section at about the same level (C–D). (E–G) Serial transverse sections from the cervical to the lumbar spinal cord of the same fetus. 4, fourth ventricle; me, metencephalon; my, myelencephalon; sg, spinal (dorsal root) ganglion; vh, ventral horn of the spinal cord.

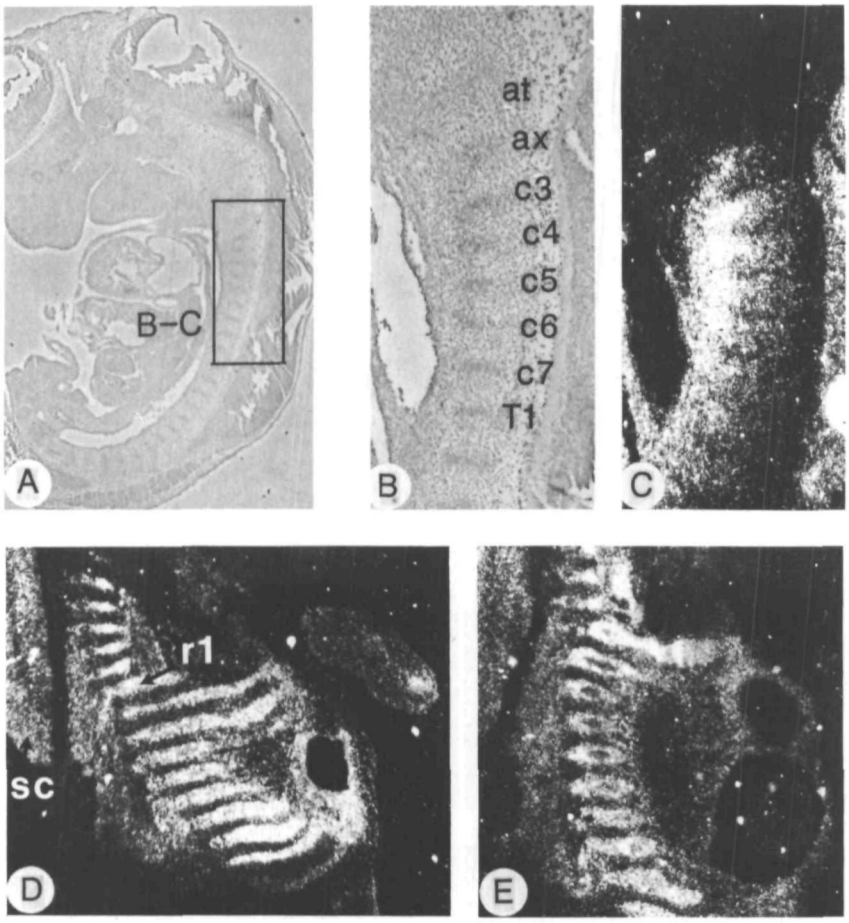


**Fig. 12.** Expression of Hox-1.4 in the prevertebral column and the ribs. (A) Sagittal section of a 12.5-day fetus hybridized with the Hox-1.4 antisense probe. (B–C) Enlargement of the region of the cervical prevertebrae (boxed in A) showing strong hybridization from the third cervical prevertebra to the first thoracic prevertebrae. A weaker but clear signal is found in the axis (second cervical prevertebra) and in more posterior thoracic prevertebrae. (D–E) Two serial parasagittal sections of a 13.5 day fetus focused on the region of the ribs, and viewed under dark-field illumination. at, atlas (pv1); ax, axis (pv2); C3 to C7, Third to seventh cervical pv; sc, spinal cord; r1, first presumptive rib.

though the intensity of the signal gradually decreases from the upper thoracic level to the most posterior ganglia (see Fig. 11 E).

*Expression of Hox-1.4 in the sclerotomic derivatives*

The presumptive vertebral column at day 12 of fetal development consists of metameric condensations derived from the segmented sclerotomes (somitic mesoderm) and positioned around the notochord. Each condensation is the anlagen of an intervertebral disk, while the adjacent loosely packed regions will give rise to the bodies of the vertebrae (Verbout, 1985).

Contrary to a previous report (Toth *et al.* 1987), Hox-1.4 transcripts are detected in the prevertebral column, with the exception of the first cervical prevertebra (pv), the atlas, which is negative (Fig. 12 C). The axis (second cervical pv) is weakly labelled, the intensity of the signal increasing in more posterior cervical pv. The labelling begins to weaken at the second thoracic pv but remains slightly above background until the most posterior identified pv (caudal pv). A strong hybridization signal is also present in the presumptive ribs which, like the vertebrae, are derived from the sclerotome (Fig. 12D). At day 13 of development, the future ribs consist of mesodermal cells differentiating into cartilaginous cells. The ossification centers will only appear by day 14 (Theiler, 1972). As seen in Fig. 12 E, the central parts of the future ribs are not labelled, whereas the cells surrounding the cartilaginous cells display a high level of Hox-1.4 expression. At day 15 p.c., hybridization will then be restricted to thin crowns around developing ribs (visible on the top of Fig. 13F). In this case, the cells expressing Hox-1.4 are likely to be mesodermal cells in the course of cartilage differentiation, the expression of Hox-1.4 being turned off in differentiated chondroblasts.

