

Human cathepsin W, a putative cysteine protease predominantly expressed in CD8⁺ T-lymphocytes

C. Linnevers, S.P. Smeekens, D. Brömme*

Arris Pharmaceutical Corp., 180 Kimball Way, South San Francisco, CA 94080, USA

Received 21 December 1996

Abstract A 750-bp fragment of a novel human cysteine protease has been identified from the dbEST databank. PCR cloning and DNA sequencing yielded a 1.38-kb full-length cDNA which encodes a polypeptide of 376 amino acids. The protein consists of a putative 21-residue signal peptide, a 106-residue propeptide and a 252-residue mature protein. The deduced amino acid sequence contains the highly conserved residues of the catalytic triad of papain-like cysteine proteases: cysteine, histidine, and asparagine. Furthermore, the protein sequence possesses two potential *N*-glycosylation sites: one in the propeptide and one in the mature protein. Comparison of the amino acid sequence of human cathepsin W with other human thiol-dependent cathepsins revealed a relatively low degree of similarity (21–31%). In contrast to cathepsins L, S, K, B, H and O, cathepsin W contains a 21-amino acid peptide insertion between the putative active site histidine and asparagine residues and an 8-amino acid C-terminal extension. This unique sequence may indicate that cathepsin W belongs in a novel subgroup of papain-like proteases distinct from that of cathepsin L- and B-like proteases. Northern blot analysis indicates a specific expression of cathepsin W in lymphatic tissues. Further analysis revealed predominant levels of expression in T-lymphocytes, and more specifically in CD8⁺ cells. The expression of the protease in cytotoxic T-lymphocytes may suggest a specific function in the mechanism or regulation of T-cell cytolytic activity.

© 1997 Federation of European Biochemical Societies.

Key words: Cathepsin W; Cysteine protease; cDNA cloning; T-lymphocytes.

1. Introduction

Lysosomal cysteine proteases such as the cathepsins B, L, H and C play an important role in intracellular protein turnover. This function, and the proteases' ubiquitous tissue distribution may reflect a 'housekeeping' role of these cathepsins [1]. In contrast to cathepsins B, L, H and C, the recently characterized cathepsins K and S exhibit a tissue-specific expression profile. Cathepsin K (also known as cathepsin O, O2, X and OC-2) has been described as a cysteine protease which plays a key role in bone remodeling and which is predominantly expressed in osteoclasts [2–9]. On the other hand, cathepsin S has been identified as the first cathepsin specifically expressed in lymphatic tissues [10].

Proteases of different classes have been implicated in specific functions of the immune system. A variety of proteases are selectively expressed by immune cells. The best characterized examples are the granzymes, a family of related serine

proteases, which are highly expressed in CTL and NK cells. Granzymes have been implicated in the CTL-mediated cytolytic event [11,12]. Evidence suggests that granzyme B initiates the apoptotic pathway of targeted cells by activation of CPP32 [13,14]. Cathepsin C (synonymous with dipeptidyl peptidase I) is thought to activate the precursor molecules of granzymes by cleaving the N-terminal activation dipeptide [15].

Lysosomal cathepsins such as the aspartate protease, cathepsin D, and the cysteine protease cathepsin B have been implicated in the degradation of exogenous antigen [16–19]. Processed antigen can be presented by two different types of MHC molecules, MHC class I and II. Cytoplasmic antigen is processed by the proteasome, a multicatalytic protease complex, and then presented by MHC class I molecules [20,21]. Alternatively, exogenous antigen is taken up by endocytosis, transported to lysosomal-like compartments and assembled with MHC class II molecules. Before MHC class II molecules can bind to an antigenic peptide, further proteolytic processing is necessary. Class II molecules are associated with the invariant chain *Ii*, which has to be proteolytically processed in the acidic lysosomal compartment. It has been shown that the lysosomal cysteine protease cathepsin S specifically degrades *Ii* and generates the CLIP peptide, a prerequisite in rendering class II molecules competent for binding peptides [22].

In this paper we report a novel, immune cell-specific, putative cysteine protease, cathepsin W. Cathepsin W has been identified from the dbEST database as a new member of the papain family. The protease is predominantly expressed in T-lymphocytes.

2. Materials and methods

2.1. Identification of cathepsin W from the dbEST database

To identify novel papain-related cysteine proteases, we conducted a search of the dbEST database using the tblastn algorithm and a consensus amino acid sequence. The sequence was derived from the highly conserved active site region of the papain family. The query sequence (with the active site asparagine residue shown in bold) was as follows (single letter amino acid code): YWLIKNSWGTGWGENGYIRI. Amongst the retrieved human sequences was one that showed potential similarity to the papain family of cysteine proteases (accession number R56701). Although the deposited sequence was indicated as having a low sequence quality, when the clone was obtained and sequenced, it was found to be a novel member of this family of proteases.

