JOURNAL OF BONE AND MINERAL RESEARCH
Volume 14, Number 4, 1999
Blackwell Science, Inc.

Characterization of Mouse Cathepsin K Gene, the Gene Promoter, and the Gene Expression

YI-PING LI[1,2] and WEI CHEN[1]

ABSTRACT

Cathepsin K, a lysosomal cysteine protease, is abundantly and selectively expressed in osteoclasts and has a specialized role in osteoclast-mediated bone resorption. In contrast to function studies, transcription regulation of cathepsin K remains largely unknown. In this study, the gene encoding mouse cathepsin K and the promoter have been isolated and completely sequenced. In addition, the temporal and spatial expressions of cathepsin K have been characterized. Intrachromosomal mapping studies revealed that the gene contains eight exons and seven introns spanning ~10.6 kb of genomic DNA, a genomic organization that was highly conserved with respect to its human homology. Analysis of the 9 kb 5′ flanking region indicates that this gene lacks canonical TATA and CAAT boxes and contains multiple putative transcription regulatory elements which are also present in the comparable position of 5′ flanking region of human cathepsin K gene. Mouse cathepsin K was found to be a single-copy gene. Northern blot analysis of RNAs from a number of mouse tissues revealed that cathepsin K mRNA is selectively expressed in osteoclast. The selective expression of cathepsin K was confirmed by anticathepsin K immunohistochemical staining. The sequence of cathepsin K expression was linked to osteoclast differentiation in vivo and in vitro by a tartrate-resistant acid phosphatase–anticathepsin K dual immunostaining technique. Cathepsin K is initially expressed at the preosteoclast stage and throughout the mature osteoclast stage. The primer extension assay indicated a major transcription start site 58 bp upstream of the initiator Met codon. The characterization of the cathepsin K gene, its promoter, and the temporal and spatial expression may provide valuable insights into its osteoclast-specific expression and the molecular mechanisms responsible for osteoclast activation. (J Bone Miner Res 1999; 14:487–499)

INTRODUCTION

RECENTLY, A cDNA ENCODING a novel human cysteine protease was cloned independently by several groups including ours,[1–4] and was variously named cathepsin O2, cathepsin K, cathepsin X, and cathepsin O. Cathepsin K is used in this paper. All four sequences are identical and represent the human equivalent of the OC2 gene cloned from rabbit osteoclasts.[5] High-level expression of this enzyme has been detected in osteoclasts, but the level was low in kidney, and cathepsin K was absent from other human cell lines and tissues.[3] In contrast, mRNAs for cathepsin L, B, and S were either absent or expressed at very low levels in osteoclasts.[6] Bromme et al.[1] demonstrated that cathepsin K is a highly active cysteine protease that is capable of hydrolyzing extracellular matrix proteins at pH 5.5.

The high-level expression of the cathepsin K protein in osteoclasts, as well as its enzymatic properties, implies that it has a key role in normal bone remodeling and in pathological processes, such as osteoporosis and osteoarthritis. The key role of cathepsin K in osteoclast function is further underscored by the finding that mutations in this gene cause pycnodysostosis, a human disorder characterized by short stature, osteosclerosis, bone fragility, and abnormal bone and tooth development.[7,8] Inui et al.[9] provided further evidence that cathepsin K plays a crucial role in osteoclastic function by employing cathepsin K antisense oligonucleotides to inhibit bone resorption.

[1]Department of Cytokine Biology, Forsyth Dental Center, Boston, Massachusetts, U.S.A.
[2]Harvard-Forsyth Department of Oral Biology, Harvard School of Dental Medicine, Boston, Massachusetts, U.S.A.

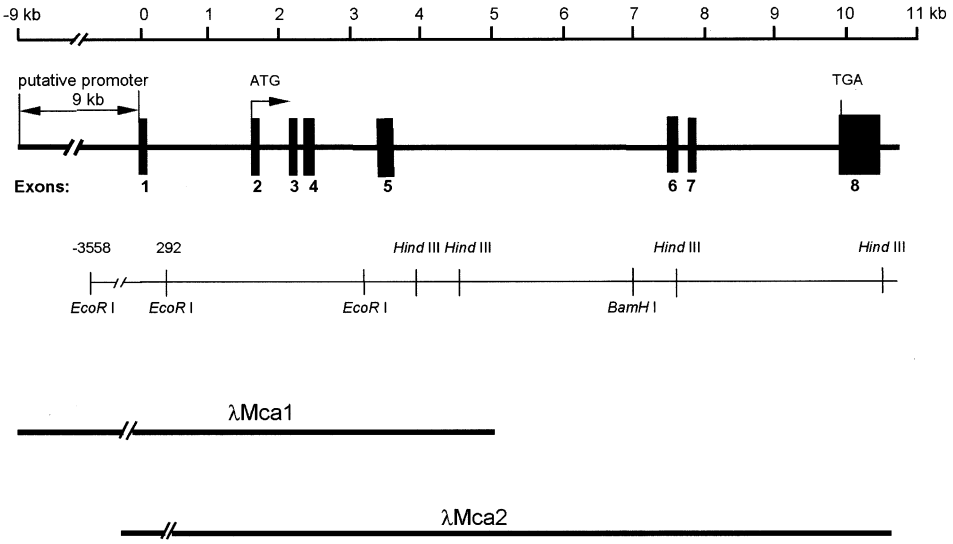

FIG. 1. Genomic organization and restriction map of mouse cathepsin K. The cathepsin K gene is composed of eight exons and seven introns distributed over ~10.6 kb. Solid bars represent exons. The locations of the initiator MET code and stop code TGA are indicated.

There has been much interest in cathepsin K, since the enzyme has been strongly implicated as a major protease responsible for bone resorption. However, the genomic sequence of cathepsin K has not been characterized, and the cathepsin K gene expression regulation in the osteoclast remains unknown. To approach an understanding of the factors modulating cathepsin K gene expression and to better define the function of cathepsin K, the present studies were initiated to characterize the mouse cathepsin K gene together with its regulatory region and the temporal and spatial expression of the cathepsin K gene.

MATERIALS AND METHODS

Isolation of mouse cathepsin K gene

A mouse 129/SVJ genomic library, constructed on lambda FIX II (Stratagene, La Jolla, CA, U.S.A.), was used for cloning mouse cathepsin K gene using a 1600-bp human cathepsin K cDNA ^{32}P-labeled (DuPont NEN, Boston, MA, U.S.A.) by random primer labeling kit. About 1×10^6 plaques were screened with a ^{32}P-labeled probe derived from the human full-length cathepsin K cDNA. The positive clones for mouse cathepsin K were isolated and subcloned into pBluescript (SK) vector (Stratagene). The structure of the isolated genes was determined by restriction endonuclease mapping and Southern blot analysis. Restriction endonuclease–digested DNA was fractionated on 0.8% agarose gel and transferred to nylon membranes. The DNA fragments containing portions of cathepsin K cDNA were identified by hybridization with nick-translated DNA probes from different portions of the cathepsin K cDNA. Standard methods were used for library screening, purification of phage DNA inserts, mapping of restriction enzyme cut sites, and Southern blotting.[10]

DNA sequencing

A 14-kb fragment containing the 5′ end of the mouse cathepsin K gene and the first 5 exons, and a 15 kb fragment containing the rest of the exons were isolated from a bacteriophage clone (Fig. 1) and subcloned into pBluescript SKII (Stratagene). All of these subclones were sequenced as double-stranded templates with Sequenase in combination with dideoxynucleotide chain terminators according to the supplier's recommendation (U.S. Biochemical Corp., Cleveland, OH, U.S.A.) or with a 377 sequencer (ABI Advanced Biotechnologies, Inc., Columbia, MD, U.S.A.). Initial DNA sequencing was performed with T7 and T3 primers. Subsequent reverse orientation sequencing was performed using insert-specific primers. The DNA sequence was assembled and analyzed for the location of exons and introns by comparison with the sequence of mouse cathepsin K cDNA using the DNAsis II program (Hitachi, Tokyo, Japan). The 5′-flanking region of the gene was searched for transcriptional factor binding sites using the DNAsis II program.