*Expression of Hox-1.4 in other mesodermal derivatives*

Hox-1.4 transcripts are also detected in structures partially derived from lateral plate or para-axial mesoderm, including fetal lung, intestine and kidney. In addition, some cervical and thoracic regions of mesodermal origin hybridize to the Hox-1.4 antisense probe. These include the mesenchymal tissue located between the aortic arch and the base of the heart (Fig. 13 B), and the mesenchyme surrounding the base of the larynx and the trachea (Fig. 13 B). In this latter region, as well as in the diaphragm (Fig. 14 B), the signal is particularly strong.

In the embryonic lungs, Hox-1.4 expression is restricted to mesenchyme. The endodermally derived cells forming the epithelium of the bronchi are not labelled (Fig. 13 C–F). At day 12.5 p.c., the hybridization is stronger in the peripheral parenchyma than in the peribronchial regions. Expression in the stomach and

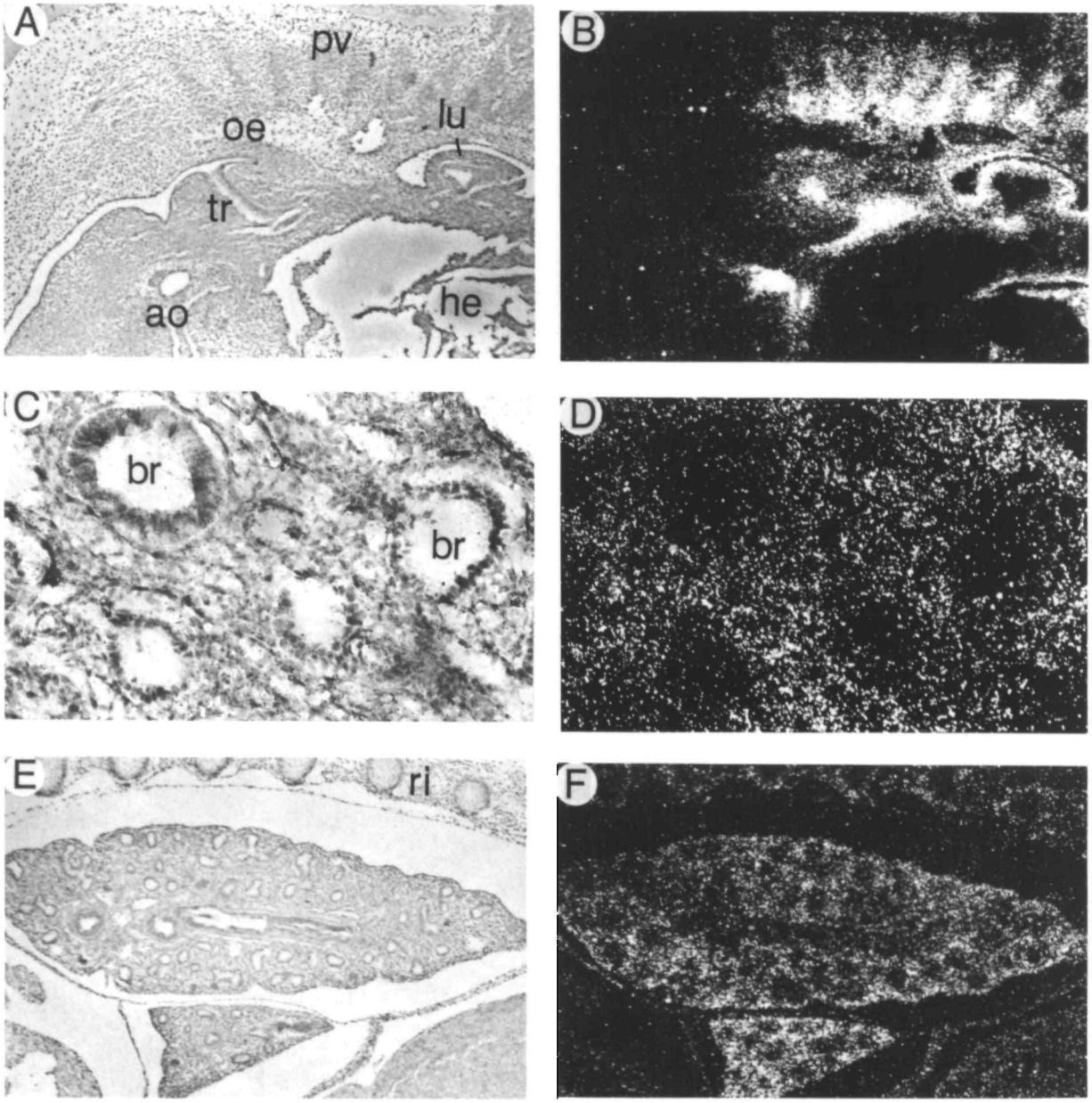


**Fig. 13.** Expression of Hox-1.4 in the cervical and the respiratory region. (A,B) Parasagittal section of a 12.5 days p.c. fetus: detail of the pharyngo-laryngeal region. (C,D) Cross section through a 12.5-day-old lung. (E,F) Sagittal section through the lung of a 15.5-day-old fetus. pv, prevertebrae; lu, lung; oe, oesophagus; tr, trachea; ao, aorta; he, heart; br, lobar bronchus; ri, ribs.

gut is also limited to mesenchymal cells. The epithelium is again negative (Fig. 14 B–F). Hox-1.4 transcripts show an AP restriction along the alimentary canal, since no hybridization is found in the oesophagus whereas a clear signal is present from the stomach wall to that of the hindgut, presumably until the anterior end of the gut. By day 15.5, the intestinal tissue shows a double layer of Hox-1.4 positive cells which likely correspond to the differentiation of the muscle layers of the digestive tract (Fig. 14 F). However, the resolution of this technique does not allow us to exclude an expression of Hox-1.4 in other cell types such as the myenteric plexus, which is part of the peripheral nervous system.

