Cathepsin J, a novel murine cysteine protease of the papain family with a placenta-restricted expression

Kai Tisljar[a,1], Jan Deussing[a,b,1], Christoph Peters[a,*]

[a] *Medizinische Molekularbiologie, Abteilung Hämatologie-Onkologie, Klinikum der Albert-Ludwigs-Universität Freiburg, Hugstetter Strasse 55, 79106 Freiburg, Germany*

[b] *Institut für Biologie III, Albert-Ludwigs-Universität, Schänzlestrasse 1, 79106 Freiburg, Germany*

Received 6 July 1999; received in revised form 6 September 1999

Abstract A novel mouse cysteine protease of the papain family was identified by searching the dbEST database. A 1.28 kb full-length cDNA was obtained which contains an open reading frame of 999 nucleotides and encodes a predicted polypeptide of 333 amino acids. The deduced polypeptide exhibits features characteristic of cysteine proteases of the papain type including the highly conserved residues of the catalytic triad, and was hence named cathepsin J. Cathepsin J represents the murine homologue of a previously described rat cathepsin L-related protein. Mature cathepsin J shows 59.3% identity to mouse cathepsin L and contains the characteristic ER(F/W)NIN motif within the propeptide indicating that this protease belongs to the subgroup of cathepsin L-like cysteine proteases. Northern blot analysis of various tissues revealed a placenta-restricted expression. This expression pattern may suggest a role of cathepsin J in embryo implantation and/or placental function. *Ctsj* was mapped to mouse chromosome 13 in the vicinity of cathepsin L suggesting that cathepsin J may have arisen by gene duplication from cathepsin L or a common ancestral gene.

© 1999 Federation of European Biochemical Societies.

Key words: Cysteine protease; Cathepsin J; cDNA cloning; Expressed sequence tag

1. Introduction

The C1 family of papain-like cysteine peptidases represents a major component of the endosomal/lysosomal proteolytic system. All members of the C1 family share the characteristic catalytic residues Cys^{25}, His^{159} and Asn^{175} (numbering according to mature papain) forming a 'catalytic triad'. To date 11 members of the C1 family have been identified and characterized at the molecular level in human and mouse ([1–5], http://www.bi.bbsrc.ac.uk/Merops/Merops.htm). The C1 proteases of mammals can be subdivided into two functional groups according to their respective tissue distribution. The first group exhibits ubiquitous expression and comprises the highly abundant cathepsins B, H and L [6–8] as well as the recently discovered cathepsins C, O, F and Z [9–12]. These ubiquitously expressed proteases are believed to play an essential role in unspecific terminal protein degradation [13]. Additional more specific functions in physiological and pathophysiological processes such as prohormone processing [14], antigen presentation [15], rheumatoid arthritis [16], pulmonary emphysema [17], cancer invasion and metastasis [18], muscular dystrophy [19] and Alzheimer's disease [20] have been postulated. Recently it has been shown that cathepsin L is necessary for processing of the invariant chain (Ii) of major histocompatibility complex class II molecules in cortical thymic epithelial cells and thus for positive selection of $CD4^+$ T-lymphocytes [21]. Furthermore, cathepsin L was shown to be essential for regulation of proliferation of basal keratinocytes and hair follicle epithelial cells [22].

The members of the second functional group of C1 proteases comprising cathepsins K, L2, S and W [23–26] exhibit a tissue-restricted expression pattern. Cathepsin K is highly expressed in osteoclasts and is a major player in bone resorption [27]. Mutations in the human cathepsin K gene are the molecular basis of pycnodysostosis, an autosomal recessive osteochondrodysplasia [28]. Cathepsin S is expressed in lymphatic tissues and is essential for the degradation of Ii in peripheral antigen presenting cells [29,30]. Furthermore, a role of cathepsin S in macrophage-mediated tissue destruction has been postulated [25]. Cathepsin L2 [24], which is highly related to cathepsin L, is expressed in thymus and testis, whereas cathepsin W is predominantly expressed in $CD8^+$ T-lymphocytes [26].

The first four mammalian C1 proteases, cathepsin B, H, L and S [31–34], were discovered with conventional biochemical techniques. The enzyme activities were purified from tissues and characterized by standard enzymological approaches. Increasing numbers of entries in expressed sequence tag (EST) databases accelerated the identification of novel C1 proteases considerably [11,12,24,26,35].

Here we describe the identification of a novel murine C1 protease, named cathepsin J (CTSJ), its expression pattern and the chromosomal localization of the cathepsin J gene.

2. Materials and methods

2.1. Identification and sequencing of cathepsin J cDNAs

A dbEST database [36] search for novel C1 proteases using an alignment of the C1 protease family [4] as a search matrix was performed with SearchWise [37] at the Abteilung Theoretische Bioinformatik, DKFZ, Heidelberg. Identified mouse ESTs were aligned using GelStart, GelEnter, GelMerge and GelView from the Heidelberg Unix Sequence Analysis Resources (HUSAR). Resulting contigs were aligned with all known human and murine C1 proteases. Contigs representing known C1 proteases were discarded. Remaining contigs were searched for the presence of motifs specific for C1 proteases [4]. ESTs of contigs resembling putative novel cysteine proteases were obtained from the Resource Center of the German Human Genome Project (DHGP), Berlin. Here we report analysis of contig J containing ESTs: J1 (AA096626, IMAGp998L121330) and J2 (AA013726, IMAGp998C171042). Both strands of the two ESTs were sequenced using an Applied Biosystems model 377 DNA sequencer. Sequences were assembled and analyzed using DNASTAR 1.3 (DNASTAR Inc.). Multiple sequence alignments with previously described mouse

*Corresponding author. Fax: (49) (761) 270-7177.
E-mail: petersc@mm11.ukl.uni-freiburg.de

[1] These authors contributed equally to the work.

