Alternative splicing of the HOX 2.2 homeobox gene in human hematopoietic cells and murine embryonic and adult tissues

Wei-fang Shen, Kristina Detmer, Teresa A.Simonitch-Eason, H.Jeffrey Lawrence and Corey Largman*

Martinez VA Medical Center, 150 Muir Rd, Martinez, CA 94553 and Department of Internal Medicine, UC Davis School of Medicine, Davis, CA, USA

Received October 22, 1990; Revised and Accepted December 31, 1990

ABSTRACT

The HOX 2.2 homeobox gene is expressed in human hematopoietic cell lines with erythroid features (W.-F. Shen, et al, *Proc. Natl. Acad. Sci.* 86, 8536–8540, 1989). Both human and murine Hox 2.2 genes contain a single 1 kb intron which interrupts the sequence encoding the proposed homeobox protein. Four human erythroleukemia cell lines express the spliced, homeobox-coding transcript as the major form of message, and variable low amounts of unspliced HOX 2.2 mRNAs. Murine embryonic tissues and adult kidney and uterus contain approximately equal amounts of transcripts containing this intron and mRNAs from which the intron has been excised. The spliced transcript encodes a 224 amino acid homeobox protein, while the unspliced transcript would potentially encode a 140 residue protein containing the same N-terminal sequence but lacking the homeodomain.

INTRODUCTION

Homeobox-containing genes are transcription factors which regulate programs of embryonal development (1, 2) and cellular differentiation (3–8). The Hox 2.2 gene is expressed in spatially and temporally specific patterns during murine spinal cord development (9). In terms of cellular differentiation, we have previously reported that the HOX 2.2 gene is expressed in an erythroid specific pattern in a survey of human hematopoietic cell lines (10). We have recently demonstrated that there is a marked decrease in HOX 2.2 mRNA in an erythropoietin responsive human leukemia cell line following hormone-mediated erythroid differentiation (Shen et al, submitted). Furthermore, we have also demonstrated that over-expression of a HOX 2.2 coding mRNA alters the phenotype of the erythroleukemia cell line K562 (Shen et al, submitted). These studies suggest that this homeobox protein may play a role in determination of the erythroid phenotype.

In order to investigate the role(s) of HOX 2.2 homeobox gene expression in red cell differentiation and murine development,

we have characterized these genes and their transcription products in four human erythroid cell lines, as well as in murine embryonic and adult tissues. Since many homeobox genes are alternatively spliced to form transcripts either containing or missing the homeobox domain (11–14), we have also characterized the splicing patterns of the HOX 2.2 gene products. Comparison of the human and murine data demonstrate that, while the relative amounts of alternatively spliced products vary considerably, the same splice sites are utilized in all human erythroleukemia cell lines as well as in murine adult kidney and embryonic tissues. RNase protection and cDNA cloning show the presence of two HOX 2.2 transcripts—a spliced mRNA which encodes a 224 amino acid homeobox protein, and an unspliced message which would potentially yield a 140 amino acid protein lacking the homeodomain.

MATERIALS AND METHODS

Human Cell Lines

HEL (15), K562 (16), OCIM1 (17), and MB02 (18) are human leukemic cell lines with erythroid features. The cells were propagated in RPMI medium or IMDM medium (OCIM1) with 10% fetal calf serum or 10% human serum plus GM-CSF (MB02), at 37° in a 5% CO₂ incubator.

cDNA Library Construction And Screening

Poly A⁺ RNA was isolated by the guanidinium thiocyanate procedure (19), followed by chromatography on oligo dT cellulose. cDNA libraries for HEL cells (15) were prepared from poly A⁺ RNA by the method of Gubler and Hofman (20). A series of human genomic HOX 2.2 sub-clones (21); a murine 12 day embryo λGT-11 cDNA library; and a murine genomic cosmid library were all gifts from C. Hauser and R. Tjian, UC Berkeley.

Human HOX 2.2 Cloning. The human HOX 2.2 genomic sequence was determined from fragments originally isolated by Levine et al (21). cDNA libraries were constructed using either

* To whom correspondence should be addressed at Biochemistry Research, Martinez VA Medical Center, 150 Muir Road, Martinez, CA 94553, USA

oligo dT primers, or in some cases specific human HOX 2.2 primers (nt 1917 to 1901, antisense; nt 2508 to 2491, antisense; and nt 1811 to 1791, antisense). In attempts to enhance the synthesis of 5' regions of human HOX 2.2 cDNAs, first strand cDNAs were synthesized in the presence of methyl mercury. Initial cDNA library screenings were performed with a 2.1 kb EcoRI HOX 2.2 homeobox-containing genomic probe at 42°C, in 50% formamide (22). Subsequent screens were performed with gel purified DNA fragments as shown in Fig. 3. All probes were labeled with the random primer method (23). Positive clones were purified and subcloned into M13 vectors for dideoxy sequence analysis using the ³⁵S modified procedure (24). A partial human cDNA clone was obtained from the oligo-dT primed library which consisted of bp 710–722 and 1783–2700 and which thus lacked the intron which spans bp 723–1782 (Figure 3). PCR amplification of HEL cell mRNA was used to generate cDNAs lacking the intron but containing all of the 5' protein coding region and overlapping with HEL-7 (Figure 3). cDNAs were prepared from HEL cells using methyl mercury (22) and were amplified with the following sets of overlapping primers:

Primer 1: AAGCTTCCTCCCAATGAGTTCCT (nt 309–328)
 Primer 2: CTGCATCCACGGGTAGA (nt 701–685, antisense)
 Primer 3: GAATTCGACAAGAGCGTGTTCGG (nt 646–662)
 Primer 4: AAGCTTACAGGGCGTGCGCGATC
 (nt 1917–1901, antisense)