Northern blotting

Mouse tissue including brain, liver, heart, lung, kidney, spleen, and skeletal muscle total RNA and tartrate-resistant acid phosphatase ($TRAP^+$) multinuclear osteoclast-like cells (OCLs) total RNA derived from mouse bone marrow cocultures were prepared as described.[11] Northern blot analysis of murine cathepsin K was performed as described.[12] Fifteen micrograms of total RNA was fractionated by electrophoresis through 1% agarose/formaldehyde gels followed by Northern blot transfer to nitrocellulose membranes presoaked in 20× SSC and then covalently bound by baking for 1 h at 80°C. Visualization by ethidium bromide was used to evaluate the integrity of

mRNA and equal loading per lane. The nylon membrane was hybridized with radioactively labeled human cathepsin K as a probe. After washing, the filters were exposed to X-ray film (Fuji, Tokyo, Japan) at −80°C for autoradiography.

Identification of the transcription start site of mouse cathepsin K gene in osteoclast

The initiation sites for the transcription of mouse cathepsin K genes in osteoclasts were analyzed by primer extension as described.[13] Two synthetic 30-base oligomers were used. They are located at −1 to −30 (5′-CCTGCTGTTGAGAATCTGTTCGCTAGGCTC-3′) and −47 to −76 (5′-GTGGCTACTGTGAGCGGAAGACTAAGGGTG-3′) of the 5′-untranslated region of the mouse cathepsin K cDNA according to our genomic sequences and cDNA sequences.[14] The primers were end-labeled with [λ-^{32}P]ATP using T4 polynucleotide kinase (Life Technologies, Inc., Grand Island, NY, U.S.A.). The ^{32}P-labeled primers were hybridized with 10 μg of total RNA from mouse OCLs prepared as described[11] and extended using avian myeloblastosis virus (AMV) reverse transcriptase. The extended cDNA product was analyzed on a sequencing gel with DNA sequence reactions as size markers.

TRAP staining and cathepsin K immunostaining

TRAP staining was performed using a Kit from Sigma according to the manufacturer's instructions. Cathepsin K was expressed and purified from *Escherichia coli* using the QIA expression system (Qiagen, Inc., Chatsworth, CA, U.S.A.). This protein was then used to generate an anticathepsin K polyclonal antibody in rabbits. The antibody selectively reacts with human and mouse osteoclast. To link expression of cathepsin K to the process of osteoclast differentiation, a dual staining technique, i.e., anticathepsin K immunostaining after TRAP staining, was employed. Freshly isolated metatarsals from mouse embryonic day 15.5. (E 15.5) through embryonic day 18.5 (E 18.5) were used for testing the sequence of cathepsin K expression during osteoclast differentiation in vivo by the dual staining. Mouse metatarsals were fixed and stained by TRAP. The same section was used for anticathepsin K immunostaining. Endogenous peroxidase activity was quenched with 0.3% hydrogen peroxide in methanol for 20 minutes. Nonspecific background staining was blocked by incubating sections in 10% goat serum in PBS for 20 minutes. Primary antibody (anti-human cathepsin K, 1:100) was applied for 1 h at room temperature. The horseradish peroxidase avidin-biotin complex system (Rabbit Elite ABC Kit; Vector Laboratories, Burlingame, CA, U.S.A.) 3,3′-diaminobenzidine was used to visualize bound antibody.

RESULTS

Isolation of the mouse cathepsin K genomic clone

To establish the mouse cathepsin K gene copy number, Southern analysis of mouse genomic DNA was performed with a panel of restriction enzymes. Hybridization, using a 1.6-kb human cathepsin K DNA as the probe, revealed the restriction enzyme fragments that were consistent with the restriction map of the genomic clones shown in Fig. 1. The absence of additional bands for most digestion indicated that there is a single gene for mouse cathepsin K (data not shown).

Eight lambda bacteriophage clones were identified and isolated from the Sv129 strain mouse genomic library using a full-length human cathepsin K cDNA probe. The DNA from the eight lambda clones was hybridized with different portions of the cDNA by Southern analysis. This resulted in the isolation of two nonidentical bacteriophage clones (λMca1 and λMca2) with overlapping inserts spanning ~30 kb of genomic DNA (Fig. 1). The approximate location of the 5′ end of the mouse gene and exons within the cloned region was determined by Southern blot analysis using different portions of the human cathepsin K cDNA as probes. The restriction map generated for these clones by Southern blot analysis was consistent with the genomic restriction fragments observed with the genomic Southern blot described above (Fig. 1).

Gene structure and sequence

The *Eco*RI genomic DNA fragments from λMca1 and λMca2 were subcloned into pBluescript IIKS (+) (Stratagene) for sequencing. A primer walking strategy was employed to obtain 19.6-kb contiguous sequences covering all exons, introns, 9.0 kb 5′ flanking sequences of the gene to the transcription start site and 0.5 kb 3′ sequences to the first polyadenylation signal (Figs. 2 and 3). The genomic sequence was divided into eight exons ranging from 57 to 561 nucleotides (Table 1). Exon 1 contains the 5′ untranslated region. Exon 2 contains the translation initiation site, all codons for the signaling peptide, and a portion of the prepeptide (Fig. 3). Exon 7 encodes the C terminus of the propeptide, and exon 3 encodes the N terminus to the mature enzyme. Exon 8 contains the stop codon and 3′ untranslated region (Fig. 3). The seven introns ranged in size from 84 to 3809 nucleotides, with the entire gene ~10.6 kb (Figs. 1 and 3 and Table 1). Sequences of the genomic clones were used to construct a partial restriction map of the entire cathepsin K gene (Fig. 1). The restriction map was confirmed by Southern blot using different portions of human cathepsin K cDNA as the probes.

We have determined the exon–intron boundaries of mouse cathepsin K. As shown in Table 1, the exon–intron boundary sequences conformed to the GT/AG rule with a single exception. The 5′ splicing donor site of exon 3 was nonconforming, containing the sequence TGgt. A TGgc pattern was observed in the 5′ splicing donor site of exon 3 of the human gene.[15,16] The boundary sequences were generally consistent with the 5′ and 3′ splice consensus sequences. Variations from the consensus sequences were all permissible for primate exon–intron boundary sequences. Intron 1 is present in the 5′ noncoding region of the primary transcript. Introns 2–5 disrupt each exon between amino acids. Intron 6 interrupts a codon between the first and the second nucleotide, and intron 7 disrupts a code between the

FIG. 2. Nucleotide sequence of the 5′ flanking region of the mouse cathepsin gene. Numbers on the left are the nucleotide positions relative to the transcriptional initiation site. Putative regulatory elements are underlined.

