The EMBO Journal vol.6 no.10 pp.2977–2986, 1987

Hox-1.6: a mouse homeo-box-containing gene member of the *Hox-1* complex

Agnès Baron, Mark S.Featherstone, Robert E.Hill[1], Alison Hall[1], Brigitte Galliot and Denis Duboule

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Faculté de Médecine, 11, rue Humann, 67085 Strasbourg Cédex, France, and [1]Clinical and Population Cytogenetic Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

Communicated by P.Chambon

***Hox-1.6*, a mouse homeo-box-containing gene member of the *Hox-1* complex, is described. The *Hox-1.6* homeo-box shows more divergence than the other members of the complex with the *Drosophila* Antennapedia-like homeo-box class. This previously undescribed gene was studied with respect to its transcription pattern and was found to be expressed during mouse fetal development in an intestine-specific manner in adults, and in tumours or cell types exhibiting early endodermal-like differentiation. The study of embryonic partial *Hox-1.6* cDNA clones revealed structural features common to other *Drosophila* and vertebrate homeo-box-containing genes, but also indicated that *Hox-1.6* transcripts might display splicing patterns more complex than those known for other vertebrate homeo-genes. One of these cDNA clones contains a rather short open reading frame which would encode a protein of ~14.5 kd. The use of this clone as a probe for S1 nuclease mapping confirmed that different *Hox-1.6* transcripts were present both in embryonic total RNA and in embryonal carcinoma cell cytoplasmic RNA. These various transcripts are probably generated by an alternative splicing mechanism and may thus encode a set of related proteins.**

Key words: Mus musculus/homeo-box/*Hox-1* complex/development

Introduction

Following recent reports that many genes implicated in the early development of *Drosophila* share a 180-bp DNA sequence, the homeo-box (McGinnis *et al.*, 1984a,b; Scott and Weiner, 1984), several similar sequences have been isolated from different vertebrates such as amphibians (Carrasco *et al.*, 1984; Muller *et al.*, 1984; Harvey *et al.*, 1986), rodents (McGinnis *et al.*, 1984b; Colberg-Poley *et al.*, 1985a,b; Jackson *et al.*, 1985, Hart *et al.*, 1985; Hauser *et al.*, 1985; Joyner *et al.*, 1985; Awgulewitch *et al.*, 1986; Wolgemuth *et al.*, 1986; Duboule *et al.*, 1986) and humans (Levine *et al.*, 1984; Boncinelli *et al.*, 1985; Simeone *et al.*, 1986; Mavilio *et al.*, 1986). In the house mouse, the number of sequences showing high homology with the *Drosophila* Antennapedia-like (*Antp*-like) homeo-box is estimated at 25–30 according to Southern blot experiments on genomic DNA or by the frequency of positive clones in library screenings using various *Drosophila* homeo-boxes as probes (see references above and our unpublished results). The homology observed between *Drosophila* and *Mus musculus* homeo-box sequences is often >80% at the amino acid level. Moreover, as in *Drosophila* which shows two major clusters of homeotic genes (reviewed in Lewis, 1978), two main homeo-box-containing clusters have been described in the mouse. These are the *Hox-1* and *Hox-2* complexes (Colberg-Poley *et al.*, 1985b; Duboule *et al.*, 1986; Hart *et al.*, 1985) located on chromosomes 6 and 11, and containing six and at least four homeo-boxes, respectively. [A new nomenclature for murine homeo-box-containing genes was recently accepted by the international Committee for standardized Genetic Nomenclature (Martin *et al.*, 1987; Lyon, 1987). This new nomenclature is used in this paper regardless of all previously reported appellations (see also Figure 2A).] In all cases so far reported, these murine sequences were shown to be parts of genes whose expression patterns, while exhibiting individual variations, are specific for given stages of embryogenesis or for restricted numbers of adult tissues or cell lines (see references cited above concerning the murine homeo-boxes). It thus seems likely that this set of murine genes (homeo-genes) has an important role during the course of vertebrate development, although their functions are not yet established. Recent results suggest that some of these genes might be coordinately regulated, since they appear to be expressed in the same embryonic structures but at slightly different times or positions (Gaunt *et al.*, 1986; K.Mahon, C.Dony, P.Gruss and H.Westphal, personal communication). Accumulating evidence (White and Wilcox, 1984; Beachy *et al.*, 1985; DiNardo *et al.*, 1985; Desplan *et al.*, 1985; Harvey *et al.*, 1986) supports the suggestion that the homeo-box might correspond to a DNA-binding domain (Shepherd *et al.*, 1984; Laughon and Scott, 1984) and that such genes might therefore play regulatory roles, as proposed for *Drosophila* (Lewis 1978; Garcia-Bellido *et al.*, 1977).

The molecular cloning of the *Hox-1* complex has been previously reported (Colberg-Poley *et al.*, 1985b; Duboule *et al.*, 1986) and shown to be located on mouse chromosome 6 (Duboule *et al.*, 1986), at bands B3 → C (Bucan *et al.*, 1986). This complex was shown to contain four characterized homeo-genes plus two sequences (*Hox1-x* and *Hox1-y*) hybridizing to *Drosophila Antp*-like homeo-box probes. *Hox1-x* belongs to an independent homeo-gene; *Hox-1.3* (B. Zinc, personal communication). We report in this study that *Hox1-y*, now designated *Hox-1.6*, is also a homeo-gene member of the *Hox-1* complex. We show that this gene, like *Hox-1.1*, *Hox-1.2* (Colberg-Poley *et al.*, 1985a,b), *Hox-1.4* (Duboule *et al.*, 1986; Rubin *et al.*, 1986) and *Hox-1.5* (McGinnis *et al.*, 1984a; Gaunt *et al.*, 1986), is expressed during mouse embryogenesis and exhibits tissue and cell type specificity in adult organs and in teratocarcinomas propagated *in vivo*. However, the specificity of expression of *Hox-1.6* proved to be rather different from those reported for other murine homeo-genes. We also demonstrate that different transcripts are generated from the *Hox-1.6* gene by differential splicing mechanisms. We suggest that these transcripts might encode proteins of variable lengths. These data further support the assertion that vertebrate homeo-genes have important functions during development.

Results

Molecular cloning and sequence analysis of Hox-1.6

An *Eco*RI restriction fragment of ~ 8 kb, located at the 3′ end of a set of overlapping cosmid clones defining the *Hox-1* complex, weakly hybridized to various *Drosophila* homeo-box-containing probes (Figures 1 and 2). A 1.6-kb *Bam*HI−*Eco*RI subfragment was subcloned in pBR322 and its restriction map is shown in Figure 2B. Sequence analysis revealed the presence of a previously undescribed homeo-box lying within the *Hox-1* complex in the same orientation as the five other members. This box shows ~ 65% homology at the nucleotide level with the other boxes of the *Hox-1* complex, although the first 40 nucleotides show more divergence than usually observed (<30% homology; Figure 3). Immediately outside the homeo-box no significant homology was found with any other reported sequences except for the presence of polypyrimidine stretches upstream from the homeo-box (see Figure 6C) and a conserved pentapeptide (Mavilio *et al.*, 1986; see below). The only open reading frame present throughout this sequence generates the putative partial protein sequence shown in Figure 3. Overall, there is less homology between this protein sequence and the corresponding murine sequences previously reported (Figure 3). However, when the comparison is made starting at codon 14, the homology with other homeo-domains increases to >70% due to the high degree of conservation of the subregion thought to form a helix−turn−helix DNA-binding structure (Pabo and Sauer, 1984; Shepherd *et al.*, 1984; Laughon and Scott, 1984). The highly conserved Arg codon at position 43 (lying within the region that presumably interacts with the major groove of the DNA) is replaced by a Thr residue (Figure 3).

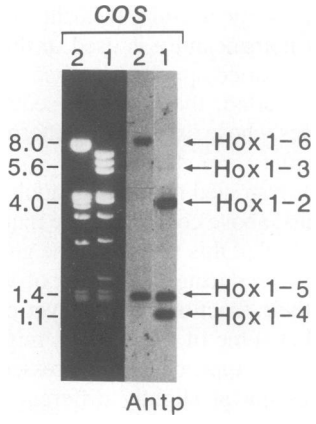

Fig. 1. The detection of the *Hox-1.6* homeo-box within the *Hox-1* complex. The two left panels show an *Eco*RI restriction digest of cos2 and cos1 DNAs, respectively (see Figure 2A). The two right panels show the same digests after transfer to a nitrocellulose membrane and hybridization to a *Drosophila Antp*-containing homeo-box probe. Five different restriction fragments contain cross-hybridizing homeo-box sequences (*Hox-1.2* to *Hox-1.6*). *Hox-1.6* is located within the upper cos2 8-kb *Eco*RI fragment, outside the region of overlap between cos1 and cos2 and is therefore lying in the 3′ part of the *Hox-1* complex. The various homeo-boxes are designated using the recently accepted terminology (Martin *et al.*, 1987).

