

Remarkable Intron and Exon Sequence Conservation in Human and Mouse Homeobox *Hox 1.3* Genes

ELIZABETH TOURNIER-LASSERVE,* WARD F. ODENWALD, JAMES GARBERN, JOHN TROJANOWSKI,
AND ROBERT A. LAZZARINI†

Laboratory of Molecular and Viral Pathogenesis, National Institute of Neurological and Communicative
Disorder and Stroke, Bethesda, Maryland 20892

Received 25 October 1988/Accepted 27 January 1989

A high degree of conservation exists between the *Hox 1.3* homeobox genes of mice and humans. The two genes occupy the same relative positions in their respective *Hox 1* gene clusters, they show extensive sequence similarities in their coding and noncoding portions, and both are transcribed into multiple transcripts of similar sizes. The predicted human *Hox 1.3* protein differs from its murine counterpart in only 7 of 270 amino acids. The sequence similarity in the 250 base pairs upstream of the initiation codon is 98%, the similarity between the two introns, both 960 base pairs long, is 72%, and the similarity in the 3' noncoding region from termination codon to polyadenylation signal is 90%. Both mouse and human *Hox 1.3* introns contain a sequence with homology to a mating-type-controlled *cis* element of the yeast Ty1 transposon. DNA-binding studies with a recombinant mouse *Hox 1.3* protein identified two binding sites in the intron, both of which were within the region of shared homology with this Ty1 *cis* element.

The development of a fertilized egg into a complex multicellular organism is an intricate process requiring cell division, cell migration, and differentiation. Our insights into these mechanisms operating during embryogenesis, particularly during *Drosophila* body pattern formation, have advanced enormously in the past few years (for a review, see reference 25). Genetic analyses of *Drosophila* mutants have defined three classes of genes that play important roles in providing positional information: the maternal effect genes, the segmentation genes, and the homeotic genes. Many but not all of these genes contain a highly conserved sequence of 180 base pairs (bp), the homeobox (17, 24), which is known to encode a DNA-binding domain (for review, see reference 10).

In an earlier communication, we described the organization, sequence, and expression of the murine *Hox 1.3* homeobox gene, a member of the *Hox 1* cluster located on mouse chromosome 6 (21). This gene has two exons separated by a 960-bp intron and encodes a 270-amino-acid homeodomain protein. Recently, we have shown that the *Hox 1.3* protein is phosphorylated and that it binds to DNA in a sequence-specific manner (20). The consensus binding motif is CPyPyNATTAT/GPy (Py = pyrimidine).

In this communication, we describe the sequence and organization of the human *Hox 1.3* gene and its location within the human *Hox 1* cluster. A high level of homology is observed between the human and mouse *Hox 1.3* genes. The nucleotide sequence analysis predicts that the human and the mouse proteins both have 270 amino acids and differ from each other at only seven positions. The homology is also very high in the region 5' to the predicted cap site, the 3' noncoding region, and, most surprisingly, the intron. We show here that a conserved region in the human and murine *Hox 1.3* introns contains two binding sites for the *Hox 1.3* protein and shares homology with the central epsilon region of the yeast Ty1 transposon.

Organization of the human *Hox 1* cluster. A human genomic λ EMBL 3 library (a gift from J. Weiss, Harvard University, Cambridge, Mass.) was screened under stringent conditions with a cDNA fragment (5' probe) encoding the amino-terminal portion of the murine *Hox 1.3* protein but not the homeodomain. Two λ clones ($\lambda 26$ and $\lambda 32$) were identified and further characterized by Southern blot analysis, using the 5' probe and two additional probes which represented the homeobox region and the 3' untranslated regions of the gene (designated *Hox 1.3* box probe and 3' probe). A 3.5-kilobase-pair (kbp) (*SalI-EcoRI*) DNA fragment from $\lambda 32$ hybridized with the 5', 3', and homeobox probes. A 6-kbp *SalI* DNA fragment from $\lambda 26$ hybridized with the 5' probe and the *Hox 1.3* box probe but not with the 3' probe, which suggested that this fragment did not contain the entire *Hox 1.3* gene. Both fragments were subcloned into the Bluescribe plasmid vector (Stratagene), and the relevant portions of each clone were sequenced on both strands by the dideoxy-chain termination sequencing technique (23). The two λ clones overlap by 400 bp in the 5' protein-encoding region of the *Hox 1.3* gene. DNA from these two clones was prepared by standard procedures (14); Southern analysis, using a *Hox 1.3* homeobox probe, identified a total of four homeobox homologies in these overlapping clones. Two of them were located in $\lambda 26$, and the other two were in $\lambda 32$. To further characterize this homeobox cluster, we probed Southern transfers with specific oligonucleotide probes containing sequences immediately flanking the 3' end of the homeoboxes of the murine genes *Hox 1.1*, *Hox 1.2*, *Hox 1.3*, and *Hox 1.4* (in the antisense orientation). The sequences of these oligonucleotide probes were, respectively:

Oligo 1.1 AACGGAGGGCACCGCTCTTCCGGGGCTGCAG
TGGGAGC
Oligo 1.2 TTTGGCCTCAGAGTCTTCCCCGCTGGCCTGCGT
Oligo 1.3 TCCTGCCGCGCCATGCTCATGCTTT
Oligo 1.4 CGAGGCAGTGTGGAAGATCGCATCTT

We also used the *Hox 1.1* pm6 probe, corresponding to the *PvuII-SacI* DNA fragment from clone pm6 (4) which con-

* Corresponding author.