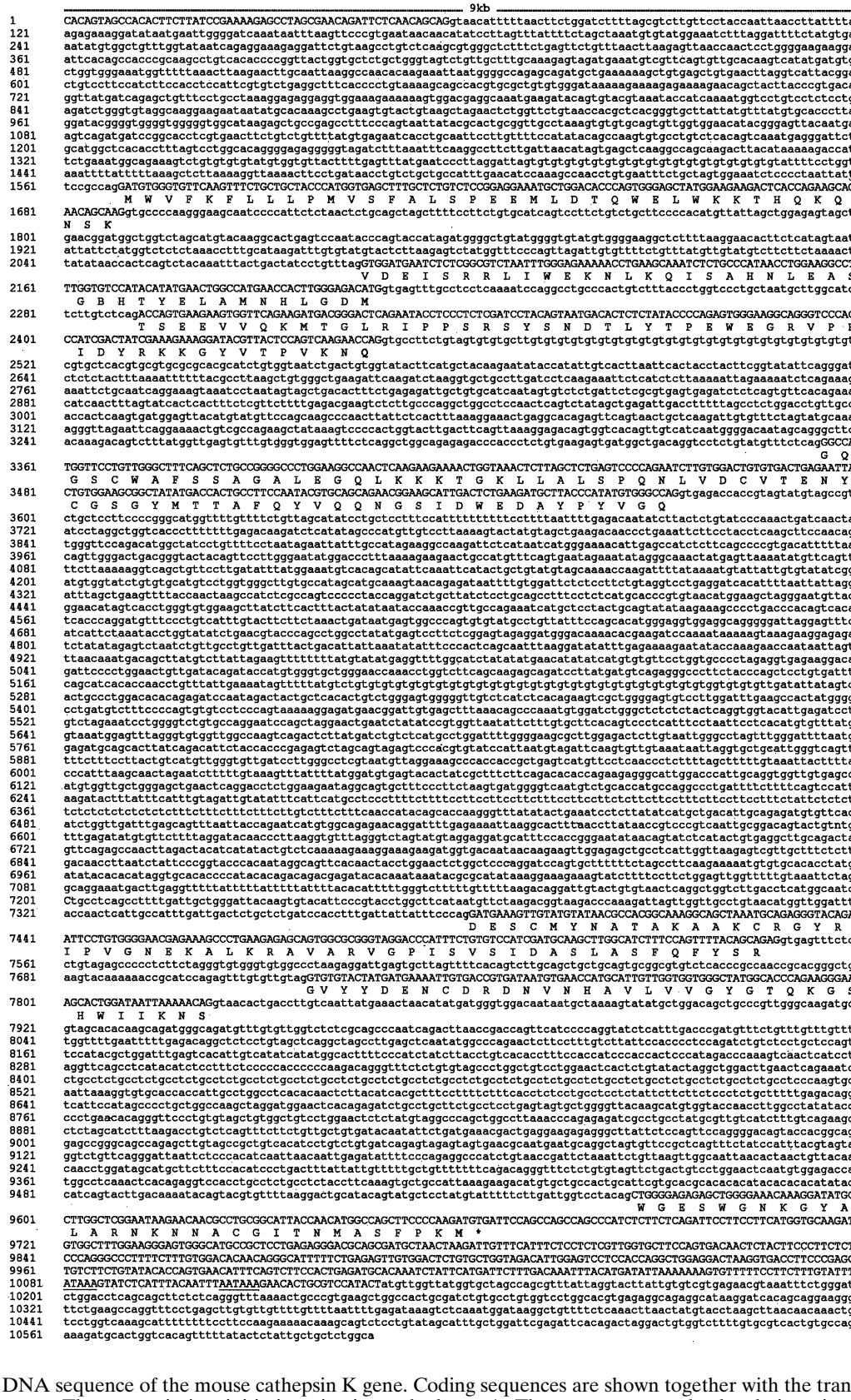

FIG. 3. DNA sequence of the mouse cathepsin K gene. Coding sequences are shown together with the translated amino acid sequence. The transcription initiation site is marked as +1. The two census polyadenylation signal sequences (AATAAA) are underlined. The 3′ flanking sequence beyond the polyadenylation site is shown below. Exons are indicated by capital letters. The 5′ flanking sequence, introns and 3′ flanking sequence are indicated by small letters.

TABLE 1. EXON–INTRON JUNCTIONS OF THE MOUSE CATHEPSIN K GENE

Exon number and (size) (nt)	*cDNA position of exon*	*5′ splicing site*	*Intron number and (size) (kb)*	*3′ splicing site*
1 (57)	78–134	CAACAGCAG gtaacattt	1 (1.52)	ttccgccag GATGTGGGT
2 (121)	135–255	ACAGCCAAG gtgccccaa	2 (0.395)	cctgtttag GTGGATGAA
3 (123)	256–378	GGAGACATG gtgagtttg	3 (0.084)	ttgtctcag ACCAGTGAA
4 (156)	379–534	AAGAACCAG gtgccttct	4 (0.905)	gtttctcag GGCCAGTGT
5 (219)	535–753	GTGGGCCAG gtgagacca	5 (3.809)	atttcccag GATGAAAGT
6 (166)	754–919	ACAGCAGAG gtgagtttc	6 (0.170)	gtgttgtag GTGTGTACT
7 (106)	920–1025	TAAAAACAG gtaacactg	7 (1.742)	gtcctacag CTGGGGAGA
8 (561)	1026–1586			
Consensus sequence:		donor: $^{A}_{C}$AG gt$^{a}_{g}$		acceptor: $^{c}_{t}$ag G

second and the third nucleotide (Fig. 3). There were two putative consensus polyadenylation signals (AATAAA) located 42 and 18 bp upstream from the poly(A) tract (Fig. 3).

Alignment of the mouse cathepsin K gene exon sequences to the human cDNA revealed that there is strong homology at both the nucleotides of code regions (86%) and amino acid (84%) levels. The sequence homology with the human gene at the 5′ and 3′ untranslated region is 71% and 56%, respectively. Substantial homology extends into the 5′-flanking region, indicating functional importance for these sequences (Fig. 4). The deduced amino acid sequence for the nucleotide following the ATG initiation code shows 97% homology with the mouse cathepsin K cDNA.[17] The nucleotide sequence of the exons shares 98% homology with that of mouse calvarial-deprived cDNA.[17] The first exon of the mouse is 9 bp longer than that of the human. Exons 2–7 in the human and mouse cathepsin K genes are identical in size. Exon 8 is 104 bp shorter than that of human[15] (Fig. 1 and Table 1). In contrast to the strong conservation of the exon sequences, intron sequences vary considerably between the human and the mouse cathepsin K genes. However, the size of introns in comparable positions were similar (Table 1 and Figs. 1 and 3).

Identification of the transcription start site of the cathepsin K gene in osteoclasts by primer extension

To determine the cathepsin K transcription initiation site, primer extension analysis was performed using a mouse cathepsin K–specific antisense oligonucleotide primer. The 5′ untranslated region of the mouse cathepsin K gene was predicted to be 135 bp, based on amplification by 5′ RACE using mRNA from fibroblast by Gelb et al.[14] Our DNA sequencing confirmed the reported first exon sequence. Based on this published exon sequence, we made the antisense primer for the primer extension experiment. Two end-labeled antisense primers, –1 to –30 (5′-CCTGCTGTTGAGAATCTGTTCGCTAGGCTC-3′) and –47 to –76 (5′-GTGGCTACTGTGA-GCGGAAGACTAAGGGTG-3′) relative to the ATG initiation codon, were annealed to total RNA isolated from OCLs. The primers were extended with Moloney murine leukemia virus reverse transcriptase and analyzed by polyacrylamide–urea gel electrophoresis. A major and two minor primer extension products of 58, 57, and 56 nucleotides were observed with the downstream primer corresponding to transcription start site at –58, –57, and –56 bp relative to the ATG initiation code, respectively. No product was observed with the upstream primer. The transcript initiation at –58 bp is therefore the major transcription initiation site in mouse OCLs (Fig. 5).

Structural analysis of the 5′ region of the mouse cathepsin K gene

To identify potential regulatory elements, the 9000 bp 5′ flanking sequence was analyzed by database searches as described in the Materials and Methods. There is no TATA or CCAAT box located 5′ of the transcriptional start site. Other putative regulatory elements, including AP1, AP3, PU.1, Ets1, GATA, and bHLH were found (Fig. 2 and Table 2).