0014-5793/99/$20.00 © 1999 Federation of European Biochemical Societies. All rights reserved.
PII: S0014-5793(99)01263-6

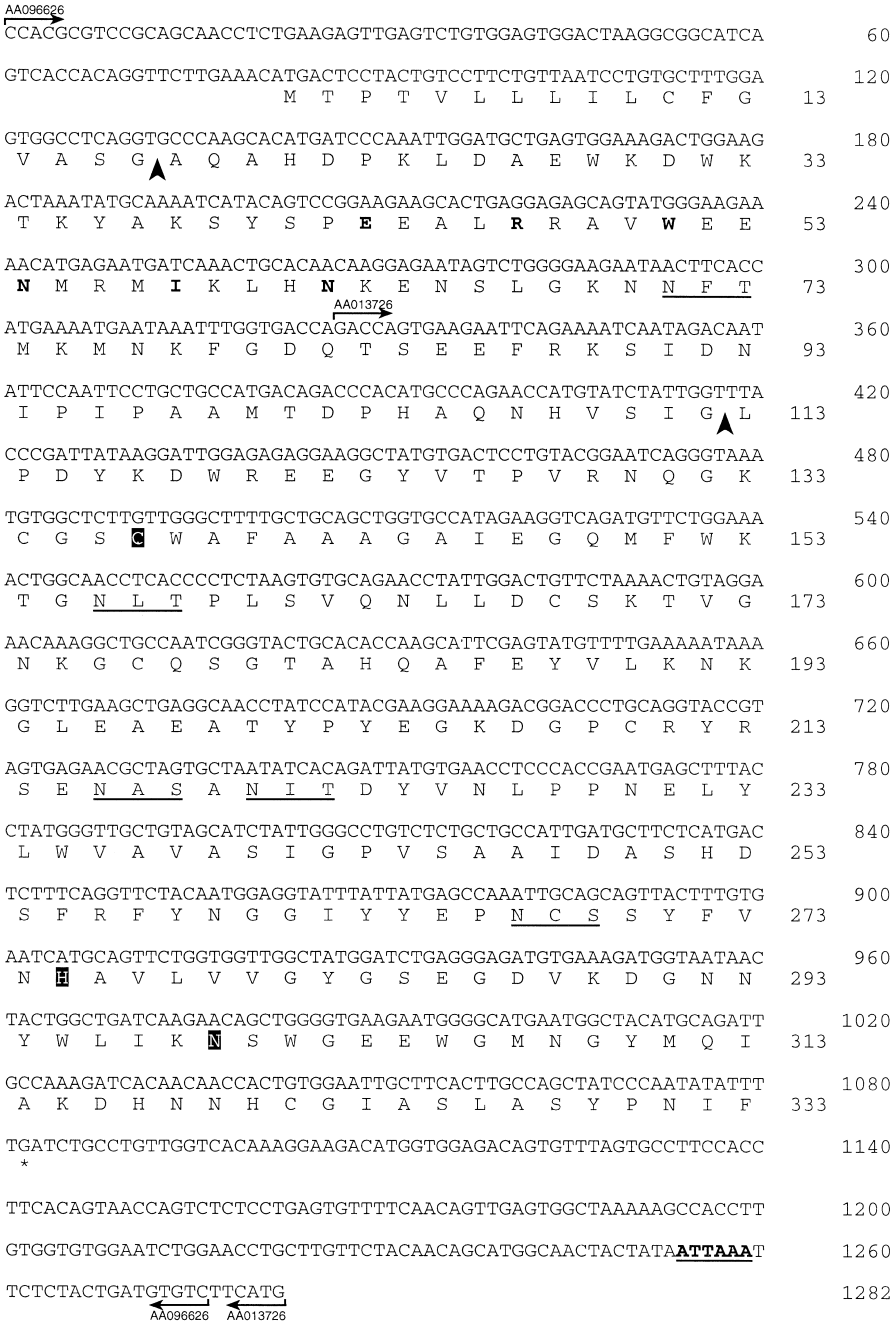

Fig. 1. Nucleotide sequence and predicted amino acid sequence of murine cathepsin J cDNA. The amino acid sequence is shown in single-letter code below the nucleotide sequence. The active site residues Cys^{137}, His^{275} and Asn^{299} of the 'catalytic triad' characteristic for the C1 family of cysteine proteases [4] are shown in black boxes. Potential *N*-glycosylation sites are underlined, arrowheads indicate the putative cleavage sites between the signal sequence and the propeptide as well as between the propeptide and the mature enzyme. Conserved amino acids in the proregion forming the ER(F/W)NIN motif are given in bold lettering. The putative polyadenylation signal is indicated in bold and underlined. The extension of the ESTs is enclosed in arrows with respective accession number.