† Present Address: Brookdale Center for Molecular Biology, Mount Sinai School of Medicine, New York, N.Y. 10029.

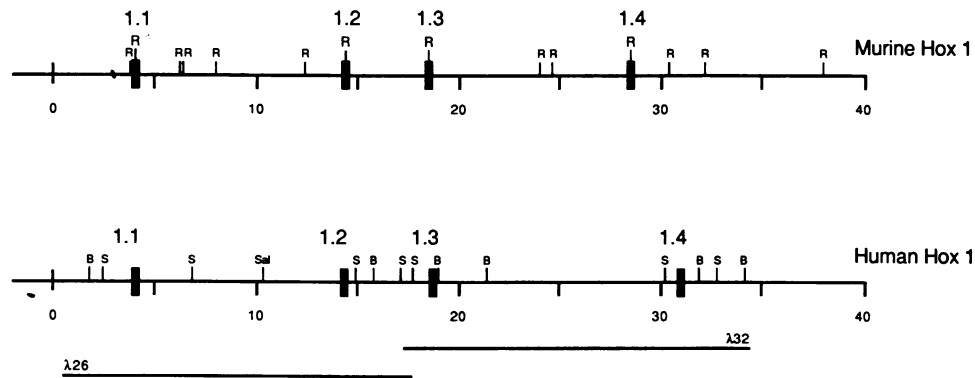


FIG. 1. Comparison of the human *Hox 1* cluster contained in the genomic clones $\lambda 26$ and $\lambda 32$ with the mouse cluster. The *EcoRI* restriction map of the murine cluster was compiled from references 4, 6, 15, and 21. The human restriction map of the *Hox 1* cluster was compiled by using the *BamHI* (B), *SacI* (S), and *Sall* (Sal) restriction enzymes. Black boxes represent homeoboxes. The two overlapping recombinant phage inserts ($\lambda 26$ and $\lambda 32$) are shown below the human *Hox 1* cluster restriction map.

tains sequences downstream of the *Hox 1.1* homeobox. These probes were chosen because of their uniqueness, and none contained homeobox sequences.

When using the oligonucleotide probes oligo 1.1, 1.2, 1.3, and 1.4, hybridizations were performed overnight at 55°C in 2× SSC (SSC is 0.15 NaCl plus 0.015 sodium citrate)–10× Denhardt solution–1% sodium dodecyl sulfate. Blots were washed at 55°C in 1× SSC–0.1% sodium dodecyl sulfate until the background was sufficiently reduced (15 to 30 min). For the *Hox 1.1* pm6 probe, hybridization was performed at 37°C in 43% formamide containing SSC, 1× Denhardt solution, and 1% sodium dodecyl sulfate. The blots were washed at 55°C in 0.2× SSC. Under these conditions, we detected two different restriction endonuclease fragments of $\lambda 26$ that hybridized strongly, one with the *Hox 1.1* and the other with the *Hox 1.2* probe. Two restriction endonuclease fragments from $\lambda 32$ gave a strong hybridization signal with the *Hox 1.3* homeobox probe, and one of them also hybridized with the *Hox 1.3* oligonucleotide probe (data not shown). The other fragment, which was positive with the homeobox probe, did not give a strong signal when probed with the *Hox 1.4* oligonucleotide, presumably because of sequence differences between the human and murine genes in the region 3' of the *Hox 1.4* homeobox.

From restriction mapping and Southern analyses of the $\lambda 26$ and $\lambda 32$ clones, we deduced the genomic map shown in Fig. 1, in which the human cluster is compared with the murine cluster. The murine *Hox 1.3* gene is a member of a cluster containing at least six homeobox genes on chromosome 6 (21). Here we show that the human *Hox 1* cluster, mapped by Rabin et al. (22) to human chromosome 7 (7p14 to p21), is organized in a very similar way, with cognate homeoboxes occupying the same positions relative to one another. Similarly, the *Hox 2* clusters of mice and humans also display conserved organization (1, 11). This conserved clustered organization is intriguing. It is possible that the order and spacing of homeobox genes in a cluster are conserved to maintain the complex transcriptional pattern of these genes. Northern (RNA) blot analysis of mRNA (prepared as described by Maniatis et al. [14]) from either cultured nonconfluent human embryonic foreskin fibroblasts or frozen autopsy specimens from a 56-year-old woman showed that the human gene is expressed in transcripts of several sizes (Fig. 2). We have previously shown that the murine *Hox 1.3* sequence is transcribed into at least four different polyadenylated transcripts of approximately 1.85,

4, 8, and 9 kbp. The abundance of these transcripts relative to one another varies from tissue to tissue (21). The human *Hox 1.3* gene is expressed in transcripts of equivalent sizes and, as in mice, most of the adult human tissues examined do not contain all four transcripts (Fig. 2). For instance, kidney contains a 1.85-kbp and a 4-kbp mRNA, whereas lung contains almost exclusively a 1.85-kbp mRNA. Interestingly, in human embryonic fibroblasts we cannot detect the 1.85- and 4-kbp transcripts but find the 8- and 9-kbp transcripts. We have also found that the 5' end of the murine 4-kbp transcript overlaps by over 1,100 bp with the 3' end of