2.2. Cloning of human cathepsin W cDNA

The identified expressed sequence tag (accession number R56701; human clone number 138300) was obtained through the Lawrence Livermore National Laboratory and sequenced using an Applied Biosystems model 373A DNA sequencer. The plasmid contained an approximately 750 bp insert which included the 3' non-translating re-

*Corresponding author. Present address: Department of Human Genetics, Mount Sinai School of Medicine, Box 1498, Fifth Avenue at 100th Street, New York, NY 10029, USA. Fax: (1) (212) 360-1809.

gion (nucleotides 640–1390 of the reported cDNA). The remainder of the sequence, including a portion of the 5' non-translating region, was obtained by PCR using the *Pfu* polymerase (Stratagene, La Jolla, CA; used as described by the supplier), a human placental lambda gt10 cDNA library as the template (Clontech, Palo Alto, CA) and primers corresponding to the lambda vector (5'-AGCAAGTTCAGCCTGGT-TAAG-3') and to the cDNA clone (3'-GTAGTACGACGTCTT-GTTGC-5'). The obtained 853-bp fragment was subcloned into a pBluescript II SK(+) vector (Stratagene) and sequenced.

2.3. Preparation of RNA blots and Northern blot analysis

Freshly isolated human blood containing an anti-coagulant was diluted in 1×PBS and placed over an isotonic Percoll solution. After centrifugation at 1000×g for 30 min, the mononuclear cells at the interface were carefully removed. The lymphocytes were then separated into CD2, CD19, CD8, CD4 and CD14 subfractions using Dynabeads (Dynabeads M-450 Pan-B CD19; M-450 Pan-T CD2; M-450 CD4; M-450-CD8; M-450 CD14; Dynal Inc., Lake Success, NY). Total mRNA was isolated using a Micro RNA Isolation Kit from Stratagene and blotted onto a Duralon UV membrane (Stratagene) following formaldehyde-agarose gel electrophoresis.

Multiple tissue (Human I, Human II, Human Immune System; Clontech) and lymphocyte Northern blots were hybridized for 60 min at 68°C in ExpressHyb hybridization buffer (Clontech) containing an [α -³²P]dCTP-labeled 500-bp fragment of either cathepsin W or cathepsin S. The blots were washed in 2×SSC/0.05% SDS for 60 min at room temperature and for 60 min at 50°C in 0.1×SSC/0.1% SDS.

3. Results and discussion

3.1. Human cathepsin W cDNA

The cDNA of human cathepsin W encodes a 376-amino acid protein (Fig. 1) with a calculated molecular weight of 42 090 Da. The open reading frame starts with an ATG initiation codon at nucleotide 115 and ends at nucleotide 1302. Preceding the typical translation initiation sequence [23] a 111-bp nucleotide is present. A relatively short non-translated 3' end follows the stop codon. A typical polyadenylation site (AATAAA) starts 99 nucleotides after the stop codon. The poly A sequence begins 15 base pairs after the polyadenylation site. The total length of the sequenced cDNA of human cathepsin W is 1390 nucleotides.

The open reading frame of cathepsin W contains a putative signal sequence, a proregion and a mature protease sequence typical of lysosomal cysteine proteases of the papain family. Using the −3, −1 rule [24] a putative cleavage site between the presignal and the proregion (Gly₂₁-Ile₂₂) has been identified. The putative processing site between the proregion and the mature enzyme has been assigned to the Ser₁₂₇-Val₁₂₈ bond followed by a proline residue which is typical for cysteine proteases.

Human cathepsin W contains two potential *N*-glycosylation sites, one at position −78 (N₇₈-R₇₇-S₇₆) in the proregion and one at position 78 (N₇₈-N₇₉-S₈₀) in the mature form.

The deduced amino acid sequence of human cathepsin W consists of a 21-amino acid residue presignal, a 106-residue proregion, and a 249-residue mature form. The calculated molecular masses are 1865, 12 060 and 26 571 Da, respectively.

Fig. 1. Nucleotide sequence and deduced amino acid sequence of human cathepsin W. The putative active site residues (C153, H291 and Asn330) are indicated by asterisks. Potential glycosylation sites and the polyadenylation initiation site are underlined. Arrowheads show the putative post-translational cleavage sites between the signal sequence and the propeptide and the propeptide and mature enzyme.

→

3.2. Sequence similarities and differences to other human cysteine proteases

To date, seven human thiol-dependent cathepsins (cathepsins L, K, S, B, H, O and C) have been characterized by their complete amino acid sequences. Multiple sequence alignments of these proteases clearly demonstrate that cathepsin W is a new member of the papain-like cysteine protease family (Fig. 2). Table 1 shows the amino acid sequence homologies of human cathepsin W with the related cathepsins L, S, K, B, O, and H. Sequence homology is relatively low, in the range of 21–31%.

All of the highly conserved residues in papain-like cysteine proteases were present in cathepsin W [25]. This includes the residues of the catalytic triad (Cys₂₅, His₁₅₉, Asn₁₇₅), Gln₁₉ of the putative oxyanion hole, Trp₁₇₇ in the S1' binding pocket as well as Gly₆₇ and Gly₆₈ (papain numbering).