Transcription of Hox-1.6

A *Bgl*II−*Eco*RI restriction fragment containing essentially the immediate 3′ sequence flanking the homeo-box (Figure 2, solid bar) was hybridized under high stringency conditions to a Southern blot containing mouse genomic DNA digested with both *Eco*RI and *Bam*HI. As expected, a 1.6-kb fragment (see Figure 2) containing the *Hox-1.6* homeo-box was detected. Other bands of higher mol. wts hybridized only weakly (Figure 4A). This specific *Hox-1.6* 3′ probe was therefore used in combination with different end-labelled probes (see later) to study the transcription of *Hox-1.6*.

Northern blots of total RNA extracted from day 9−day 13 mouse fetuses, separated on methyl mercury gels and transferred onto DBM paper were examined using either the above probe labelled by nick translation or anti-sense RNA synthesized in pGEM1. Both probes detected the same two bands which gave their strongest signals in day 9 fetal RNA and which were virtually absent by day 13 (Figure 4B). The smaller of these bands (4.5 kb) disappeared under very stringent hybridization and washing conditions (last washes in 0.1% SSC, 80°C; data not shown) and may not correspond to a genuine *Hox-1.6* transcript. The larger band, however, was still detected at high stringency. Thus, the *Hox-1.6* homeo-box-containing gene is transcribed in a stage-specific manner during mouse embryogenesis (see also S1 nuclease mapping results below). 28S and 18S rRNAs were shown to cross-hybridize weakly but repeatedly with these probes as noted by others (Hart *et al.*, 1985; Rubin *et al.*, 1986),

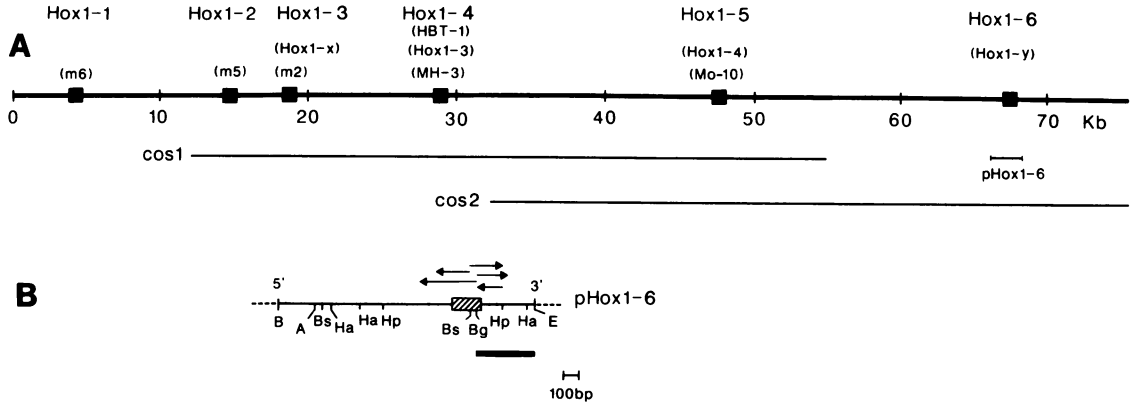

Fig. 2. Position of *Hox-1.6* within the *Hox-1* complex and restriction map of the pHox-1.6 sub-clone. (**A**) The exact position of *Hox-1.6* along the *Hox-1* complex was established using single and multiple restriction digests of the cosmid clone cos2 with the enzymes *Eco*RI,. *Bam*HI and *Xho*I. *Hox-1.6* is located ~20 kb downstream from *Hox-1.5*. All six genes have the same direction of transcription, and are presented 5′ to 3′ from left to right. The names under which these genes previously appeared in the literature are within brackets (m6, m5 and m2; Colberg-Poley *et al.*, 1985; Mo-10, McGinnis *et al.*, 1984; *Hox1-1* to *Hox1-y*, Duboule *et al.*, 1986; HBT-1, Wolgemuth *et al.*, 1986; MH-3, Rubin *et al.*, 1986). (**B**) A fine restriction mapping was carried out on a 1.6-kb *Eco*RI/*Bam*HI subclone containing the Hox-1.6 homeo-box (hatched box) using the following enzymes: *Hae*III (Ha), *Hpa*II (Hp), *Bgl*II (Bg), *Bss*HII (Bs) and *Ava*I (A). The sequencing strategy is outlined on the top by arrows. The solid bar on the bottom represents the *Bgl*II/*Eco*RI restriction fragment which was used for parts of the transcription studies (see Figure 4).

regardless of the labelling protocol or hybridization conditions.

Various adult tissues were examined for *Hox-1.6* expression using the same approach. Total RNA from eight adult organs was screened with the probes described above. The u.v. pattern after electrophoresis confirmed that the RNA was intact (Figure 5). Two prominent bands of 1.4 and 1.0 kb were detected in intestine-derived RNA (Figure 5). Both bands remained after extensive high-stringency washes (as above) and could not be detected in any other tissues analyzed. These two intestine-specific transcripts are much smaller than those observed in day 9 mouse fetuses (Figure 4B). The possibility that these two bands are degradation products was excluded by an additional control involving hybridization of the same DBM-paper with a probe from the triose phosphate isomerase (TPI) gene. This probe detected a unique transcript of the expected size in intestine and some other adult tissues (Figure 5). Thus, there is specific expression of *Hox-1.6* in mouse adult intestinal tissue. This hybridization was confirmed by S1 nuclease mapping analysis (not shown). Weaker hybridizing bands (~2 and 4.5 kb) were also detected in testis, brain, kidney and liver. They might represent transcripts hybridizing to either *Hox-1.6* or *Hox-1.6*-related sequences. These bands might also be artefactually produced by cross-hybridization with 28S and 18S rRNAs as noted before (Figure 4B).

In the previously reported characterization of another member of the *Hox-1* locus, *Hox-1.4* (Duboule *et al.*, 1986), we showed that a RNA splice acceptor site was located immediately upstream from the homeo-box. In order to determine whether *Hox-1.6* is similarly arranged and to verify that the homeo-box is part of the RNA sequence, a series of S1 nuclease protection analyses were performed. The two probes used in these experiments are diagrammed in Figure 6C. Probe A is a 630-bp fragment end-labelled at the *Bgl*II site within the homeo-box. Probe B is 965 bp long and was end-labelled at the *Eco*RI site located 3′ to the homeo-box. These probes were hybridized with total RNA derived from day 12 mouse embryos, digested with S1 nuclease and the products run on 8% polyacrylamide−urea sequencing gels. Using day 12 embryonic RNA, 177-bp and 512-bp S1 nuclease-resistant fragments were obtained from probes A and B, respectively (Figure 6A). Both of these results map the 5′ end of a S1 nuclease-resistant region to a point 34 nucleotides upstream of the *Hox-1.6* homeo-box (−34, Figure 6C). This site falls within a consensus sequence PyAG located downstream of a polypyrimidine stretch, both of which are characteristic of splice sites (Breathnach *et al.*, 1978). This is comparable to the splice acceptor region in *Hox-1.4*. (Duboule *et al.*, 1986). Since Northern blot analysis suggested that *Hox-1.6* is expressed as early as day 9 of gestation but that transcripts disappear by days 13−15, we used this highly sensitive nuclease protection assay to confirm these results. Day 10 and 15 total embryonic RNAs were hybridized with probe A (Figure 6C) and digested as before. The expected 177-bp S1 nuclease-resistant product was obtained with day 10 RNA, whereas no strong protection was observed with RNA from day 15 fetuses (Figure 6B), confirming the results obtained by Northern blot analysis.