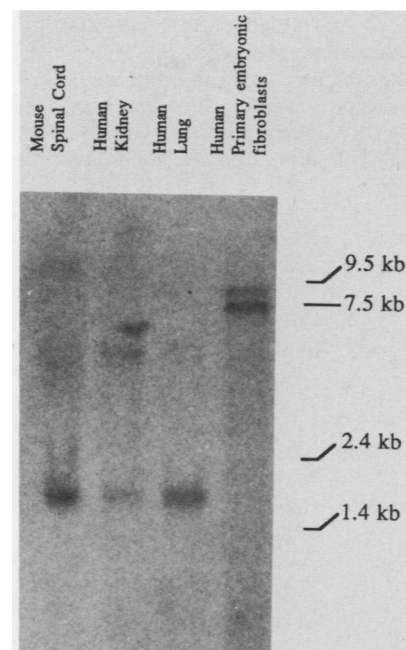


FIG. 2. Northern blot analysis of mouse and human *Hox 1.3* transcripts. Autoradiographic pattern of 5 μ g of total RNA from day 15 embryonic mouse spinal cord and 2 μ g of poly(A)⁺ selected mRNA from lung and kidney of a 56-year-old woman or from primary human embryonic fibroblasts. A *Sall*-*BglIII* restriction fragment (positions 116 to 524) from the human *Hox 1.3* gene, devoid of any homeobox sequence, was labeled and used as a probe. Positions of the bands from an RNA ladder (Bethesda Research Laboratories, Inc., Gaithersburg, Md.) are indicated on the right. Autoradiography exposure time was 10 days.

-248 ACCCCCATCGCCTCCACCCAACCTCCCTATTAGTGCACGAGTTTACCTCTAGAGGTCATCAGGCAGGATTTACGACTGGACAAACAAAAGCACGTGATTCC

-148 AAGTCGTACCCCATATTTGGGTGCCTACGTAGGAGGGAACCAAGTACATGTCCCAGTCATTTCCATAATTCATCATAAATTTGCAAGGGTGTCTATAGAC

-48 GCACAAAGCAGCCGGAGCCACAAATCAAGCACACATATCAAAAACAA ATG AGC TCT TAT TTT GTA AAC TCA TTT TGC GGT CGC TAT 13
Met Ser Ser Tyr Phe Val Asn Ser Phe Cys Gly Arg Tyr

40 CCA AAT GGC CCG GAC TAC CAG TTG CAT AAT TAT GGA GAT CAC AGT TCC GTG AGC GAG CAA TTC AGG GAC TCG GCG 38
Pro Asn Gly Pro Asp Tyr Gln Leu His Asn Tyr Gly Asp His Ser Ser Val Ser Glu Gln Phe Arg Asp Ser Ala

115 AGC ATG CAC TCC GGC AGG TAC GGC TAC GGC TAC AAT GGC ATG GAT CTC AGC GTC GGC CGC TCG GGC TCC GGC CAC 63
Ser Met His Ser Gly Arg Tyr Gly Tyr Gly Tyr Asn Gly Met Asp Leu Ser Val Gly Arg Ser Gly Ser Gly His

190 TTT GGC TCC GGA GAG CGC GCC CGC AGC TAC GCT GCC AGC GCC AGC GCG GCG CCC GCC GAG CCC AGG TAC AGC CAG 88
Phe Gly Ser Gly Glu Arg Ala Arg Ser Tyr Ala Ala Ser Ala Ser Ala Ala Pro Ala Glu Pro Arg Tyr Ser Gln

265 CCG GCC ACG TCC ACG CAC TCT CCT CAG CCC GAT CCG CTG CCC TGC TCC GCC GTG GCC CCC TCG CCC GGC AGC GAC 113
Pro Ala Thr Ser Thr His Ser Pro Gly Pro Asp Pro Leu Pro Cys Ser Ala Val Ala Pro Ser Pro Gly Ser Asp

340 ACG CAC CAC GGC GGG AAA AAC TCC CTA AGC AAC TCC AGC GGC GCC TCG GCC GAC GCC GGC AGC ACC CAC ATC AGC 138
Thr His His Gly Gly Lys Asn Ser Leu Ser Asn Ser Ser Gly Ala Ser Ala Asp Ala Gly Ser Thr His Ile Ser

415 AGC AGA GAG GGG GTT GGC ACG GCG TCC GGA GCC GAG GAG GAC GCC CCT GCC AGC AGC GAG CAG GCG AGT GCG CAG 163
Ser Arg Glu Gly Val Gly Thr Ala Ser Gly Ala Glu Glu Asp Ala Pro Ala Ser Ser Glu Gln Ala Ser Ala Gln

490 AGC GAG CCG AGC CCG GCG CCG CCC GCC CAA CCC CAG ATC TAC CCC TGG ATG CGC AAG CTG CAC ATA AGT CAT [▼]GGTA 187
Ser Glu Pro Ser Pro Ala Pro Pro Ala Gln Pro Gln Ile Tyr Pro Trp Met Arg Lys Leu His Ile Ser His