Murine Hox 2.2 Cloning. A 2.1 kb human HOX 2.2 Eco RI genomic fragment was used to screen a murine 12 day embryonic cDNA library at reduced stringency (42°C, 40% formamide). One class of the resulting cDNAs appeared to represent partial clones lacking their respective 5' regions. PCR amplification was used to confirm the presence of mRNAs which contain the 5' protein coding region contiguous with the adjacent intron, using the following primers:

Primer 5: ATGCTCTAGACCTCCCAATGAGTTCCT
 (nt 216–232)
 Primer 6: CGGGAAGTCTCACTCACTGT (nt 655–636, antisense)

Figure 1. Northern Gel Analysis of HOX 2.2 Transcripts. A. Expression of HOX 2.2 transcripts in mRNA isolated from human HEL cells (5 µg). Lane 1 was probed with a specific homeobox probe (Fig. 3, probe B). Lane 2 was probed with the cDNA 5' flanking region (Fig. 3, probe A). B. Expression of murine Hox 2.2 in adult kidney tissue mRNA (5 µg). The blot was probed with a partial cDNA clone containing the homeobox and 3' flanking region. Hybridization to an actin control probe is shown below lanes 1 and 2.

RNase Protection Analysis

In order to prevent contamination of cytoplasmic mRNA with unprocessed heteronuclear RNA, cytoplasm was routinely separated from nuclei prior to RNA isolation. Human leukemia cells were lysed in NP40 buffer, while murine embryonic tissues were disrupted by Dounce homogenization in the presence of NP40 (25). Cytoplasmic RNA samples were prepared by separation from nuclei by sucrose pad centrifugation followed by proteinase K treatment and chromatography on oligo dT cellulose (19, 25). DNA fragments shown in Fig. 3 were subcloned into Bluescript vectors and RNA probes synthesized with ³²P UTP as described (25). Probes were purified by electrophoresis in 4% polyacrylamide/urea gels, hybridized to poly A+ RNA at 65°C, digested with RNase A and T1, and the protected bands electrophoresed in 4% polyacrylamide/urea gels as described (25). Each protection experiment was repeated with at least two separate preparations of cytoplasmic mRNA from each cell line.

Northern Gel Analysis

RNA samples were prepared by the guanidium thiocyanate procedure (19). mRNAs were fractionated on agarose/formaldehyde gels, transferred to Gene screen, cross-linked by UV irradiation, and probed as previously described (10). Filters were probed sequentially with double-stranded probes prepared by the random primer method (23), or with single stranded probes prepared by primer extension of cDNA fragments cloned into M13.

RESULTS

HOX 2.2 Expression In Human Erythroleukemia Cell Lines and in Murine Embryonic and Adult Tissues

Northern gel analysis using a genomic probe containing the entire HOX 2.2 region (21) shows expression in HEL cells as a single broad band centered on 1.4 kb (10). A probe containing only the homeobox (B, Fig. 3) hybridizes to an identical broad band (Fig. 1A, lane 1). A probe containing only the 5' coding region (probe A, Fig. 3), reveals a tighter band at 1.4 kb (Fig. 1A, lane 2). In contrast to HEL cells, murine Hox 2.2 is expressed in

Figure 2. Expression of Hox 2.2 in Murine Embryonic and Adult Tissues. Northern gel analysis of mRNA isolated from mouse embryo tissues: day 10 (lane 1), day 12 (lanes 2 and 3), and day 16 (lane 4), adult kidney (lanes 5 and 6), spinal cord (lane 7), and adult uterus (lanes 8 and 14). Adult heart (lane 9), spleen (lane 10), liver (lane 11), brain (lane 12), and testes (lane 15) were negative for Hox 2.2 mRNA. Filters were probed with a partial murine Hox 2.2 cDNA clone which contains the homeobox and 3' untranslated region. The bottom panel shows hybridization to an actin control probe.

adult kidney, uterus and spinal cord and in embryonic tissues as two distinct transcripts of 1.6 and 2.3 kb, respectively (Fig. 1B and Fig. 2). The 2.3 kb band is barely detectable in adult murine spinal cord, while heart, pancreas, brain, spleen, liver, and testis are all negative for Hox 2.2 expression. Northern gel analysis of the human erythroleukemia line OC1MI reveals a pattern of 1.4 and 2.2 kb mRNA bands similar to that seen for murine tissues. MB02 cells show a single 1.4 kb band of HOX 2.2 mRNA, while no signal is detected in K562 cells by northern gel analysis (10), but can be detected by RNase protection (Figure 5).

Sequence of the Human HOX 2.2 Gene

A series of overlapping clones spanning the human HOX 2.2 gene (21) were analyzed to yield a 2.8 kb genomic sequence

Figure 3. Structure of the Human and Murine HOX 2.2 Genes and cDNAs. **A.** Organization of the human HOX 2.2 gene and cDNAs. The drawing labeled genomic corresponds to the sequenced region of the human HOX 2.2 gene reported in Figure 4. The homeobox is filled, while the stippled boxes represent protein coding regions. HEL represents a composite cDNA structure consisting of the partial cDNAs isolated from an HEL cell library extending from the poly A tail to the R1 site immediately 5' to the intron splice site (e.g. clone HEL-7) combined with cDNA clones established by PCR amplification of two overlapping sets of oligonucleotide primers (arrows 1 and 2; 3 and 4). Fragments D and E represent intron donor probes and fragment C represents the intron acceptor probe used to detect alternative splicing by RNase protection. Fragments A and B were used for northern gel probes. **B.** Organization of the murine Hox 2.2 gene and cDNAs. The drawing labeled genomic corresponds to the region of the murine Hox 2.2 gene reported in Figure 4. The hatched areas correspond to the regions of the cDNA which would encode protein, while the dimpled box represents an additional 5' transcribed region. The striped box represents the 'intron' which is present in alternatively spliced transcripts, while the closed box represents the homeobox. Two types of cDNA clones are depicted—those which lack an intron sequence (WF-10), and clones containing the apparent intron sequence (WF-14). Probes a and b were used for RNase protection experiments described in Figure 7. Arrows 5 and 6 indicate PCR primers used for generation of cDNA clones containing the 5' coding region and the intron. These clones and the partial cDNA clones (WF-14), together with RNase protection data were combined to form a composite intron-containing transcript depicted as Hox 2.2 IC. The dashed 5' portion of Hox 2.2 IC represents proposed transcription initiation at the same site as was used in WF-10.