At 9.5 days p.c., Hox-1.4 transcripts are detected in the nephrotomes which are metamerized structures derived from the intermediate (para-axial) mesoderm. Three days later, the mesonephros is weakly and homogeneously labelled with no apparent restriction to the mesonephrotic tubules. Labelling in the mesonephros is only slightly above background, and is thus not clearly visible in the section shown in Fig. 14 B. Labelling of metanephric tissue, the precursor of the adult kidney, is again homogeneous, with no enhanced

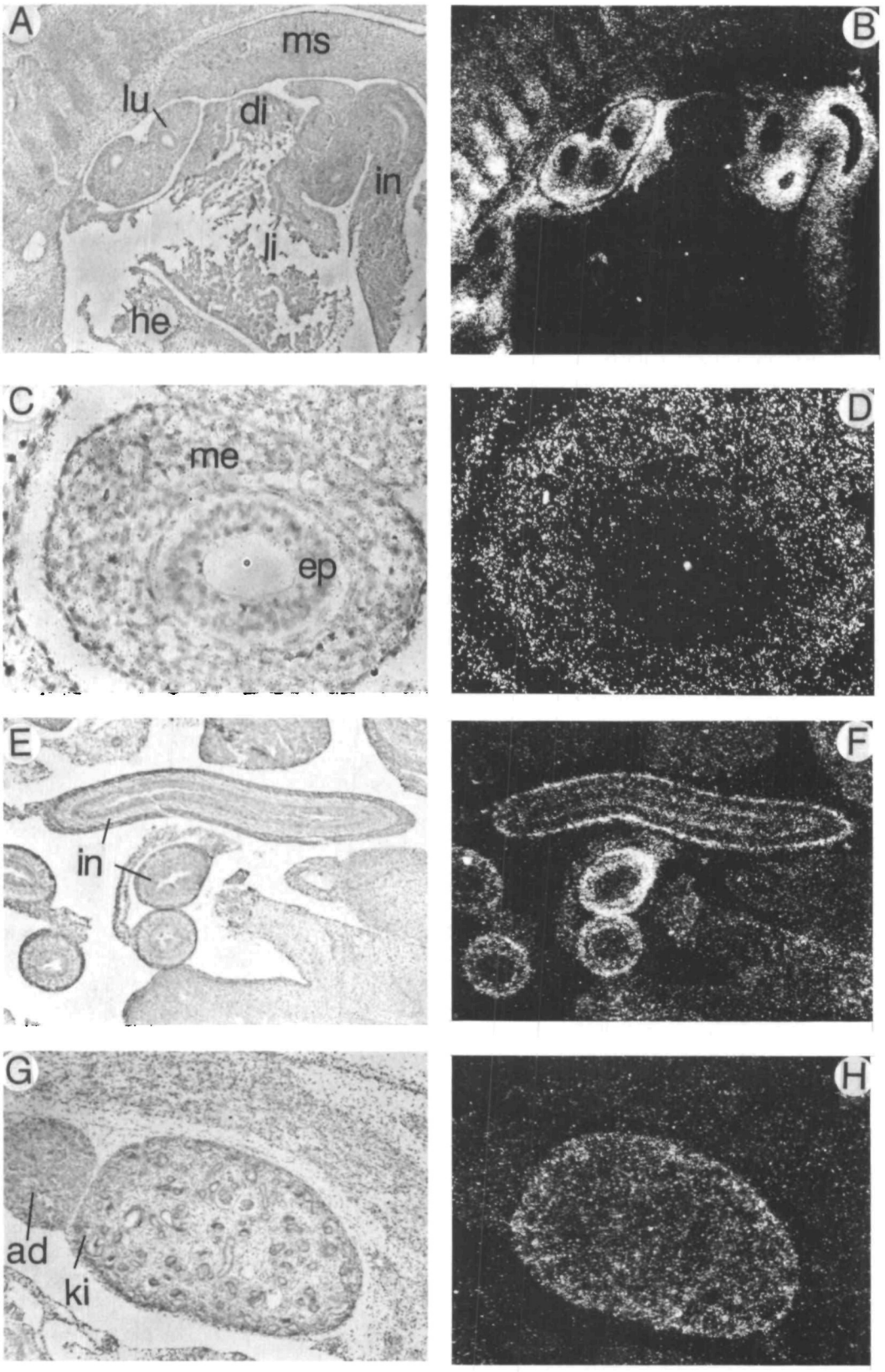


**Fig. 14.** Expression of Hox-1.4 in the alimentary tract and fetal kidney. (A–B) Parasagittal section of 12.5-day p.c. fetus through the fetal lung and intestine. (C–D) Transverse section of the primitive intestine of a 12.5-day p.c. fetus. (E–F) Enlargement of the intestinal region in a sagittal section of a 15.5-day fetus. (G–H) Sagittal section of the kidney and adrenal gland of a 15.5-day p.c. fetus. lu, lung; in, intestine; di, diaphragm; ms, mesonephros; ao, aorta; he, heart; li, liver; ep, intestinal epithelium; me, intestinal mesenchyme; ad, adrenal gland; ki, kidney.

hybridization to either the glomeruli or the tubular epithelium. The adrenal glands and fetal gonads do not hybridize to the Hox-1.4 probe (Fig. 14 H and data not shown).