Although the important functional role of transcription factors *c-fos*, PU.1, and NF-κB in osteoclast differentiation has been evaluated in vivo,[18–20] the signals that regulate cathepsin K expression have not been delineated. Since the expression of cathepsin K is related to osteoclast differentiation, putative AP-1, PU.1, and NF-κB response elements were specifically sought in this genomic sequence. Several of AP-1 and PU.1 were found, but only one NF-κB response element, located at –8886, was found. Additional studies are required to evaluate fully the possible role of these *cis* elements in regulating the expression of cathepsin K in osteoclasts. The potential GATA and bHLH binding sites located in the 5′-flanking regulatory region of the mouse cathepsin K gene (Fig. 2 and Table 2) indicate that members of bHLH and GATA transcription factors families may also play an important role in osteoclast development.

Previously, Rood et al.[16] have sequenced 1100 bp 5′ flanking sequence of the human cathepsin K gene. Alignment of the 5′ flanking sequence of the mouse cathepsin K gene with the 1100 bp 5′ flanking sequence of the human gene revealed limited sequence identity, which was re-

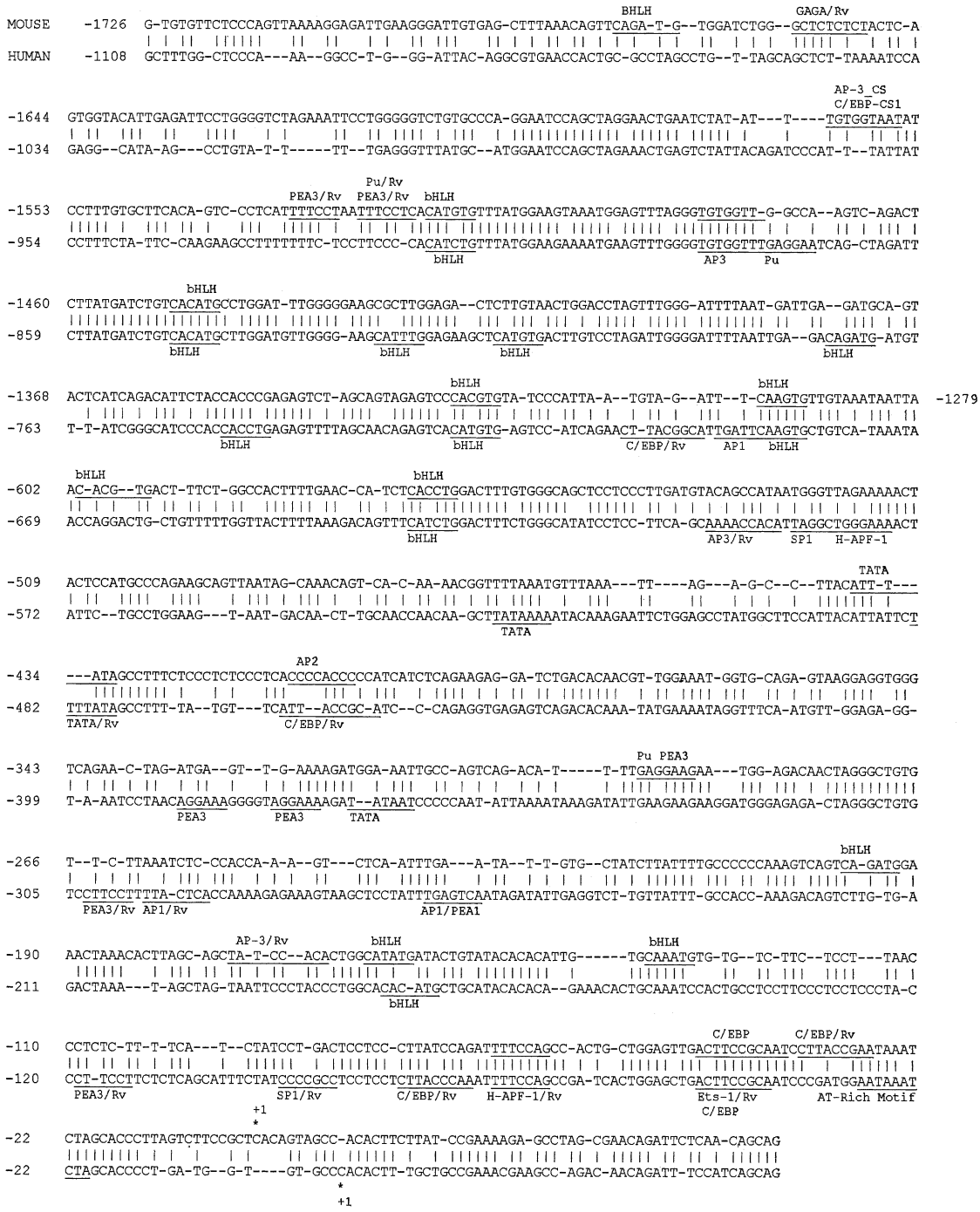

FIG. 4. Cross-species alignment of the sequence representing the 5′-flanking region of the cathepsin K gene. Vertical lines between bases indicate regions of identity between human and mouse. Consensus elements for binding known transcription factors are underlined. The start site of transcription (base +1) for the mouse and human genes is indicated by * located above (mouse, –58 to initiation code ATG) and below (human, –49 to initiation code ATG) in the alignment, respectively. DNA sequences closely corresponding to several consensus sequences are underlined.

stricted primarily to a proximal promoter region –1 to –602 and more distal promoter region –1279 to –1726 (Fig. 4). The two regions are ~63% and 67% conserved, respectively, and contain a few regions of 30–50 bp exhibiting much greater identity (Fig. 4).

Tissue distribution

The tissue and cellular distribution of cathepsin K has been characterized.[3] Mouse tissue distribution of cathepsin K mRNA was investigated by Northern blot analysis,

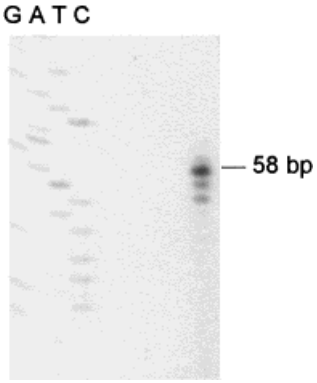

FIG. 5. Identification of the mouse cathepsin K transcription initiation site by primer extension. The primer extension product was determined on denaturing polyacrylamide gel electrophoresis using the adjacent DNA sequencing reaction (GATC) as a size marker. The major primer extension products was 58 bp long, from nucleotide –1 to the putative transcription start site.

using the human full-length cDNA fragment as a hybridization probe. The results demonstrated that the mouse cathepsin K was specifically and highly expressed in OCLs and was undetectable in other tissues including brain, liver, heart, lung, kidney, spleen, and skeletal muscle (Fig. 6). The arrow in Fig. 6 indicates that the 1.6 kb of band cross-hybridizing with human cathepsin K cDNA. This result is consistent with our human data,[3] i.e., the mouse cathepsin K gene is regulated in a tissue- and cell type– specific manner. The mouse tissue distribution of cathepsin K protein was investigated by immunostaining mouse tibia and metatarsals with anti-human cathepsin K polyclonal antibodies. Human cathepsin K was expressed and purified from *E. coli* using QIA expression system (Qiagen). A 38 kDa protein band was visible on SDS-PAGE gel by Coomassie Blue staining (Fig. 7). An anticathepsin K polyclonal antibody was generated in rabbit using the purified cathepsin K protein as antigen. Mouse embryo day-19 sections were incubated with rabbit Polyclonal antibodies and then developed with peroxidase-coupled secondary antibodies as described in the Materials and Methods. Antibodies did not react with extraskeletal tissues or cells, including muscle, intestine, skin, spleen, stromal cells, and other bone marrow cells. In contrast, osteoclasts in the lumbar vertebrae demonstrated intense cathepsin K expression (Fig. 8M).