→

Fig. 2. Multiple protein alignment of murine cathepsin J with rat cathepsin-related protein and mouse cathepsins B, C, F, H, K, L, S, W, and Z. The amino acid sequences of previously described rat or murine C1 cysteine proteases were extracted from SwissProt and GenBank databases (accession numbers: rat cathepsin L related protein (rCLRP), L14776; mouse cathepsin B (mCTSB), P10605; mouse cathepsin C (mCTSC), P97821; mouse cathepsin H (mCTSH), P49935; mouse cathepsin K (mCTSK), P55097; mouse cathepsin L (mCTSL), P06797; mouse cathepsin S (mCTSS), AF038546; mouse cathepsin W (mCTSW), P56203; mouse cathepsin F (mCTSF), AF136280; mouse cathepsin Z (mCTSZ), AF136277. Multiple sequence alignment of mature proteases was performed with CLUSTAL X [37]. Active site residues are indicated by an asterisk. Conserved residues which are common to all sequences are shadowed, residues which are conserved in ≥70% of the sequences are in bold.

```
                                                  *
mCTSJ   LPDYKDWREE G......YVT PVRNQGK... CGSCWAFAAA GAIEGQMFWK TG...NLTPL SVQNLLDCSK    58
rCLRP   LPNFKDWRKE G......YVT PVRNQGK... CGSCWAFAAV GAIEGQMSLK TG...NLTPL SAQNLLDTKS    58
mCTSL   LPKSVDWREK G......CVT PVKNQGQ... CGSCWAFSAS GCLEGQMFLK TG...KLISL SEQNLVDCSH    58
mCTSK   VPDSIDYRKK G......YVT PVKNQGQ... CGSCWAFSSA GALEGQLKKK TG...KLLAL SPQNLVDCVT    58
mCTSS   LPDTVDWREK G......CVT EVKYQGS... CGACWAFSAV GALEGQLKLK TG...KLISL SAQNLVDCSN    58
mCTSF   APPEWDWRKK G......AVT EVKNQGM... CGSCWAFSVT GNVEGQWFLN RG...TLLSL SEQELLDCDK    58
mCTSW   VPRTCDWRKA KN.....IIS SVKNQGS... CKCCWAMAAA DNIQALWRIK HQ...QFVDV SVQELLDCER    59
mCTSH   YPSSMDWRKK GN.....VVS PVKNQGA... CASCWTFSTT GALESAVAIA SG...KMLSL AEQQLVDCAQ    59
mCTSC   LPESWDWRNV QG...VNYVS PVRNQES... CGSCYSFASM GMLEARIRIL TNN.SQTPIL SPQEVVSCSP    63
mCTSZ   LPKNWDWRNV NG...VNYAS VTRNQHIPQY CGSCWAHGST SAMADRINIK RKGAWPSILL SVQNVIDCGN    67
mCTSB   LPETFDAREQ WSNCP..TIG QIRDQGS... CGSCWAFGAV EAISDRTCIH TNG.RVNVEV SAEDLLTCCG    64

mCTSJ   TVG..NKGCQ SGTAHQAFEY VLKNKGLEAE ......ATYP YE..... ..... GKDGPCR...   100
rCLRP   EG....IGLP WGTAHQAFNY VLKNKGLEAE ......ATYP YE..... ..... GKDGPCR...    98
mCTSL   AQG..NQGCN GGLMDFAFQY IKENGGLDSE ......ESYP YE..... ..... AKDGSCK...   100
mCTSK   EN....YGCG GGYMTTAFQY VQQNGGIDSE ......DAFP YV..... ..... GQDESCM...    98
mCTSS   EEKYGNKGCG GGYMTEAFQY IIDNGGIEAD ......ASYP YK..... ..... AMDEKCH...   102
mCTSF   VD....KACL GGLPSNAYAA IKNLGGLETE ......DDYG YQ..... ..... GHVQTCN...    98
mCTSW   CG....NGCN GGFVWDAYLT VLNNSGLASE ......KDYP FQ..... .....GD RKPHRCL...   101
mCTSH   AFN..NHGCK GGLPSQAFEY ILYNKGIMEE ......DSYP YI..... ..... GKDSSCR...   101
mCTSC   YA....QGCD GGFPYLIAGK YAQDFGVVEE ......SCFP YT..... ..... AKDSPCK...   103
mCTSZ   AG.....SCE GGNDLPVWEY AHKHG.IPDE ......TCNN YQ..... ..... AKDQDCDK..   106
mCTSB   IQCG..DGCN GGYPSGAWSF WTKKGLVSGG VYNSHVGCLP YTIPPCEHHV NGSRPPCTGE GDTPRCNKSC   132

mCTSJ   ...YRSENAS ANITDYVNLP P..... .....NELYL WVAVASIGPV SAAIDASHDS FRFYNGGIYY   152
rCLRP   ...YHSENAS ANITGFVNLP P..... .....NELYL WVAVASIGPV SAAIDASHDS FRFYSGGVYH   150
mCTSL   ...YRAEFAV ANDTGFVDIP Q..... .....QEKAL MKAVATVGPI SVAMDASHPS LQFYSSGIYY   152
mCTSK   ...YNATAKA AKCRGYREIP VG..... .....NEKAL KRAVARVGPI SVSIDASLAS FQFYSRGVYY   151
mCTSS   ...YNSKNRA ATCSRYIQLP FG..... .....DEDAL KEAVATKGPV SVGIDASHSS FFFYKSGVYD   155
mCTSF   ...FSAQMAK VYINDSVELS R..... .....NENKI AAWLAQKGPI SVAINAFGM. .QFYRHGIAH   148
mCTSW   ...AKKYKKV AWIQDFTMLS N..... .....NEQAI AHYLAVHGPI TVTINMKLL. .QHYQKGVIK   151
mCTSH   ...FNPQKAV AFVKNVVNIT LN..... .....DEAAM VEAVALYNPV SFAFEVTED. FLMYKSGVYS   153
mCTSC   ...PRENCLR YYSSDYYYVG GFYGG..... ....CNEALM KLELVKHGPM AVAFEVHDD. FLHYHSGIYH   159
mCTSZ   ...FNQCGTC TEFKECHTIQ NYTLWRVGDY GSLSGREKMM AEIYAN.GPI SCGIMATEM. MSNYTGGIYA   170
mCTSB   EAGYSPSYKE DKHFGYTSYS VSN..... ....SVKEIM AEIYKN.GPV EGAFTVFSD. FLTYKSGVYK   189

                          *                                          *
mCTSJ   ..EPNCSSY. ....FVNHAV LVVGYGSE.. ..... .GDVKDGNNY WLIKNSWGEE WGMNGYMQIA   202
rCLRP   ..EPNCSSY. ....VVNHAV LVVGYGFE.. ..... .GNETDGNNY WLIKNSWGEE WGINGFMKIA   200
mCTSL   ..EPNCSSK. ....NLDHGV LLVGYGYE.. ..... .GTDSNKNKY WLVKNSWGSE WGMEGYIKIA   202
mCTSK   ..DENCDRD. ....NVNHAV LVVGYG.... ..... ...TQKGSKH WIIKNSWGES WGNKGYALLA   197
mCTSS   ..DPSCTG.. ....NVNHGV LVVGYG.... ..... ...TLDGKDY WLVKNSWGLN FGDQGYIRMA   200
mCTSF   PFRPLCSPW. ....FIDHAV LLVGYG.... ..... ...NRSNIPY WAIKNSWGSD WGEEGYYYLY   196
mCTSW   ATPSSCDPR. ....QVDHSV LLVGFGKKKE GMQTGTVLSH SRKRRHSSPY WILKNSWGAH WGEKGYFRLY   216
mCTSH   ..SKSCHKTP ...DKVNHAV LAVGYG.... ..... ...EQNGLLY WIVKNSWGSQ WGENGYFLIE   201
mCTSC   ..HTGLSDPF NPFELTNHAV LLVGYG.... ..... .KIPVTGIKY WIIKNSWGSN WGESGYFRIR   212
mCTSZ   EHQDQA.... ....VINHII SVAGWG.... ..... ..VSNDGIEY WIVRNSWGEP WGEKGWMRIV   216
mCTSB   ..HEAGDM.. ....MGGHAI RILGWG.... ..... ...VENGVPY WLAANSWNLD WGDNGFFKIL   234

mCTSJ   KDH....... NNHCGIASLA S.....YPN IF                                       221
rCLRP   KDR....... NNHCGIASQA S.....FPDI F                                       219
mCTSL   KDR....... DNHCGLATAA S.....YPVV N                                       221
mCTSK   RNK....... NNACGITNMA S.....FPKM                                         215
mCTSS   RNN....... KNHCGIASYC S.....YPEI                                         218
mCTSF   RG........ SGACGVNTMA S.....SAVV N                                       214
mCTSW   RG........ NNTCGVTKYP FTAQVDSPVK KARTSCPP                                246
mCTSH   RG........ KNMCGLAACA S.....YPIP QV                                      220
mCTSC   KG........ TDECAIESIA VAA...IPIP KL                                      233
mCTSZ   TSTYKGGTGD SYNLAIESAC TFG...DPIV                                         243
mCTSB   RG........ ENHCGIESEI VAG...IPRT DQYWGRF                                 260
```

