A novel murine homeobox gene isolated by a tissue specific PCR cloning strategy

Michael J.Kern, David P.Witte, M.Todd Valerius, Bruce J.Aronow and S.Steven Potter
Childrens Hospital Research Foundation, Basic Science Research, Elland and Bethesda Avenues,
Cincinnati, OH 45229, USA

Received April 13, 1992; Revised and Accepted July 29, 1992

EMBL accession nos X59725, X59726

ABSTRACT

We have identified a novel homeobox gene, designated K-2, using a reverse transcription PCR cloning strategy. Sequence analysis reveals that the homeobox of K-2 is 77.6% homologous at the nucleotide level and 97% identical at the amino acid sequence level to another murine gene, S8. Homeodomain sequence comparisons indicate that K-2 and S8 represent a distinct subclass of paired type homeobox genes. Northern blot analysis of RNA from murine embryos and adult tissues identified multiple transcripts that are expressed in a developmentally specific and tissue restricted manner. Alternate splicing of K-2 at the 3'-coding region leads to the inclusion of a chain terminating sequence. In addition, the developmental expression pattern of this gene at day 12 of gestation was determined by in situ hybridization. Expression was observed in diverse mesenchymal cells in craniofacial, pericardial, primitive dermal, prevertebral, and genital structures.

INTRODUCTION

Genes carrying the conserved 180 bp homeobox are known to play a major role in establishing segmentation pattern and in determining segment identity in *Drosophila* (1,2). Increasing evidence suggests that homologous homeobox containing genes in vertebrates are equally important in coordinating developmental processes (for review see ref. 3). The homeobox encodes a DNA binding homeodomain with a helix-turn-helix motif (4,5). The homeodomain proteins can function as transcription factors that auto-regulate (6–8) as well as trans-regulate other homeobox genes (7,9) and batteries of not yet identified target genes (10). At least some homeobox genes appear to function as developmental 'master control switches' in the genetic circuitry of development (11).

A number of strategies have been used to isolate homeobox genes, including hybridization with probes from other homeobox genes (12,13), degenerate oligonucleotide hybridization (14–17), and PCR (18–20). In this report we describe a specific PCR screening approach that is particularly useful in the identification of a wide range of homeobox gene types expressed during the development of an individual tissue of interest. Using this procedure we have isolated a novel homeobox gene, designated

K-2, that is expressed during cardiogenesis. An initial characterization of the pattern of K-2 gene expression in murine embryos and adult tissues is described based on Northern and in situ analyses.

MATERIALS AND METHODS

RNA isolation and Northern blot analysis

Tissues for RNA isolation were obtained from CD-1 embryos (Charles River Laboratories) which were housed in an AAALAC approved facility. Noon of the day of finding a vaginal plug was designated as day 0.5 post-coitum (p.c.). Total RNA was isolated by homogenization of tissue or embryos in guanidinium isothiocyanate (21). Total RNA (20 µg) was electrophoresed through a 1% (w/v) agarose gel with 6% formaldehyde using a 1×(12mM Tris pH=7.5, 6mM sodium acetate, and 0.3mM EDTA) gel buffer and blotted onto a Genescreen Plus nylon membrane (Dupont). Hybridization was according to the manufacturers specifications.

A multiple tissue Northern blot of adult mouse RNA (Clontech Laboratories) contained 2 µg of poly A+ selected RNA from mouse heart, brain, spleen, lung, liver, skeletal muscle, kidney, or testis. Hybridization was according to the manufacturers specifications.

PCR amplification and oligonucleotides

Routine PCR amplifications were performed with 30 cycles in a Perkin-Elmer/Cetus DNA Thermal Cycler, using a cycle of: 1 min at 94°C, 1 min at 45–60°C depending on the primers, and 2–3 min at 72°C. The 100 µl reaction mixture contained: 100 pmoles of each primer, 2 units of Amplitaq DNA polymerase (Perkin-Elmer/Cetus), 10mM Tris (pH=8.3), 50mM KCl, 1.5mM MgCl₂, 0.2mM dNTPs, and 200µg/ml gelatin (22). The primers were synthesized at the University of Cincinnati core facility on an Applied Biosystems Model 394 DNA synthesizer and are: K2-UT3 5'-CAGAACAGGATAGGTGTGTTATG-G-3', K2-EXON3' 5'-GCTATTGTTGGCACATGTGG-3', K2-EXON5' 5'-TCTCCTCAAGTCCTACTCTG-3'. The dT primer 5'-TCAGAAAGCTT(T)₂₄-3' contains a HindIII site in the 5' end and the HOX 34 primer 5'-AGCTAGAATTCCA-A/GA/GTIAAA/GA/GTITGGTTICAA/GAA-3' contains an EcoRI site in the 5' end.