566 AAGCCAGCCTTTTCTAATCCACGGACGCGGGGACGGGCTCTCGTCCGCCCTCTCTTCTCTGCGCCCTCTCCAGTCTCTCTGGTCTCATTTC

666 TCCCAGCCCTGCGGAGCTCTCCTTCCGCTTCTCCCTCCTGCTTCTCCTCTCTTCTTTCGAGCACCCCTGCGCTCACAAATAGTGGGGAAA

766 TGGCGTCTCTGGACAGTTTAGACGTTGGAAGGGGGAAGCAAAACCCCTCTGGAACCCACGCCTTGGACCGCTCCCGGTGAGGGCAGCC

866 GAGCAAGGCGCAGAGAGGTAGGATGGCTGTAGCAGCGGTGAATCGGGCTGTGTCAGGGCGATAATTTATGAGGGGTACGCTGGGGAAACAGCGTT

966 ACTAATTACAGCCCCAAAAGGGGCTTGGGGAAAGAATCGAGGCCGAGAGCCTGCAGGATCTGAATTTGGGGGAGGAGGAGAGAGAAAGGG

1066 AAGABABAGAAAACAGGCTCCCAACCCCTGCAAGGCTGAAACGGGACGGCGCTCTCGGGGTGGAACCTTGTAGGGAGGGTGCACCGAAGGCCACTTGGGC

1166 GCTCAGAAAGGGCCTTGTCTCTGGGTTTCTGTGCGGTGGGCAGCTGGGAGGGCTGTGCTCCCGATCGGGGCGCCCGGGGAGGGCGGGGCGAG

1266 GAGAGGGGCCAGGGAAGCCGAGTCCGCCGGGACACCGCCACGCTCAGATGGGCAGATGTGTTCCAGGGTCCAAATCGTATTGTTTTCTTCTAGAA

1366 ACGAAGAGAGAAGGAATTCGGGAGGGGTGTGGCGCTGGTGAACGAACTTGTGTAGCTTTTCCGCTGGGTCCCTGTCTATGACCCAAAGCTTGTCCCC

1466 CTGGCGGACTTTGGAAGACAGGAGTGGTGGCTAAACCGCTGACTTTTCTATTGCAG [▼]AC AAC ATA GGC GGC CCG GAA GGC AAA AGG GCC 197
Asp Asn Ile Gly Gly Pro Glu Gly Lys Arg Ala

1555 CGG ACG GCC TAC ACG CGC TAC CAG ACC CTG GAG CTG GAG AAG GAG TTC CAC TTC AAC CGT TAC CTG ACC CGC AGA 223
Arg Thr Ala Tyr Thr Arg Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe His Phe Asn Arg Tyr Leu Thr Arg Arg

1630 AGG AGG ATT GAA ATA GCA CAT GCT CTT TGC CTC TCC GAG AGA CAA ATT AAA ATC TGG TTC CAA AAC CGG AGA ATG 248
Arg Arg Ile Glu Ile Ala His Ala Leu Cys Leu Ser Glu Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met

1705 AAG TGG AAA AAA GAT AAT AAG CTG AAA AGC ATG AGC ATG GCC GCG GCA GGA GGG GCC TTC CGT CCC TGAGTATCTGAG 270
Lys Trp Lys Lys Asp Asn Lys Leu Lys Ser Met Ser Met Ala Ala Ala Gly Gly Ala Phe Arg Pro **