C1 proteases were performed with CLUSTAL X [38] and homologies were calculated with Gap (HUSAR). The nucleotide sequence of mouse cathepsin J was submitted to GenBank (accession number: AF136272).

2.2. Northern blot analysis

Total RNA of various tissues from 8 week old mice and of placenta from mice at day 18 of gestation was prepared as described [39]. Total RNA of blood was isolated using the QIAamp RNA Blood Mini Kit (Qiagen) according to the recommendations of the manufacturer. RNA was separated in a 1% formaldehyde agarose gel, transferred to Hybond-N membrane (Amersham) and hybridized with an [α-^{32}P]dCTP-labeled full-length CTSJ cDNA as described [39]. Filters were washed at high stringency as described [40]. Membranes were stripped by boiling in 0.1% SDS for 10 min and rehybridized with a 540 bp cDNA fragment of murine β-actin [41].

2.3. Chromosomal localization

Mapping of *Ctsj* using an interspecific backcross panel between C57BL/6J and *Mus spretus* (BSS) was performed as described [40]. Briefly, a *Xba*I restriction fragment length polymorphism (RFLP) was identified by Southern blot analysis of genomic DNA from *Mus musculus* C57BL/6J and *M. spretus* digested with multiple restriction endonucleases (New England Biolabs) and probed with the full-length CTSJ cDNA. For mapping of *Ctsj* a mouse interspecific backcross panel derived from 94 N2 animals of the backcross [(C57BL/6JEi×SPRET/Ei)F_1×SPRET/Ei] (BSS) was used [42]. Southern blot filters of this panel digested with *Xba*I were purchased from The Jackson Laboratory. The presence or absence of C57BL/6J-specific *Xba*I fragments was followed in the panel. Calculations of results was performed at The Jackson Laboratory.