The mesenchyme of the forelimb buds is homogeneously labelled by day 9.5 p.c. Later on, at day 12 and 13, a weak signal is detected in some regions of both limbs (Figs 8B and 12D). No precise histological identity could be assigned to the labelled structures of probable mesodermal origin (see Discussion).

## Discussion

*Structure of the Hox-1.4 gene*

In this paper, we report an extensive analysis of the structure and expression pattern of the Hox-1.4 gene. The sequence of this gene revealed a structure comparable to that of most described murine Hox genes. A major upstream exon encoding the hexapeptide is separated by a relatively short intron from a major downstream exon which encodes the homeodomain. Preliminary evidence suggests that Hox-1.4 RNAs could be further spliced so that differences in the *N*-terminal part of the Hox-1.4 protein might account for the different transcript sizes present in embryos *versus* adult testes (~1.5 and 1.3 kb, respectively; Rubin *et al.* 1986; Wolgemuth *et al.* 1986; Duboule *et al.* 1986). This additional splicing event would occur in a region extremely rich in GC (greater than 80 %), which explains the difficulty in isolating full-length cDNAs clones corresponding to the expected different transcripts. This very high GC content correlates with a high amount of proline and alanine as has already been observed for genes belonging to the same subfamily (i.e. Hox-5.1, Featherstone *et al.* 1988). The overall protein sequence shows blocks of similarities with the other members of this subfamily especially in the vicinity of the homeodomain, the hexapeptide and at the *N*-terminus. Some of these highly conserved regions are found in other species such as in the translation products of the *Xenopus* XHox1A gene and the *Drosophila* homeotic gene *deformed* (*Dfd*). This sequence conservation between species may reflect a conservation of function(s) at the molecular level. Nonetheless, at the cellular and organismal levels their function might be different. Because Hox-1.4 and the *Drosophila Dfd* gene are related by structure, position within their respective clusters, and by the relative position of their expression domains along the embryonic anteroposterior axis, it is tempting to speculate that aspects of the *Drosophila* homeotic gene network have been conserved in vertebrates (Duboule and Dollé, 1989; Graham *et al.* 1989).

The combination of both S1 nuclease and primer extension analysis using mRNAs extracted from embryos and untransfected or transfected cell lines did not reveal a clearcut transcription start site but rather suggested the presence of multiple starts within 10 to 20 bp (around position +1 in Fig. 2). These apparent discrepancies might be due to the presence of stretches of A and T residues immediately upstream of this putative start site. Alternatively, multiple functional start sites may really exist due to the structure of the Hox-1.4 promoter; in fact, though this putative site is preceded by a short AT-rich region, no genuine TATA or CAAT motifs are found at the expected positions, which may explain the lack of accurate transcription initiation and the 250 bp preceding this site are extremely GC rich. In order to see whether this putative promoter region had transcriptional activity, Hox-1.4/rabbit β-globin hybrid constructs were assayed in different cell lines. The results of these experiments showed that a DNA fragment extending 360 bp upstream from this putative start site was capable of promoting the rabbit β-globin transcription in all the cell lines that were used. In this case, both primer extension and S1 nuclease analysis specifically map a unique start site next to that mapped for embryonic mRNA. An increase in β-globin transcripts in response to RA was only obtained when a larger 2.75 kb Hox-1.4 fragment was placed in front of the rabbit β-globin gene. This suggests that upstream sequences that map outside of the 360 bp fragment could be responsible for the RA responsiveness of the Hox-1.4 gene. The 360 bp fragment could therefore represent a GC-rich minimal promoter fragment. GC-rich promoters have been described for viral and cellular housekeeping genes (Dynan *et al.* 1986), genes involved in growth control e.g. EGF and insulin receptors, c-*myc*, Ha-*ras* (Ishii *et al.* 1985, 1986; Araki *et al.* 1987; Lobanenkov *et al.* 1986) or specifically expressed during fetal development (theta-globin; Leung *et al.* 1987). The structures found upstream of the Hox-1.4 start codon suggested that this gene might be controlled by such a class of promoter elements. Such promoters often contain hexanucleotide 5′-GGGCGG-3′ binding sites for the transcription factor Sp1 (Dynan and Tjian, 1983; reviewed by Kadonaga *et al.* 1986). DNAse I footprint experiments were carried out using nuclear extracts from those cell lines where this fragment had been shown to function. In all cases, 6 strong protections were observed over GC-rich regions. Three of these protected sequences include genuine Sp1-binding sites and were shown to bind this transcription factor. Further studies using gel retardation and competition with different consensus oligonucleotides confirmed that Sp1 was involved in some of these protections. They also revealed that other factor(s) are present which, alone or in combination with Sp1, recognize sequences with high GC content. GC-rich promoters have been shown to exhibit bidirectional activity (Hoffman *et al.* 1987) and GC box sequences can also function in both orientations (Everett *et al.* 1983; Baty *et al.* 1984). This might explain the band we detect after reverse transcription of mRNA extracted from cells transfected with the 360 bp fragment in the reverse orientation (upper band in panel PMT5AS360b of Fig. 5).