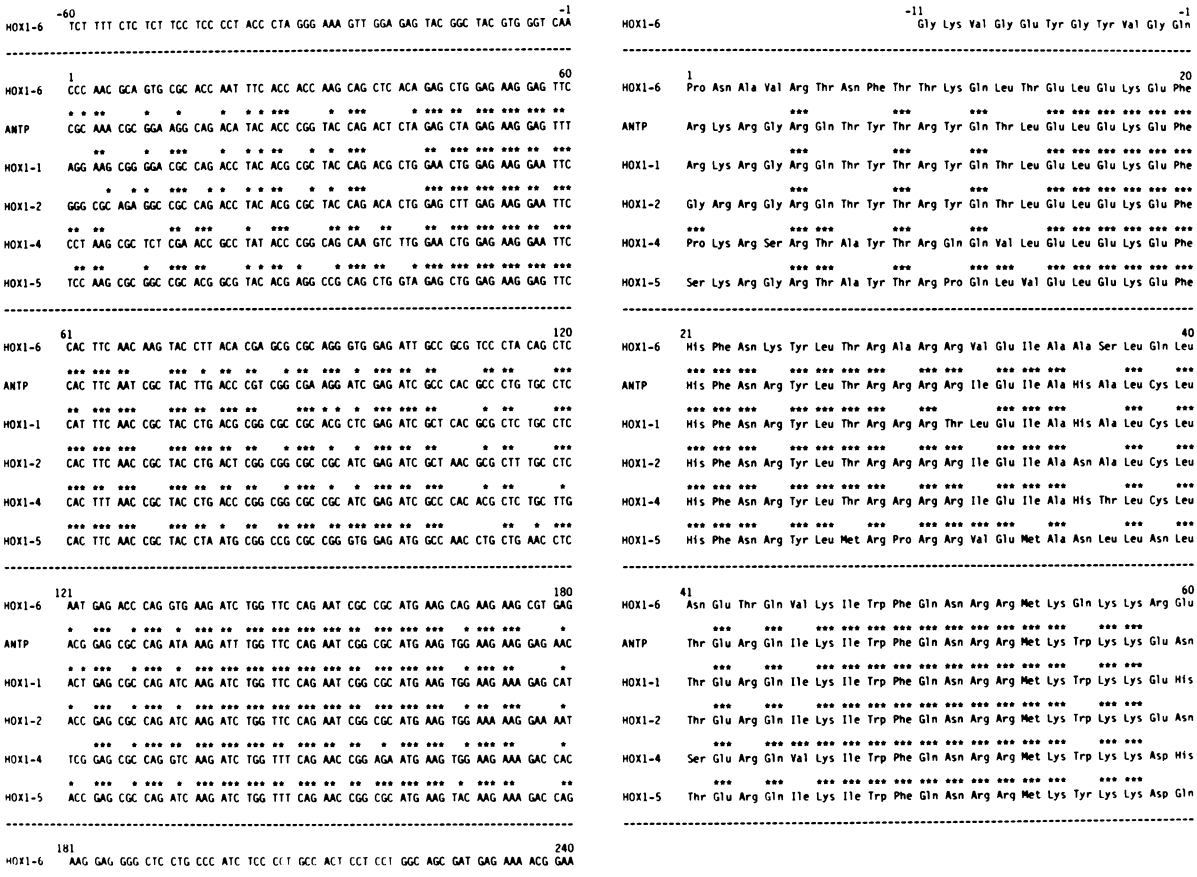

Fig. 3. The nucleotide and predicted amino acid sequences of the *Hox-1.6* homeo-box. **Left panel**: DNA sequence of *Hox-1.6* aligned 5′ to 3′ with homologous homeo-box regions from either *Drosophila* (*Antp*) or the mouse *Hox-1* complex (*Hox-1.1, Hox-1.2, Hox-1.4, Hox-1.5*). Conserved nucleotides are indicated by a star. Nucleotide 1 corresponds to the start of the homeo-box. **Right panel**: putative amino acid sequence of the only open reading frame extending over the *Hox-1.6* homeo-box region aligned with other homeo-domains deduced from sequences shown in the left panel. Homologous amino acids are indicated by three stars. Amino acid 1 corresponds to the start of the homeo-box.

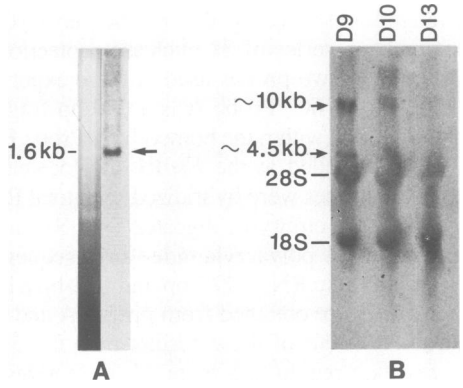

Fig. 4. The transcription of *Hox-1.6* in mouse fetuses. (**A**) Southern blot of mouse genomic DNA. Mouse genomic DNA (10 μg) was restricted with both *Eco*RI and *Bam*HI, gel-separated and transferred onto a Hybond N membrane. When hybridized with the nick-translated *Bgl*II/*Eco*RI fragment (Figure 2B, solid bar), a single strong band of the expected size (1.6-kb) was detected. (**B**) Northern blot analysis of fetal mouse RNAs. 30 μg of total RNA extracted from mouse fetuses at days 9 (D9), 10 (D10) and 13 (D13) were separated on a methyl mercury gel, transferred to DBM-paper and hybridized to the nick-translated *Bgl*II/*Eco*RI fragment. Two major bands (~4.5 and ~10 kb) were detected in RNA from day 9 which were still present at day 10, but undetectable by day 13. The same result was obtained when similar Northern blots were hybridized with *Bgl*II/*Eco*RI-labelled anti-sense RNA prepared from subclones in the Gemini vector (not shown). In both cases, cross-hybridization was observed with 18S and 28S rRNA. The quality of the various RNAs was controlled by the u.v. pattern observed before gel blotting (not shown), and proved to be intact.

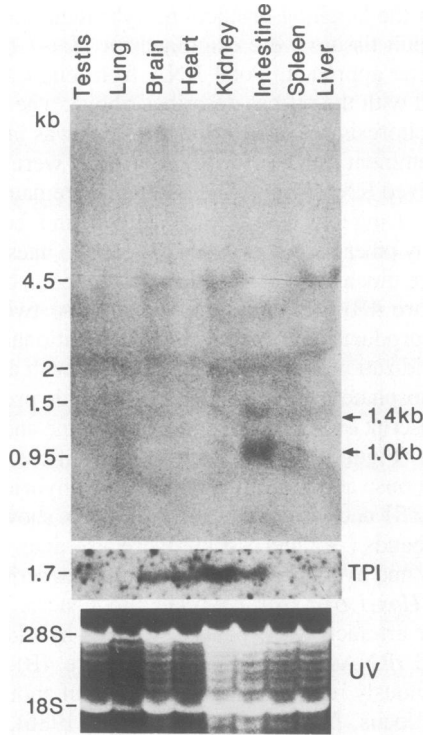

Fig. 5. Expression of the *Hox-1.6* gene in adult tissues. Various adult tissues were analysed for the presence of *Hox-1.6* transcripts using both the same method and probes as for Figure 4. Size markers are given on the left. Among the tissues analysed (top panel), the intestine showed two transcripts (1 and 1.4 kb, arrows), both much shorter than those observed using fetal RNAs (Figure 4). These two transcripts were not detected in RNA extracted from other adult tissues which, in some cases (testis, brain, kidney, liver), exhibited two larger hybridizing bands (~2 and ~4.5 kb). The middle panel shows a control hybridization of the same DBM-paper with a TPI probe. The bottom panel shows the u.v. pattern of the various RNA samples before their transfer onto DBM-paper. The middle and bottom panels clearly show that the RNA extracted from the intestine is not degraded.

We also examined *Hox-1.6* expression in teratocarcinoma tumours and in embryonal carcinoma (EC) cells. S1 nuclease analysis was performed with RNA extracted from either solid or ascitic teratocarcinoma tumours showing various developmental capacities (see Blühtmann *et al.*, 1983; Sassone-Corsi *et al.*, 1985 for details). The strongest signals were obtained with two ascitic tumours, LT-113 (Figure 6A) and OTT2158 (data not shown). The former is composed exclusively of endodermal vesicles, whereas the latter contains embryoid bodies (Blühtmann *et al.*, 1983) which are structures composed of pluripotent embryonic cells surrounded by a layer of primitive endodermal cells. In both cases, the protection observed was identical to those obtained when fetal RNA was used. Signals were much weaker or not detected with RNA extracted from tumours exhibiting mesodermal or ectodermal-like differentiation (e.g. TDR602, TDN2283. described in Blühtmann *et al.*, 1983).

Similar experiments were carried out using the established nullipotent EC cell line F9 (Bernstine *et al.*, 1973). Cytoplasmic RNA was extracted from F9 cells before and after exposure to retinoic acid (RA) which induces differentiation of these cells into parietal endoderm (Strickland and Madhavi, 1978). Whereas a weak protection of probe A was obtained with RNA extracted from untreated F9 cells, a strong signal (177 bp) was obtained with RNA from cells treated for 24 h with RA. In the latter case, the level of *Hox-1.6* transcripts increased by at least 20-fold. A weak protection was also observed (C.Stewart and D.Duboule, unpublished results) when the same probe was hybridized with RNA extracted from the pluripotent EC cells PC13 (Bernstine *et al.*, 1973) and P19 (McBurney and Rogers, 1982).

The Hox-1.6 gene contains multiple exons

Two partial cDNA clones were independently isolated from lambda libraries made from poly(A)$^+$ RNA obtained from mouse embryos at either day 8.5 of gestation (a gift of B.Hogan and K.Fahrner, cDNA 1; Figure 7Ab) or day 9.5−10 of gestation (Duboule *et al.*, 1987, cDNA 2; Figure 7Ac). The cDNA 1 clone appeared as one in ~ 40 000 plaques after screening with a *Drosophila* fushi-tarazu (*ftz*) homeo-box-containing probe (gift of W.Gehring) whereas cDNA 2 appeared as one in ~4 500 000 plaques, using the whole genomic pHox-1.6 insert as a probe (see Figure 2), several screenings being required to isolate these rare clones. The sequence of cDNA 1 (677 bp) revealed the presence of the *Hox-1.6* homeo-box plus 5′- and 3′- flanking sequences (Figure 7Ab; Figure 7B underlined with thick line). Sequence comparison with the genomic clone and additional restriction enzyme mapping showed that the *Hox-1.6* transcription unit contains at least three exonic sequences (Figure 7A and B). The homeo-box is lying in the last included exon which starts at the exact position of the acceptor splice site previously proposed from the results of S1 nuclease digestion (−34 in Figure 6C, 512 in Figure 7B) and extends until the end of this clone (*Eco*RI site in Figure 7A and B). Therefore, cDNA 1 represents only a restricted portion of a *Hoc-1.6* transcript and does not contain the 3′ part of the *Hox-1.6* gene. In addition, cDNA 1 revealed the presence of at least two other exonic sequences located further upstream. A 95-bp mini-exon (from position 417 to 511 in Figure 7B) is located ~450 bp upstream from the homeo-box exon and 182 bp upstream of a more 5′ located 40-bp mini-exon (from position 377 to 416 in Figure 7B). The remaining 5′-located 18 bp of cDNA 1 were not localized on the cosmid