3. Results and discussion

3.1. Mouse cathepsin J cDNA

For identification of novel members of the C1 family of papain-like cysteine proteases, a dbEST database [36] search using an alignment of cysteine proteases [4] as search matrix was performed. Two overlapping murine ESTs (AA096626 and AA013726), potentially encoding the conserved active site sequence motifs of C1 proteases but differing from all previously described murine C1 proteases, were identified. Sequence analysis of these two ESTs revealed that EST AA096626 is a full-length cDNA of 1276 nucleotides, whereas the 5′ end of EST AA013726 is located at nucleotide 327 and its 3′ end extends six nucleotides 3′ of AA096626 (Fig. 1). The nucleotide sequence depicted in Fig. 1 exhibits one long open reading frame (ORF) of 999 nucleotides starting at nucleotide 82 with the putative initiator ATG codon and ending with a stop codon at position 1081. The ORF potentially encodes a polypeptide of 333 amino acids with a predicted molecular weight of 37.2 kDa. This putative C1 protease was tentatively named cathepsin J (CTSJ). The initiator ATG codon lacks a typical Kozak consensus sequence [43] but three nucleotides 5′ of the initiator codon a characteristic and highly conserved nucleotide A is present. The stop codon is followed by a 199 nucleotide 3′ untranslated region which contains a putative polyadenylation signal (ATTAAA) at position 1254, which parallels the position of the polyadenylation signal of the previously described rat homologue of CTSJ [44]. A poly(A) tract not depicted in Fig. 1 was identified in both ESTs. CTSJ contains a putative signal sequence of 17 amino acids according to the −3, −1 rule for signal sequence cleavage sites [45]. The putative proregion of CTSJ consists of 95 amino acids. The potential processing site between proregion and mature enzyme has been assigned to the Gly^{112}-Leu^{113} peptide bond followed by a proline residue, which is typical for cysteine proteases of the C1 family [4]. Mature CTSJ comprises 221 amino acids. CTSJ contains five putative *N*-glycosylation sites at positions 71, 156, 216, 220 and 266 (Fig. 1). The presence of a signal sequence and potential glycosylation sites may suggest targeting of CTSJ to the endosomal/lysosomal compartment by mannose 6-phosphate receptors.

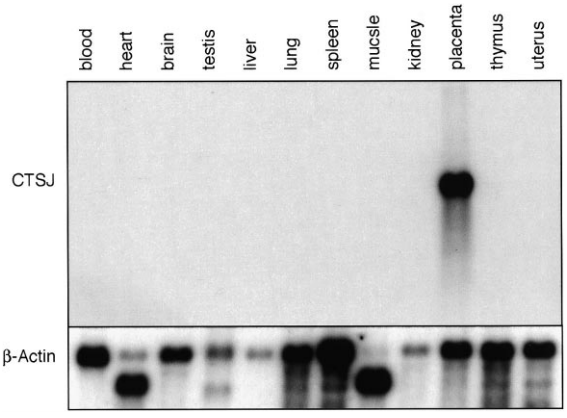

Fig. 3. Northern blot analysis of murine cathepsin J in different mouse tissues. Total RNA (8 μg) from various organs was hybridized to the [α-^{32}P]dCTP-labelled full-length CTSJ cDNA. Filters were rehybridized after stripping with a 540 bp cDNA fragment from murine β-actin [41].

3.2. Sequence similarities to other murine cysteine proteases

At present eight murine cathepsins of the C1 family have been characterized at the molecular level: cathepsins B, C, F, H, K, L, S, and Z. A multiple sequence alignment of the mature forms of these proteases with CTSJ clearly suggests that CTSJ is a novel member of the papain-like C1 family of cysteine proteases (Fig. 2). All highly conserved sequence motifs characteristic of C1 proteases have been identified, including residues Cys^{137}, His^{275} and Asn^{299} forming the 'catalytic triad' (Fig. 2 [4]). CTSJ shares the highest identity at the amino acid level (85%) with the previously described rat cathepsin L-related protein (rCLRP [44]) which was deduced from a partial cDNA. This would indicate that CTSJ and rCLRP resemble homologous proteins of different species. Alignment of mouse CTSJ and rCLRP demonstrates a high degree of

Table 1
Sequence comparison of murine cathepsin J with murine cathepsins L, S, K, H, F, W, C, Z and B[a]

	Percent identity and similarity to murine cathepsin J								
	mCTSL	mCTSS	mCTSK	mCTSH	mCTSF	mCTSW	mCTSC	mCTSZ	mCTSB
Prepro-enzyme	52.9 (58.9)	47.5 (54.6)	45.5 (55.1)	37.9 (45.7)	36.1 (45.0)	31.3 (39.9)	29.7 (34.9)	29.7 (38.1)	28.8 (36.0)
Mature enzyme	59.3 (65.6)	55.8 (62.3)	52.3 (61.7)	45.6 (54.0)	41.1 (50.5)	34.6 (44.2)	39.5 (44.6)	34.6 (43.1)	32.4 (39.0)

[a]Sequence identities and similarities (in parentheses) were calculated with the program Gap (HUSAR) using the Needleman and Wunsch algorithm [56].