Besides possessing typical papain-like features, the sequence of human cathepsin W contains unique properties compared to other mammalian cathepsin sequences. The two most prominent differences are (1) a 21-amino acid peptide insertion between the active site histidine and asparagine residue and (2) an 8-amino acid residue C-terminal extension. The peptide insertion contains a serine-glutamine-proline rich stretch (SSQSPPQPP). A similar Ser, Gln, Gly rich sequence insertion at the same site has been described for cysteine proteases isolated from *Dictyostelium discoideum* [26]. However, the sequence homology to the *Dictyostelium* cysteine proteases is only 27% or less. The function of this insertion is unknown.

The low degree of similarity to cathepsin B- and cathepsin L-like proteases as well as the insertion of the SPQ rich peptide segment suggests that cathepsin W constitutes a subgroup of cathepsins distinct from the cathepsin L- and B-like subgroups. This assertion is also supported by a sequence comparison of the propeptides. Karrer et al. [27] described two subgroups of cysteine proteases based on a highly conserved motif in cathepsin L-like proteases. This so-called ERFNIN motif is present in cathepsin L-like but not in cathepsin B-like proteases. In cathepsin W, this motif is partially conserved (ERFNAQ, see Fig. 2).

A further peculiarity of the sequence of cathepsin W is found around the active site cysteine residue. Usually, in human and other mammalian cathepsins, the third residue downstream from the catalytic cysteine residue is a phenylalanine. In cathepsin W, however, this aromatic residue is replaced by a methionine followed by three alanine residues

Table 1
Sequence homologies of human cathepsin W with the cathepsins L, S, K, H, B and O

% to hCatW	hCatL	hCatS	hCatK	hCatH	hCatB	hCatO
Preprocathepsin	27.6	29.9	24.6	31.6	21.5	29.8
Mature cathepsin	33.6	34.1	33.5	33.6	25.4	30.4
Proregion	22.7	20.4	17.2	20.7	16.1	22.9

```

                20                40                60
tggattttccg ggctttttcta agctggcatt cacaccccca cctatcatac taagactcct

                80                100                120
ttgttcactt tccacaacat cagagagaca ctaccaactc cagactgcac cggcatggca
                                     M A
                140                160                180
ctgactgccc acccctcctg cctcctggcc ctgttggtgg caggcctagc ccaaggcctc
L T A H P S C L L A L L V A G L A Q G ▲ I
                200                220                240
agaggccccc ttagggccca ggacctaggt cccagccgc tagagctgaa agaggccttc
R G P L R A Q D L G P Q P L E L K E A F
                260                280                300
aagtgtttcc agatccagtt caaccggagt tacctgagcc cagaagagca tgctcaccgc
K L F Q I Q F N R S Y L S P E E H A H R
                320                340                360
ctggacatct ttgccacaa cctggcccag gctcagagcc tgcaggagga ggacttgggc
L D I F A H N L A Q A Q R L Q E E D L G
                380                400                420
acagctgaat ttgggtgac tccattcagt gacctcacag aggaggagtt tgcccagctc
T A E F G V T P F S D L T E E E F G Q L
                440                460                480
tatggctatc ggagggcagc tggagggtgc cccagcatgg gcagagaaat aaggtctgaa
Y G Y R R A A G G V P S M G R E I R S E
                500                520                540
gagccagagg agtcagtacc ttccagctgt gactggcgga aggtggccgg cgccatctca
E P E E S V ▲ P F S C D W R K V A G A I S
                560                580                600
cccatcaagg accagaaaaa ctgcaactgc tgctgggcca tggcagcggc aggcaacata
P I K D Q K N C N C C W A M A A A G N I
                620                640                660
aaaaccctgt ggcccatcag tttctgggat tttgtggacg tctccgtgca ggaactgctg
K T L W R I S F W D F V D V S V Q E L L
                680                700                720
gactgtggcc gctgtgggga tggctgccac ggtggcttcg tctgggacgc gttcataact
D C G R C G D G C H G G F V W D A F I T
                740                760                780
gtcctcaaca acagcggcct ggccagtga aaggactacc cgttccaggg caaagtccaga
V L N N S G L A S E K D Y P F Q G K V R
                800                820                840
gccacaggt gccaccccaa gaagtaccag aaggtggcct ggatccagga cttcatcatg
A H R C H P K K Y Q K V A W I Q D F I M
                860                880                900
ctgcagaaca acgagcacag aattgcgcag tacctggcca cttatggccc catcacctg
L Q N N E H R I A Q Y L A T Y G P I T V
                920                940                960
accatcaaca tgaagccctc tcagctatac cggaaggtg tgatcaaggc cacaccacc
T I N M K P L Q L Y R K G V I K A T P T
                980                1000                1020
T I N M K P L Q L Y R K G V I K A T P T
acctgtgacc cccagcttgt ggaccactct gtctgtctgg tgggttttgg cagcgtcaag
T C D P Q L V D H S V L L V G F G S V K
                1040                1060                1080
tcagaggagg ggatatgggc agagacagtc tcatcgagc ctcagcctca gcctccacac
S E E G I W A E T V S S Q S Q P Q P P H
                1100                1120                1140
cccaccccat actggatcct gaagaactcc tggggggccc aatggggaga gaagggtat
P T P Y W I L K N S W G A Q W G E K G Y
                1160                1180                1200
ttccggctgc accgagggag caatacctgt ggcacacca agttcccgct cactgcccgt
F R L H R G S N T C G I T K F P L T A R
                1220                1240                1260
gtgcagaaac cggatatgaa gcccggagtc tctgcccctc cctgaacca cctggccccc
V Q K P D M K P R V S C P P *
                1280                1300                1320
tcagttctgt cctgttaggc caactgcctc cttgccagcc ccaccccccag gtttgtgcca
                1340                1360                1380
tcttcccaat ctcaatacag tctgaataaaa ccaagacaag acctaaaaaa aaaaaaaaaa
aaaaaaaaaa