(Figs. 3A and 4). The presence of stop codons in the homeobox reading frame and of consensus splice junctions, as well as comparison with cDNA clones suggest that there is a 1 kb intron immediately upstream from the homeobox, similar to those observed for other homeobox genes (26). Examination of the HOX 2.2 genomic sequence also revealed several consensus TATA and CCAAT boxes as well as a consensus PU binding site (GAGGAA) (27) upstream from a proposed AUG translation start site (nt 315).

Isolation Of HOX 2.2 cDNAs From Human Erythroleukemia Cells (HEL)

Screening an HEL cell cDNA library with a HOX 2.2 genomic probe yielded a series of partial 3' homeobox-containing clones. The longest clone (HEL-7) appears to be missing approximately 600 bp of 5' protein coding sequence when compared to the human genomic sequence or mouse Hox 2.2 cDNAs (Figure 3B). This clone did, however, contain sequences 5' to the splice junction of the 1 kb intron. Although a genomic probe representing the 5' region missing in these cDNAs hybridizes strongly to HEL mRNA (Fig. 1A, lane 2), attempts to produce 5' enriched cDNA libraries using specific primers did not yield additional clones. We employed a series of oligonucleotide primers designed from the genomic sequence to complete the proposed full-length cDNA, using polymerase chain amplification (PCR) to clone overlapping regions of HEL mRNA. These PCR derived cDNA clones were combined with the HEL-7 partial cDNA clone to yield a 1400 bp composite cDNA corresponding to the major HOX 2.2. coding transcript (Figure 3, HEL).

Expression of Alternatively Spliced mRNAs In Human Erythroid Cell Lines

A series of genomic probes have been used in RNase protection experiments to investigate the HOX 2.2 transcripts in cytoplasmic mRNAs isolated from four human erythroleukemia cell lines, HEL (15), MB-02 (18), OC1M1 (17), and K-562 (16). In order to confirm the presence of transcripts encoding the putative HOX 2.2 N-terminal flanking region, we used a genomic probe containing 25 nt of sequence 5' to the proposed translation start site, the entire proposed N-terminal protein coding sequence (407 nt), and 50 nt of the intron (probe D, Fig. 3). mRNA from all four cell lines protects a major band of approximately 410 nt which appears to represent the 5' flanking region from which the intron has been spliced. Unexpectedly, small amounts of a second, 460 nt band presumably representing unspliced transcripts are also observed (Fig 5).

To confirm that the 1 kb intron is differentially spliced in these human cell lines, RNase protection was performed on cytoplasmic mRNAs using a second probe which spans the splice donor site but starting at a different point in the coding region while ending at the same point in the intron (probe E, Fig. 3). As shown in Figure 6, both HEL (lanes 2 and 3) and K562 (lane 5) cell mRNA protect a major 217 nt band which represents spliced transcripts and a minor 267 nt band which represents unspliced transcripts.

A probe which spans the splice acceptor site (probe C, Fig. 3) was used to assess splicing at the other side of the proposed intron. HEL and K562 cell mRNA protect a major 98 nt band representing the spliced transcripts and a weak 167 nt band representing unspliced mRNA (Fig.6, lanes 4 and 6). Similar results have been obtained with probes C and E for OC1M1 and MB02 cells (data not shown).