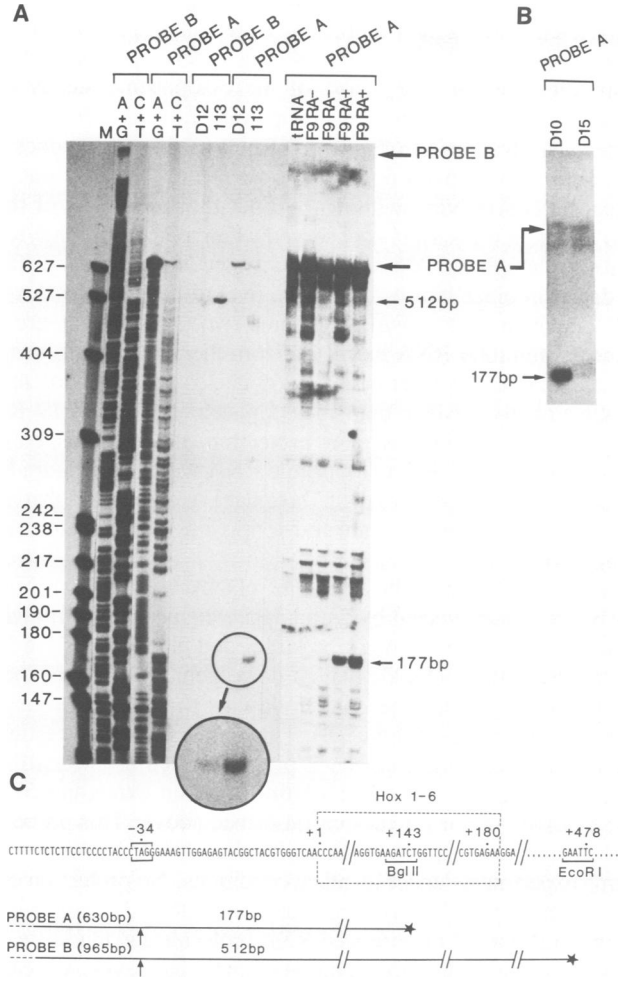

Fig. 6. S1 nuclease protection analysis. (**A**) 50−100 μg of total fetal RNA at day (D12), total RNA extracted from the endodermal tumour LT-113 (113), and 100 μg or 200 μg of cytoplasmic RNA extracted from F9 teratocarcinoma cells before (F9RA$^-$) and after (F9RA$^+$) treatment of the cells with RA were hybridized to either the end-labelled probes A (630 bp, see under **C**) or B (965 bp, see under **C**) and digested with S1 nuclease. The resistant hybrids were loaded on a sequencing gel and analyzed in comparison with the sequences of the two probes (A+G and C+T reactions). The positions of the two undigested probes (probe A; probe B) as well as the sizes of the protected fragments are given on the right hand side. Day 12 fetal RNA (D12) and 113 RNA (113) protect 177 bp of probe A from S1 nuclease digestion. When the larger probe B is used under the same conditions, a protection of 512 bp is observed (in both cases much stronger with the 113 RNA than with the fetal RNA). The difference in size between the two protected fragments is equal to the difference in the sizes of the probes. Both of these protections start 34 nucleotides upstream from the *Hox-1.6* homeo-box (see under **C**). Whereas a weak protection of the expected size (177 bp) is detected when 100 or 200 μg of RNA from untreated F9 cells (F9RA$^-$) are hybridized with probe A and S1 digested, strong signals appear when the same cells are treated with RA for 24 h (F9RA$^+$) before RNA extraction (100 and 200 μg of RNA). Note the essentially identical intensities of the artefactual bands in the F9 lanes as well as the absence of a specific signal (177 bp) when tRNA was treated in the same manner. (**B**) Identical amounts of fetal RNA extracted from fetuses at day 10 (D10) and 15 (D15), hybridized with probe A and treated as in **A**. The position of the probe and the size of the protection (177 bp) are shown on the left-hand side. The strong protection detected when day 10 RNA is used is barely detected with RNA extracted from fetuses 5 days older (D15) (see also Figure 4B). (**C**) Sequence of the *Hox-1.6* region 5′ to the homeo-box (dashed lines starting at position +1). Polypyrimidine stretches end at position −40 (also visible in **A**. probe A, A+G and C+T above the 177-bp protection). The boxed TAG triplet represents a consensus acceptor splice site located by the start of the different protections observed using either probe A (177 bp) or probe B (512 bp). See figure 7 for more details.

DNA and are therefore indicated as coming from an upstream region, possibly from an additional 5′-located exon (Figure 7A; see also the 18 nucleotides in the top left of the sequence in Figure 7B). This short sequence might also have been produced by a cDNA cloning artefact (see below).

cDNA 2 (Figure 7Ac; Figure 7B, underlined with interrupted line) overlaps in its 3′ part with cDNA 1 since it contains 39 out of the 40 bp of the cDNA 1 mini-exon 2, ending one nucleotide before the occurrence of the donor splice site preceding the 182-bp intron (position 416 on Figure 7B). Surprisingly, this cDNA clone diverges with cDNA clone 1 at the apparent 5′ boundary of the cDNA 1 mini-exon 2 (position 377 in Figure 7B, see also Figure 7A) since it makes an unbroken match with the genomic sequence and extends 38 nucleotides upstream of the *Bam*HI site in the pHox-1.6 genomic clone (Figure 7A and B).

The putative amino acid sequence of cDNA 1 revealed that, like most of the homeo-box-containing genes so far reported, *Hox-1.6* contains a conserved pentapeptide (Mavilio *et al.*, 1986; from position 469 to 483 in Figure 7B; dashed box). However, this sequence does not show the typical Ile/Val-Tyr-Pro-Trp-Met consensus but instead, a Thr-Phe-Asp-Trp-Met amino acid sequence, only the last two codons being conserved. This divergent pentapeptide is closely homologous to the Tyr-Phe-Asp-Trp-Met sequence present at the same position in the *Drosophila* gene Caudal (*cad*; Mlodzik *et al.*, 1985). Fourteen out of 15 nucleotides are identical between *Hox-1.6* and *cad* within this region. No other striking homology is detected.

The splicing pattern of cDNA 1 brings a termination codon (TAA, boxed in Figure 7B, position 400) located within the 40-bp mini-exon 2, (Figure 7Ab) into phase with the unique extended open reading frame that correctly translates the homeo-domain. The first usable Met codon (Figure 7B, asterisk) lies within the conserved pentapeptide of the 95-bp mini-exon 3, (Figure 7A, exon 3), 10 codons before the occurrence of the exon containing the homeo-box (exon 4). The putative protein sequence contains 128 amino acids and ends 48 residues downstream of the homeo-domain. This would correspond to a protein of ~ 14.5 kd, rich in lysine (11%) and threonine (11%) residues distributed throughout the polypeptide. Serine residues (11%) are clustered in the C-terminal part of the sequence (11 of the last 26 amino acids are serines), as is the case for the human C13 gene (Mavilio *et al.*, 1986) and its murine homologue (our unpublished work).

The putative translation of cDNA clone 2 did not reveal any extended open reading frame. Two additional termination codons (positions 76 and 220, boxed in Figure 7B; see also Figure 7Ac) occur in the frame corresponding to that of cDNA 1 in the region where these two clones overlap.

The termination codon lying within the 40-bp mini-exon 2 of cDNA 1 (position 400, Figure 7B) necessitates translation initiation from the ATG within the sequence encoding the conserved pentapeptide of the 95-bp mini-exon 3. This would give rise to a surprisingly small protein when compared with the size of the RNA. However, a larger protein including the pentapeptide could be generated from transcripts lacking the TAA-containing exon 2. In this case, translation could initiate at some upstream Met codon and read through. In order to investigate this possibility, a cDNA 1 probe was prepared for S1 nuclease protection analysis. This probe was a cDNA 1 fragment 5′-end-labelled at its *Bgl*II site (position 682 in Figure 7B; Figure 8A) and extending within the vector (pUC) to avoid contaminating signals produced by the probe reannealed on itself. Transcripts corresponding to the entire cDNA 1 should protect 324

nucleotides of the probe, 306 nucleotides if the 5'-located 18 nucleotides are missing, and 266 nucleotides if mini-exon 2 is absent (see Figure 8A). When this cDNA 1/pUC hybrid probe is hybridized to either embryonic total RNA or F9 and P19 cells cytoplasmic RNA and treated with S1 nuclease, two major resistant hybrids are clearly and repeatedly detected (Figure 8B). These two protected fragments are 306 and 266 nucleotides long and exactly map to the 5' ends of both the 40-bp mini-exon 2 and the mini-exon 3, respectively (triangles in Figure 7B; Figure 8A; compare also with the sequence in Figure 8B). Surprisingly, no detectable full length protection (324 nucleotides Figure 8A) of cDNA 1 was observed suggesting that the 5'-located 18 bp (called exon 1 in Figure 7Ab) are either present in very few *Hox-1.6* transcripts or due to a cloning artefact. As illustrated in Figure 8B, most of the cytoplasmic steady-state *Hox-1.6* RNAs contain the 40-bp mini-exon 2. However, a significant proportion of *Hox-1.6* RNA (~20–30%) does not contain this mini-exon nor, as a consequence, the in-frame termination codon lying within it. About the same ratio of these two different transcripts is observed when total embryonic RNA is used, though the 266-nucleotide protection becomes difficult to detect in older fetuses due to the overall decrease in *Hox-1.6* RNA content (see above). The same two protections were obtained when total RNA extracted from the 113 tumour was used, although in this latter case the 266-nucleotide protection seems proportionally weaker than in cytoplasmic F9 and P19 RNA (Figure 8B). A much weaker protection of ~175 nucleotides was detected when P19 cytoplasmic RNA was used (Figure 8B). This would correspond to the 5' boundary of exon 4 and therefore suggests that some very rare transcripts might not even contain mini-exon 3.