1783 CGTTTAAAGTACTGAGCAGTATTAGCGGATCCCGGTAGTGTGACTACTAAGGTGACTTTCTGAACTCCCTTGTGTTCTCTGTGAAGAAGCCCTGTT

1883 CTCGTGCCCTAATTCATCTTTTAAATCATGACCTGTTTATGOCATTATAGCGCCTGTATAAGTAGACTGCTTTCTGTTCATCTTTGTCTCTGAATG

1983 GCTTTGTCTTGAAAAAATAGATGTTTTAACTTATTTATATGAAAGCAAGCTGTGTTACTTGAAGTAACATAACAAAAAAGAAAAGAGAAAAA

2083 CACACAAAAGTCCCCTTCAATCTCGTTAGTGCCAATGTTGTGTGTGCACTCAAGTGTGTTAACTGTGCATGTGCGTGGAGTGTCTCTGTCTCAAT

2183 AGCTCCAAGCTGTAAAGATATTTTATTCAAACCTAATATTCCTGTGTAATTAAGCTGTGTGAGGGTACTTGTGATGAGACAACTGTGTCGAC

2283 GTGTAGTGACTAGTGACTCTGTGATGAAAAGTGTGACTCCAAGCGGTGTGCTCCCTGCGTCCCTTATAGGACCCCTTGACCAACTCTGGAAAGTGCTCT

2383 TATAAGCGCAGCTTCAGTGATGTATGTTTTGTGAACAAAGTTACAAATATTGTCCAAGTCTGGCTGTTTTAAGCAAAGCTGTGATCAGCTTTTTTTTTT

2483 TTTTTTTTTTGTATTGTTTTTAAAGAAAAATACTGACTGGAACAAA AATAA TTTCTATTGTAAGTTCTCTTGTGCTGGTTTGTGCCAAATAGTG

2583 AGCGGCTCTGTCTGTTTTCTGTCTGTCTGCAAGTCTGTGAAAGCTGTTGGGTCTGAGGCTAAGTGCAGATGACCTGTGACGGGAGACCTCATACCAA

2683 CACTGTCCCATCGCTTCCCTACCTCTGACCCATTGCAAAGTTCAGGCATAAGTGGAAAAAGCTGTAGGCTGTTCAAAGCCCAAGACCTGTCCAT

2783 CTCTGAGGAACCAAGTTAACTTGCTGGGTACAAAAAGAGAGAAGAGC

FIG. 3. Primary structure of the human *Hox 1.3* gene. Numbering begins at the predicted translation start codon (nucleotide numbering is on the left; amino acid numbering is on the right). The seven differences in the predicted amino acid sequence between the human and the murine genes are circled. Boundaries of the predicted intron (defined by identity with the murine gene) are indicated by arrowheads. Stretches of 9 bp or longer that are identical in the two introns are underlined. The homeobox is underlined with a solid line; a consensus polyadenylation signal is boxed.

the *Hox 1.2* transcript (J. Garbern, unpublished data). The similarity in size and sequence of human *Hox 1.2* and *1.3* transcripts suggests that a similar situation exists in humans as well.

Comparison of the human and murine *Hox 1.3* primary structures. Sequences of the murine and human *Hox 1.3* genes were analyzed by using the software package of the University of Wisconsin Genetics Computer Group (developed by J. Devereux et al.) on a VAX computer. The sequence of the human *Hox 1.3* gene (Fig. 3) is very similar to the murine *Hox 1.3* sequence (9, 21). This similarity allowed us to assign the putative exon boundaries and to predict that the human *Hox 1.3*, like its murine counterpart, is organized into two exons. The degrees of similarity between the two genes are 98% in the 250 bp upstream the start codon, 94% in the coding region, 72% in the intron, and 90% in the 3' untranslated region. The predicted proteins encoded by the murine and human genes differ in only 7 of the 270 residues. These seven residues lie in the central region of the protein and are encoded by exon 1 (Fig. 3). A similar degree of conservation has been seen between the human and mouse *Hox 2.3* genes (18, 26) and the human and mouse *Hox 5.1* genes (8, 16). Thus, it appears that the cognate proteins are highly similar not only in the homeo-domain but also in the regions outside this domain. In contrast, related proteins of the *Hox 1* and *Hox 2* clusters from the same species do not show such a high degree of similarity.

The high degree of homology between the coding regions of the murine and the human *Hox 1.3* gene extends to the 5' noncoding regions, where the two genes are 98% identical in the 250 bp upstream of the predicted initiation codon. The murine *Hox 1.3* gene has a potential 5' extension of the reading frame of 333 bp (Garbern et al., unpublished data). We have not fully sequenced the human gene across the entire corresponding region, but there are only five nucleotide differences within the first 250 bp upstream of the designated initiation codon, and the reading frame remains open. The five nucleotide changes occur at positions -243, -242, -241, -221, and -107 and would result in four amino acid changes as compared with the murine sequence. Thus, in both human and murine sequences, an open reading frame extends upstream of the first ATG and the 5' end of the major 1.85-kbp murine transcript. This highly conserved region is transcribed as part of a 4-kbp transcript in mouse spinal cord and may encode another form of the *Hox 1.3* protein.

Conservation has also been observed in the 5' noncoding regions of the murine *Hox 2.3* gene and its human counterpart, where the 100 bp upstream of the initiation codon are 97% identical (18). However, these upstream regions do not have extended open reading frames. Alternatively the 5' noncoding regions in the human and murine *Hox 1.3* genes may have been conserved because they are *cis* regulatory elements for the 1.85-kbp transcript expressed in lung and other tissues. Indications for such a role come from the following observations. DNase I protection and electrophoretic mobility shift assays experiments (preparation of the *Hox 1.3* protein and the method used for gel shift are described below) with the recombinant *Hox 1.3* protein identify a protected region between positions -157 and -142, which lie upstream from the cap site of the 1.85-kbp transcript. Thirty-base-pair oligonucleotides containing this region or its human counterpart bind to the *Hox 1.3* protein (20). The homologous region in the human cognate is identical to the murine sequence except for a T instead of a C at

position -221 (Fig. 3). This single base-pair change does not severely influence binding to the *Hox 1.3* protein. It is possible that these binding sites are necessary *cis* elements for regulation of the *Hox 1.3* gene by the *Hox 1.3* protein. Such autoregulation has been shown to occur in a *Drosophila* homeobox gene, *fushi tarazu*, whose protein product is required for the functioning of its own transcriptional enhancer (12).

The 3' untranslated sequences retain 90% similarity with the corresponding murine sequence. Differences between the sequences of the two species are due to short insertions or deletions between long stretches of perfectly conserved sequences. The high similarity observed in the 3' untranslated regions of the human and murine *Hox 1.3* genes is unusual. The *Hox 2.3* murine gene and its human counterpart, both highly homologous in coding sequences and 5' noncoding regions, display only 67% homology in the 3' untranslated regions. The high degree of similarity between the human and murine *Hox 1.3* cognates in the 3' regions drops to 35% 30 bp downstream to the polyadenylation signal (data not shown). This suggests that the region between translation termination and the polyadenylation signal is functionally important. Such functional importance of the 3' untranslated regions, particularly in influencing mRNA half-lives, has recently been demonstrated for other messages (for review, see reference 2). The 3' untranslated sequences retains 90% similarity with the corresponding murine sequence. Differences between the sequences of the two species are due to short insertions or deletions between long stretches of perfectly conserved sequences.