Human1	gctggcggttggtgtgcccgcgcgcgggagagagagaggaagcggaagagagaaggagga	H1339	cccAAcTGATGTCCATTAGCcGGACTCGAAAGTGAGGAGCcGTTAATGtGGACTATGgAT
H61	aaaaaaaaagttagttgtgcgcggtcctcgtgcgctgCGCTCCTCTGGGTGCTA. TT	M1275	ttt-t-t-----a-----a-----g-t-----c-----a-----
Mouse1	-t-----	H1399	cGAtcTAcgtcAttacgGattaacggccttgattatcattGGGTtGgggggatgacggg
H120	GACAAATCTCGCTCTGCCATTGGTCAGTGTGGATCAGATGGTTGATTTCCTCTTGCC	M1335	c--gg-cgagt-gggta-gga-----
M23	-----g-----c-----a-----	H1459	gggtgggaacaaGATGGATGGAAAGGgAGGGAAGAGcAAGATgCAaCTgAGaGAGACAC
H180	CCTCACT.GCACCTCGCCATCCCCCTCAG.CTAAACCCAATCTCGGATATACTACTA.	M1362	-a-----a-----g--c-t-g--g--
M83	-----g-----c-----a-----	H1519	ACCTGACCAG..CCCCCCCCcAgGCTCAGGTGgGAGt.TCCAACtG...CTCCTCcC
H237	TA.GCGCGCGCTCGGACTATAAAACACAACAAT.ATAAACCCGCGGAGCAGCAGCGG	M1410	-----g--tc-----ag-----g-----cccc-----t-
M143	--g-----c-----	H1572	ctCCCCcAtctAgAGTCTAcagAcg.....GCACAGGCTcAGGAGcAGGAGGAATC
Human	MetSerSerTyrPheValAsnSerThrPheProValThr	M1469	-----ctg-a-----ggc-gaaagaaa-----c-----t-c-----
H295	CCGCGcGCGCTCCCCCTCCcATGAGTTCCTATTTCGTGAATCCACCTCCCGCTCACT	H1626	TGGAGGGGCGCTGGAGGAGTgcGAAACGGGGgagGGagcGGcAGaAtagAgGcCatt
M203	-----g-----	M1527	-----ag-----acca-gaat-g-gggga-aga-tac
	GlnLeu	H1686	cCCgTactGgctGgagTctGTGtTcGaGggtcgactAGGGAGGGGtTcctGGgcccGg
H355	LeuAlaSerGlyGlnGluSerPheLeuGlyHisValProLeuTyrSerSerGlyTyrArg	M1587	t-c-acag-cta-ctt-ac-----ga-g-ccgagtg.....-c-atg-cttg-t
M262	CTGGCCAGCGGGCAGGAGTCTCTGGGCCAcgTaCCGCTCTATTCGTcGGGTATcGg		-----gc-g-----c-----gc-----
	Ala	H174	tgaccG...CAGGCTCAGcATCCcACTCTGCTGTAACAGGTCTCTcTTTGGCCCC
H405	AspProLeuArgHisTyrProAlaProTyrGlyProGlyGlnAspLysGlyPhe	M1645	gtgat-acc-----tt-----t-----
M322	GACCCGCTGaGAcATACCCCGCGCCTACGGGCCAGGcCGCGGCCAGGACAAAGGCTTT		-----SerGlyArgArgGlyArgGlnThrTyrThrArgTyrGlnThrLeuGluLeuGluLysGlu
	Ala	H1800	AGCGGcCGCGAGGCCCGCAGAcTACACACGTtTACCAGAcGCTGGAGCTGGAGAGGAg
H465	AlaTyrSerSerTyrTyrProProAlaGlyGlyTyrGlyArgAlaAlaProCysAsp	M1703	-----c-----c-----c-----A
M382	GCcAtCTCTCTATTACCGcCGcGgGGGTGCTACGGCCGcAGCGCCCTGcGcc		-----PheHisTyrAsnArgTyrLeuThrArgArgArgArgIleGluIleAlaHisAlaLeuCys
	AspAla	H1860	TTTCACTAcAATCGCTACCTGACgCGCGCGCGCATcGAGATCGCGcAGCCCTGTGC
H525	TyrGlyProAlaProAlaPheTyrArgGluLysGluSerAlaCysAlaLeuSerGlyAla	M1763	-----t-----c-----a-----
M442	TAcGgcCGCGCCGCGCTTCTACCGcGAGAAAGAGtCGGCTCGCCaCTCTCCGGCGCC		-----LeuThrGluArgGlnIleLysIleTrpPheGlnAsnArgArgMetLysTrpLysLysGlu
	Pro	H1920	CTGACgGAGAGGCAGATCAAGATaTGTTTgcTACAAACCGcCGCATGAAGTGGAAAGGAG
H585	AspGluGlnProProPheHisProGluProArgLysSerAspCysAlaGlnAspLysSer	M1823	-----g--c-----t-----t-----g-----
M502	GACGAGcAGCCCGCTTCCACCCGAGCGCGcAAGTcGACTCGCGcCAGGACAAGAGC		-----SerLysLeuLeuSerAlaSerGlnLeuSerAlaGluGluGluGluGluGluGluTRM 224
	ValPheGlyGluThrGluGluGlnLysCysSerThrProValTyrProTrpMetGlnArg	H1980	AGCAAACTGCTCAGcGCGCTCTCAGCTCAGTGCcGAGGAGGAGGAaGAAAAcAgGCCAGTGA
H645	GTGTTTCGCGcGAGAcAaGAGCAGAAGTGTCCACTCCgGTCTACCCGTGGATGCAGCGg	M1883	-----t-----g-----g-----c-----g--ca-----
M562	-----a-----c--g-----g--c-----c-----	H2043	AGGTGCTGgAAAGGGAGGAGGAGCGcGAGGGAaAGGcCTGTGGGgagcCacgGGcgtCa
	MetAsnSerCysAsnS- * * * # #	M1946	-----g-----ga-----g--t-----gaa-cta-gca-c
H705	ATGAATTCGTGCAACAGTGTAGTGTCCCGTcGcC...GTCgccccGgcTccccG	H2103	gagAgaCCCGggaagGAAGGCTCTcGgGT.GGGGAGcCaGgacAcTGTCTcCCgGcGc
M622	-----c-gt--cgt--t-gtgt--tg-gtgtg--	M200	cga-ag--ca...-gc-----c-----t-t-gg--t--g--a-a--
H762	GGcGccacc...CGGgAcACAGAAActAgTGAgcGcccCctg...CCCCAA.....TC	H2162	A.GACaGcGGGCCCAGcGCTctCCTGGAAGCCCCCgCgcacagctcCcgGgGtgcT
M682	--g-ggggggggt--c-ag--ag-c--aa-tgt-agcatc-----ccccaat--	M2062	-c--g--a-----t--tg-----t-----t-tgcagagct-ttc-t--gcgc
			(g)
H810	TCCCAcAGGgTgCTgGGTGGGCATCTCAGttagGATAGAagAAGATTGgGcGcGgCC	H2221	cTGAAGcTcaCtActCgAGCcCaCCAGCAtCcCgctCgtCCctctctcCgAGgaa
M742	-----tt--aa--a-----c-----gc-----cc-c--cc--	M2122	t--gc-ga-gc-c--c-a--a--c--c--c--c--cgaagaa--c--cctt
H870	gggGgtctCtgcgtgtgtCcCtTattagCGCAGGAGTtCAAAAGTTTCAAAAGTCCAGCc	H2271	ctGcCCTCagcctgATCAGGCTtCCTGGTGAAGTCTGAGgAGCGGACTCACTTGTGtT
M798	ccc-cctc-agatcagtc-t-t-gccc.....-g-----	M2174	..-g--.....-ct-----ct-----a-t-----c-----
H930	GCACtGggAGcCAtGGGAGcAAGTGTGCTCTCCTggggcgccTggccagAGccggGgTt	H2341	CCTGGAAGCAGAGCAaAagTtCTCTcTGTCcTGTCGcGTCTCAATTTtGTCATGTCCCC
M858	---a-ta-g-g-g---at-----g-----ctacttaag-caggct--gatt-t-g	M2233	-----g--g--a-----t-----c-----
H990	ccagGacAGGgAagGgAGcAGgagcctgcaGtcaCTggcTTggcTTtactTtgagCccc	H2401	GTGCACGGTTCAATGTTAGTGTCTGT...CCTCAGCGGGGcCTtGAAGACTCCCTGAt
M918	atcc-ga--a-gtta-g-g-ctcagctctct-gtc--gtc--cct--atg--g-cct--tag-	M2293	-----cc-----a--a--ga-----c
H1050	aaccccCctctcCagccctgGtaGGGTtCCCAcCAAGAGCggCTCCATTAAAAACg.	H2459	CCCAGACtggctGTCTCTCCcACCCCCCTCCCAAGGCCACTGGAAGGAGCACATACTAC
M977	tcttgt--ggat--tgaggggt-gg--g--gc-.....ca-----aa	M2349	-----tt.....t-----
H1109	GAGTGGCTCATGcAAGGGGgtaGgGgG..TGACGCTGaagAGGaaGAGTtAgTctgtct	H2519	CTAGAA3tAAGAAGAGAGCCCTCAGAAGAAACAAAGTTCTATTTTATTAATTTCTATG
M1037	-----a-----agg-t--ca-----c-gca--gc-----c--t-gggtc	M2409	-----g-----
H1167	gg.GAcTGTGTTTCCCTGTgGt..GGGGAGctGgggcaaGatgaGgc.ACCAGGAaaggG	H2579	TGTTGTGTTTGTAGTCTTGTCTTAGCTCTGGAGcGAAACTACTCGaTgATgATgAT
M1097	ctg-t-----c-gt--ga-aactgg-gaag--gg-----g-gct--	M2469	-----c-----c-----g-a--a--at-a--
H1227	gtgTgagtcAgaggaGGGagGaggAaTaaGgGAAGGAAG.AGAGAGGgAgCAGCGGAG	H2639	gaTgaTAATAATaATaATaAcaAcaAcaAcaAcaAATATAAARGATGTAAACCTCGa
M1155	tga-acact-ccata--tc-tat-g-ct-a-----a-----	M2529	at-at-----g--t--g--g--tg-tg-t--t-----c
H1272	cGAGaGAGgagAaAcAggcGCG...GGGTCTCAGGTTGGATTATTGTtGGCtTtAAGt	H2699	a-cgctcggtcacctccaatcctccctgcccatttttctctctcctcaacccccagcg
M1215	g--t-t--aaga-c-a-ctg--tcct-----g--t-c--a	M2586	cgaaaaa (agctcagtcacatccaatcttctcctgtcatct)
		H2759	tctccatctttctgtgccagaattcaaaaaaaaaa