Sequence of cathepsin K gene expression during osteoclast differentiation in vivo and in vitro

Despite extensively functional studies, the sequence of cathepsin K gene expression still remains unknown. TRAP serves as a molecular marker of osteoclast differentiation and is expressed in both preosteoclasts and osteoclasts. To link expression of cathepsin K to osteoclast differentiation, a dual staining technique, i.e., anti–cathepsin K immunostaining after TRAP staining, was employed.

We confirmed Scheven et al.'s results[21] that $TRAP^+$ preosteoclasts and osteoclasts first appeared in metatarsals on embryonic days 16.5 and 17.5, respectively (Figs. 8A–8C). The same sections that had been used for TRAP staining were utilized for anti–cathepsin K immunostaining. As shown in Figs. 8E and 8F, the black immune deposits for cathepsin K were localized $TRAP^+$ preosteoclast and osteoclast. No antibody reaction was observed on embryo day 15.5 sections (Fig. 8D). Our results indicated that cathepsin K is initially expressed at the preosteoclast stage (E16.5) and throughout mature osteoclast stage (E 17.5) in metatarsals. Sequence of cathepsin K expression also was determined by dual-staining technique in vitro in the MOCP-5 (osteoclast precursor cells) coculture system.[11] In vitro results indicate that TRAP is expressed at the onset of the preosteoclast stage (culture day 4) and through the mature osteoclast stage (culture day 8) (Figs. 8H and 8I) and cathepsin K is expressed at the same osteoclast differentiation stages (Figs. 8K and 8L). The result is consistent with that from metatarsals staining in vivo.

The sequence of cathepsin K expression was further analyzed in a bone marrow coculture system and during mouse embryonic development by Northern blot. Total RNA was isolated from MOCP-5 cocultures on days 1–9 and from undifferentiated MOCP-5 cells. Hybridization was carried out with a ^{32}P-labeled probe from human cathepsin K. As shown in Figs. 9A–9C, the mRNA level of cathepsin K was closely correlated to the number of $TRAP^+$ OCLs in MOCP-5 cocultures. The sequence of cathepsin K expression during mouse embryo development was also analyzed by Northern blot. There is no detectable expression of cathepsin K until E 15.5 (Figs. 9D and 9E). This result was correlated to osteoclast development during mouse development,[22] indicating the unique function of cathepsin K in bone remolding.

DISCUSSION

Cathepsin K has been identified as a critical enzyme in bone degradation and remodeling. The expression of cathepsin K appears to be linked to differentiation states of osteoclasts. In this study, we aimed to characterize the mouse gene with their regulatory regions and to analyze the temporal and spatial cathepsin K expression in vivo and in vitro as a starting point for the investigation of the regulation of the gene. The mouse cathepsin K gene, including ~9 kb of the 5′ flanking region, was sequenced and characterized. We find that the cathepsin K gene has a similar exon–intron arrangement to the recently reported human cathepsin K gene organization.[15,16] The cathepsin K gene is highly conserved between mouse and human. Northern analysis of various mouse tissues demonstrated that cathepsin K is expressed selectively and predominantly in osteoclasts, a finding consistent with a previous observation of the human cathepsin K gene.[3] Homology at the 5′ flanking region between the human and mouse cathepsin K genes indicates that the two genes may have a similar mechanism of transcription regulation. These data suggest that cathepsin K plays a similar function in bone resorption in the human and mouse and that the mouse may provide a

TABLE 2. THE POTENTIAL RECOGNITION SEQUENCES FOR TRANSCRIPTION FACTORS IN THE REGULATION REGION

Name of cis-*element*	*Position*		*Sequence*
	Mouse	*Human*	
C/EBP	–5450		TTTCGTAAG
	–4579		TGAAGCAAT
	–2752		T TC TGCAAT
	–1564		TG TGGTAAT
	–46	–46	TT CCGCAA T
SP1	–2458		GAGGCTGAG
	–2212		GGGGCGGGG
	–882		TGGGCTGGT
	–2209		GGGCGG
		–587	TAGGCTGGG
AP-2	–2483		CCCCAGGC
	–3120,–872		CCCCAGCC
	–871		CCCAGCCG
	–708		CCCCTGCC
	–411		CCCCACCC
	–410		CCCACCCC
AP2/Rv	–4074		GGGTGGGGGA
	–3887		GGGTGGGGGG
	–2213		GGGGGGGCGG
PEA3	–5751,–4647,–4321,–4174,–294,		AGGAAG
	–5422,–5339,–5138,–5018,–4195,–4156,–3999,–3550,–1209	–388,–377	AGGAAA
AP-1	–5100		TGAATCA
	–1192	–264	TGAGTCA
		–692	TGATTCA
AP-1/Rv	–5736		TTAATCA
	–5080		TGAATCA
	–1192	–264	TGAGTCA
		–296	TTACTCA
		–692	TGATTCA
Pu_box	–4648,–4196,–4175,–4154,–3551,–295	–876	GAGGAA
H-APF-1	–1016	–583	CTGGGAA
H-APF-1/Rv	–70,–3738	–71	TTTCCAG
AP-3	–1564		TGTGGTAA
		–884	TGTGGTTT
AP-3/Rv	–172		TATCCACA
		–596	AAACCACA
Ets-1	–4598		GAGGATGC
Ets-1/Rv	–5517		ACTTCCGG
	–1747		GCATCCTG
	–638		GCTTCCTG
		–48	ACTTCCGC
GATA	–5429		AGATAAC
	–5068		TGATAG
	–4170		AGATAG
	–2669		TGATAA
bHLH	–1026	–810,–720	CATGTG
	–160		CATATG
	–1044,–198	–773	CAGATG
	–569	–747	CACCTG
	–601		CACGTG
	–133		CAAATG
PEA1	–1192	–264	TGAGTCA
NFκB	–8886		GGGACTGCCC

good model to study the role of cathepsin K in normal bone remolding and in pathological processes.

This is the first report describing the genomic sequence of the cathepsin K gene. The genomic organization of cathepsin K gene was reported,[14–16] but the genomic sequence of the human and mouse cathepsin K gene have not been reported. Gelb et al.[14] reported a partial characterization of the mouse cathepsin K genomic organization determined by polymerase chain reaction. The exon–intron junctions and the exact sizes of exons and introns of the

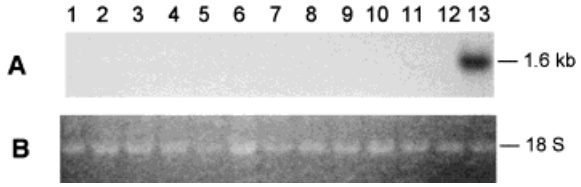

FIG. 6. Northern hybridization of human cathepsin K cDNA to total cell RNA from cell lines and mouse tissues. Fifteen micrograms of total RNA from each mouse tissue were blotted onto a nylon filter: lane 1, testes; lane 2, brain; lane 3, skeletal muscle; lane 4, liver; lane 5, spleen; lane 6, intestine; lane 7, kidney; lane 8, thymus; lane 9, lung; lane 10, skin; lane 11, MS12 stromal cells; lane 12, undifferentiated MOCP-5; lane 13, purified OCLs differentiated from MOCP-5. (A) Autoradiography after hybridization with human cathepsin K cDNA probe; the arrow indicates the 1.6 kb band. (B) Ethidium bromide–stained gel showing equivalent RNA loading.