Reverse transcription PCR (RT-PCR) cloning of K-2

Five μ g of total RNA from each one of the embryonic time points 9.5–18.5 days p.c. was used in reverse transcription reactions with Superscript (BRL) according to the manufacturers specifications. Total RT reactions were diluted 1:5 and subjected to PCR amplification with the Hox 34 and dT primers. The PCR products were analyzed by gel electrophoresis and major bands were gel purified, PCR amplified, and then subcloned into M13 (23). Single-stranded sequencing was performed by the dideoxy method using Sequenase 2.0 (USB). Sequence analysis was performed using the Fasta program of the online service of Genbank 71.0 and EMBL 30.0 (24). An NcoI-HindIII fragment in which the NcoI site was part of the K-2 sequences (nucleotides 744–749 figure 2a) and the HindIII site was on the 5' end of the dT primer was utilized as a probe to screen an 8.5 day mouse embryo cDNA library (25) generously provided by Dr. Brigid Hogan (23). The resulting cDNA clones were subcloned into pBSISK⁺ (Stratagene) and subjected to double-stranded sequencing.

Reverse transcription PCR amplification

For the nested RT-PCR which evaluated the K-2 alternative exon pattern of expression 1 μ g of total RNA was used in a 50 μ l reaction volume with Retrotherm RT (Epicentre Technologies) according to the manufacturers specifications. The Retrotherm enzyme is a thermostable RNA and DNA dependant DNA polymerase. The primers Hox 34 and K2-UT3 were utilized in the reverse transcription and primary PCR reaction. The products of that reaction were then diluted 1:20 and further amplified with the K2-EXON3' and the K2-EXON5' primers in a secondary PCR reaction with Taq polymerase and standard PCR conditions. Some of the products were TA subcloned (Invitrogen) and sequenced to confirm that this procedure was amplifying the correct RNA targets.

45	46	47	48	49	50	51	52	53	54	55	56
GLN	ILE	LVS	ILE	TRP	PHE	GLN	ASN	ARG	ARG	MET	LVS
Q	I	K	I	W	F	Q	N	R	R	M	K
CA ^A _G	ATC ^T _A	AA ^A _G	ATC ^T _A	TGG	TT ^T _C	CA ^A _G	AA ^T _C	C ^C _G	C ^C _G	ATG	AA ^A _G
CA ^A _G	AT ^T _I	AA ^A _G	AT ^T _I	TGG	TT ^T _I	CA ^A _G	AA	HOX 34			

Figure 1. The design of primer HOX 34. The position within the 60 amino acid homeodomain is noted by the numbers above the top. The black dots denote the nearly invariant amino acid residues between the various homeobox genes of different species (5). The 3-letter and 1-letter amino acid codes are directly below the homeodomain numbers. The amino acids shown are the most conserved and the nucleotide sequence of the codons is listed below. The HOX 34 primer is actually a mix of 32 different primers. The inclusion of inosine in homeodomain positions 46, 48, and 50 has kept the degeneracy lower than just using the nucleotide sequence of the codons. The HOX 34 mix is also capable of binding to more divergent homeoboxes because of the extra guanine nucleotides at codon positions 46 and 48. For non-Antennapedia class homeodomains valine is sometimes in these positions and the addition of the guanines allows HOX 34 to bind these type of homeobox sequences. The HOX 34 primer has an EcoRI site on the 5' end in order to facilitate cloning (not shown).

In situ hybridization

Labeled sense and antisense RNA probes were prepared with α^{35} S-UTP and RNA polymerase T7 from plasmids linearized at convenient restriction sites located after the insert. Four probes were generated: TA # 10 was an antisense probe from nucleotides 780–630 (figure 2) plus the 72 base alternative exon, TA # 3 was a sense probe from nucleotides 630–780 (figure 2) plus the 72 base alternative exon, TA # 4 was an antisense probe from nucleotides 457–207 (figure 2), and TA # 2 was a sense probe from nucleotides 132–460 (figure 2). The probes all lacked the conserved homeodomain.

CD-1 fetuses in anesthetized dams were perfused for 5 min with 4% paraformaldehyde in phosphate buffered saline (PBS) via their placentas. Fetuses were removed and fixed in the same solution overnight at 4°C, transferred to 20% sucrose in PBS overnight at 4°C, then embedded in O.C.T. compound (Miles Laboratories) and snap frozen. Cryostat sections (6–8 μ) were prepared and processed previously described (26). Slides were coated with NTB2 emulsion (Kodak), exposed for 10–14 days and developed in D19 developer (Kodak). Sections were counterstained in hematoxylin and eosin and photographed under dark and brightfield illumination.