Figure 4. Nucleotide Sequence of Human and Murine HOX 2.2 Genes and cDNA Clones. The human (H) and murine (M) HOX 2.2 genomic sequences are shown. Nucleotides which are identical are capitalized in the human sequence and shown as dashes in the murine sequence. Dots denote gaps placed in the sequence to maximize homology. The amino acid sequence derived from the open reading frame of the corresponding human cDNA is presented. Amino acids which are different in the murine Hox 2.2 protein are shown above the human sequence. Asterisks (*) mark the splice donor and acceptor sites of an intron which interrupts some of the cDNA clones (see text). A TGA codon which would terminate protein synthesis in the intron-containing transcripts is marked by # # #. The conserved hexapeptide 5' to the homeobox and a polyadenylation signal are underlined, and the end of the human cDNA clone at nt 2701 is marked by a +. The homeobox domain is bracketed. A composite murine Hox 2.2 genomic region sequence was constructed from the WF-10 cDNA clone (bp 1–637), genomic sequence (bp 638–733), and the WF-14 cDNA clone (bp 734–2605). The 5' portion of the murine cDNA which extends beyond data reported by Schughart et al (29) is underlined. The nucleotides below the murine sequence in the 3' untranslated portion are differences between our sequence and that of Schughart et al.

Human HOX 2.2 Transcripts Encode Potential Proteins Containing or Lacking a Homeodomain

The major HOX 2.2 transcript, from which the intron has been spliced (Fig. 3, HEL), contains a long open reading frame which would encode a 224 amino acid homeobox protein when the first

in frame AUG is used for initiation (Fig. 4). The sequence obtained for the homeobox domain is altered in the first six amino acids from that previously reported by Levine *et al* (21). The protein contains a hexapeptide (residues 127–132) which is conserved in several other homeobox proteins (28, 29). The N-terminal half of the proposed HOX 2.2 protein is rich in proline

Figure 5. 5' Region of Human HOX 2.2 Transcripts Contains Spliced and Unspliced Structures. RNase protection analysis was performed on cytoplasmic mRNA isolated from HEL (lane 1), K562 (lane 2), MB02 (lane 3) and OCIM1 (lane 4) cells. A 550 nt probe containing the 5' coding region and 50 nt of the adjacent intron (Fig. 3, probe D) was used to establish the degree of splicing of the HOX 2.2 transcripts. In all four human cell lines, a major protected band of 410 nt represents spliced mRNA, while a minor band at 460 nt represents unspliced transcripts. 5 μ g of Poly A⁺ cytoplasmic RNA was used from each cell line except K562 (20 μ g). The 460 nt band is visible in MB02 cells (lanes 3) upon longer exposure.

residues (16 of 92). In addition, the C terminal of the protein, which terminates shortly after the homeobox, is very negatively charged, containing a stretch of 5 contiguous glutamic acid residues.