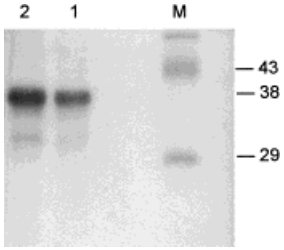

FIG. 7. Expression and purification of human cathepsin K from *E. coli*. Lanes 1 and 2: first and second eluted fractions from the Ni-NTA resin, respectively. Note 38 kDa translation product in fractions. Lane M: MW standards.

mouse cathepsin K gene were not reported. The size of intron 5 of mouse cathepsin K gene is 3.809 kb (Fig. 1). The intron size is consistent with the size of the intron 5 of human cathepsin K gene,[16] but is different from that of Gelb et al.'s report (~0.7 kb).[14]

Restricted tissue distribution of cathepsin K has been previously reported in humans.[3] Our results indicated that there is no notable difference between tissue distribution of human cathepsin K and that of mouse cathepsin K. Using in situ hybridization approach, Dodds et al.[22] reported recently that in the 11–13 day mouse fetuses, cathepsin K mRNA was not expressed in any extraskeletal tissue. However, they found that developmental stage-dependent pattern of cathepsin K expression was observed in osteoclast and preosteoclasts at sites of cartilage and bone modeling in the 15–17 day fetuses; no expression was observed in any other nonskeletal tissue at these time points. We used Northern blot analysis of cathepsin K expression to reveal the limited expression in osteoclast and sequence of cathepsin K expression during embryo development (Figs 6 and 9), and reached the conclusions similar to those in Dodds et al.'s report.[22] The results of the immunohistochemical analysis at cathepsin K protein level (Fig. 8) add further support to these conclusions. To define the sequence of cathepsin K gene expression, we have used the TRAP/anti–cathepsin K dual-staining technique to link directly cathepsin K expression to TRAP expression. The high and selective gene expression in osteoclasts of both human and mouse suggests that cathepsin K plays an important role in bone resorption in the species.

Primer extension reactions employing OCLs RNA as a template were used to determine the transcription start site of the mouse cathepsin K gene. The transcription start site was identified 58 bp up 5′ to the initiator ATG, similar to the transcription start site of human cathepsin K gene at human osteoclast.[16] Interestingly, Gelb et al.[14] reported a mouse fibroblast cathepsin K cDNA, which contains 135 bp before the first encoded methionine. This suggests that the first exon should be at least 135 bp in length. We have confirmed our results through additional experiments. An antisense primer positioned –47 to –76 (5′-GTGGCTACTGTGAGCGGAAGACTAAGGGTG-3′) to the start ATG did not yield primer extension products when OCLs RNA was used as a template. Since the 135 bp 5′ cDNA sequence reported by Gelb et al.[14] is identical to that observed upstream in our genomic clone (Fig. 2), all of these data are consistent with the conclusion that different tissues utilize different transcription start sites for the initiation of cathepsin K gene expression. Whether the differential initiation sites in different tissues reflect different elements promoting gene expression remains to be determined.

The signals that regulate the expression of cathepsin K are unknown. However, recently genetically engineered osteopetrotic mutant mice have yielded important insights into the regulation of osteoclast differentiation. Disruption of the c-*fos* proto-oncogene, which is a component of the AP-1 complex, blocks osteoclast development at the point of divergence from the common MO-CSF precursor, prior to the expression of osteoclast-gene (e.g., cathepsin K, and TRAP), and leads to an osteopetrotic phenotype.[18] PU.1 knockout mice are osteopetrotic and are devoid of both osteoclasts and macrophages.[20] Iotsova et al.[19] reported that mice lacking NF-κB1 and NF-κB2 (double-knockout mice) developed osteopetrosis because of a defect in osteoclast differentiation. Since expression of cathepsin K is correlated to osteoclast differentiation, the osteopetrotic phenotype of these knockout mice suggests that AP-1, PU.1, and NF-κB sequences may exist in the promoter region of cathepsin K and may be critical in regulating expression of the cathepsin K gene. Therefore, AP-1, PU.1, and NF-κB binding site sequences were specifically sought by computer search and several consensus sequences for AP-1 and PU.1 binding site were found (Table 2). However, only a NF-κB sequence was found at the more distal promoter region (–8886). Further studies are required to characterize and examine the function of these putative regulatory elements in controlling cathepsin K gene expression. Like the putative promoter region for human cathepsin B, S,[23,24] and K,[15,16] the mouse cathepsin K 5′-flanking region also shows no conventional TATA or CAAT boxes typical of most RNA polymerase II transcriptional units. The 5′-flanking regulation region of the cathepsin K gene contains a low GC content (~45%) and multiple regulatory elements (Table 2), indicating that cathepsin K is not a constitutive gene product. The restricted tissue and cellular distribution