Since the 18 bp at the 5' end of cDNA 1 (exon 1 in Figure 7Ab) were not detected by S1 analysis, the possibility remained that the 40-bp exon 2 might be larger and be contained, at least partially, within cDNA 2 which does include 39 out of the 40 mini-exon 2 bp and extends 5' following the genomic sequence for 376 nucleotides (Figure 7B). This possibility was investigated by S1 analysis with a cDNA 2 probe labelled at the *Ava*II site (position 128 in Figure 7B; Figure 8A) and extending 5' into the pEMBL vector for reasons described above. This probe was hybridized to the same RNA as for the cDNA 1 probe, in the same experiments under identical conditions. No protection could be detected (not shown) suggesting that cDNA 2 was generated from an abnormal or unspliced RNA molecule and therefore does not represent part of the genuine exon 2. However, we cannot rule out the possibility that exon 2 extends slightly more 5' (between the *Ava*II site and box 2 in Figure 8A) from its apparent boundary which was fixed according to cDNA 1 structure.

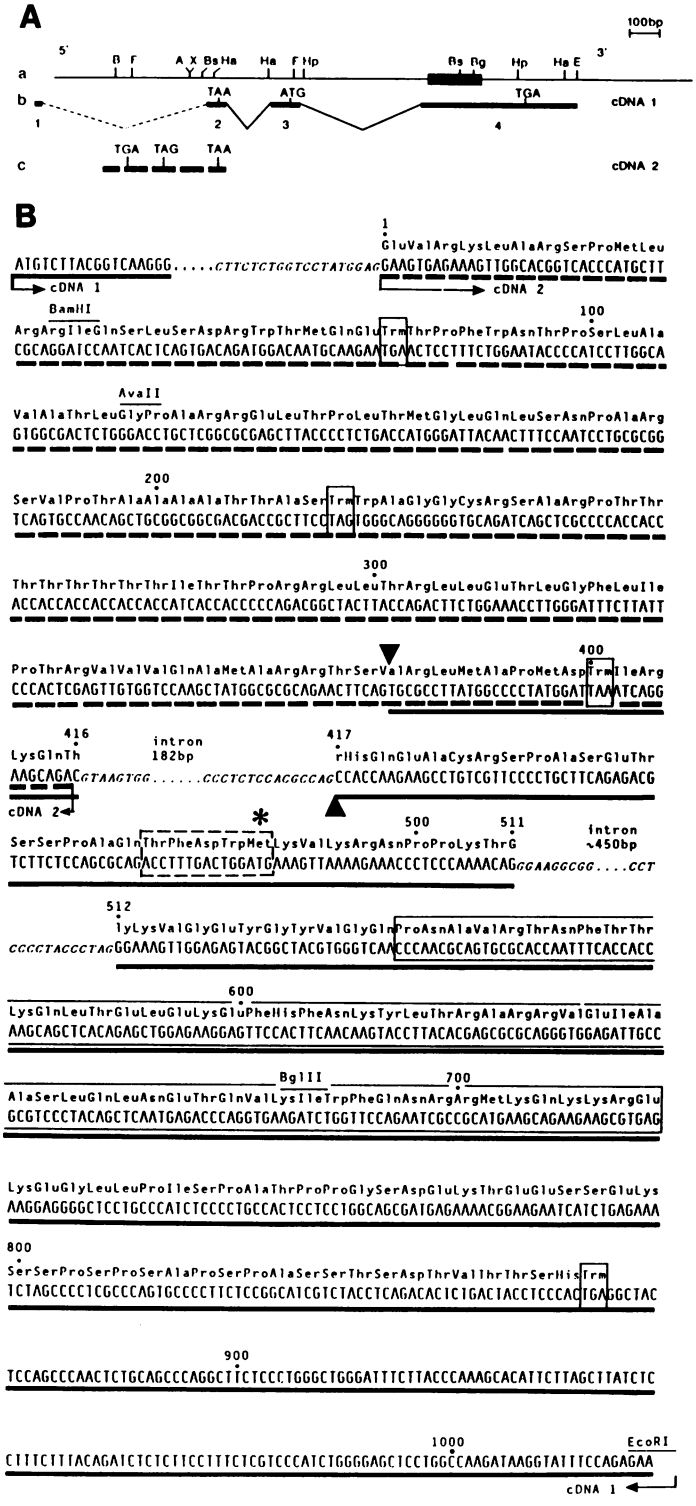

Fig. 7. Structures and sequences of partial *Hox-1.6* embryonic cDNA clones. **(A) a:** restriction map of a *Hox-1.6* genomic subregion. The restriction sites are: F, *Fok*I; A, *Ava*I; X, *Xho*I; the others as for Figure 2. The homeo-box is represented by a black box. **b:** structure of cDNA 1: the thick lines represent the locations of the different exons (see text) present in cDNA 1. The position of the 5'-located 18 nucleotides (called exon 1) was not determined on the genomic clone. This short sequence is therefore linked to exon 2 by a dashed line. **c:** structure of cDNA 2. The cDNA clone 2 is represented by a thick dashed line (for clarity in **B**). The location of an initiation codon (ATG) as well as terminators (TGA, TAG, TAA) are indicated on the top of each cDNA clone. **(B)** Sequence of cDNA clones 1 and 2. The sequences of both cDNAs were concatemerized at their overlapping parts. The sequence of cDNA 1 is underlined with a continuous thick line whereas cDNA 2 sequence is underlined with an interrupted thick line (as for **A**). The starts and the ends of both cDNAs are indicated by arrows underneath the sequences. The top left 18 nucleotides present in cDNA 1 were not mapped (see above) and are therefore separated from the rest of the sequence. As a consequence, position No. 1 is the first nucleotide of cDNA clone 2. Italicized nucleotides represent genomic sequences located either at the exon/intron boundaries or 5' from the start of cDNA clone 2. The various termination codons as well as the homeo-box are boxed. The conserved pentapeptide is indicated by a dashed line box. The asterisk marks the Met codon possibly used as initiator in some alternative transcripts (see text). The two triangles show the exact 5' boundaries of the two main S1 nuclease protections obtained with the cDNA 1 probe (see Figure 8). They exactly coincide with the 5' boundaries of exon 2 and 3 (see the thick line underlying). Only the first and last several nucleotides of the intronic sequences (182 bp and ~450 bp) are shown in small type.