```

hKatW	MALTAHPSCILLAVLAGAGGIRGPIRAQLGQPLEKEAFKLFQIQNRSYLS-PEEHAHRL	63
hKatK	M--WG-LKVL-----L--PVVSFALYP-EETLOT-H---WEIWKTHRKQYNNKVTETSRRL	48
hKatS	M--KR-LVCVL-----LVC-SSAVAGLHKDPTLDH-H---WHLWKTYGQYKENEAEVRRRL	50
hKatL	M--NP-TLILA-----AFCIGTASATLTFDHSLEA-Q---WTKWAMNRLYGM-NEEGARRA	50
hKatH	M--WA-TLPILCAGAWLLCVPCGAELCVNSLEKFH---FKSWMSKHRTYST--EEYHRL	55
hKatO	M--DVRALPWLFWLLWLLCR-GGDA---DSRAPPT---PIWRSREREAAFR-ESLNRH	51
hKatB	MWQLWASIC---CLLVL---ANA-----RSRPSFHPVS--DEL-VNY	32
cons.	M.....a.....w.....y....E...r.	
hKatW	DIEAHNLQAQRLQEDL-GIAEF--GVTFSDLTIEEFQQLYG-YRRAAGGVPSMGREIRSEE	123
hKatK	-IWEKNLKYISIHNLASLGVHTYLLAMNHLGDMTSEEVQRMIGLKVPLSHRS-NDITLYPE	110
hKatS	-IWEKNLKFVMIHNLHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVP--SQWQ-RNITYKSN	110
hKatL	-VWEKNMKMTEHNOEYREGKHSFTMAMAFGDMTSEEFQVMNGFQ--NRKPR-KGKVFQEP	109
hKatH	QTFAFNRKINAHN---NGNHTFKMALNQFSDMSFAELKHKY-LWSEP-QNCSA-TKSNYLRG	112
hKatO	-----RY---NSLFPSENSTAFYGINQFSYLFPEEFKATVLRSK-PSKFPFY-SAEVHMSI	104
hKatB	-VNRKNTWQAGHNFYVNDMSYLKRLGT-----FLGG--PKPQRMFTED-----	77
cons.n.....hn.....g.....n...m...ee.....p.....	
I *		
hKatW	PEESVFPSCDWRKVGAISSP--IKIQKNCNCWAMAAAGNIKILWRISFWDFVDV--SVQEL	181
hKatK	WEGRAPSDVDYRKKG---YVTPVKNGGCGSCWAFSSVGALEGQLKKTKGLINL--SPQNL	167
hKatS	PNRLPDSVDWREKG---CVIEVKYQSGCAGWAFSAVGALEQLKTKGLVSL--SAQNL	167
hKatL	LFYFAFSDVDWREKG---YVTPVKNGGCGSCWAFSAITGALEGQMFRTKGLISL--SEQNL	166
hKatH	-TGYPSPSDWRRKG---NEVSPVKNGGCGSCWTFSTTGALSAITATGMLSL--AEQQL	169
hKatO	PNVSLPLFFDWRDQ---WTVQVRNQMGCGCWAFSSVWGAESAIAIKKPLEDL--SVQQV	160
hKatB	LK--LPSFTARDQWQCPITIKETIDQSGCGSCWARGAVEAISDRICITHAHVSVEVSAQL	138
cons.P.svDwR.k.....v..vk.Qg.CgsCwafs...gale.....t.....l..s.q.l	
hKatW	LDCGR--CGDCHGFWDAFTIVINNSG-LASEKDY-----PF-----QG	219
hKatK	VDCVSE--NDGCGGGMINAFQYVQKIRG-IDSEDAY-----PY-----	203