*Expression of the Hox-1.4 gene during embryonic and fetal development*

Hox-1.4 transcripts are detected at the earliest by day

7.5 p.c. in posterior regions of the embryo, presumably in the embryonic mesoderm. Such a restriction of homeogene expression has been reported in late gastrula embryos, suggesting a function for these genes by the time the AP axis of embryo is established (Gaunt, 1987). Our detailed analysis of Hox-1.4 expression in the CNS during development is in agreement with studies previously reported (Toth *et al.* 1987; Gaunt *et al.* 1988). Briefly, Hox-1.4 is expressed in the fetal spinal cord up to an anterior boundary located within the hindbrain, corresponding, by day 12.5 p.c., to the middle portion of the floor of the myelencephalon. This anterior boundary of expression within the hindbrain (rhombencephalon) was established as early as day 9. During later stages, this anterior border lies in the very anterior portion of the medulla oblongata. Within the spinal cord, Hox-1.4 shows differential dorsoventral expression in the mantle layer at distinct metameric levels: transcripts are preferentially detected in the ventral horns and dorsal regions of the cervical spinal cord, whereas the thoracic spinal cord shows stronger hybridization only in its dorsal regions. This different distribution of Hox-1.4 transcripts concerns two zones of distinct developmental fate since ventral horns harbor the bodies of the future motor neurons whereas the dorsal region contains sensory nuclei. There is a striking similarity in the position of the anteroposterior boundary of expression of the three murine gene members of the Hox-1.4 (*Dfd*) related family; Hox-1.4 (Gaunt *et al.* 1988 and this work), Hox-5.1 (Featherstone *et al.* 1988) and Hox-2.6 (Graham *et al.* 1988). The similar relative positions of these three genes along their respective clusters is therefore accompanied by both structural and functional homologies. Comparative studies of the domains of expression of these genes on neighbour histological sections will clarify this point (S. Gaunt *et al.* submitted).

Hox-1.4 transcripts, like other Hox genes (Gaunt *et al.* 1988; Holland and Hogan, 1988), are detected in the scleretome-derived prevertebrae, with the exception of the first vertebral anlage, which is not labelled. Thus, the metameric anterior boundary of Hox-1.4 lies between metameres C1 and C2 (corresponding to the level of the 6th somite). A marked decrease in the intensity of labelling is found between the first and second thoracic prevertebrae, at the same level where the distribution of Hox-1.4 transcripts changes in the adjacent spinal cord. Between day 12 and 15, Hox-1.4 expression becomes progressively restricted to the mesenchymal cells that closely surround the cartilage primordia of the vertebral bodies. A similar restriction is found in the developing ribs, as previously described for the Hox-1.3 gene (Dony and Gruss, 1987), suggesting that Hox gene products are commonly required for the accurate development of the axial skeleton. A similar restriction to chondrocyte precursors is seen with the expression of two genes of the HOX-5 complex in developing limbs (Dollé and Duboule, 1989).

Hox-1.4 is also expressed in various nonsomitic mesodermal derivatives, including the whole mesonephric column, the metanephric parenchyma and mesenchymal components of the digestive and respiratory tracts. The anterior limit of Hox-1.4 expression in nonsegmented mesoderm is consistent with that seen between metameres C1 and C2 since no structures derived from the cephalic somitomeres are labelled. The most anteriorly labelled mesodermal structures are the trachea and the base of the larynx. The latter is derived from the mesoderm of the 5th and 6th branchial bar, located close to the 6th somite, the boundary of Hox-1.4 expression. Such a distribution of Hox-1.4 transcripts is consistent with a role in the specification of positional identity along the AP axis in an otherwise homogenous population of cells in both the segmented and nonsegmented mesoderm (Gaunt *et al.* 1986; 1988; Gaunt, 1988; Dony and Gruss, 1987; Holland and Hogan 1988).

Interestingly Hox-1.4 expression is detected in the digestive splanchnic mesoderm only from the distal part of the oesophagus and caudally. According to the proposed model (refs above), a large anterior part of the oesophagus mesoderm would thus be derived from cells originating anterior to the 6th somite and thus migrating to a more posterior position after the early elongation phenomenon typical of primitive intestine morphogenesis. Later on, the splanchnic mesenchyme that surrounds the growing tracheal gutter would derive from a more posterior cellular population (where Hox-1.4 is expressed), leading to the apparent paradox of an unlabelled oesophageal mesenchyme adjacent to a strongly labelled trachea. This illustrates the potential interest of Hox genes as markers during migratory processes.

The presence of Hox-1.4 transcripts in the gut is of particular interest since it was recently reported that a high level of expression of this gene in gut could lead to the formation of a megacolon in transgenic mice (Wolgemuth *et al.* 1989). However, it is not clear whether the expression of the transgene does exactly correspond to the endogenous 'double layer' pattern we report here. It is therefore possible that such a phenotype is obtained by ectopic expression of this transgene within a nonectopic structure. This illustrates the difficulty in using such an approach to analyse expression in single cells and the requirement for a more precise approach to answer these questions.

Hox-1.4 transcripts are clearly detected in the mesenchyme of both limb buds. However, if a clear signal is seen within the mesenchyme of the forelimb bud by day 9.5 p.c., the signal becomes very weak in the following days of development thus precluding a systematic study of Hox-1.4 expression in these structures. Expression of Hox genes in the limbs has been reported in several cases (Savard *et al.* 1988; Oliver *et al.* 1988; Dollé and Duboule, 1989; Robert *et al.* 1989) suggesting a possible role for these genes in limb morphogenesis (reviewed by Brockes, 1989).