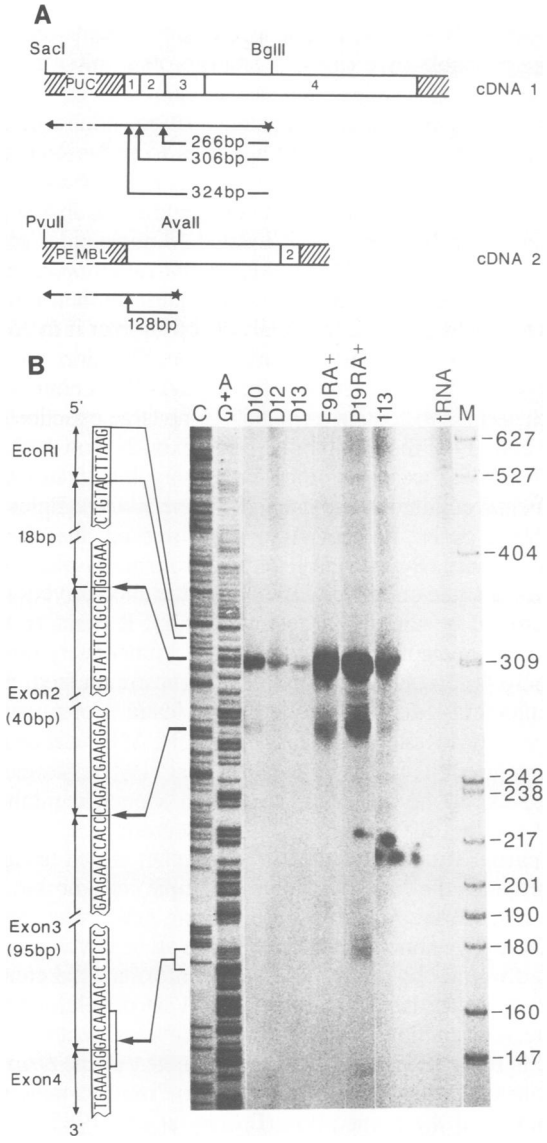

Fig. 8. Alternative transcripts for the *Hox-1.6* gene. (**A**) Schemes of the two cDNA probes. Top: open boxes (1–4) represent the different exons present in cDNA 1 (see Figure 7A). The shaded boxes are the vector (pUC). The probe was end-labelled at the *Bgl*II site and subsequently digested within the vector with *Sac*I. The sizes of the expected protections are indicated below. Bottom: the open boxes represent cDNA 2 including 39 bp of cDNA 1 exon 2. Shaded boxes as above. The probe was end-labelled at the *Ava*II site and cut at the *Pvu*II site. The protection expected if cDNA 2 corresponded to a genuine *Hox-1.6* transcript is shown below. (**B**) S1 mapping of alternative *Hox-1.6* transcripts using cDNA probe 1. 100 μg of total fetal RNAs at days 10 (D10), 12 (D12) and 13 (D13), cytoplasmic RNA extracted from F9 (100 μg) and P19 (70 μg) cells treated with RA (F9 RA^+, P19 RA^+) and total RNA extracted from the endodermal-like tumour LT-113 (T113; 100 μg) as well as tRNA (100 μg) were hybridized to cDNA 1 probe (see above) and digested with S1 nuclease as for Figure 6. The resistant hybrids were analyzed by comparison with the cDNA 1 probe sequence: (G and T+C reactions indicated as C and A+G reactions in the figure in order to allow direct comparison with the RNA sequence). Molecular markers (M) in nucleotides are in the right most panel. The left side of the figure shows the sequence and exonic structures of cDNA 1 with, at the top, the 5'-located *Eco*RI cloning site (also indicated in the sequencing lanes) and, at the bottom, the 5' end of the homeo-box-containing exon 4. Only sequences around the exons' junctions are shown. Arrows indicate the 5' starts of the two main protections which exactly match with the 5' ends of exons 2 (306 bp) and 3 (266 bp). A much weaker protection is detected in P19 cell cytoplasmic RNA which might correspond to the 5' boundary of exon 4 (~177 bp, bottom arrow). The sequencing, marker, embryonic and F9 lanes are from the same gel, while the remaining lanes are from another. Exposure times were 12 h (P19), 6 days (113 and tRNA) and 10 days (F9 and embryos).

Discussion

This paper reports the characterization of *Hox-1.6*, a previously undescribed mouse homeo-box-containing gene member of the *Hox-1* complex. Among the homeo-box sequences of the *Hox-1* complex, *Hox-1.6* shows the least homology with the *Drosophila Antp* homeo-box reference sequence due to a higher divergence in the 5' part of the box. The divergence is maintained at the level of the putative amino acid sequence since, contrary to what is normally observed, the homology of the *Hox-1.6* homeo-domain with the other *Hox-1* homeo-domains is lower (~60%) than that observed at the nucleotide level (~65%). However, the two sub-domains thought to form the two helices of the helix–turn–helix structure (Pabo and Sauer, 1984) (from codon 31 to 38 and 42 to 50, Sheperd *et al.*, 1984; Laughon and Scott, 1984; Laughon *et al.*, 1985) show high conservation with previously reported sequences; only conservative changes are observed, or substitutions at positions which usually show variability (for example, codons 36 and 37). A significant difference is the presence of a Thr codon at position 42 of the *Hox-1.6* homeo-domain, instead of the Arg codon found in most *Antp*-like homeo-boxes so far reported. This amino acid substitution has been previously observed in the case of the *Drosophila* engrailed-like (*En*-like) homeo-boxes (Fjose *et al.*, 1985; Poole *et al.*, 1985) and in the yeast a1 and α2 mating types (Nasmyth *et al.*, 1981) and occurs in a region presumably interacting with the major groove of the DNA. Variations within this region might therefore affect the binding specificity of the protein.

Specific expression of Hox-1.6 during development, in adult tissues and in teratocarcinoma cells

Trancriptional studies with *Hox-1.6* revealed that this homeo-box-containing gene exhibits ontogenetic, tissue and cell type specificity. Two large transcripts (~4.5 and ~10 kb) are recognized by the *Hox-1.6* probe in fetuses at day 9, but are not detected by day 14–15. Recent evidence obtained with other murine homeo-genes using *in situ* hybridization on early mouse embryos and fetuses (Awgulewitch *et al.*, 1986; Gaunt *et al.*, 1986) favours the hypothesis that such *Hox-1.6* homeo-box-containing transcripts are present in a restricted number of embryonic and fetal structures. We are currently investigating this question using an identical approach. Unlike other RNA species of the *Hox-1* complex (*Hox-1.1*,. *Hox-1.2* and *Hox-1.4*; Colberg-Poley *et al.*, 1985a,b; Duboule *et al.*, 1986; Wolgemuth *et al.*, 1986), the major *Hox-1.6* transcripts are detected neither in adult testes nor in kidneys. In fact, the distribution of the prominent mRNAs hybridizing to the *Hox-1.6* probe shows a specificity for the adult intestine where two shorter transcripts (~1 and ~1.4 kb) are observed. Which cell-type(s) (e.g. epithelial cells, muscle cells, nerve cells) within the intestine expresses *Hox-1.6* remains to be determined. Interestingly, homeo-box-containing transcripts in the intestine are also observed in the case of the human genes C1 and C8 (Sebastio *et al.*, in preparation). However, no particular sequence homology was found between these two human genes and *Hox-1.6*.

S1 nuclease digestion studies have confirmed that the homeo-box is part of the RNA sequence. Because of its high sensitivity, S1 nuclease digestion was also used to look for *Hox-1.6* cell type specificity among different *in vivo* propagated teratocarcinoma tumours (Blühtmann *et al.*, 1983; Sassone-Corsi *et al.*, 1985). The strongest signals were obtained with RNAs extracted from tumours showing endodermal-like differentiation (LT-113; OTT2158). The other tumours, including those showing mesodermal-like differentiation and known to contain *Hox-1.4*

transcripts (Duboule *et al.*, 1986) were negative for *Hox-1.6* expression. This observation is consistent with the very significant increase in the amount of *Hox-1.6* transcripts after treatment of F9 cells with RA which induces these cells to differentiate into parietal endoderm-like cells. It appears therefore that *Hox-1.6* shows a true specificity for expression in early endodermal-like cells. This increase of homeo-box-containing transcripts in RA-treated F9 cells was previously observed with genes which otherwise show transcription patterns different from *Hox-1.6* (Colberg-Poley *et al.*, 1985b).

Alternative splicing for the Hox-1.6 homeo-gene transcripts

The nucleotide and deduced amino acid sequences of a 677-bp partial *Hox-1.6* embryonic cDNA clone has confirmed the presence and position of the acceptor splice site which was localized by S1 nuclease mapping immediately upstream (−34) from the homeo-box. The occurrence and position of this site is remarkably conserved in most *Drosophila* (see for example: *ftz*. Kuroiwa *et al.*, 1984; *Antp*. Schneuwly *et al.*, 1986; *Ubx*, Hogness *et al.* 1985; *cad*, Mlodzik and Gehring, 1987; *Dfd*, Regulski *et al.* 1987; *eve*, Macdonald *et al.*, 1986) and vertebrate (Mavilio *et al.*, 1986; Duboule *et al.*, 1986) genes containing homeo-boxes. Such a conservation implies a function for this splice site which may be of general importance for homeo-box-containing genes of various species. In most cases reported to date, another conserved region encoding a pentapeptide (Mavilio *et al.*, 1986) is consistently linked to the homeo-box region via this splice site. Though the *Hox-1.6* pentapeptide differs from the consensus sequence, it is very much homologous to the corresponding sequence present in the *Drosophila* gene Caudal (*cad*, Mlodzik *et al.*, 1985; Mlodzik and Gehring, 1987).

The position of the C-terminal termination codon (48 amino acids downstream from the homeo-box) as well as the abundant serine residues in the 3′ end of the *Hox-1.6* protein sequence are comparable to what is observed in other vertebrate homeo-genes such as the human C13 gene (termination occurs 41 amino acids downstream from the homeo-box; Mavilio *et al.*, 1986), the *Xenopus Xhox1A* gene (32 amino acids, Harvey *et al.*, 1986) or different murine genes (our unpublished work). These similarities suggest that both the structure and length of the C terminus of the protein may be functionally important.