RESULTS

In order to isolate novel homeobox genes expressed during cardiogenesis, RNA from murine fetal heart tissue was isolated and total cellular RNA was reverse transcribed using the dT primer. The resulting cDNA was PCR amplified by the dT primer in conjunction with the HOX 34 primer mix which consists of 32 different primers corresponding to the most conserved region of the homeobox (figure 1). The resulting PCR reactions were electrophoresed and the four most prominent bands were gel purified and further PCR amplified with the same primers. The amplified bands were then subcloned into M13 and sequenced. The HOX 34 and dT primers amplified nucleotides corresponding to the 3' end of the homeobox, thus the sequences of the M13 inserts could be used to discern homeobox cDNAs from non-homeobox containing cDNAs. Of the initial four clones characterized three proved to represent non-homeobox genes resulting from artifactual priming during PCR amplification. The fourth clone however carried a portion of a genuine homeobox and is the subject of this report.

Since this original PCR clone contained only the 3' portion of the homeobox we attempted to expand the cloned region. An NcoI-HindIII fragment of this clone was used to screen an 8.5 day mouse embryo cDNA library (25) and five clones were obtained. The putative open reading frame of a 2.4 kb cDNA is listed in figure 2a along with the deduced amino acids it would encode. It is possible that this clone contains the entire open reading frame since the sequence GACCATGACC is very similar to the Kozak consensus sequence of CACCATGGCG (27,28). Nevertheless, no in-frame stop codon has been observed within the 150 bp 5' of the methionine codon sequenced to date. There is however, an in frame stop codon 735 bp downstream of this putative start site of translation and the resulting open reading frame would encode a 245 amino acid polypeptide (figure 2a). One of the five cDNA clones carried an alternative exon of 72bp which was spliced in at nucleotide position 740 (figure 2). Thus two different transcripts from the K-2 gene would be generated;

K-2a, which does not contain the exon and has already been described, and *K-2b* which contains the 72 bp exon and would encode a shorter 217 amino acid polypeptide since the alternative exon contains a stop codon (figure 2b).

A

CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCACTCCCCCCTCTTCTCTCCCACTCGG	60
CTCTCTCTCCCCCTCACGCCACAGCGTTTGGTGTGATTAGAGCGGGGAAGAGGGGGGTG	120
GGTGGGATTTGGAGGGGAGACCATTGACCTTCAGCTACGGGCACGTTCTGGAGCGGCACCGC	180
M T S S Y G H V L E R Q P I	240
CGCCTGGGCGGCCGCTTGACAGCCCGGGACAACCTCGACACCTTCGAGCGCAAAAAGAAC	13
A L G G R L D S P G N L D T L Q A K K N	33
TTCCTCGCTAGTCACCTGCTAGACACTGGAGGAGCGGGGACATGTTGGCGGCACAAGCA	300
F S V S H L L D L E E A G C D M V A Q A	360
GACAAAGTGTGGGCGAGGCGGGCGGAGCCCTGCTGGAGTCACCGGACTGACCATGGC	53
D E S V G E A G R S L L E S P G L T S G	73
AGCGACACCCCTCAGCAGGACAATGACCAGTTGAACTCTGAGGAGAAGAAGAAGAAAG	92
S D T P Q Q D N D Q L N S E E K K R K	130
CAGCGGAGAAACAGGACAACATTTCAACAGCAGCCAACTCAGGCCCTTGGAGCGTGTCTTT	480
Q R R N R T T F N S S Q L Q A L E R V F	113
GAGCGGACACATTACCCGGATCGTTTGTTCGAGAAGATCTCGCAGCTCGGGTGAACCT	540
E R T H Y P D A F T R E D L A R R V N	133
ACTGAGGCCAGAGTGCAGGTGTGGTTTTAGAACCGAAGAGCCAAAGTTCCGCAGGAATGAG	600
T E A R V Q V W F Q N R R A K F R R N E	153
CGAGCATACTGGCCAATAA AACGCTTCTCTCTCTCAAGTCTCTACTCAGGAGACGTGACT	670
R A M L A N K N A S T L L K S Y S G D V T	173
GCTGTGGAGCAACCCATCGTAGCTCTGTCTGCCAGACCAAACCGATTATCTCTCTCTGG	720
A V E Q T P I V P R P A P R P T D Y L S W	193
GGGACAGCCTCTCCGTACGCCMATGGCTACTTATTC TGCCACATGTGCCAACAATAGC	780
G T A S P Y S G C C A T T A T C A N N S	213
CCTGCACAGGGTATCAACATGGCCAACAGCATTTGCCAACCTGAGACTGGAAGGCCAAGGAA	840
P A O G I N M A N S I A N L R L K A K E	233
TATAGTTTACAGAGAACCGAGTGCCAAACAGTCAACTGAGGAAAAA AAAA	890
Y S L O R N O V P T V N *	245