Comparison of the two introns. The similarity between the human and murine *Hox 1.3* introns is highlighted in Fig. 3 and 4. The human gene contains consensus splice donor and acceptor sites at the same positions as in the mouse gene. Both introns are predicted to be 960 bp. Although the intron sequences of the murine and human *Hox 1.3* genes are more divergent than the coding sequences, the similarity is still 72%, and there are many regions of perfect conservation. In Fig. 3, we have underlined stretches of nine bases or more that are identical to the murine sequences. One conserved sequence is 111 bp long (bp 888 to 998; Fig. 3) and displays 90% similarity between the two species. It contains two consensus binding motifs for the *Hox 1.3* protein. In addition, this region contains a sequence with 59% similarity to positions 1179 to 1227 of the epsilon region of the yeast Ty1 transposon (7; Fig. 5).

To examine whether these intron sequences could bind the *Hox 1.3* protein, we used a mobility shift assay with nuclear extracts containing recombinant *Hox 1.3* protein. To obtain sufficient amounts of the *Hox 1.3* protein, we have used the baculovirus expression system which, in addition to supporting high levels of protein expression, has been shown to perform correct posttranslational modifications of nuclear proteins (19, 20). The *Hox 1.3* murine protein-encoding cDNA sequence (present in the major 1.85-kbp transcript) was inserted behind the polyhedrin promoter of the recombinant baculovirus and used to infect *Spodoptera frugiperda* cells (for details concerning shuttle vector construction and expression, see reference 20). Nuclear extracts were prepared according to the protocol of Dignam et al. (5) from *Hox 1.3* recombinant virus-infected cells. The mobility shift assay was performed as follows. Four sets of 30-bp oligonucleotides, two corresponding to the human potential binding sites (Hu intron 1 and Hu intron 2) and two corresponding to the murine sites (Mu intron 1 and Mu intron 2) were prepared. Short duplex DNA fragments with 5' overhanging

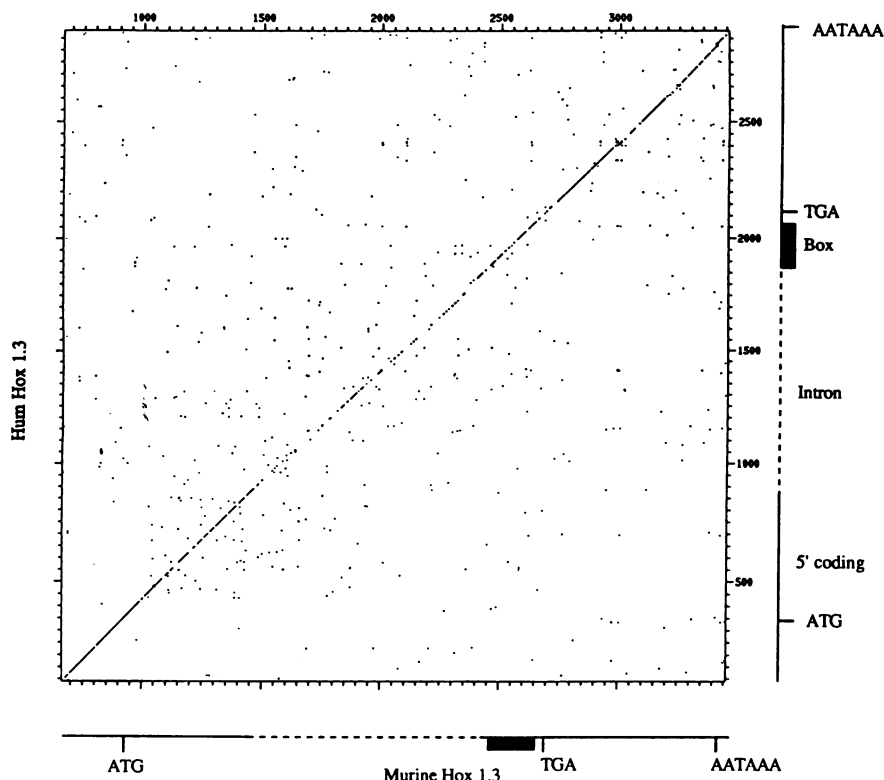


FIG. 4. Dot matrix comparison of the human *Hox 1.3* gene and its murine counterpart. The comparison was performed by using a k-tup (number of nucleotides compared at a time) of 7. The predicted translation initiation codons, termination codons, and polyadenylation signals are indicated as ATG, TGA, and AATAAA, respectively. Black boxes represent homeoboxes; broken lines indicate the respective introns.

ends were prepared by hybridizing two 27-base-long complementary oligonucleotides. The DNA fragments were end labeled by filling in the 5' protruding ends with Klenow enzyme to create 30-bp blunt-ended double-stranded fragments containing three ³²P-labeled dCTP nucleotides. The plus-sense sequences of the DNA fragments used in this study were

Mu *Hox 1.3* TGCCAACTCCCCATTAGTGCACGAGACCT
 Hu intron 1 TTGTCACGGCGGATAATTTATGAGGAGCCA
 Mu intron 1 TTGTCAGGGGAGATAATTTATGGGAAGCCA
 Hu intron 2 AGGACAGCGTTACTAATTACAGCCCCACT
 Mu intron 2 AGGACAGCGTTACTAATTAGAGCCCCACT