hKatS	VDCSTEKYGKNGCGGFMITAFQYITDNKG-IDSDASY-----PY-----	197
hKatL	VDCSGFQ-GNEGCGGLMDYAFQYVQDNG-IDSEESY-----PY-----	204
hKatH	VDCA-QDRNNGCGGLPSQAFYILYKNG-IMGEDTY-----PY-----	207
hKatO	IDCS--YNNYCGNGSILNANLWLNRMQVVKLVDSD-----PF-----	197
hKatB	LTCGSMCGD-GNGGYPAFAWNF-WTRKGL-VSGGLYESHVGRPYSLPCEHHVNGSRPCT	199
cons.	.dC.....n.CcngG...a.y...n.G...s...y.....Py.....	
hKatW	-KVRARHC--PKYQKVAWIQ--DFIMLQNN---EHRIAQVLTATYGPITVTINM--KHQLY	272
hKatK	-VQEEESCM--YNPTGKAACR--GYREIP--EGNEKALKRAVAR/GPVSVAIDASLTSQFY	259
hKatS	-KAMDQKQ--YDSKYRAATCS--KYTELP--YGREIVLKEAVANGKPVSVGVGDARHPSTFLY	262
hKatL	-EATIEESCK--YNHKYSVANDT--GFVDIP--KQ-EKALMKAVATVPISVAIDAGESFLFY	259
hKatH	-QGRDGYCK--FQFGKAIGFVK--DVANIT--IYD-EEAM/EAVALYNPVSFAFEVTD-FMY	234
hKatO	-KQNGLCH-YFSGSHSGFLK--GYSAVD--FSDQDEMAKALLTGFPPLVTVDA--VSAQDY	229
hKatB	GEGDTPKCSKICEFGYSPTYKQDKHYGNSYSVNSSEKIMAEIYKNGPVEGAFSV-YSDFLY	262
cons.C.....E.....a.a..gP.....f..Y	
* *		
hKatW	RKGVIKATPTCDRLVDHSLVLVGFSGVKSEGIWAETVSSQSQPQPHPTP-YWILKNSWG	334
hKatK	SGGVYDESC--NSDNLHNAVLAVGYG-----IQGNKHWTIKNSWG	299
hKatS	SGGVY-EPS--CTQNNHGVLVWGYG-----DLNGKEXWLVKNSWG	301
hKatL	KEGTYFEEDC--SSEMDHGVLVWGYG-----ESTESDNNKYWLKNSWG	303
hKatH	RTGIYSSISCHKTEDKVNHAVLAVGYG-----EKNGIPYWLKNSWG	304
hKatO	LGGI-IQHHC--SSGEANHAVLITGFD-----KTGSTPYWLVKNSWG	293
hKatB	KSGVYQHVT--GEMGGHAIIRILGCV-----ENGIPYWLKNSWN	301
cons.	...y.....H.vlv.G.g.....yw.vkNSWg	
hKatW	AQWGEKGYFRIHFGS-NICGITK-----FPLTARVQKPMKPRVSCP	376
hKatK	ENWNGKGYILMARNNACGTAN-LASFP-K-----M	329
hKatS	HNFGEGGYIRMARNGNHGGLAS--FPSYP-E-----I	331
hKatL	EEWNGGGYVMAKDRNHGGLAS--AASYP-T-----V	333
hKatH	PQWGMNGYFLIERGK-NMGLAA--CASYP-IP-----LV	335
hKatO	SSWGVGZYAHVMGS-NMGLAD--SVS--SI-----FV	322
hKatB	TWGVNGFVKILGGQ-DHCGIESEWAGIERTQW-EKI	339
cons.	..wG..Gy.....n.CGla.....p.....	