As shown in Figure 3, a single 1060 bp intron interrupts the human HOX 2.2 coding region. This intron contains numerous stop codons within the homeobox reading frame including a TGA which occurs immediately 3' to the splice donor junction (nt 723) (Fig. 4). The potential HOX 2.2 protein encoded by the unspliced transcript observed in RNase protection experiments (Figs. 5 and 6) would utilize the same AUG start site as the proposed full-length protein, but would terminate immediately after the intron junction, and would thus lack the homeodomain and 3' acidic region.

Isolation Of Two cDNA Classes For Murine Hox 2.2

Since the northern gel analysis suggested that multiple transcripts of the Hox 2.2 gene are expressed in murine tissues (Fig. 1), we screened a mouse embryo cDNA library with a 2.1 kb human genomic probe to isolate Hox 2.2 cDNAs. Restriction and partial sequence analysis revealed that the resulting clones consisted of approximately equal numbers of two alternatively spliced structures (Figure 3B). WF-10 represents transcripts from which the intron is excised. Comparison with the murine Hox 2.2 genomic sequence revealed that six other unique clones represented partial cDNAs consisting of the Hox 2.2 homeobox domain and 3' flanking region preceded by portions of the intron which are excised from WF-10. The longest of these clones (WF-14), which contains approximately 960 bp preceding the homeobox, is shown in Figure 3B. Since WF-14 is too short to account for the 2.3 kb message (Figure 1B), we hypothesize that it is a partial clone derived from unspliced message which might contain the 5' coding region. PCR amplification has been used to obtain cDNA clones representing this region. As shown in Figure 3B, use of primers 5 and 6 with cytoplasmic mRNA from 12 day mouse embryos resulted in isolation of cDNAs whose sequence contains the entire 5' coding region extending through the intron to overlap WF-14. These overlapping clones have been incorporated into the proposed structure for the alternatively spliced transcripts (Figure 3-Hox 2.2-IC).

Figure 6. Alternative Splicing of a 1 kb Intron in HOX 2.2 Transcripts from Human Erythroleukemia Cell Lines. A 337 nt intron donor probe which spans the 5' coding region and the intron (Fig. 3, probe E) was used in RNase protection experiments with 5 μ g of HEL cell cytoplasmic mRNA (lanes 2 and 3 [longer exposure of lane 2]) and 20 μ g of K562 cell cytoplasmic mRNA (lane 5). A strong band at 217 nt represents spliced transcripts while the band at 267 nt represents unspliced mRNAs. A 218 nt intron acceptor RNA probe which spans the intron junction and the homeobox (Fig. 3, probe C) was hybridized with HEL cell mRNA (lane 4) and K562 cell mRNA (lane 6). The major protected band of 98 nt represents spliced transcripts while the minor band at 163 nt (visible in HEL cells upon longer exposure) represents unspliced mRNAs. The upper arrow denotes undigested probe.

Murine Hox 2.2 Transcripts Encode Potential Homeobox-containing and Boxless Proteins

The spliced cDNA clone (WF-10) contains 1542 bp, which corresponds closely to the 1.6 kb message. This clone contains a long open reading frame which would encode a 224 amino acid homeobox protein. The murine Hox 2.2 homeobox region shows 92% nucleotide identity and 100% amino acid identity with the human homeobox domain. The remaining protein sequence is 96% identical to the proposed human Hox 2.2 translation product and contains the conserved hexapeptide sequence N-terminal to the box (residues 127–132, Fig. 4). The murine intron sequence also contains numerous stops in the homeobox reading frame, including the TGA located immediately 3' to the splice donor site. Thus translation of unspliced transcripts using the common AUG initiation site would yield a truncated N-terminal protein missing the homeodomain.

Demonstration Of Intron-Containing Murine mRNAs By RNase Protection

We used RNase protection to demonstrate the presence of intron containing transcripts in cytoplasmic mRNA and to delineate the relationship of the cloned cDNAs to transcripts present in murine embryonic tissues. As shown in Fig. 7, a probe which spans the splice donor site yields two protected bands: a 204 nt fragment corresponding to a spliced transcript and a 230 nt fragment corresponding to unspliced mRNA. These data show that

Figure 7. Alternatively Spliced Hox 2.2 Transcripts are Present in Murine Embryonic Tissue. **Left.** A 404 nt murine Hox 2.2 intron donor probe which spans the sequence preceding the intron and the 5' portion of the intron (Fig. 3, probe a) was hybridized with: 11, 13, and 15 day mouse embryo cytoplasmic mRNA (5 μ g per sample). Two bands of approximately 204 nt and 230 nt, representing spliced and unspliced transcripts, respectively, are noted by the arrows. **Right.** A 267 nt murine Hox 2.2 splice acceptor probe which spans the intron and the 5' region of the homeobox (Fig. 3, probe b) was hybridized with: 9, 11, 13, and 15 day mouse embryo cytoplasmic mRNA. Two bands of approximately 105 and 192 nt, representing spliced and unspliced transcripts, respectively, are noted by the arrows. Lane 1 is molecular weight markers.

approximately 40 percent of the murine Hox 2.2 transcripts contain sequences which span the 5' coding exon and the intron and thus would potentially encode 'boxless' proteins, while 60 percent of the transcripts are spliced and would encode a full-length homeobox protein. A probe which spans the splice acceptor has been used to further characterize Hox 2.2 transcripts. This probe again demonstrates the presence of two transcripts: a 105 nt band representing spliced transcripts and a 192 nt fragment corresponding to unspliced transcripts (Fig. 7 right). Given the darker autoradiographic image inherent in a larger band, these results also indicate that approximately 60 percent of the message is spliced, while 40 percent contain intron sequences. RNase protection experiments on adult murine kidney, uterus, and spinal cord mRNA show that all three tissues contain both spliced and unspliced transcripts, but that spinal cord contains relatively little of the unspliced mRNA (results not shown).