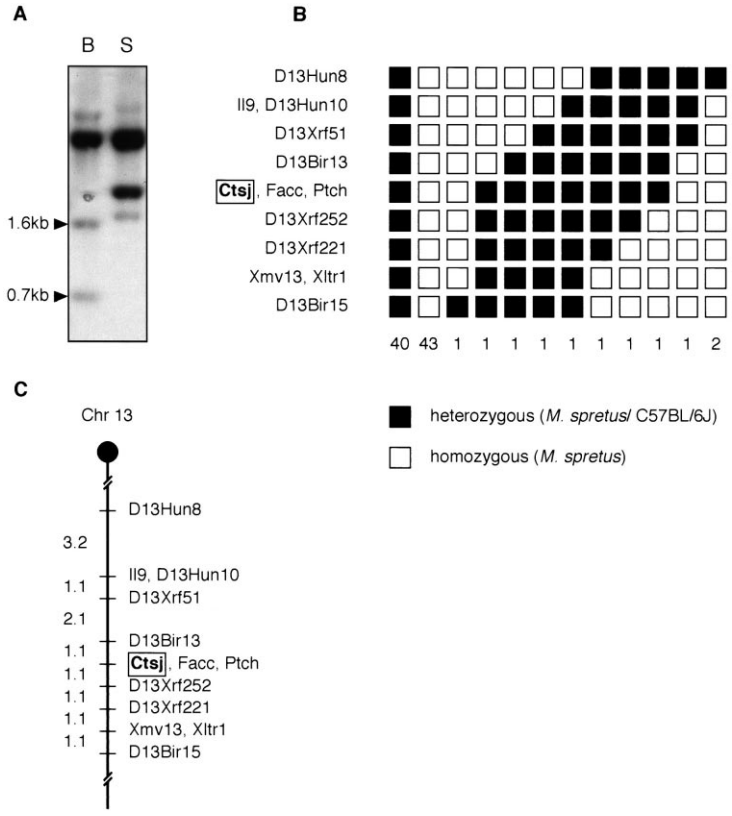

Fig. 4. Chromosomal localization of mouse cathepsin J. A: *Xba*I RFLP between *M. musculus* C57BL/6J (B) and *M. spretus* (S) detected with the full-length CTSJ cDNA. The segregation of the 1.6 kb and 0.7 kb C57BL/6J-specific DNA fragments was followed. B: Segregation pattern of *Ctsj*-specific RFLP in 94 (C57BL/6JEi×SPRET/Ei)F_1×SPRET/Ei interspecific backcross progeny. Loci linked to *Ctsj* are listed from proximal at the top to distal at the bottom. Each column represents a chromosomal haplotype, with the number of animals observed with each specific haplotype given below. Animals were scored at each locus as heterozygous for the *M. spretus* and C57BL/6J alleles (filled boxes) or homozygous for the *M. spretus* alleles (open boxes). C: Partial linkage map of mouse chromosome 13 depicting the position of *Ctsj*. The distances between the loci are given to the left in cM. *Ctsj* has been confirmed as an approved locus name by the MGD Nomenclature Support Staff.

homology throughout the whole amino acid sequence (Fig. 2). The only region with striking mismatches is located between residues Cys^{56} and Ser^{67} (Fig. 2). In contrast to all other known C1 proteases – including CTSJ – rCLRP lacks the cysteine residues Cys^{56} and Cys^{65} (Fig. 2 [4]). Among mouse C1 proteases CTSJ shares the highest identity with CTSL (Table 1). Furthermore, the proregion of CTSJ contains an ER(F/W)NIN motif (Fig. 1) characteristic of the subgroup of cathepsin L-like C1 proteases [46]. BLAST searches of the human subset of the dbEST database with the murine CTSJ cDNA as query sequence did not identify a human homologue of CTSJ, bringing the existence of a human CTSJ homologue into question.

3.3. Tissue distribution of mouse cathepsin J

Northern blot analysis of RNA from multiple organs employing the full-length CTSJ cDNA as a probe showed that CTSJ expression is restricted to the placenta only. This corresponds to the previously described expression pattern of rCLRP. rCLRP was found to be highly expressed in placenta during late gestation but not in liver or kidney [44]. The approximate size of the CTSJ mRNA was found to be 1.7 kb (Fig. 3). This tissue-specific expression of CTSJ in the placenta may suggest a role of this protease in embryo implantation and/or placental development and function. Both embryo implantation and anchoring of the placenta in the uterine wall require controlled invasion of the uterine stroma by embryonic trophoblasts. This invasion is known to be dependent upon secretion of serine and matrix metalloproteinases, which degrade the extracellular matrix of the uterine wall [47]. In vitro trophoblast invasion has been inhibited by cysteine protease inhibitors suggesting that cysteine proteinases may also play an important role in embryo implantation [48].