Fig. 2. Multiple amino acid alignment of human cathepsin W with human cathepsins K, S, L, H, O and B. *, active site residues; conserved residues of the ERFNIN motif in the propeptides are underlined. Amino acids identical in all seven proteases are assigned as upper case letters in the consensus sequence, and amino acids identical in five out of seven are assigned in lower case letters. Gaps are indicated by hyphens. Numbers indicate the position of the last amino acid in each line and arrowheads show the putative post-translational cleavage sites. Sequence informations have been taken from the following references: hCatK [5]; hCatS [31]; hCatL [32]; hCatH [33]; hCatO [34]; hCatB [35].

(CWAMAAA). A similar sequence in this region has also been described for two papain-like sequences from *Caenorhabditis elegans* [28]. In other human cathepsins, the neighboring residue upstream from the active site Cys₂₅ is either a serine,

alanine or glycine. In this position, cathepsin W contains a second cysteine residue. Since the mature part of the protease contains 11 cysteine residue it may form up to 5 disulfide bridges.

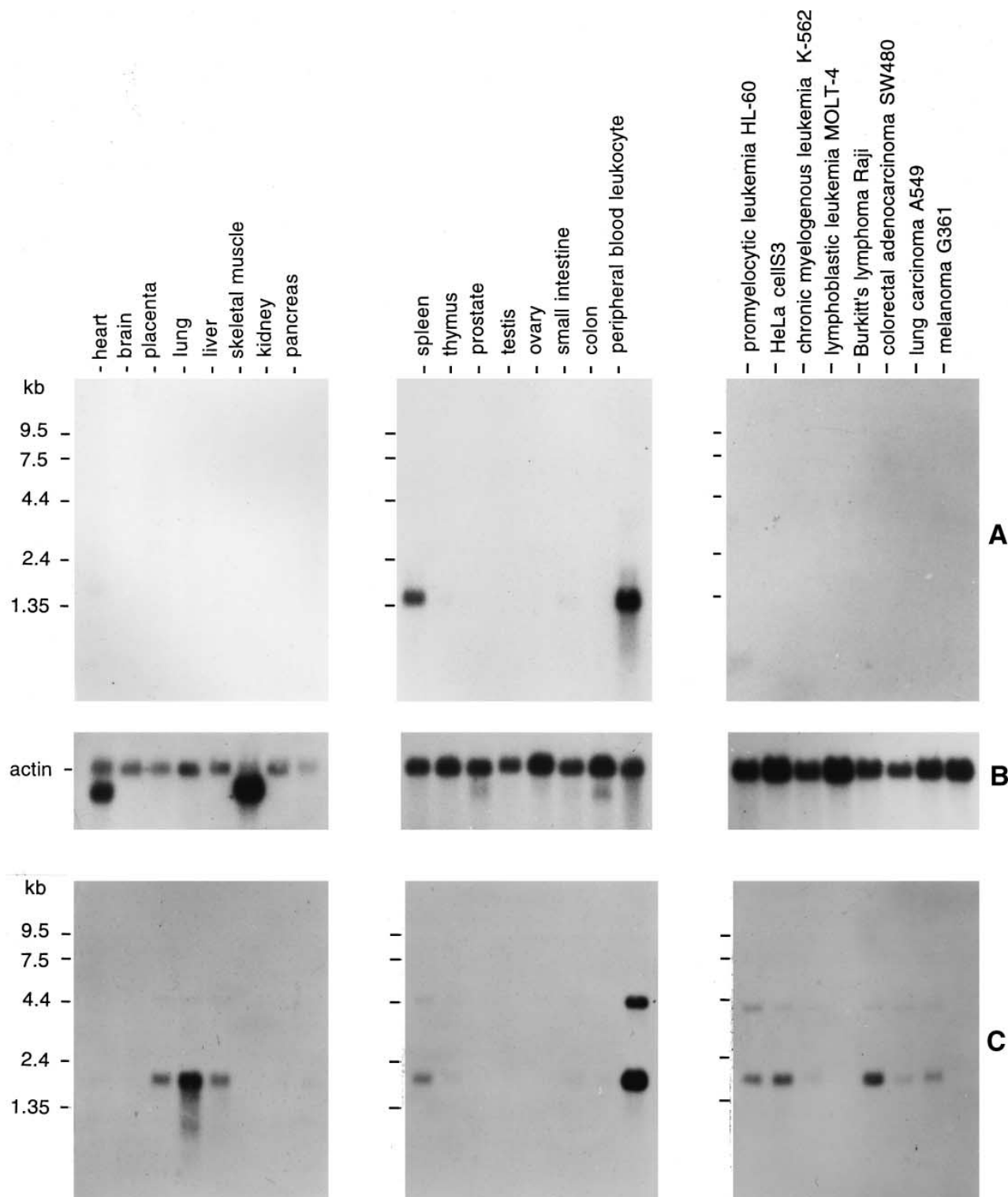


Fig. 3. Northern blot analysis of human cathepsins W and S in human tissues and tumor cell lines. Nitrocellulose blots were hybridized with 32 P-labeled probes of human cathepsins W and S. A: Cathepsin W; B: actin was used as control probe; C: cathepsin S. The additional signal at approximately 4.4 kb may represent an alternatively spliced transcript of cathepsin S. See also Figs. 4 and 5.

3.3. Tissue distribution of human cathepsin W

Cathepsin W was detected exclusively in tissues of the immune system by Northern blot analysis. The approximate size of the cathepsin W mRNA on the Northern blots is 1.45 kb which closely matches the size of the sequenced cDNA (1.38 kb). High mRNA levels were observed in peripheral blood,

spleen and lymph nodes, medium levels in bone marrow and appendix and low to very low levels in thymus (Figs. 3 and 4). Human cathepsin S, which has been previously described as a lymphoid tissue-specific cathepsin, is characterized by an almost identical expression pattern (Figs. 3 and 4). In addition, cathepsin S is also expressed in lung, liver and placenta. Im-

munohistochemical studies have assigned the expression of cathepsin S to alveolar macrophages in lung tissue [29] and to Kupffer cells in liver tissue (D. Brömme, unpublished results). No detectable levels of cathepsin W expression could be observed in various tumor cell lines whereas cathepsin S displayed expression in cell lines such as HL-60, HeLaS3, Raji, SW480 and A549 (Fig. 3).

In order to distinguish between the expression patterns of cathepsins W and S, a differential lymphocyte Northern blot was prepared. Cathepsin W was detected specifically in the T-cell population, and in particular, at very high levels in cytotoxic T-cells (CD8⁺). A low signal was also observed in CD4⁺ T-cells (Fig. 5).

In contrast, cathepsin S was almost exclusively detected in the monocytic cell population (CD14) and at very low levels in B-cells. Monocyte-derived macrophages and B-lymphocytes