Both the structure and expression data reported in this study support currently available models concerning the structural and functional organization of the murine HOX gene family (Gaunt, 1988; Gaunt *et al.* 1988; Holland and Hogan 1988; Duboule and Dollé,

1989; Graham *et al.* 1989). Indeed the Hox-1.4 gene belongs to a subfamily of homeogenes expressed from a rather anterior level in a subset of cellular types within different organs. The origin and nature of these different Hox-1.4-positive structures reinforces the proposal that murine homeogenes may serve as positional cues during vertebrate development.

We thank Professor Pierre Chambon for his constant interest and support, Dado Boncinelli for making data available prior to publication and helpful discussions, F. Dreyer and P. Sondermeyer (Transgene) for their advice, Bernard Boulay, Christiane Werlé and Petra Riedinger for artwork and Hilary Davies for preparing the manuscript. B. G. and M. S. F. are recipients of fellowships from the INSERM and the Medical Research Council of Canada, respectively. This work was supported by grants from the CNRS, INSERM and ARC and the Fondation pour la recherche médicale.

## References

ARAKI, E., SHIMADA, F., UZAWA, H., MORI, M. AND EBINA, Y. (1987). Characterization of the promoter region of the human insulin receptor gene. *J. biol. Chem.* **262**, 16186–16191.

AUFFRAY, C. AND ROUGEON, F. (1980). Purification of mouse immunoglobulin heavy-chain messenger RNAs from total myeloma tumor RNA. *Eur. J. Biochem.* **107**, 303–314.

AWGULEWITSCH, A., UTSET, M. F., HART, C. P., MCGINNIS, W. AND RUDDLE, F. H. (1986). Spatial restriction in expression of a mouse homeo box locus within the central nervous system. *Nature, Lond.* **320**, 328–335.

BARRERA-SALDANA, H., TAKAHASHI, K., VIGNERON, M., WILDEMAN, A., DAVIDSON, I. AND CHAMBON, P. (1985). All six GC-motifs of the SV40 early upstream element contribute to promoter activity *in vivo* and *in vitro*. *EMBO J.* **4**, 3839–3849.

BATY, D., BARRERA-SALDANA, H. A., EVERETT, R. D., VIGNERON, M. AND CHAMBON, P. (1984). Mutational dissection of the 21 bp repeat region of the SV40 early promoter reveals that it contains overlapping elements of the early-early and late-early promoter. *Nucl. Acid. Res.* **12**, 915–932.

BONCINELLI, E., SOMMA, R., ACAMPORA, D., PANNESE, M., D'ESPOSITO, M., FAIELLA, A. AND SIMEONE, A. (1988). Organization of human homeobox genes. *Human. Repr.* **3**, 880–886.

LE BOUC, Y., DREYER, F., JAEGER, M., BINOUX, M. AND SONDERMEYER, P. (1986). Complete characterization of the human IGF-1 nucleotide sequence isolated from a newly constructed adult liver cDNA library. *FEBS Lett.* **186**, 108–112.

BRIGGS, M. R., KADONAGA, J. T., BELL, S. P. AND TJIAN, R. (1986). Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. *Science* **234**, 47–52.

BROCKES, I. P. (1989). Retinoids, homeobox genes, and limb morphogenesis. *Neuron* **2**, 1285–1294.

BUCAN, M., YANG-FENG, T., COLBERG-POLEY, A. M., WOLGEMUTH, D. J., GUENET, J. L., FRANCKE, U. AND LEHRACH, H. (1986). Genetic and cytogenetic localization of the homeo box containing genes on mouse chromosome 6 and human chromosomes 7. *EMBO J.* **5**, 2899–2905.

BUCHER, P. AND TRIFONOV, E. N. (1986). Compilation and analysis of eukaryotic POL II promoter sequences. *Nucl. Acid Res.* **14**, 10009–10026.

CHEN, E. Y. AND SEEBURG, P. H. (1985). Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA* **4**, 165–170.

COLBERG-POLEY, A. M., VOSS, S. D., CHOWDHURY, K., STEWART, C. L., WAGNER, E. F. AND GRUSS, P. (1985). Clustered homeo boxes are differentially expressed during murine development. *Cell* **43**, 39–45.

DAVIDSON, I., FROMENTAL, C., AUGEREAU, P., WILDEMAN, A., ZENKE, M. AND CHAMBON, P. (1986). Cell-type specific protein binding to the SV40 enhancer in Hela and lymphoid cell nuclear extracts. *Nature, Lond.* **323**, 544–548.

DENTE, L., CESARENI, G. AND CORTESE, R. (1983). pEMBL: a new family of single stranded plasmids. *Nucl. Acid. Res.* **11**, 1645–1655.

DOLLÉ, P. AND DUBOULE, D. (1989). Two gene members of the murine Hox-5 complex show regional and cell-type specific expression in developing limb and gonads. *EMBO J.* **8**, 1507–1515.

DONY, C. AND GRUSS, P. (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.

DUBOULE, D., BARON, A., MÄHL, P. AND GALLIOT, B. (1986). A new homeo-box is present in overlapping cosmid clones which define the mouse Hox-1 locus. *EMBO J.* **5**, 1973–1980.