3.4. Mouse Ctsj maps to chromosome 13

The chromosomal localization of the murine CTSJ gene was determined by typing the BSS interspecific backcross panel [(C57BL/6JEi×SPRET/Ei)F_1×SPRET/Ei] purchased from The Jackson Laboratory [42]. A *Xba*I RFLP between *M. musculus* C57BL/6J and *M. spretus* was detected with the CTSJ cDNA (Fig. 4A). Segregation of the C57BL/6J-specific 1.6 kb and 0.7 kb *Xba*I fragments was followed in 94 N2 mice. Comparison of the distribution pattern of this *Ctsj*-specific RFLP with those for loci already located on the backcross map [42] by minimizing mismatches resulted in localization of mouse *ctsj* to chromosome 13, colocalizing with Fanconi anemia (*Facc* [49]) and patched (*Ptch* [50]; Fig. 4B,C). Mouse *Ctsl* had been previously mapped to chromosome 13 in the same area 4.0±1.7 cM proximal of *Facc* [40,51,52]. These findings and the high degree of homology between CTSJ and CTSL (Fig. 2; Table 1) suggest that CTSJ has evolved from CTSL or a common ancestral gene by gene duplication. Furthermore, the genes of the cytotoxic T-lymphocyte-associated proteins 2 alpha (*Ctla-2α*) and beta (*Ctla-2β*) map to the

same region of mouse chromosome 13 [53–55]. These proteins of unknown function show high homologies to the proregion of cathepsin L-like cysteine proteases suggesting at least one additional partial gene duplication in this region of chromosome 13. These duplication events may have resulted in a cluster of genes encoding cathepsin L-like proteases and related proteins such as CTLA-2α and CTLA-2β.

In conclusion,cathepsin J is a novel mammalian member of the C1 family of cysteine proteases with tissue-specific expression restricted to placenta. This may suggest physiological in vivo functions in embryo implantation and/or placental function.

Acknowledgements: We thank Martin Vingron, DKFZ, Heidelberg for performing the dbEST database search and Michael Behrens, Freiburg for computational support. EST clones were kindly provided by the Resource Center of the German Human Genome Project (DHGP), Berlin. We thank Marie Follo for critical comments on the manuscript. J. Deussing was supported during part of this study by the Fonds der Chemischen Industrie. This study was supported by a grant of the Fonds der Chemischen Industrie.