Human And Murine HOX 2.2 Genomic Sequences Are Highly Homologous

Comparison of the human and murine genomic and cDNA sequences show that in addition to the highly conserved coding region, there are other extremely conserved regions (Figure 4). The 5' non-coding sequences are more highly conserved than the protein coding region. In the 3' untranslated region there is an extremely conserved (99%) 135 bp region nested within a larger segment of high identity (91% over 327 bp). A search of Genbank did not reveal any significant homology for this region with other reported sequences. Immediately following this highly homologous section, both genes contain repetitive, AT-rich sequences: the human gene contains (TGA)₆(TAA)₆(TCA)₅(TAA)₂ immediately prior to the polyadenylation signal; the murine cDNA in the same region consists of a similar repeat: (TAA/TAT)₈(TGA)₄(TAA)₄. Although the intron regions show

much lower overall homology, there is a 150 bp section (1186–1335 in the murine sequence) which is 82% identical to that of human.

DISCUSSION

We have previously reported the lineage-restricted expression of a number of homeobox genes in human hematopoietic cell lines (10). HOX 2.2 expression appears limited to human hematopoietic cells with erythroid potential. Although the HOX 2.2 gene was described several years ago (30), our results represent the initial characterization of the human gene products. Schugart *et al* have described isolation of spliced murine Hox 2.2 coding transcripts (29), but data on 'intron'-containing murine transcripts have not been reported. Since these investigators employed spinal cord as a tissue source for cDNA synthesis, and our northern gel and RNase protection experiments show spinal cord to be relatively low in the intron-containing transcripts, it is possible that tissue differences account for the disparity of observed transcripts. Our murine cDNA extends an additional 115 bp 5' upstream compared to that of Schugart *et al*. Recently Njolstad *et al* reported sequence data for a zebrafish Hox 2.2 gene (31). The proposed zebrafish homeobox protein is highly homologous to the human and mouse proteins, including the box region, which we find to differ from a previously published sequence (21). Unfortunately, the zebrafish Hox 2.2 gene transcripts were not characterized, so that the occurrence of unspliced transcripts could not be evaluated.

We have presented sequence and RNase protection data for two types of human and murine HOX 2.2 transcripts. A single 1 kb intron interrupts the coding sequence of the HOX 2.2 gene. Excision of this intron results in one type of mRNA which encodes a potential 224 residue protein containing an extended proline-rich N-terminal flanking region, a homeodomain and a short 3' acidic cluster. Proline rich regions (32) and acidic clusters (32) have been shown to function as activation domains. This spliced transcript is the predominant message in human erythroleukemia cell lines and is a major transcript in the mouse embryo and adult tissues.

A second form of mRNA is expressed in nearly equal amounts in murine adult kidney and uterus and embryonic tissues. This mRNA appears to contain at least portions of the unspliced intron contiguous to the extended N-terminal flanking region. Since we have not isolated a complete cDNA clone which extends from the translation start site through the intron to the homeobox, it is unclear whether the entire intron remains in the intron-containing clones. The 2.3 kb band of mRNA observed by northern gel analysis is slightly smaller than would be expected for a transcript containing the entire intron. However, the significance of the intron containing transcript is not its size but the fact that the intron interrupts the protein coding sequence. Thus we have shown by RNase protection that minor amounts of human HOX 2.2 transcripts in hematopoietic cells extend from the AUG initiation codon in the 5' coding exon across the splice junction into the proposed intron. Furthermore, we have shown by RNase protection and PCR amplification that a major fraction of the murine transcripts extend from the 5' coding region into the intron-containing cDNAs. In addition, the 2.3 kb murine mRNA band observed in northern gels hybridizes to both 5' exon and intron probes suggesting that these transcripts contain these two elements. Both the human and murine introns contain an in frame TGA stop codon 6 nucleotides past the splice junction.

Translation initiation of these transcripts at the common AUG would yield 140 amino acid proteins consisting of the N-terminal flanking region, but lacking the homeobox and acidic cluster.

Alternative splicing of homeobox genes resulting in potential protein products lacking either the homeodomain or the extended flanking region appears to be a common event. The *Drosophila* homeobox genes *Antp* (12), *Ubx* (14), and *Abd-B* (13) are all differentially spliced to yield transcripts which would encode full-length homeobox proteins as well as either altered or truncated proteins. The alternative *Abd-B* transcript, which would encode a homeobox lacking the N-terminal amino acid region, was proposed to be a transcription repressor (13). We have recently demonstrated that a novel homeobox gene expressed in human myeloid cells (PL1), is similarly expressed as transcripts which would encode a full-length homeobox protein and a truncated protein containing only the homeobox (Lowney et al, submitted).

In contrast to these homeobox genes, the murine retinoic acid responsive gene (*ERA-1*) is expressed as alternative transcripts which would encode a homeobox protein or a protein containing only the amino terminal flanking region, lacking the homeodomain (33). Our results suggest that the *HOX 2.2* gene forms an alternatively spliced transcript with the potential to code for a similar 'boxless' protein product. It is likely that the target gene specificity of the homeobox protein depends, in part, on the interaction of the flanking regions with other nuclear factors. A truncated *HOX 2.2* protein without a homeodomain might compete for these nuclear factors, thereby influencing transcription. Future studies will be directed towards expression of these partial proteins in transfected cell lines to investigate potential influence on *HOX 2.2* function.