B

ATCTTCGTCCTCCCAAGATGTTGTTTACACGAGGGGCTT	780
R S S S L P R C C L H E G L	213
CATAACGGATTCTAACGGAAGACACTGAAAAG	812
H N G F *	217

Figure 2. Nucleotide and deduced amino acid sequences of *K-2a* and *K-2b*. (A) The open reading frame of the *K-2a* transcript is shown in full. The homeobox is in bold type. The nucleotide sequence of the underlined portion would be the same in the *K-2b* transcript but would not encode any amino acids since there is a termination codon in the alternatively spliced exon. The splice site is denoted by ∇ above the intron/exon boundary. (B). The nucleotide and deduced amino acid sequence of the alternatively spliced 72 bp exon which is included in the *K-2b* transcript. The EMBL databank accession numbers are X59725 and X59726.

The *K-2a* and *K-2b* transcripts are identical in the homeobox region and would encode a homeodomain that is similar to the paired class of homeodomains. The homeodomains of *K-2* and *S8*, a previously described homeobox gene, are nearly identical with only two conservative changes throughout the 60 amino acid domain (figure 3). Sequence comparison with other homeobox genes indicates that *K-2* and *S8* are most closely aligned with the paired class of homeobox genes, but these two represent a distinct subclass of paired type homeobox genes (figure 3). The *K-2* and *S8* homeodomains differ significantly from the bona fide members of the paired class in positions 9–11, 28–30, and 55–60, whereas the paired class members are all very similar to each other at these positions. The *Pax* genes and the *Drosophila* genes all have approximately 50/60 amino acid similarity with each other within the homeodomain. In contrast, the *K-2* and *S8* genes appear to encode a distinct subgroup of paired type homeodomains, with only 40/60 amino acid identity with the *prd* gene, 40/60 with the *Pax-3* gene, 38/60 with the *Pax-7* gene, 38/60 with the *Pax-6* gene, and even less homology with other paired type homeodomains (figure 3). In comparison the *K-2* and *S8* homeodomain identity with the other classes of homeodomains is quite low, with for example 20 of 60 amino acids the same for the *Antennapedia* homeodomain.

The *K-2* and *S8* genes encode very similar though not identical proteins. When 5 amino acids 5' and 15 amino acids 3' of the homeodomain are included with the homeodomain comparison the identity is 73/80 amino acids; 78/80 when conservative changes are considered (figure 4). Since the homeodomain and regions flanking it are important for DNA binding specificity it is feasible that *K-2* and *S8* might bind the same DNA sequence. The known portion of the open reading frame of *S8* is shown in figure 4 as it aligns with the deduced amino acid sequences of *K-2a* and *K-2b*. The polypeptide encoded by the *K-2b* transcript however diverges from that encoded by the *S8* and *K-2a* transcripts at the point of the alternative exon. It is interesting that the homology between *S8* and *K-2* appears to extend to the gene structure since the splice site of the alternative exon of *K-2* is also the same spot in which *S8* has a splice site (29).

The expression of the 72bp alternative exon of *K-2* was evaluated by nested RT-PCR (figure 5). Multiple repetitions of

	1	10	20	30	40	50	60							
	B	B	H	PH	P	HH	HH	H	H	BH	B	B	P	P
	*	*	*	*	*	*	*	*	*	*	*	*	*	*
K-2	QRRNRTTFS	SQLQALERV	ERTHYPDÄVF	REDLARVRNL	TEARVQVWFQ	NRRAKFRRNE								
S8				-E-	S-									
PRD	-C-	SA	-DE-A-	-Q-	IYT	-E-Q-T-		I-	S-		RL	-KQ		
PAX-3	-S-	TA	E-E-A-		IYT	-E-Q-AK-		S-		RW	KHA			
PAX-7	-S-	TA	E-EE-KA-		IYT	-E-Q-TK-		F-	S-		RW	KQ		
PAX-6	LQ-	-S-TQ	E-IE-	KE-		V-A	-R-AKID-	P-	I-	S-		W	-E-	
GSB-p	-S-	TA	E-E-	A-	S-Q-	VYT	-E-QTTA-		I-	S-		RL	KHS	
GSB-d	-S-	SN	D-ID-	I-	A-Q-	VYT	-E-QTTA-				S-	RL	L-KQS	
GOOSCD	K-H-	I-TD	E-E-	NL	Q-K-	VGT	-Q-	H-	R-EK-	E-K		W	-QK	
OTD	-GP-	TR	A-DV-	AL	GK-R-	I-M	-EV-LKI-		P-S-		K	-C	QQL	
MIX 1	-K-	K-F	Q-T	-DI-QF	QTNM-	IHM	-E-HIYI		P-S-	I-		V	-QG	
CEH-10	K-K-	H-I	Q-T	Y-IDE-KE	QDS-	IYA	-V-GKTE-		Q-D-I-			-C	QQL	
ANTPD	RK-G-Q-YTR	Y-TLE-KE	HFNRL	LTRRR	IEI	HALC-			QDIKI			-M	WKEN	