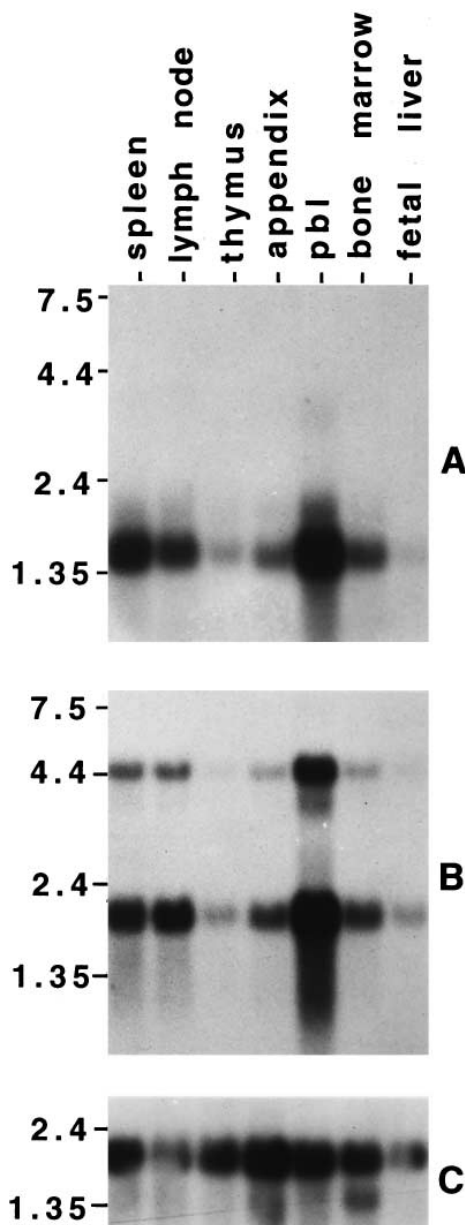


Fig. 4. Northern blot analysis of human cathepsin W and S in human immune tissues. A: Cathepsin W; B: cathepsin S; C: actin was used as control probe. pbl, peripheral blood leukocytes.

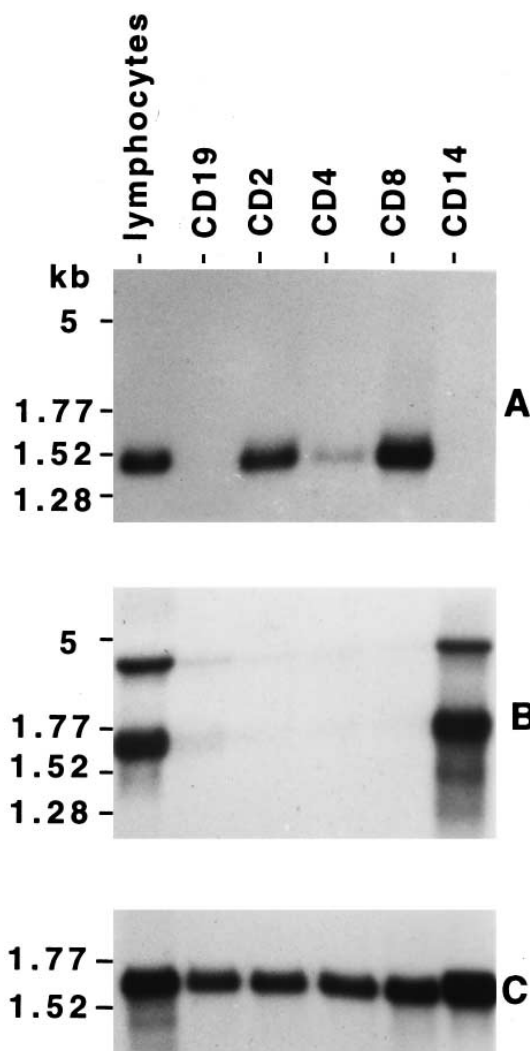


Fig. 5. Northern blot analysis of human cathepsins W and S in human lymphocytes. A: Cathepsin W; B: cathepsin S; C: GAPDH was used as control probe.

are known to be involved in MHC class II antigen presentation. It has been demonstrated that cathepsin S is directly involved in MHC class II antigen presentation by specifically cleaving the invariant chain [22].

It can be assumed that the expression of cathepsins W and S in spleen, lymph nodes, appendix and bone marrow is attributed to resident monocytes and CD8⁺ cells in these organs.

Interestingly, cathepsin W is only very weakly expressed in the thymus. Since the vast majority of thymic T cells are represented by immature CD4⁺8⁺ thymocytes and since cathepsin W is almost exclusively expressed in CD8⁺ cells, it can be concluded that the induction of cathepsin W is in some way coupled with the process of thymocyte selection in the thymus. Furthermore, the specific expression of cathepsin W in CD8⁺ cells suggests a specific function of this protease in the cytolytic activity of cytotoxic T cells. So far, only the granzymes, a subclass of serine proteases, have been described as specific to CTL-lymphocytes.

In conclusion, cathepsin W is the third example of a cell type-specific cathepsin of the papain family. So far only cathepsin S has been described as a monocyte/macrophage-spe-

cific protease [10,30] and cathepsin K as an osteoclast-specific cathepsin [2]. The cell-specific expression of these proteases raises the questions: (1) Do other cell- and tissue-specific cathepsins exist? and 2) What makes them functionally different from the ubiquitously expressed cathepsins B, L and H?