References

[1] Rawlings, N.D. and Barrett, A.J. (1994) Methods Enzymol. 244, 461–486.
[2] Rawlings, N.D. and Barrett, A.J. (1999) Nucleic Acids Res. 27, 325–331.
[3] Barrett, A.J., Rawlings, N.D. and Woessner, J.F. (Eds.) (1998) Handbook of Proteolytic Enzymes, pp. 543–643, Academic Press, London.
[4] Berti, P.J. and Storer, A.C. (1995) J. Mol. Biol. 246, 273–283.
[5] Chapman, H.A., Riese, R.J. and Shi, G.-P. (1997) Annu. Rev. Physiol. 59, 63–88.
[6] Chan, S.J., Segundo, B.S., McCormick, M.B. and Steiner, D.F. (1986) Proc. Natl. Acad. Sci. USA 83, 7721–7725.
[7] Fuchs, R. and Gassen, H.G. (1989) Nucleic Acids Res. 17, 9471.
[8] Troen, B.R., Gal, S. and Gottesmann, M.M. (1987) Biochem. J. 246, 731–735.
[9] Paris, A., Strukelj, B., Pungercar, J., Renko, M., Dolenc, I. and Turk, V. (1995) FEBS Lett. 369, 326–330.
[10] Velasco, G., Ferrando, A.A., Puente, X.S., Sánchez, L.M. and Lopez-Otin, C. (1994) J. Biol. Chem. 269, 27136–27142.
[11] Wang, B., Shi, G.-P., Yao, P.M., Li, Z., Chapman, H.A. and Brömme, D. (1998) J. Biol. Chem. 273, 32000–32008.
[12] Santamaria, I., Velasco, G., Pendas, A.M., Fueyo, A. and Lopez-Otin, C. (1998) J. Biol. Chem. 273, 16816–16823.
[13] Barrett, A.J. (1992) Ann. NY Acad. Sci. 674, 1–15.
[14] Shinagawa, T., Do, Y.S., Baxter, J.D., Carilli, C., Schilling, J. and Hsueh, W.A. (1990) Proc. Natl. Acad. Sci. USA 87, 1927–1931.
[15] Vidard, L., Rock, K.L. and Benacerraf, B. (1991) J. Immunol. 147, 1786–1791.
[16] Duffy, J.M., Walker, B., Guthrie, D., Grimshaw, J., McNally, G., Grimshaw, J.T., Spedding, P.L. and Mollan, R.A.B. (1994) Eur. J. Clin. Chem. Clin. Biochem. 32, 429–434.
[17] Chapman, H.A., Munger, J.S. and Shi, G.-P. (1994) Am. J. Respir. Crit. Care Med. 150, 155–159.
[18] Mignatti, P. and Rifkin, D.B. (1993) Physiol. Rev. 73, 161–195.
[19] Takeda, A., Jimi, T., Wakayama, Y., Misugi, N. and Kumagai, T. (1992) Biochem. J. 288, 643–648.
[20] Marks, N., Berg, M.J., Sapirstein, V.S., Durrie, R., Swistok, J., Makofske, R.C. and Danho, W. (1995) Int. J. Peptide Protein Res. 46, 306–313.
[21] Nakagawa, T., Roth, W., Wong, P., Nelson, A., Farr, A., Deussing, J., Villadangos, J.A., Ploegh, H., Peters, C. and Rudensky, A.Y. (1998) Science 280, 450–453.
[22] Roth, W., Deussing, J., Botchkarev, V.A., Pauly-Evers, M., Saftig, P., Hafner, A., Schmidt, P., Schmahl, W., Scherer, J., Anton-Lamprecht, I., von Figura, K., Paus, R. and Peters, C. (1999) submitted.
[23] Inaoka, T., Bilbe, G., Ishibashi, O., Tezuka, K.-i., Kumegawa, M. and Kokubo, T. (1995) Biochem. Biophys. Res. Commun. 206, 89–96.
[24] Santamaria, I., Velasco, G., Cazorla, M., Fueyo, A., Campo, E. and Lopez-Otin, C. (1998) Cancer Res. 58, 1624–1630.
[25] Shi, G.-P., Munger, J.S., Meara, J.P., Rich, D.H. and Chapman, H.A. (1992) J. Biol. Chem. 267, 7258–7262.
[26] Linnevers, C., Smeekens, S.P. and Brömme, D. (1997) FEBS Lett. 405, 253–259.
[27] Brömme, D. (1995) Biol. Chem. Hoppe-Seyler 376, 379–384.
[28] Gelb, B.D., Shi, G.-P., Chapman, H.A. and Desnick, R.J. (1996) Science 273, 1236–1238.
[29] Shi, G.-P., Villadangos, J.A., Dranoff, G., Small, C., Gu, L., Haley, K.J., Riese, R., Ploegh, H.L. and Chapman, H.A. (1999) Immunity 10, 197–206.
[30] Nakagawa, T.Y., Brissette, W.H., Lira, P.D., Petrushova Stock, J., McNeish, J.D., Eastman, S.E., Howard, E.D., Clarke, S.R.M., Rosoloniec, E.F., Elliott, E.A. and Rudensky, A.Y. (1999) Immunity 10, 207–217.
[31] Takio, K., Towatari, T., Katunuma, N., Teller, D.C. and Titani, K. (1983) Proc. Natl. Acad. Sci. USA 80, 3666–3670.
[32] Kirschke, H., Langner, J., Wiederanders, B., Ansorge, S., Bohley, P. and Broghammer, U. (1976) Acta Biol. Med. Germ. 35, 285–299.
[33] Kirschke, H., Langner, J., Wiederanders, B., Ansorge, S. and Bohley, P. (1977) Eur. J. Biochem. 74, 293–301.
[34] Kirschke, H., Wiederanders, B., Brömme, D. and Rinne, A. (1989) Biochem. J. 264, 467–473.
[35] Nägler, D.K. and Menard, R. (1998) FEBS Lett. 434, 135–139.
[36] Boguski, M.S., Lowe, T.M. and Tolstoshev, C.M. (1993) Nature Genet. 4, 332–333.
[37] Birney, E., Thompson, J.D. and Gibson, T.J. (1996) Nucleic Acids Res. 24, 2730–2739.
[38] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) Nucleic Acids Res. 25, 4876–4882.
[39] Chirgwin, J.M., Przybla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 241–245.
[40] Deussing, J., Roth, W., Rommerskirch, W., Wiederanders, B., von Figura, K. and Peters, C. (1997) Mamm. Genome 8, 241–245.
[41] Alonso, S., Minty, A., Bourlet, Y. and Buckingham, M.E. (1986) J. Mol. Evol. 23, 11–22.
[42] Rowe, L.B., Nadeau, J.H., Turner, R., Frankel, W.N., Letts, V.A., Eppig, J.T., Ko, M.S.H., Thurston, S.J. and Birkenmeier, E.H. (1994) Mamm. Genome 5, 253–274.
[43] Kozak, M. (1986) Cell 44, 283–292.
[44] Conliffe, P.R., Ogilvie, S., Simmen, R.C.M., Michel, F.J., Saunders, P. and Shiverick, K.T. (1995) Mol. Reprod. Dev. 40, 146–156.
[45] von Heijne, G. (1986) Nucleic Acids Res. 14, 4683–4690.
[46] Karrer, K.M., Pfeiffer, S.L. and DiThomas, M.E. (1993) Proc. Natl. Acad. Sci. USA 90, 3063–3067.
[47] Cross, J.C., Werb, Z. and Fisher, S.J. (1994) Science 266, 1508–1518.
[48] Babiarz, B., Romagnano, L. and Kurilla, G. (1992) In Vitro Cell. Dev. Biol. 28A, 500–508.
[49] Wevrick, R., Clarke, C.A. and Buchwald, M. (1993) Hum. Mol. Genet. 2, 655–662.
[50] Goodrich, L.V., Johnson, R.L., Milenkovic, L., McMahon, J.A. and Scott, M.P. (1996) Genes Dev. 10, 301–312.
[51] Pilz, A., Woodward, K., Povey, S. and Abbott, C. (1995) Genomics 25, 139–149.
[52] Blake, J.A., Richardson, J.E. and Eppig, J.T. (1997) Nucleic Acids Res. 25, 85–91.
[53] Denizot, F., Brunet, J.-F., Roustan, P., Harper, K., Suzan, M., Luciani, M.-F., Mattei, M.-G. and Golstein, P. (1986) Eur. J. Immunol. 19, 631–635.
[54] Ko, M.S.H., Wang, X., Horton, J.H., Hagen, M.D., Takahashi, N., Maezaki, J. and Nadeau, J.H. (1994) Mamm. Genome 5, 349–355.
[55] Stephenson, D.A. and Lueders, K.K. (1998) Mamm. Genome 8, 258–274.
[56] Needleman, S.B. and Wunsch, C.D. (1970) J. Biol. Mol. 48, 443–453.