Figure 3. Comparison of paired type homeodomains from mouse: *K-2*, *S8* (29), *Pax-3* (30), *Pax-7* (31), *Pax-6* (32); *Drosophila*: *prd* (33), *gsb-p* (34), *gsb-d* (34), *otd* (35), and *Antpd* (5); *Xenopus*: *Mix-1* (36) and *C.elegans*: *Ceh-10* (37). Horizontal lines indicate amino acid identity to *K-2*. The numbers across the top represent the amino acid position within the 60 amino acid homeodomain. B=base pair contact point with DNA, H=hydrophobic core of the homeodomain, P=phosphate backbone contact point of the homeodomain, and *=invariant or relatively invariant as defined by Laughon (4).

	79	88	98	108	118	128
K-2a	QDNDQLNSEE	KKKKRQRNRK	TTFNSSQLQA	LERVFERTHY	PDADFVREDA	
S8	VTVPAPAVGT	-R-KR-				
K-2b	-----	-----	-----	-----	-----	-----
	138	148	158	168	178	
K-2a	RRVNLTEARV	QVMFQNRRAK	FRNRERAMLA	NKNASLLKSY	SGDVTAVEQP	
S8	-S-			TRS	-QEA-I-	
K-2b	-----	-----	-----	-----	-----	-----
	188	194	204	214	223	
K-2a	IVPRPAPRPT	DYLSWG	TASFSYAMAT	YSATCANNSP	AQ GIMMANS	
S8	-	TMSP----	-----SVFP	--PGGS--	-TP-V----	
K-2b	-----	-----	-----RSSLL	PRCCLHEGLH	NG F*	
	233	243				
K-2a	IANLRRLKAKE	YSLQRNQVPT	VN*			
S8	-S-	F-HHS-	-*			

Figure 4. Amino acid comparison between *K-2a*, *K-2b*, and *S8*. The numbers represent the amino acid position within the *K-2a* protein. The differential splicing of *K-2a* and *K-2b* results in different amino acid sequence beginning at position 200. All of the available sequence of *S8* is shown. A few spaces were inserted to achieve optimal alignment between the *K-2* encoded proteins and *S8*. The homeodomain is in bold type. The asterisk denotes a stop codon and horizontal lines indicate identity with the *K-2a* sequence.

Figure 5. RT-PCR analysis of the alternative splicing of *K-2* transcripts. Total cellular RNA was used in the RT-PCR from the time points 8.5–18.5 days post-coitum from the whole murine embryo. The nested RT-PCR was performed in two stages with the reverse transcription and initial PCR performed by the Retrotherm enzyme in the presence of the K2 UT3 and HOX34 primers. The first stage was diluted 1:20 and one ml was amplified in the presence of Taq polymerase and the K2-exon5' and K2-exon3' primers. '–' represents the negative control sample which did not contain DNA or RNA but was treated and PCR amplified the same as the other samples.

this experiment showed some variability in both the time points at which the alternative exon expression was detected, and in the ratios of the two transcripts. Nonetheless, the data clearly demonstrate the presence of transcripts carrying the alternate exon at multiple phases in embryonic development. We determined the sequence of the RT-PCR transcript containing the exon by TA subcloning (Invitrogen) the top band from three different time points and sequencing five different isolates (data not shown). This confirmed that the RT-PCR procedure was amplifying the correct RNA targets. In addition RNase and DNase controls confirmed that only RNA and not DNA was initially present and amplified (data not shown). This 72bp sequence cannot represent an unprocessed intron because consensus splice sequences, GT at the splice donor and AG at the splice acceptor termini, are not present at the junctions of the 72 bp alternative exon. Furthermore, PCR amplification of pure genomic DNA with the same primers described above did not generate the 150 or 220 bp segments (data not shown), suggesting the presence of introns flanking the alternative exon in the genomic DNA too large to effectively PCR amplify.