Mu *Hox 1.3* contains the *Hox 1.3* binding site located in the murine *Hox 1.3* promoter region (20). Hu intron 1 and Hu intron 2 contain the two potential *Hox 1.3* binding sites within the human *Hox 1.3* gene. Mu intron 1 and Mu intron 2 are almost identical to Hu intron 1 and Hu intron 2, respectively, and contain homologous sequences in the murine intron. Protein-DNA complexes were resolved in low-ionic-strength 9.5% polyacrylamide gels according to Carthew et al. (3), with the following modifications. For binding reactions, labeled fragments (20,000 cpm) were incubated for 2 h at 4°C with 2 μl of *Hox 1.3* protein-containing extracts (~4 μg of total protein) and 4 μg of poly(dI-dC) in 200 mM KCl binding buffer (final volume, 20 μl). The reaction mixtures were loaded on a 9.5% polyacrylamide gel in 1× TAE (0.04 M Tris acetate plus 0.002 M EDTA) buffer and electrophoresed at 15 V/cm for 240 min. The gel was dried, and radioactivity was visualized by autoradiography.

All four oligonucleotides were able to bind the *Hox 1.3*

protein (Fig. 6). For both the human and the mouse sequences, the site containing the TAATTA sequence showed stronger binding. Multiple *Hox 1.3* protein-DNA complexes could be resolved in 9.5% native polyacrylamide gels (Fig. 6), as we had previously observed with the putative murine *Hox 1.3*-binding site. These multiple complexes may be due to binding of the multiple forms of the *Hox 1.3* protein (20). The highly conserved sequences in the noncoding regions may be critical to various transcriptional or posttranscriptional processes common to the human and murine genes. In this regard, interesting conservation of noncoding sequences has also been observed between the *engrailed* genes of two *Drosophila* species, *melanogaster* and *virilis*, which diverged about 60 million years ago, at about the time of the rodent-primate divergence. Kassis et al. (13) have found blocks of sequence homology in several areas, including the introns, which flank the exons of the *engrailed* genes of these two species. Recently, J. A. Kassis (personal communication) has shown that the 5' untranscribed sequences alone are not sufficient for directing reporter gene expression in embryos in an *engrailed*-like pattern. However, the 5' region together with sequences from the *engrailed* first intron of *D. virilis* does direct reporter gene expression in *engrailed*-like

955	GAAACAGCGTTACTAATTACAGCC	978	Hu <u><i>Hox 1.3</i></u>
1206	*****G****	1229	Mu <u><i>Hox 1.3</i></u>
1204	*****AAAC*****GA	1227	Ty1

FIG. 5. Homology of the human and murine *Hox 1.3* introns with the Ty1 activation region. Sequences between the Ty1 activation region (positions 1204 to 1227; 7) and the human and mouse *Hox 1.3* introns are aligned. *, Nucleotide identity. DNA core recognition for the *Hox 1.3* protein is underlined.

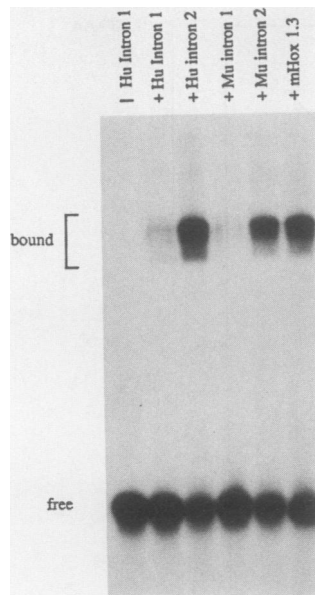


FIG. 6. *Hox 1.3* binding sites within the intron of the human and murine *Hox 1.3* genes. Shown are results of a gel retardation assay of *Hox 1.3* protein-DNA complexes. Double-stranded end-labeled oligonucleotide probes, each 30 bp long and corresponding to sequences located within the promoter region (mHox 1.3) or within the murine and human *Hox 1.3* introns, were incubated without (–) or with (+) 4 μ g of nuclear extract containing *Hox 1.3* protein. The complexes were resolved on a 9.5% nondenaturing polyacrylamide gel and visualized by autoradiography of the dried gel.

fashion in *D. melanogaster*. This implies that intron sequences may influence the tissue-specific expression of the *engrailed* gene and that these *cis*-active sequences are conserved between distantly related species. The murine and human *Hox 1.3* introns are even more closely related than are the *engrailed* introns of the two *Drosophila* species, tempting speculation that regulatory sequences also exist in the *Hox 1.3* intron.

We are grateful to Heinz Arnheiter for helpful discussions and critical reading of the manuscript.

E.T.L. was, during the first part of this work, a Fondation pour la Recherche Medicale grantee and later was partially funded by the NATO fellowship program.