DUBOULE, D. AND DOLLÉ, P. (1989). The structural and functional organization of the murine Hox gene family resembles that of *Drosophila* Homeotic genes. *EMBO J.* **8**, 1497–1505.

DUBOULE, D., GALLIOT, B., BARON, A. AND FEATHERSTONE, M. S. (1989). Murine homeo-genes: some aspects of their organisation and structure. In *Cell to Cell Signals in Mammalian Development.* NATO ASI Series (ed. S. de Laat, J. G. Bluemink, and C. L. Mummery), 97–108, Springer Verlag, Heidelberg.

DYNAN, W. S., SAZERT, S., TJIAN, R. AND SCHIMKE, R. T. (1986). Transcription factor Sp1 recognizes a DNA sequence in the mouse dihydrofolate reductase promoter. *Nature, Lond.* **319**, 246–248.

DYNAN, W. S. AND TJIAN, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* **35**, 79–87.

EVERETT, R. D., BATY, D. AND CHAMBON, P. (1983). The repeated GC-rich motifs upstream from the TATA box are important elements of the SV40 early promoter. *Nucl. Acid. Res.* **11**, 2447–2464.

FEATHERSTONE, M. S., BARON, A., GAUNT, S. J., MATTEI, M. G. AND DUBOULE, D. (1988). Hox-5.1 defines a homeo-box containing gene locus on mouse chromosome 2. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4760–4764.

GAUNT, S. J. (1987). Homeobox gene expression in mouse embryos: Earliest detection by in situ hybridization is during gastrulation. *Development* **101**, 51–61.

GAUNT, S. J. (1988). Mouse homeo-box gene transcripts occupy different but overlapping domains in embryonic germ layers and organs: a comparison of Hox-3.1 and Hox-1.5. *Development* **103**, 135–144.

GAUNT, S. J., MILLER, J. R., POWELL, D. J. AND DUBOULE, D. (1986). Homeo-box gene expression in mouse embryos varies with position by the primitive streak stage. *Nature, Lond.* **324**, 662–664.

GAUNT, S. J., SHARPE, P. T. AND DUBOULE, D. (1988). Spatially restricted domains of homeo-gene transcripts in mouse embryos: relation to a segmented body plan. *Development* **104 Supplement** 169–181.

GEHRING, W. J. (1985). The homeo-box: a key to the understanding of development? *Cell* **40**, 3–5.

GEHRING, W. J. (1987). Homeo-boxes in the study of development. *Science* **236**, 1245–1252.

GRAHAM, A., PAPALOPULU, N. AND KRUMLAUF, R. (1989). The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. *Cell* **57**, 367–378.

GRAHAM, A., PAPALOPULU, N., LORIMER, J., MCVEY, J.-H., TUDDENHAM, E. G. J. AND KRUMLAUF, R. (1988). Caracterisation of a murine homeobox gene, Hox-2.6, related to the *Drosophila* Deformed gene. *Genes and Dev.* **2**, 1429–1438.

GROUDINE, M., PERETZ, M. AND WEINTRAUB, H. (1981). Transcriptional regulation of hemoglobin switching in chicken embryos. *Molec. cell. Biol.* **1**, 281–288.

HALL, C. V., JACOB, P. E., RINGOLD, G. M. AND LEE, F. (1983). Expression and regulation of E.Coli lac z gene fusions in mammalian cells. *J. molec. appl. Genet.* **2**, 101–109.

HARDING, K., WEDDEEN, C., MCGINNIS, W. AND LEVINE, M. (1985). Spatially regulated expression of homeotic genes in *Drosophila*. *Science* **229**, 1236–1242.

HART, C. P., AWGULEWITSCH, A., FAINSOD, A., MCGINNIS, W. AND RUDDLE, F. H. (1985). Homeo box gene complex on mouse chromosome 11: molecular clonig expression in embryogenesis and homology to a human homeobox locus. *Cell* **43**, 9–18.

HART, C. P., FAINSOD, A. AND RUDDLE, F. H. (1987). Sequence analysis of the murine Hox-2.2, -2.3, and -2.4 homeo-boxes: evolutionary and structural comparisons. *Genomics* **1**, 182–195.

HARVEY, R. P., TABIN, C. J. AND MELTON, D. A. (1986). Embryonic expression and nuclear localization of *Xenopus* homeo box gene products. *EMBO J.* **5**, 1237–1244.

HOFFMAN, E. K., TRUSKO, S. P., FREEMAN, N. AND GEORGE, D. L. (1987). Structural and functional characterization of the promoter region of the mouse c-Ki-ras gene. *Mol. and Cell. Biol.* **7**, 2592–2596.

HOLLAND, P. W. H. AND HOGAN, B. L. M. (1988). Expression of homeo-box genes during mouse development: a review. *Genes and Dev.* **2**, 773–782.

ISHII, S., KADONAGA, J. T., TJIAN, R., BRADY, J. N., MERLINO, G. T. AND PASTAN, I. (1986). Binding of the Sp1 transcription factor by the Human Harvey ras-1 proto-oncogene promoter. *Science* **232**, 1410–1413.

ISHII, S., XU, Y.-H., STRATTON, R. H., ROE, B. A., MERLINO, G. T. AND PASTAN, I. (1985). Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4920–4924.