Northern blot analysis was used to examine the *K-2* gene expression pattern (figure 6). The fact that all of the bands are larger than the 1.4 kb transcript of *S8* confirms that this probe, at 62% nucleotide homology with *S8*, does not cross hybridize. RNA isolated from murine fetal heart and pericardium was analyzed and a very faint hybridization signal was detected for a transcript of approximately 4.5 kb (data not shown). The same size transcript appears to be expressed at higher levels in the whole embryo (figure 6). A transcript of approximately 2 kb is also observed in some of the whole embryo samples (14.5 and 15.5 days p.c.). In addition, faint bands (approximately 3.0–3.5 kb) are observed most prominently at 10.5 and 14.5 days p.c. in the whole embryo.

To examine *K-2* gene expression in murine adult tissues we probed a Northern blot (figure 7) with a PCR generated probe from the 3' section of the *K-2* cDNAs. This probe did not contain the homeobox nor did it contain the alternative exon. In the heart and skeletal muscle there were abundant 4.5 kb transcripts and fewer 6 and 3 kb transcripts. The brain, spleen, lung, kidney, and testis showed low levels of these same transcripts but for the liver there was no observed hybridization signal. In the heart and skeletal muscle samples the 4.5 kb transcript was more

Figure 6. Northern blot analysis of *K-2* gene expression. Twenty μ g of total cellular RNA, as determined by spectrophotometric analysis, was loaded in each lane. The RNA was isolated from either whole mouse embryos or heart tissue at 8.5–18.5 days post-coitum. The blot was probed with an NcoI-HindIII fragment from the original PCR *K-2* clone. The 28S and 18S ribosomal bands are marked. The ethidium bromide stained gel is shown below the autoradiogram for comparison of RNA loaded in each lane. The X-ray film was exposed to the Northern blot for four days.

abundant than the 6 kb transcript, yet in the brain, spleen, lung, kidney, and testis samples equal amounts of the 4.5 and 6 kb transcripts were detected.

Finally, to analyze *K-2* gene expression throughout the murine embryo we performed in situ analysis on the day 12 mouse embryo. At this developmental stage, *K-2* signal was highly restricted to mesenchymal derived tissue. In parasagittal sections, high level expression occurred in a thin, uniform band around the entire developing embryo, just beneath the outer primitive ectodermal covering (figure 8a). On cross section of the embryo this signal localized to a broader zone of lateral mesenchymal tissue particularly around the developing neural tube (figure 8b).

Figure 7. Northern blot analysis of RNA isolated from murine adult tissue. Each lane contains 2 μ g of poly A⁺ RNA from either heart (H), brain (B), spleen (S), lung (Lu), liver (Li), skeletal muscle (SM), kidney (K), or testis (T). The probe utilized was PCR generated from the 3' section of *K-2* (nucleotides 630–780 figure 2a) and did not contain either the homeobox or the alternative exon. The X-ray film was exposed to the Northern blot for three days.

This distribution corresponds to the developing dermatomal tissue. The more medial mesenchymal tissue adjacent to the neural tube did not show any expression (figure 8b). There was also strong signal in the mesenchymal tissue of the developing facial structures. This was most notable in the developing maxilla, the mandible, and the tongue. Strong signal was also present in the anterior paravertebral mesenchymal tissue as well as in the pericardium, but not in the heart at this stage of development. Mesenchymal cells within the genital tubercle were strongly positive. No signal was evident in epidermal cells that overlay the expressing mesenchymal cells. None of the developing endodermal-derived tissues of the abdominal or thoracic organs such as the intestinal tract, liver, or lung exhibited *K-2* expression. No expression was observed in developing neural cells. Identical sets of mesenchymal cells were strong *K-2* expressers using either the TA #4 or TA #10 antisense probes. Although the alternative exon is included in TA #10 probe, but not #4, a considerable portion of clone #10 also includes sequences common to both and is therefore not likely a valid test of where the alternative exon is expressed.

DISCUSSION

An RT-PCR cloning strategy using a degenerate primer was used to isolate *K-2*, a novel homeobox gene expressed during heart development. This procedure should facilitate cloning of additional homeobox genes expressed during the development of specific organs of interest. This RT-PCR approach is faster and more sensitive than the conventional method of constructing and screening a cDNA library. Moreover, the HOX 34 primer mix can perfectly complement the homeoboxes of many different classes, such *Antennapedia*, *labial*, *deformed*, *abd-b*, *sex combs reduced*, *fushi tarazu*, *even skipped*, *msh*, and *engrailed* (4). In

addition with the 3' end of the HOX 34 located within the most conserved portion of the homeodomain, it is a functional PCR primer for even more diverse homeoboxes since mismatches in the 5' portion of a primer do not drastically affect primer binding and PCR amplification (38). The *K-2* cDNA, for example, was isolated despite its divergent homeodomain.