The *Hox-1.6* gene contains at least three exons which are present in the cDNA clone 1 reported here (exons 2, 3,4). This cDNA would link a putative 5′ exon to two mini-exons of 40 bp and 95 bp, followed by the last exon 4 harbouring the homeo-box and the region coding for the C terminus of the putative protein. A termination codon in-phase with the homeo-box is found in mini-exon 2. As a consequence, translation would begin at the initiation codon lying within the conserved pentapeptide, toward the end of mini-exon 3. The resulting protein would thus not include most of the amino acids of the conserved pentapeptide. This is surprising because this conserved sequence is present in the products of most homeo-box-containing genes. One possibility is that this *Hox-1.6* cDNA clone represents only one of the several mRNAs generated by an alternative splicing mechanism. For instance, a splicing pattern excluding exon 2 by directly linking a 5′ exon to exon 3 would remove the termination codon and thus allow the conserved pentapeptide to be translated as part of a larger protein initiated further upstream. The existence of such an alternative splicing mechanism was demonstrated by using the cDNA 1 as a probe for S1 mapping. The results showed that at least two different species of *Hox-1.6* transcripts were present in all RNA samples used (fetal or cellular cytoplasmic RNA). The major species do contain exon 2 and therefore probably give rise to a small protein initiating within the conserved pentapeptide whereas the minor species do not contain exon 2 and, consequently, the termination codon located just upstream. Further work should allow us to prove or disprove the hypothesis that this second RNA species might encode a much larger protein. Since no full-length protection was obtained with this cDNA 1 probe, the most 5′-located 18 nucleotides are likely not part of a genuine *Hox-1.6* RNA but rather were artefactually added to exons 2, 3 and 4. The question then remained open as to whether exon 2 was really 40 bp large or if these 40 bp represented only the 3′ part of a larger exon. The latter possibility was supported by the structure of the second cDNA clone (cDNA 2) which seemed to contain a unique 415-bp large exon including, at its 3′ end, 39 of the 40 bp of cDNA 1 exon 2. However, when this cDNA 2 was used as a probe for S1 mapping under the same experimental conditions and using the same RNA samples as for the cDNA 1 probe, no protection was observed. It is therefore very likely that cDNA 2 was generated from an unspliced RNA. This does not rule out the possibility that the alternatively spliced exon 2 might be slightly larger than 40 bp. It is interesting to note that the nucleotide sequence located immediately upstream from these 40 bp on the genomic DNA (upstream from the upper triangle in Figure 7B) presents some features of a consensus acceptor spliced site (. . . CAGAACTTCAG/) although the overall sequence is rather atypical (Mount, 1982). Should exon 2 indeed be 40 bp larger, this atypical sequence might well account for the differential splicing observed.

The partial structure of the *Hox-1.6* gene, which is reported here, provides the first example of multiple intronic sequences within a vertebrate homeo-box-containing gene. The existence of an alternative splicing pattern which may be developmentally regulated would obviously increase the informational content of this homeo-gene since it may generate a set of related proteins with specific functions in the course of development or in adult tissues. In this respect, it is interesting to note that the *Drosophila* gene Ultrabithorax (*Ubx*) also has two 'mini-exons' located upstream from the homeo-box (Beachy *et al.*, 1985), that are differentially spliced to yield alternative transcripts (Lipshitz *et al.*, 1987). Moreover, a nonsense mutation within the second mini-exon impairs *Ubx* function in the developing epidermis but not in the central nervous system (Weinzierl *et al.*, 1987). In the case of *Hox-1.6* different proteins carrying the same putative DNA binding site (the homeo-domain) but differing in the composition of their amino-terminal regions could also be produced and have different functions.

Materials and methods

Screening of the genomic cosmid library, subcloning and mapping of cos2

The cosmid clone cos2 was one of a set of overlapping cosmid clones whose isolation and characterization were previously reported (Duboule *et al.*, 1986). The mouse cosmid library Pcos2EMBL (Poutska *et al.*, 1984) was provided by H.Lehrach. The various *Drosophila* homeo-box-containing probes (*Antp*, ftz, *Ubx*) were obtained from W.Gehring. The 1.6-kb *Eco*RI/*Bam*HI restriction fragment containing the *Hox-1.6* homeo-box was subcloned in pBR322. This subclone (pHox-1.6) was used for fine restriction mapping and sequencing by the Maxam and Gilbert technique (1977) with the strategy outlined in Figure 2. The restriction mapping of cos2 which provided both the exact location and the orientation of the subclone, was carried out using a series of partial restriction digests, subsequent transfers (Southern, 1975) and hybridizations under high stringency conditions (2 × SSC, 2 × Denhardt's, 10 μg/ml salmon sperm DNA, 0.1% SDS at 68°C for 12 h) to specific cos2 subfragments electroeluted from agarose gels and labelled by nick-translation using ^{32}P-labelled αCTP, αTTP and αATP (Amersham) to a specific activity of ~ 10^8 c.p.m./μg DNA. The nitrocellulose membranes (Schleicher and Schull) were then washed in 0.1 × SSC for 1 h at 68°C, and exposed using Kodak X-Omat films plus intensifying screens at −70°C.

Hybridization to mouse genomic DNA

10 μg of mouse genomic DNA from strain 129 were restricted with *Eco*RI and *Bam*HI, separated on 0.8% agarose gels and transferred as before onto a HY-BOND N membrane (Amersham). As a probe, a *BglII/Eco*RI fragment (solid bar in Figure 2) was subcloned in a Gemini vector (pGEM1; Promega, Biotech). the fragment was labelled either by nick-translation after electroelution from a 5% acrylamide gel or by synthesis of an anti-sense RNA molecule from the T7 promoter in the presence of ^{32}P-labelled ribonucleotide triphosphate (Amersham).

Collection of mouse fetuses, adult tissues, teratocarcinoma tumours and extraction of RNAs

As previously described (Duboule *et al.*, 1986).

EC cells

F9 cells were cultured in Dulbecco's modification of Eagle's medium (DMEM) supplemented with 10% fetal calf serum in a 37°C environment, 7% in CO_2. Cells were plated at a density of 2×10^5 cells per 100 mm culture dish. 48 h later, the medium was changed and RA (Sigma) at 5×10^{-7} M in ethanol was added to half the plates. Cells were harvested 24 h later, lysed with 0.3% Nonidet-P40 and total cytoplasmic RNA was purified. P19 cells (gift of M. W. McBurney) were treated similarly except for a 5% CO_2 environment.

Northern blot analysis

20–30 μg of total RNA extracted from all sources mentioned above were separated on agarose slab gels containing 10 mM methyl mercury hydroxide (Bailey and Davidson, 1976), photographed under u.v. illumination and further transferred to freshly prepared DBM-papers (Alwine *et al.*, 1977). Hybridizations were carried out according to established procedures (Whal *et al.*, 1979), using the *BglII/Eco*RI fragment labelled as before. The last wash was done in 0.1% SSC, 0.1% SDS at 80°C. After dehybridization in boiling water (15 min), the same DBM-paper containing RNA from various adult tissues was re-hybridized with a mouse genomic clone encoding TPI (gift of M.Metali) to check the quality of the RNA. Films were exposed for 5–7 days as before.

S1 nuclease analysis

The two different genomic double-stranded 5′ end-labelled probes as well as the cDNA 1 and cDNA 2 end-labelled probes were prepared according to Maniatis *et al.* (1982). Hybridizations were carried out using 50–200 μg of total RNA in the presence of an excess of probe in 10 μl of 10 mM Pipes pH 6.5, 80% formamide, 400 mM NaCl at 55°C overnight. Prior to hybridization, the nucleic acids were denatured for 10 min at 70°C. The samples were then diluted in 300 μl of 30 mM NaOAc pH 4.5, 3 mM ZnCl, 400 mM NaCl containing 50–100 units of S1 nuclease (Appligene) and incubated for 2 h at 25°C. The results for F9 cells presented in Figure 6 were obtained after 3 h of hybridization at 55°C and digestion for 1 h at 25°C with 200 units of S1 nuclease. Resistant hybrids were analysed on 8% acrylamide, 8.3 M urea sequencing gels. The probes were sequenced using the Maxam and Gilbert technique (1977) and run on the same gels in order to determine the exact positions of the protection. Control hybridizations were carried out with either no RNA or non-specific RNA (tRNA) under the same conditions.

Screening of the cDNA libraries, subcloning and sequencing

The cDNA libraries were prepared in lambda phage λgt10 with RNA isolated from mouse embryos at either day 8.5 (library kindly provided by B.Hogan, MRC NIMR Mill Hill and K.Fahrner, Biogen) or day 9.5–10 (described in Duboule *et al.*, 1987). The former was screened using a *Drosophila* probe containing the fushi-tarazu (*ftz*) homeo-box (gift of W.Gehring). The 677-bp partial *Hox-1.6* cDNA clone 1 appeared as one in ~40 000 plaques. The insert was subcloned in M13 phage (Messing *et al.*, 1981) and sequenced using the chain termination technique (Sanger *et al.*, 1977). Restriction enzyme sites deduced from the DNA sequence (*Hae*III, *Fok*I) as well as a synthetic oligonucleotide probe (from nucleotides 375 to 415 on the cDNA clone; Figure 7B) were used to localize the positions of the two mini-exons whose corresponding genomic sequences were determined on both strands by the Maxam and Gilbert method (1977). The latter library was screened using the entire pHox-1.6 genomic insert (see Figure 2). The 415-bp partial *Hox-1.6* cDNA clone 2 appeared as one in ~ 4 500 000 plaques. The insert, subcloned in a pEMBL vector (Dente *et al.*, 1983) as well as the corresponding genomic sequence were entirely sequenced using the Maxam and Gilbert technique.