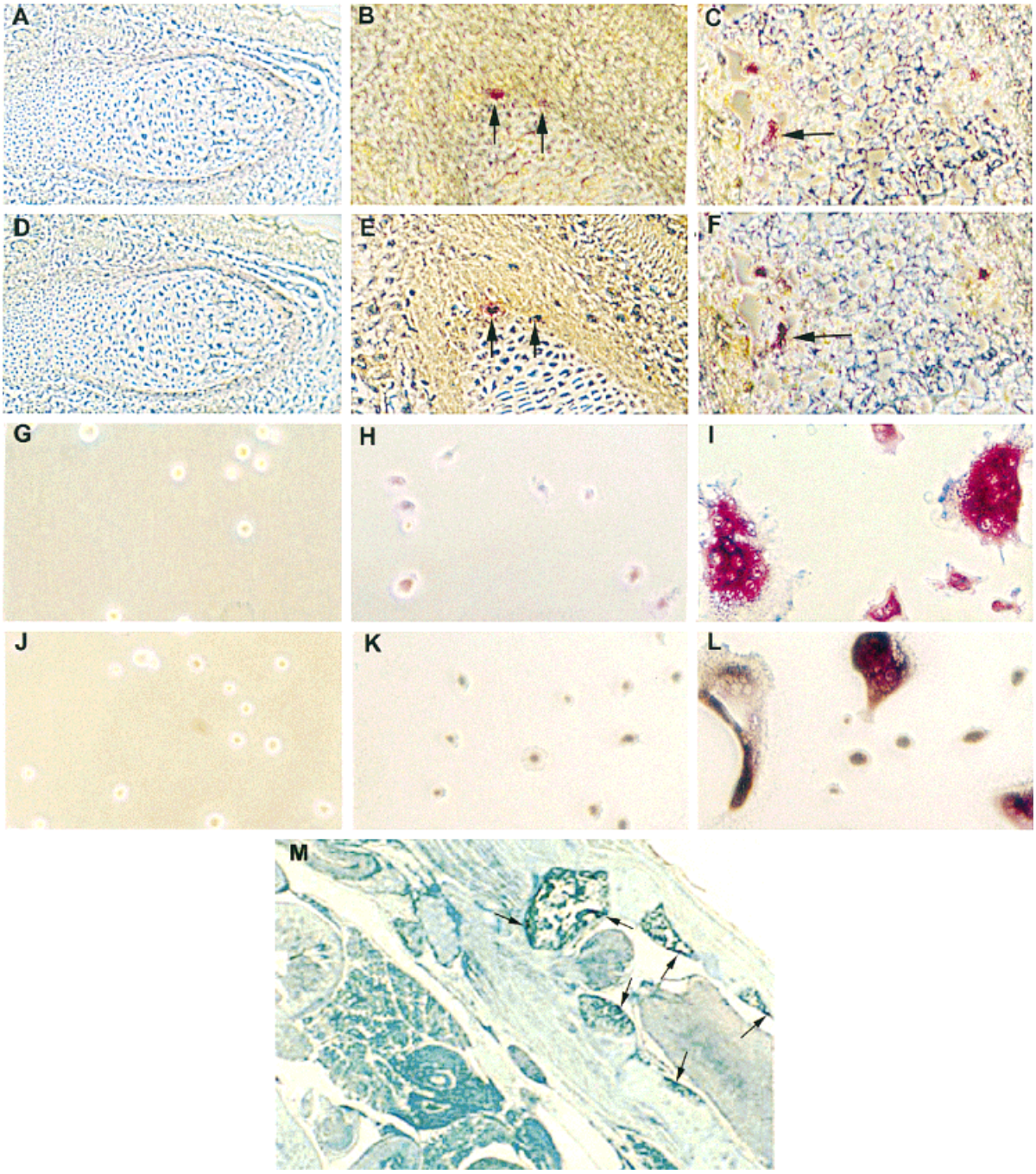

FIG. 8. Characterization of the cathepsin K gene expression by immunostaining mouse tissue and definition of sequence of cathepsin K gene expression during osteoclast differentiation in vivo and in vitro by TRAP/anticathepsin K dual staining. (A–F) Freshly isolated E15.5 (A), E16.5 (B) and E17.5 (C) metatarsals were used to examine the sequence of cathepsin expression. On E15.5 there are no TRAP positive cells (A); expression of TRAP in preosteoclasts first appeared in the periosteum on E16.5 (B, arrow). TRAP positive osteoclasts appeared in bone marrow on E17.5 (C, arrow). Same sections were used to assess cathepsin K expression by immunolocalization. Black immunodeposits for cathepsin K are localized to the $TRAP^+$ mononuclear cells (E, arrow) in the periosteum of metatarsals on E16.5, and to osteoclasts on E17.5 (F, arrow). No staining was seen on E15.5 (D). Original magnification (A, D) ×100; (B, C, E, F) ×200. (G–L) TRAP staining of undifferentiated MOCP-5 cells (G); four day differentiated MOCP-5 cells (H, mostly mononuclear cells); and 7-day differentiated MOCP-5 cells (I). Undifferentiated MOCP-5 cells were $TRAP^-$ (G). Both OCLs (I) and preosteoclasts (H) were $TRAP^+$. After TRAP staining, the cells were used for anticathepsin K staining. Black immunodeposits for cathepsin K were localized to $TRAP^+$ preosteoclasts (K) and osteoclasts (L), but not to undifferentiated MOCP-5 (J). Original magnification: ×200. (M) Mouse embryo day 19 sections were incubated with rabbit Polyclonal antibodies and then developed with peroxidase-coupled secondary antibodies. Antibodies did not react with extraskeletal tissues or cells, including muscle, intestine, skin, spleen, stromal cells, and other bone marrow cells. In contrast, osteoclasts in the lumbar vertebrae demonstrated intense cathepsin K expression (M, arrow). Original magnification: ×15.

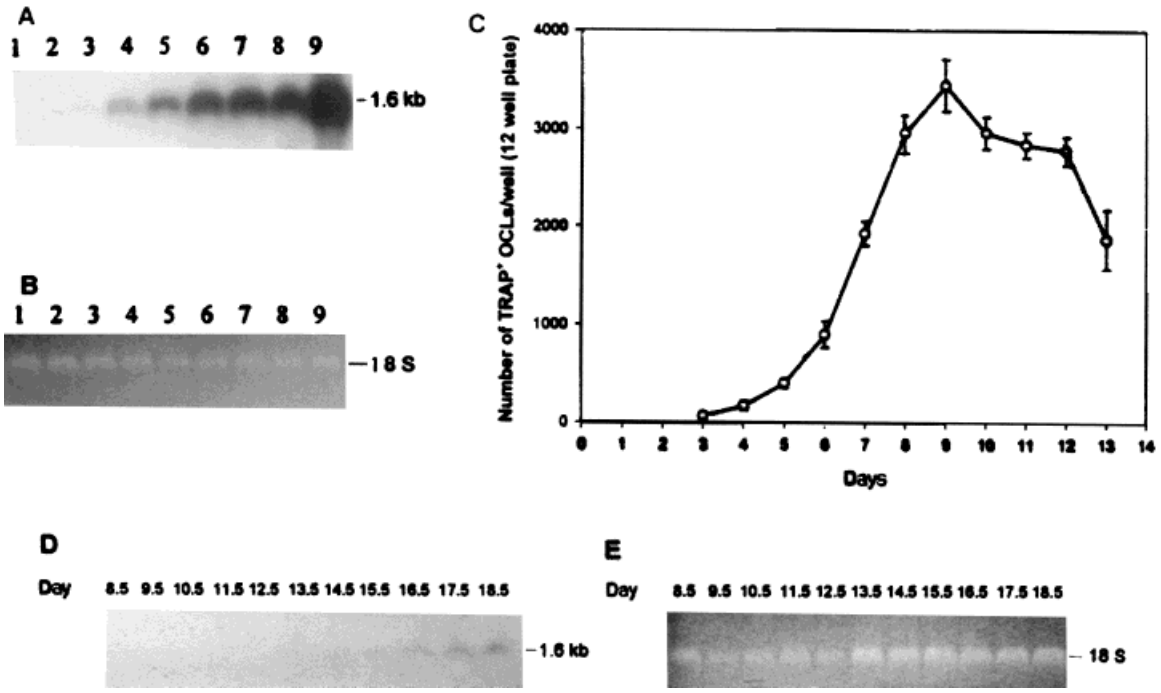

FIG. 9. Northern blot analysis of time course of cathepsin K mRNA expression in vitro osteoclast coculture system (A–C) and in vivo mouse development (D, E). (A) Cathepsin K mRNA is first detected on day 3 and increases to highest levels on day 9. (B) Ethidium bromide–stained gel indicates equivalent loading. (C) Time course for the formation of $TRAP^+$ OCLs. OCLs formation is therefore correlated to the level of cathepsin K mRNA. Data are expressed as the mean ± SEM of three cultures. (D) The sequence expression of cathepsin K was analyzed during mouse development by Northern blot. (E) Ethidium bromide–stained gel indicates equivalent loading.

of the cathepsin K observed by Northern analysis (Fig. 6) is consistent with this conclusion.

In addition, several putative regulatory *cis*-elements for Ets-1, C/EBP, AP-2, H-APF-1, AP-3, PEA1, and PEA3 were observed. Recently, a clustering of PEA3/Ets and AP1 sequences in the 92-kDa Type IV collagenase was found to be essential for transcriptional activation upon induction of *ras* in ovarian tumor-derived OVCAR 3 cells.[25] Type IV collagenase, 92 kDa, is an enzyme highly expressed in osteoclasts.[26] The involvement of interleukin-6 in stimulation of osteoclast activity has been described in numerous reports.[27,28] Consensus H-APF-1 sequences for interleukin-6 are located at –1016. H-APF-1 regulatory elements are also observed in the human promoter regions of the cathepsin K[16] and TRAP gene.[29,30] Besides the *cis* elements that were described on the human cathepsin K gene, C/EBP, AP-2, PEA1, NF-κB, GATA, and bHLH binding site sequences elucidated in this study have not been previously identified in the 5′-flanking region of human cathepsin K gene by Rood et al.[16]

In summary, we have isolated and sequenced the complete mouse cathepsin K gene, including the 9 kb 5′-flanking region. Multiple regulatory elements in the cathepsin K gene were identified. The temporal and spatial expression of cathepsin K has been determined. These findings will provide direction toward future studies to define the mechanism of cathepsin K transcription control.

ACKNOWLEDGMENTS

We thank Dr. Martin Taubman for his critical reading of the manuscript. We thank Wenjie Deng and Roger McCarrick for assistance with DNA sequence analysis. We thank Wenjie Deng and Susan Orlando for assistance with manuscript preparation. This work was supported by Grant No. DE-10887 from the National Institutes of Health to Yi-Ping Li.