LITERATURE CITED

1. Acampora, D., M. Pannese, M. D'Esposito, A. Simeone, and E. Boncinelli. 1987. Human homeobox-containing genes in development. *Hum. Reprod.* 2:407–414.
2. Birnstiel, M. L., M. Busslinger, and K. Strub. 1985. Transcription termination and 3' processing: the end is in site! *Cell* 41:349–359.
3. Carthew, R. W., L. A. Chodos, and P. A. Sharp. 1985. An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. *Cell* 43:439–448.
4. Colberg-Poley, A. M., S. D. Voss, K. Showdhury, C. L. Stewart, E. F. Wagner, and P. Gruss. 1985. Structural analysis of murine genes containing homeobox sequences and their expression in embryonal carcinoma cells. *Cell* 43:39–45.
5. Dignam, J. D., R. M. Lebowitz, and R. G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* 11:1475–1489.
6. Duboule, D., A. Baron, P. Mähl, and B. Galliot. 1986. A new homeobox is present in overlapping cosmid clones which define the mouse *Hox-1* locus. *EMBO J.* 5:1973–1980.
7. Errede, B., M. Company, J. D. Ferchak, C. A. Hutchison, and W. S. Yarnell. 1985. Activation regions in a yeast transposon have homology to mating type control sequences and mammalian enhancers. *Proc. Natl. Acad. Sci. USA* 82:5423–5427.
8. Featherstone, M. S., A. Baron, S. J. Gaunt, M. G. Mattei, and D. Duboule. 1988. *Hox 5.1* defines a homeobox-containing gene locus on mouse chromosome 2. *Proc. Natl. Acad. Sci. USA* 85:4760–4764.
9. Fibi, M., B. Zink, M. Kessel, A. M. Colberg-Poley, S. Labeit, H. Lehrach, and P. Gruss. 1988. Coding sequence and expression of the homeobox gene *Hox 1.3*. *Development* 102:349–359.
10. Gehring, W. 1987. Homeoboxes in the study of development. *Science* 236:1245–1252.
11. Hart, C. P., A. Fainsod, and F. Ruddle. 1987. Sequence analysis of the murine *Hox 2.2*, *2.3* and *2.4* homeoboxes: evolutionary and structural comparisons. *Genomics* 1:182–195.
12. Hiromi, Y., and W. J. Gehring. 1987. Regulation and function of the *Drosophila* segmentation gene *fushi tarazu*. *Cell* 50:963–974.
13. Kassis, J. A., S. J. Poole, D. K. Wright, and P. H. O'Farrell. 1986. Sequence conservation in the protein coding and intron regions of the engrailed transcription unit. *EMBO J.* 5:3583–3589.
14. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. Martin, G. R., E. Boncinelli, D. Duboule, P. Gruss, R. Jackson, R. Krumlauf, P. Lonai, W. McGinnis, F. Ruddle, and D. Wolgemuth. 1987. Nomenclature for homeobox-containing genes. *Nature (London)* 325:21–22.
16. Mavilio, F., A. Simeone, A. Giampaolo, A. Faiella, V. Zappavigna, D. Acampora, G. Poiana, G. Russo, C. Peschle, and E. Boncinelli. 1986. Differential and stage related expression in embryonic tissues of a new human homeobox gene. *Nature (London)* 324:664–668.
17. McGinnis, W., M. S. Levine, E. Hafen, A. Kuroiwa, and W. S. Gehring. 1984. A conserved DNA sequence in homeotic genes of the *Drosophila* Antennapedia and Bithorax complexes. *Nature (London)* 308:428–433.
18. Meijlink, F., R. Delaaf, P. Verrijzer, O. Destree, V. Kroezen, J. Hilkens, and J. Deschamps. 1987. A mouse homeobox containing gene on chromosome 11: sequence and tissue-specific expression. *Nucleic Acids Res.* 15:6773–6786.
19. Miyamoto, C., G. E. Smith, J. Farrell-Towt, R. Chizzonite, M. D. Summers, and G. Ju. 1985. Production of human *c-myc* protein in insect cells infected with a baculovirus expression vector. *Mol. Cell. Biol.* 5:2860–2865.
20. Odenwald, W. F., J. Garbern, H. Arnheiter, E. Tournier-Lasserre, and R. A. Lazzarini. 1989. The *Hox 1.3* homeobox protein is a sequence specific DNA-binding phosphoprotein. *Genes Dev.* 3:158–172.
21. Odenwald, W. F., C. F. Taylor, F. S. Palmer-Hill, V. Friedrich, M. Tani, and R. A. Lazzarini. 1987. Expression of a homeodomain protein in non contact-inhibited cultured cells and postmitotic neurons. *Genes Dev.* 1:482–496.
22. Rabin, M., A. Ferguson-Smith, C. Hart, and F. Ruddle. 1986. Cognate homeo-box loci mapped on homologous human and mouse chromosomes. *Proc. Natl. Acad. Sci. USA* 83:9104–9108.
23. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463–5467.
24. Scott, M., and A. Weiner. 1984. Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax and *fushi tarazu* loci of *Drosophila*. *Proc. Natl. Acad. Sci. USA* 81:4119–4125.
25. Scott, M. P., and S. B. Carroll. 1987. The segmentation and homeotic gene network in early *Drosophila* development. *Cell* 51:689–698.
26. Simeone, A., F. Mavilio, D. Acampora, A. Giampaolo, A. Faiella, V. Zappavigna, M. D'Esposito, M. Pannese, G. Russo, E. Boncinelli, and C. Peschle. 1987. Two human homeobox genes *c1* and *c8*: structure analysis and expression in embryonic development. *Proc. Natl. Acad. Sci. USA* 84:4914–4918.