Homeodomain sequence comparisons suggest that the *K-2* gene loosely fits within the paired class of homeobox genes, although *K-2* and *S8* appear to represent a distinct subcategory. One of the main differences between typical paired class members and the *K-2/S8* grouping is at position 50 in the homeodomain. Sequence compilation (4), site directed mutagenesis (39,40), and genetic studies (41) have suggested that this position may be the principal source of differences in the DNA binding specificity between the classes of homeodomains. In the paired type homeodomain a serine, or less frequently a lysine, is found in this position but for the *K-2/S8* group it is a glutamine. Altering this position in the paired homeodomain from a serine to a glutamine alters DNA binding specificity *in vitro* (40). It is interesting that all of the genes which have a paired type homeodomain with a serine in position 50 also contain a paired box. Conversely, those without a serine in position 50 of the homeodomain do not contain a paired box. This suggests that *K-2* and *S8* represent the first identified mammalian paired type homeobox genes without a paired box. Then in mammals, as in other organisms, the paired box and the paired type homeodomain can be present each alone or in combination.

The *K-2* gene is alternatively spliced with a 72 bp exon 3' of the homeobox. This splicing event generates two different transcripts, but with a difference of only 72 bp this would not be detected on Northern blots. These two transcripts encode identical homeodomains and only differ at the carboxy terminus. The functional significance of this alternative processing remains to be determined.

The Northern blot analysis of *K-2* gene expression demonstrated that *K-2* is expressed in adult and embryonic tissues. In embryonic tissue two major transcripts (2 and 4.5 kb) are observed whereas in adult tissue one transcript predominates (4.5 kb) with two minor transcripts also detected at 6 and 3 kb. The heart and skeletal muscle of adult mice appear to express high levels of *K-2* relative to other tissues.

In situ hybridizations revealed that the *K-2* gene expression pattern at day 12.5 of development is quite similar to that described for *S8*. In contrast with most homeobox genes, *K-2* and *S8* do not appear to be expressed in the developing nervous system. High levels of expression were, however, detected in the the developing mandible, tongue, facial structures, and genital tubercles. Nevertheless, *K-2* and *S8* gene expression patterns were not identical, with *K-2* expressed in the pericardium but not in the heart at day 12.5 of development, when *S8* is expressed in the mesenchymal regions of the heart but not in the pericardium.

Mutations in several paired type genes have determined that the genes are critically important in developmental processes. For example, mutation in the murine *Pax-1* gene has been associated with the undulated phenotype (42), mutations in the murine *Pax-6* gene and its human homologue are associated with the small eye phenotype in mice (43) and aniridia in humans (44), and mutations in the *Pax-3* gene and its human homologue are associated with the splotch phenotype in mice (45) and Wardenburg syndrome in humans (45). *Pax-3* (30) and *Pax-7* have very similar amino acid sequences and expression patterns,

A

B

yet mutation in the *Pax-3* gene generates a specific phenotype (45). This suggests that although *Pax-3* and *Pax-7* are very similar they are at least somewhat functionally distinct. The *K-2* and *S8* genes also encode very similar amino acid sequences and our initial expression studies suggest that *K-2* and *S8* have similar gene expression patterns. Mutational analysis, by targeted modification in ES cells, or by the identification of pre-existing mutants, will help to elucidate the functions of the *K-2* and *S8* genes in development.

ACKNOWLEDGMENTS

We thank Dr Brigid Hogan for the generous gift of the 8.5 day mouse embryo cDNA library. M.J.K. is a post-doctoral fellow in the P.O.E. Heart and Lung Training program. This work was supported by NIH grants HD 24517 and HL 41496 (SSP), the Pediatric Pathology Society (DAW), and the March of Dimes Birth Defects Foundation (1-0346, BA).