REFERENCES

1. Bromme D, Okamoto K 1995 Human cathepsin O2, a novel cysteine protease highly expressed in osteoclastomas and ovary molecular cloning, sequencing and tissue distribution. Biol Chem Hoppe Seyler **376:**379–384.
2. Inaoka T, Bilbe G, Ishibashi O, Tezuka K, Kumegawa M, Kokubo T 1995 Molecular cloning of human cDNA for cathepsin K: Novel cysteine proteinase predominantly expressed in bone. Biochem Biophys Res Commun **206:**89–96.
3. Li YP, Alexander M, Wucherpfennig AL, Yelick P, Chen W, Stashenko P 1995 Cloning and complete coding sequence of a novel human cathepsin expressed in giant cells of osteoclastomas. J Bone Miner Res **10:**1197–1202.
4. Shi GP, Chapman HA, Bhairi SM, DeLeeuw C, Reddy VY, Weiss SJ 1995 Molecular cloning of human cathepsin O, a novel endoproteinase and homologue of rabbit OC2. FEBS Lett **357:**129–134.

5. Tezuka K, Tezuka Y, Maejima A, Sato T, Nemoto K, Kamioka H, Hakeda Y, Kumegawa M 1994 Molecular cloning of a possible cysteine proteinase predominantly expressed in osteoclasts. J Biol Chem **269:**1106–1109.
6. Drake FH, Dodds RA, James IE, Connor JR, Debouck C, Richardson S, Lee-Rykaczewski E, Coleman L, Rieman D, Barthlow R, Hastings G, Gowen M 1996 Cathepsin K, but not cathepsins B, L, or S, is abundantly expressed in human osteoclasts. J Biol Chem **271:**12511–12516.
7. Gelb BD, Shi GP, Chapman HA, Desnick RJ 1996 Pycnodysostosis, a lysosomal disease caused by cathepsin K deficiency. Science **273:**1236–1238.
8. Johnson MR, Polymeropoulos MH, Vos HL, Ortiz de Luna RI, Francomano CA 1996 A nonsense mutation in the cathepsin K gene observed in a family with pycnodysostosis. Genome Res **6:**1050–1055.
9. Inui T, Ishibashi O, Inaoka T, Origane Y, Kumegawa M, Kokubo T, Yamamura T 1997 Cathepsin K antisense oligodeoxynucleotide inhibits osteoclastic bone resorption. J Biol Chem **272:**8109–8112.
10. Sambrook J, Fritsch EF, Maniatis T 1989 Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Laboratory Press, New York, NY, U.S.A.
11. Chen W, Li YP 1998 Generation of mouse osteoclastogenic cell lines immortalized with SV40 large T antigen. J Bone Miner Res **13:**1112–1123.
12. Li YP, Chen W, Stashenko P 1995 Characterization of a silencer element in the first exon of the human osteocalcin gene. Nucleic Acids Res **23:**5064–5072.
13. Feng ZM, Li YP, Chen CL 1989 Analysis of the 5′-flanking regions of rat inhibin alpha– and beta-B-subunit genes suggests two different regulatory mechanisms. Mol Endocrinol **3:**1914–1925.
14. Gelb BD, Moissoglu K, Zhang J, Martignetti JA, Bromme D, Desnick RJ 1996 Cathepsin K: Isolation and characterization of the murine cDNA and genomic sequence, the homologue of the human pycnodysostosis gene. Biochem Mol Med **59:**200–206.
15. Gelb BD, Shi GP, Heller M, Weremowicz S, Morton C, Desnick RJ, Chapman HA 1997 Structure and chromosomal assignment of the human cathepsin K gene. Genomics **41:**258–262.
16. Rood JA, Van Horn S, Drake FH, Gowen M, Debouck C 1997 Genomic organization and chromosome localization of the human cathepsin K gene (CTSK). Genomics **41:**169–176.
17. Rantakokko J, Aro HT, Savontaus M, Vuorio E 1996 Mouse cathepsin K: cDNA cloning and predominant expression of the gene in osteoclasts, and in some hypertrophying chondrocytes during mouse development. FEBS Lett **393:**307–313.
18. Grigoriadis AE, Wang ZQ, Cecchini MG, Hofstetter W, Felix R, Fleisch HA, Wagner EF 1994 c-Fos: A key regulator of osteoclast–macrophage lineage determination and bone remodeling. Science **266:**443–448.
19. Iotsova V, Caamano J, Loy J, Yang Y, Lewin A, Bravo R 1997 Osteopetrosis in mice lacking NF–kappaB1 and NF–kappaB2 [see comments]. Nat Med **3:**1285–1289.
20. Tondravi MM, McKercher SR, Anderson K, Erdmann JM, Quiroz M, Maki R, Teitelbaum SL 1997 Osteopetrosis in mice lacking haematopoietic transcription factor PU.1. Nature **386:**81–84.
21. Scheven BA, Kawilarang-De Haas EW, Wassenaar AM, Nijweide PJ 1986 Differentiation kinetics of osteoclasts in the periosteum of embryonic bones in vivo and in vitro. Anat Rec **214:**418–423.
22. Dodds RA, Connor JR, Drake F, Feild J, Gowen M 1998 Cathepsin K mRNA detection is restricted to osteoclasts during fetal mouse development. J Bone Miner Res **13:**673–682.
23. Berquin IM, Cao L, Fong D, Sloane BF 1995 Identification of two new exons and multiple transcription start points in the 5′–untranslated region of the human cathepsin B–encoding gene. Gene **159:**143–149.
24. Shi GP, Webb AC, Foster KE, Knoll JH, Lemere CA, Munger JS, Chapman, HA 1994 Human cathepsin S: Chromosomal localization, gene structure, and tissue distribution. J Biol Chem **269:**11530–11536.
25. Gum R, Lengyel E, Juarez J, Chen JH, Sato H, Seiki M, Boyd D 1996 Stimulation of 92–kDa gelatinase B promoter activity by ras is mitogen–activated protein kinase kinase 1–independent and requires multiple transcription factor binding sites including closely spaced PEA3/ets and AP–1 sequences. J Biol Chem **271:**10672–10680.
26. Wucherpfennig AL, Li YP, Stetler-Stevenson WG, Rosenberg AE, Stashenko P 1994 Expression of 92 kD type IV collagenase/gelatinase B in human osteoclasts. J Bone Miner Res **9:**549–556.
27. Manolagas SC, Bellido T, Jilka RL 1995 New insights into the cellular, biochemical, and molecular basis of postmenopausal and senile osteoporosis: Roles of IL–6 and gp130. [Review]. Int J Immunopharmacol **17:**109–116.
28. Udagawa N, Takahashi N, Katagiri T, Tamura T, Wada S, Findlay DM, Martin TJ, Hirota H, Taga T, Kishimoto T, et al. 1995 Interleukin (IL)–6 induction of osteoclast differentiation depends on IL–6 receptors expressed on osteoblastic cells but not on osteoclast progenitors. J Exp Med **182:**1461–1468.
29. Reddy SV, Kuzhandaivelu N, Acosta LG, Roodman GD 1995 Characterization of the 5′–flanking region of the human tartrate–resistant acid phosphatase (TRAP) gene. Bone **16:**587–593.
30. Reddy SV, Scarcez T, Windle JJ, Leach RJ, Hundley JE, Chirgwin JM, Chou JY, Roodman GD 1993 Cloning and characterization of the 5′–flanking region of the mouse tartrate–resistant acid phosphatase gene. J Bone Miner Res **8:**1263–1270.

Address reprint requests to:
Yi-Ping Li
Forsyth Dental Center
104 Fenway
Boston, MA 02115 U.S.A.

Received in original form July 23, 1998; in revised form November 5, 1998; accepted December 1, 1998.