References

- [1] Kirschke H, Barrett AJ, Rawlings ND. In: Sheterline P, editor. *Protein Profiles*, Vol. 2, Issue 14. London: Academic Press, 1995:1588–1643.
- [2] K. Tezuka, Y. Tezuka, A. Maejima, T. Sato, K. Nemoto, H. Kamioka, Y. Hakeda, M. Kumegawa, *J Biol Chem* 269 (1994) 1106–1109.
- [3] G.-P. Shi, H. Chapman, S. Bhairi, C. DeLeeuw, V.Y. Reddy, S.J. Weiss, *FEBS Lett* 357 (1995) 129–134.
- [4] T. Inaoka, G. Bilbe, O. Ishibashi, K. Tezuka, M. Kumegawa, T. Kokubo, *Biochem Biophys Res Commun* 206 (1995) 89–96.
- [5] D. Brömme, K. Okamoto, *Biol Chem Hoppe-Seyler* 376 (1995) 379–384.
- [6] Y.P. Li, M.B. Alexander, A.L. Wucherpfennig, P. Yelick, W. Chen, P. Stashenko, *J Bone Mineral Res* 10 (1995) 1197–1202.
- [7] D. Brömme, K. Okamoto, B. Wang, S. Biroc, *J Biol Chem* 271 (1996) 2126–2132.
- [8] F.H. Drake, R.A. Dodds, I.E. James, J.R. Connor, C. Debouck, S. Richardson, E. Lee-Rykaczewski, L. Coleman, D. Rieman, R. Barthlow, G. Hastings, M. Gowen, *J Biol Chem* 271 (1996) 12511–12516.
- [9] B.G. Gelb, G.-P. Shi, H.A. Chapman, R.J. Desnick, *Science* 273 (1996) 1236–1238.
- [10] H. Kirschke, B. Wiederanders, D. Brömme, A. Rinne, *Biochem J* 264 (1989) 467–473.
- [11] G. Berke, *Annu Rev Immunol* 12 (1994) 735–773.
- [12] Haddad P, Jenne DE, Krähenbühl O, Tschopp J. in: Sikovsky M, Henkart P, editors. *Cytotoxic Cells. Recognition, Effector, Function, Generation, and Methods*. Boston, MA: Birkhäuser, 1993: 251–262.
- [13] M.J. Smyth, J.A. Trapani, *Immunol Today* 16 (1995) 202–206.
- [14] A.J. Darmon, R.C. Bleakley, *Nature* 377 (1995) 446–448.
- [15] J.A. Kummer, A.M. Kamp, F. Citarella, A.J. Horrevoets, C.E. Hack, *J Biol Chem* 271 (1996) 9281–9286.
- [16] G.M. Rodriguez, S. Diment, *Eur J Biochem* 25 (1995) 1823–1827.
- [17] G. Bushell, C. Nelson, H. Chiu, C. Grimley, W. Henzel, J. Burnier, S. Fong, *Mol Immunol* 30 (1993) 587–591.
- [18] N. Katunuma, Y. Matsunaga, T. Saibara, *Adv Enzyme Regul* 34 (1994) 145–158.
- [19] Y. Matsunaga, T. Saibara, H. Kido, N. Katunuma, *FEBS Lett* 324 (1993) 325–330.
- [20] M.T. Michalek, E.P. Grant, C. Gramm, A.L. Goldberg, K.L. Rock, *Nature* 363 (1993) 552–554.
- [21] J.J. Monaco, *J Leukocyte Biol* 57 (1995) 543–547.
- [22] R.J. Riese, P.R. Wolff, D. Brömme, L.R. Natkin, J.A. Villadangos, H.L. Ploegh, H.A. Chapman, *Immunity* 4 (1996) 357–366.
- [23] M. Kozak, *Cell* 44 (1986) 283–292.
- [24] G. von Heijne, *Nucleic Acids Res* 14 (1986) 4683–4690.
- [25] P. Berti, A.C. Storer, *J Mol Biol* 246 (1995) 273–283.
- [26] G.M. Souza, J. Hirai, D.P. Mehta, H.H. Freeze, *J Biol Chem* 270 (1995) 28938–28945.
- [27] K.M. Karrer, S.L. Pfeifer, M.E. DiThomas, *Proc Natl Acad Sci USA* 90 (1993) 3063–3067.
- [28] C.G. Larminie, I.L. Johnston, *DNA Cell Biol* 15 (1996) 75–82.
- [29] G.P. Shi, J.S. Munger, J.P. Meara, D.H. Rich, H.A. Chapman, *J Biol Chem* 267 (1992) 7258–7262.
- [30] H.A. Chapman, J.S. Munger, G.P. Shi, *Am J Respir Crit Care Med* 150 (1994) 155–159.
- [31] B. Wiederanders, D. Brömme, H. Kirschke, K. von Figura, B. Schmidt, C. Peters, *J Biol Chem* 267 (1992) 13708–13713.
- [32] L.J. Joseph, L.C. Chang, D. Stamenkovich, V.P. Sukhatme, *J Clin Invest* 81 (1988) 1621–1629.
- [33] R. Fuchs, H.G. Gassen, *Nucleic Acids Res* 17 (1989) 9471.
- [34] G. Velasco, A.A. Ferrando, X.S. Puente, L.M. Sanchez, C. Lopez-Otin, *J Biol Chem* 269 (1994) 27136–27142.
- [35] S.J. Chan, B. San Segundo, M.B. McCormick, D.F. Steiner, *Proc Natl Acad Sci USA* 83 (1986) 7721–7725.