Acknowledgements

We wish to thank P.Chambon for continuous interest in the course of this work and helpful discussions. We thank M.LeMeur, A.Vidal and M.Metali for their expertise with DBM-paper, C.Sime for excellent technical assistance, A.Staub and F.Ruffenach for synthesizing the oligonucleotide, M.Gilbert and C.Marfin for the cell culture work, G.Richards for critical reading of the manuscript, C.Werlé and B.Boulay for the illustrations and the secretarial staff for typing the manuscript. This work was funded by grants from the INSERM, CNRS, ARC and FRMF. M.S.F. is supported by a MRC of Canada post-doctoral fellowship.

References

Alwine,J.C., Kemp,D.J. and Stark,G.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5350–5354.
Awgulewitsch,A., Utset,M.F., Hart,C.P., McGinnis,W. and Ruddle,F.H. (1986) *Nature*, **320**, 328–335.
Bailey,J.M. and Davidson,N. (1976) *Anal. Biochem.*, **70**, 75–85.
Beachy,P.A., Helfand,S.L. and Hogness,D.S. (1985) *Nature*, **313**, 545–551.
Bernstine,E.G., Hooper,M.L., Grandchamp,S. and Ephrussi,B. (1973) *Proc. Natl. Acad. Sci. USA*, **70**, 3899–3903.
Blühtmann,H., Vogt,E., Hoesli,P., Stevens,L.C. and Illmensee,K. (1983) *Differentiation*, **24**, 65–75.
Boncinelli,E., Simeone,A., La Volpe,A., Faiella,A., Fidanza,V., Acampora,D. and Scotto,L. (1985) *Cold Spring Harbor Symp. Quant. Biol.*, **50**, 301–307.
Breathnach,R., Benoist,C., O'Hare,K., Gannon,F. and Chambon,P. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 4853–4857.
Bucan,M., Yang-Feng,T., Colberg-Poley,A.M., Wolgemuth,D.J., Guenet,J.L., Francke,U. and Lehrach,H. (1986) *EMBO J.*, **5**, 2899–2905.
Carrasco,A.E., McGinnis,W., Gehring,W.J. and DeRobertis,E.M. (1984) *Cell*, **37**, 409–414.
Colberg-Poley,A.M., Voss,S.D., Chowdhury,K. and Gruss,P. (1985a) *Nature*, **314**, 713–718.
Colberg-Poley,A.M., Voss,S.D., Chowdhury,K., Stewart,C.L., Wagner,E.F. and Gruss,P. (1985b) *Cell*, **43**, 39–45.
Dente,L., Cesareni,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **6**, 1645–1655.
Desplan,C., Theis,J. and O'Farrell,P.H. (1985) *Nature*, **318**, 630–635.
DiNardo,S., Kuner,J.L., Theis,J. and O'Farrell,P.H. (1985) *Cell*, **43**, 59–69.
Duboule,D., Baron,A., Maehl,P. and Galliot,B. (1986) *EMBO J.*, **5**, 1973–1980.
Duboule,D., Haenlin,M., Galliot,B. and Mohier,E. (1987) *Mol. Cell. Biol.*, **7**, 2003–2006.
Fjose,A., McGinnis,W.J. and Gehring,W.J. (1985) *Nature*, **313**, 284–289.
Garcia-Bellido,A. (1977) *Am. Zool.*, **17**, 613–629.
Gaunt,S.J., Miller,J.R., Powell,D.J. and Duboule,D. (1986) *Nature*, **324**, 662–664.
Hart,C.P., Awgulewitsch,A., Fainsod,A., McGinnis,W. and Ruddle,F.H. (1985) *Cell*, **43**, 9–18.
Harvey,R.P., Tabin,C.J. and Melton,D.A. (1986) *EMBO J.*, **5**, 1237–1244.
Hauser,C.A., Joyner,A.L., Klein,R.D., Learned,T.K., Martin,G.R. and Tjian,R. (1985) *Cell*, **43**, 19–28.
Hogness,D.S., Lipshitz,H.D., Beachy,P.A., Peattie,D.A., Saint,R.B., Goldschmidt-Clermont,M., Harte,P.J., Gavis,E.R. and Helfand,S.L. (1985) *Cold Spring Habor Symp. Quant. Biol.*, **50**, 181–194.
Jackson,I.J., Schofield,P. and Hogan,B. (1985) *Nature*, **317**, 745–748.
Joyner,A.L., Lebo,R.V., Kan,Y.W., Tjian,R., Cox,D.R. and Martin,G.R. (1985) *Nature*, **314**, 173–175.
Kuroiwa,A., Hafen,E. and Gehring,W.J. (1984) *Cell*, **37**, 825–831.
Laughon,A. and Scott,M.P. (1984) *Nature*, **310**, 25–31.
Laughon,A., Carroll,S.B., Storfer,F.A., Riley,P.D. and Scott,M.P. (1985) *Cold Spring Harbor Symp. Quant. Biol.*, **50**, 153–262.
Levine,M., Rubin,G.M. and Tjian,R. (1984) *Cell*, **38**, 667–673.
Lewis,E.B. (1978) *Nature*, **276**, 565–570.
Lipshitz,M.D., Peattie,D.A. and Hogness,D.S. (1987) *Genes Dev.*, **1**, 307–322.
Lyon,M. (1987) *Nature*, **325**, 22.
Macdonald,P.M., Ingham,P. and Struhl,G. (1986) *Cell*, **47**, 721–734.
Maniatis,T., Fristch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
Martin,G.R., Boncinelli,E., Duboule,D., Gruss,P., Jackson,I.J., Krumlauf,R., Lonai,P., McGinnis,W., Ruddle,F.H. and Wolgemuth,D.J. (1987) *Nature*, **325**, 21–22.
Mavilio,F., Simeone,A., Giampaolo,A., Faiella,A., Zappavigna,V., Acampora,D., Poiana,G., Russo,G., Peschle,C. and Boncinelli,E. (1986) *Nature*, **324**, 664–668.
Maxam,A.M. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 499–560.
McBurney,M.W. and Rogers,B.J. (1982) *Dev. Biol.*, **89**, 503–508.
McGinnis,W., Levine,M., Hafen,E., Kuroiwa,A. and Gehring,W.J. (1984a) *Nature*, **308**, 428–433.
McGinnis,W., Hart,C.P., Gehring,W.J. and Ruddle,F.H. (1984b) *Cell*, **38**, 675–680.
Messing,J., Crea,R. and Seeburg,P.H. (1981) *Nucleic Acids Res.*, **9**, 309–321.
Mlodzik,M., Fjose,A. and Gehring,W.J. (1985) *EMBO J.*, **4**, 2961–2969.
Mlodzik,M. and Gehring,W.J. (1987) *Cell*, **48**, 465–478.
Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.
Müller,M.M., Carrasco,A.E. and DeRobertis,E.M. (1984) *Cell*, **39**, 157–162.
Nasmyth,K.A., Tatchell,K., Hall,B.D., Astell,C. and Smith,M. (1981) *Cold Spring Harbor Symp. Quant. Biol.*, **45**, 961–981.
Pabo,C.O. and Sauer,R.T. (1984) *Annu. Rev. Biochem.*, **53**, 293–321.
Poole,S.J., Kauvar,L.M., Drees,B. and Kornberg,T. (1985) *Cell*, **40**, 37–43.

Poustka,A., Rackwitz,M.R., Frischauf,A.M., Hohn,H. and Lehrach,H. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4129–4133.
Regulski,M., McGinnis,N., Chadwick,R. and McGinnis,W. (1987) *EMBO J.*, **6**, 767–777.
Rubin,M.R., Toth,L.E., Patel,M.D., D'Eustachio,P. and Nguyen-Huu,M.C. (1986) *Science*, **233**, 663–667.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Sassone-Corsi,P., Duboule,D. and Chambon,P. (1985) *Cold Spring Harbor Symp. Quant. Biol.*, **50**, 747–752.
Schneuwly,S., Kuroiwa,A., Baumgartner,P. and Gehring,W.J. (1986) *EMBO J.*, **5**, 733–739.
Scott,M.P. and Weiner,A.J. (1984) *Proc. Natl. Acad. Sci. USA*, **78**, 1095–1099.
Shepherd,J.C.W., McGinnis,W., Carrasco,A.E., DeRobertis,E.M. and Gehring,W.J. (1984) *Nature*, **310**, 70–71.
Simeone,A., Mavilio,F., Bottero,L., Giampaolo,A., Russo,G., Faiella,A., Boncinelli,E. and Peschle,C. (1986) *Nature*, **320**, 763–765.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
Strickland,S. and Mahdavi,V. (1978) *Cell*, **15**, 393–403.
Wahl,G.M., Stern,M. and Stark,G.R. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3683–3687.
Weinzierl,R., Axton,J.M., Ghysen,A. and Akam,M. (1987) *Genes Dev.*, **1**, 386–397.
White,R.A.H. and Wilcox,M. (1984) *Cell*, **39**, 163–171.
Wolgemuth,D.J., Engelmyer,E., Duggal,R.N., Gizang-Ginsberg,E., Mutter,G.L., Ponzetto,C., Viviano,C. and Zakeri,Z.F. (1986) *EMBO J.*, **5**, 1229–1235.

Received on February 23, 1987; revised on June 30, 1987