KADONAGA, J. I., JONES, K. A. AND TJIAN, R. (1986). Promoter specific activation of polymerase II transcription by Sp1. *Trends Biochem.* **11**, 20–23.

LEUNG, S., PROUDFOOT, N. J. AND WHITELAW, E. (1987). The gene for theta-globin is transcribed in human fetal erythroid tissues. *Nature, Lond.* **329**, 551–554.

LEWIS, E. B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature, Lond.* **276**, 565–570.

LOBANENKOV, V. V., NICOLAS, R. H., PLUMB, M. A., WRIGHT, C. A. AND GOODWIN, G. H. (1986). Sequence-specific DNA-binding proteins which interact with (G+C)-rich sequences flanking the chicken c-myc genes. *Eur. J. Biochem.* **159**, 181–188.

MARILIO, F., SIMEONE, A., GIAMPAOLO, A., FAIELLA, A., ZAPPAVIGNA, V., ACAMPOVA, D., POIANA, G., RUSSO, G., PESCHLE, C. AND BONCINELLI, E. (1986). Differential and stage-specific expression in embryonic tissues of a new human homeobox gene. *Nature* **324**, 664–668.

MAXAM, A. M. AND GILBERT, W. (1977). A new method for sequencing DNA. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564.

MCGINNIS, W., HART, C. P., GEHRING, W. J. AND RUDDLE, F. H. (1984). Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of *Drosophila*. *Cell* **38**, 675–680.

NOMIYAMA, H., FROMENTAL, C., XIAO, J. H. AND CHAMBON, P. (1987). Cell specific activity of the constituent elements of the Simian virus 40 enhancer. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7881–7885.

OLIVER, G., WRIGHT, C. E. V., HANDWICKE, I. AND DE ROBERTIS, E. M. (1988). A gradient of homeodomain protein on developing forelimbs of *xenopus* and mouse embryos. *Cell* **55**, 1017–1024.

REGULSKI, M., MCGINNIS, N., CHADWICK, R. AND MCGINNIS, W. (1987). Developmental and molecular analysis of Deformed: a homeotic gene controlling *Drosophila* head development. *EMBO J.* **6**, 767–777.

ROBERT, B., SASSOON, D., JACQ, B., GEHRING, W. AND BUCKINGHAM, M. (1989). Hox-7, a mouse homeobox gene with a novel pattern of expression during embryogenesis. *EMBO J.* **8**, 91–100.

RUBIN, M. R., TOTH, L. E., PATEL, M. D., D'EUSTACHIO, P. AND NGUYEN-HUU, M. C. (1986). A mouse homeo box gene is expressed in spermatocytes and embryos. *Science* **233**, 663–667.

SANGER, F. S. AND COULSON, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467.

SAVARD, P., GATES, P. B. AND BROCKES, J. P. (1988). Positon dependent expression of a homeobox gene transcript in relation to amphibian limb regeneration. *EMBO J.* **7**, 4275–4282.

SIMEONE, A., PANNESE, M., ACAMPORA, D., D'ESPOSITO, M. AND BONCINELLI, E. (1988). At least three human homeoboxes on chromosome 12 belong to the same transcription unit. *Nucl. Acid Res.* **16**, 6379.

THEILER, K. (1972). *The House Mouse*. Springer-Verlag, Berlin, Heidelberg, New York.

TOTH, L. E., SLAWIN, K. L., PINTAR, J. E. AND NGUYEN-HUU, C. M. (1987). Region-specific expression of mouse homeo-box genes in the embryonic mesoderm and central nervous system. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6790–6794.

VERBOUT, A. J. (1985). The development of the vertebral column. *Adv. Anat. Embryol. Cell. Biol.* **90**.

WOLGEMUTH, D. J., VIVIANO, C. M., GIZANG-GINSBERG, E., FROHMAN, M. A., JOYNER, A. AND MARTIN, G. R. (1987). Differential expression of the mouse homeobox-containing gene Hox-1.4 during male germ cell differentiation and embryonic development. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5813–5817.

WOLGEMUTH, D. J., ENGELMYER, E., DUGGEL, R. N., GIZANG-GINSBERG, E., MUTTER, G. L., PONZETTO, C., VIVIANO, C. AND ZAKERI, Z. F. (1986). Isolation of a mouse cDNA coding for a developmentally regulated, testis-specific transcript containing homeo box homology. *EMBO J.* **5**, 1229–1235.

WOLGEMUTH, D. J., BEHRINGER, R. R., MOSTOLLER, M. P., BRINSTER, R. L. AND PALMITER, R. D. (1989). Transgenic mice overexpressing the mouse homeobox-containing gene Hox-1.4 exhibit abnormal gut development. *Nature, Lond.* **337**, 414–167.

XIAO, J. H., DAVIDSON, I., FERRANDON, D., ROSALES, R., VIGNERON, M., MACCHI, M., RUFFENACH, F. AND CHAMBON, P. (1987). One cell-specific and three ubiquitous nuclear proteins bind *in vitro* to overlapping motifs in the domain B1 of the SV40 enhancer. *EMBO J.* **6**, 3005–3013.

ZENKE, M., GRUNDSTRÖM, T., MATTHES, H., WINTZERITH, M., SCHATZ, C., WILDEMAN, A. AND CHAMBON, P. (1986). Multiple sequence motifs are involved in SV40 enhancer function. *EMBO J.* **5**, 387–397.

(*Accepted 19 June 1989*)