REFERENCES

- Gehring, W.J. (1987) *Science*, **236**, 1245–1252.
- Peifer, M., F. Karch, and W. Bender (1987) *Genes and Development*, **1**, 891–898.
- Kessel, M. and P. Gruss (1990) *Science*, **249**, 374–379.
- Laughon, A. (1991) *Biochemistry*, **30**, 11357–11367.
- Scott, M.P., J.W. Tamkun, and G.W. Hartzell (1989) *Biochimica et Biophysica Acta*, **989**, 25–48.
- Hoey, I. and M. Levine (1988) *Nature*, **322**, 858–861.
- Zappavigna, V., et al. (1991) *EMBO J.*, **10**, 4177–4187.
- Kuziora, M.A. and W. McGinnis (1989) *Cell*, **59**, 563–571.
- Hayashi, S. and M.P. Scott (1990) *Cell*, **63**, 883–894.
- Budd, P.S. and I.J. Jackson (1991) *TIG*, **7**, 74–76.
- Gehring, W.J. (1985) *Sci. Amer.*, **253**, 153–162.
- De Robertis, E.M., G. Oliver, and C.V.E. Wright (1990) *Sci. Amer.*, July, 46–52.
- Monaghan, A.P., et al. (1991) *Development*, **112**, 1053–1061.
- Kim, Y. and M. Nirenberg (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7716–7720.
- Garcia-Fernández, J., J. Baguñó, and E. Salù (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 7338–7342.
- Singh, G., et al. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 10706–10710.
- Burglin, T.R., et al. (1989) *Nature*, **341**, 239–243.
- Murtha, M.T., J.L. Leckman, and F.H. Ruddle (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 10711–10715.
- Wedeen, C.J., et al. (1990) *Nucleic Acids Res.*, **18**, 1908.
- Kamb, A., et al. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 4372–4376.
- Chomczynski, P. and N. Sacchi (1987) *Anal. Biochem.*, **162**, 156–159.
- Saiki, R.K., et al. (1988) *Science*, **239**, 487–491.
- Sambrook, J., E.F. Fritsch, and T. Maniatis, *Molecular cloning: A laboratory manual*. 2nd ed ed. 1989, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Pearson, W.R. and D.J. Lipman (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2444–2448.
- Krumlauf, R., et al. (1987) *Development*, **99**, 603–617.
- Witte, D.P., et al. (1991) *J. Cell. Biol.*, **115**, 179–190.
- Kozak, M. (1984) *Nucleic Acids Res.*, **12**, 857–872.
- Cavener, D.R. (1987) *Nucleic Acids Res.*, **15**, 1353–1361.
- Opstelten, D.J., et al. (1991) *Mech. Dev.*, **34**, 29–42.
- Goulding, M.D., et al. (1991) *EMBO J.*, **10**, 1135–1147.
- Jostes, B., C. Walther, and P. Gruss (1991) *Mech. Dev.*, **33**, 27–38.
- Walther, C. and P. Gruss (1991) *Development*, **113**, 1435–1449.
- Bopp, D., et al. (1986) *Cell*, **47**, 1033–1040.
- Baumgartner, S., et al. (1987) *Genes and Development*, **1**, 1247–1267.
- Finkelstein, R., et al. (1990) *Genes and Development*, **4**, 1516–1527.
- Rosa, F.M. (1989) *Cell*, **57**, 965–974.
- Hawkins, N.C. and J.D. McGhee (1990) *Nucleic Acids Res.*, **18**, 6101–6106.
- Sommer, R. and D. Tautz (1989) *Nucleic Acids Res.*, **17**, 6749.
- Hanes, S.D. and R. Brent (1989) *Cell*, **57**, 1275–1283.
- Treisman, J., et al. (1989) *Cell*, **59**, 553–562.
- Hanes, S.D. and R. Brent (1991) *Science*, **251**, 426–430.
- Chalepakakis, G., et al. (1991) *Cell*, **66**, 873–884.
- Hill, R.E., et al. (1991) *Nature*, **354**, 522–525.
- Ton, C., et al. (1991) *Cell*, **67**, 1059–1074.
- Epstein, D.J., M. Vekemans, and P. Gros (1991) *Cell*, **67**, 767–774.

Figure 8. Expression of *K-2* in the 12 day mouse embryo is restricted to mesenchymal cell types. White grains indicate the presence of hybridizing antisense probe (TA #10) to *K2* mRNA. Panel A: Composite low power darkfield illumination of a parasagittal section through the mouse embryo. High level expression occurred in a thin, uniform band of mesenchymal cells immediately beneath the outer primitive ectodermal covering (triple arrows). Similar localization occurred in pericardial cells (solid arrowhead). Deeper mesenchymal cells were also strongly positive in the developing maxilla (MX), the mandible (M), tongue (T), genital eminence (G), and prevertebral mesenchyme (P). No signal was evident in developing abdominal or thoracic organs such as the intestinal tract, liver, lung and heart (H). Panel B: Cross section of the embryo showing *K-2* expression in lateral mesenchymal tissue around the developing neural tube (NT). This distribution corresponds to the developing dermatome tissue (D). The more medial mesenchymal tissue is negative. Parallel sections hybridized to sense probes TA2 & 3 exposed for the same length of time were indistinguishable from background. The sections were taken from 7mm long embryos.