ACKNOWLEDGEMENTS

We wish to thank C. Hauser for many helpful discussions and providing the human genomic clones and the mouse embryo library; and K. Schughart and F. Ruddle for kindly providing a partial murine *Hox 2.2* genomic clone and a pre-print of their manuscript, identifying the murine *Hox 2.2* intron splice donor site. We also thank D. Morgan for sending us the MB-02 erythroleukemia cells prior to publication of the properties of this line, and T. Papayannopoulou for providing the HEL and OCIM1 cell lines. Finally we thank F. Hack for growing all of the cells used in this study. Supported in part by the Veterans Affairs Research Service, and NIH grant HL-25408 and an American Cancer Society grant to C.L., NIH Grant CA-47866 and a Leukemia Research Foundation grant to H.J.L. K. D. is a University of California Cancer Research Coordinating Committee Fellow.

REFERENCES

1. Driever, W., and C. Nusslein-Volhard. 1989. *Nature* **337**: 138–143.
2. Scott, M. P., and A. J. Weiner. 1984. *Proc. Natl. Acad. Sci. USA* **81**: 4115–4119.
3. Bodner, M., J.-L. Castrillo, L. E. Theill, T. Deerinck, M. Ellisman, and M. Karin. 1988. *Cell* **55**: 505–518.
4. Dony, C., and P. Gruss. 1988. *Differentiation* **37**: 115–122.
5. Finney, M., G. Ruvkun, and R. H. Horvitz. 1988. *Cell* **55**: 757–769.
6. Ingraham, H. A., R. Chen, H. J. Mangalam, H. P. Elsholtz, S. E. Flynn, C. R. Lin, D. M. Simmons, L. Swanson, and M. G. Rosenfeld. 1988. *Cell* **55**: 519–529.
7. Ko, H.-S., P. Fast, W. McBride, and L. M. Staudt. 1988. *Cell* **55**: 135–144.
8. Scheidereit, C., J. A. Cromlish, T. Gerster, K. Kawakami, C.-G. Balmaceda, R. A. Currie, and R. G. Roeder. 1988. *Nature* **336**: 551–557.
9. Graham, A., N. Papalopulu, and R. Krumlauf. 1989. *Cell* **57**: 367–378.
10. Shen, W.-F., C. Largman, P. Lowney, J. C. Corral, K. Detmer, C. A. Hauser, T. A. Simonitch, F. M. Hack, and H. J. Lawrence. 1989. *Proc. Natl. Acad. Sci. USA* **86**: 8536–8540.
11. Kornfeld, K., R. B. Saint, P. A. Beachy, P. J. Harte, D. A. Peattie, and D. S. Hogness. 1989. *Genes Develop.* **3**: 243–258.
12. Schneuwly, S., A. Kuroiwa, P. Baumgartner, and W. J. Gehring. 1986. *EMBO J.* **5**: 733–739.
13. Zavortink, M., and S. Sakonju. 1989. *Genes Dev.* **3**: 1969–1981.
14. O'Connor, M. B., R. Binari, L. A. Perkins, and W. Bender. 1988. *EMBO J.* **7**: 435–445.
15. Martin, P., and T. Papayannopoulou. 1982. *Science* **216**: 1233–1235.
16. Lozzio, B. B., C. B. Lozzio, E. G. Bamberger, and A. S. Feliu. 1981. *Proc. Soc. Exp. Biol. Med.* **166**: 546–550.
17. Papayannopoulou, T., B. Nakamoto, S. Kurachi, M. Tweeddale, and H. Messner. 1988. *Blood* **72**: 1029–1038.
18. Perrine, S. P., B. A. Miller, and D. A. Morgan. 1989. *Biochem. Biophys. Res. Commun.* **164**: 857–862.
19. Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. *Biochem.* **18**: 5294–5299.
20. Gubler, U., and B. J. Hoffman. 1983. *Gene* **25**: 263–269.
21. Levine, M., G. M. Rubin, and R. Tjian. 1984. *Cell* **38**: 667–673.
22. Molecular Cloning, a Laboratory Manual, Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
23. Feinberg, A. P., and B. Vogelstein. 1983. *Anal. Biochem.* **132**: 6–13.
24. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. *Proc. Natl. Acad. Sci.* **74**: 5463–5467.
25. Current Protocols in Molecular Biology 1987. Edited by F. M. Ausubel. Greene Pub. Assoc. and Wiley Interscience.
26. Scott, M. P., J. W. Tamkun, and G. W. Hartzel. 1989. *Biochim. Biophys. Acta* **989**: 25–48.
27. Klemsz, M. J., S. R. McKercher, A. Celada, C. Van Beveren, and R. A. Maki. 1990. *Cell* **61**: 113–124.
28. Mavilio, F., A. Simeone, A. Giampaolo, A. Faiella, V. Zappavigna, D. Acampora, G. Poiana, G. Russo, C. Peschle, and E. Boncinelli. 1986. *Nature* **324**: 664–668.
29. Schughart, K., M. F. Utset, A. Awgulewitsch, and F. H. Ruddle. 1988. *Proc. Natl. Acad. Sci. USA* **85**: 5582–5586.
30. Hauser, C. A., A. L. Joyner, R. D. Klein, T. K. Learned, G. R. Martin, and R. Tjian. 1985. *Cell* **43**: 19–28.
31. Njolstad, P. R., A. Molven, J. Apold, and A. Fjose. 1990. *EMBO J.* **9**: 515–524.
32. Mermod, N., E. D. O'Neill, T. J. Kelly, and R. Tjian. 1989. *Cell* **58**: 741–753.
33. LaRosa, G. J., and L. J. Gudas. 1988. *Mol. Cell Bio.* **8**: 3906–3917.