Human cathepsin X: A novel cysteine protease of the papain family with a very short proregion and unique insertions

Dorit K. Nägler, Robert Ménard*

Biotechnology Research Institute, National Research Council of Canada, 6100 Avenue Royalmount, Montréal, Qué. H4P 2R2, Canada

Received 25 June 1998

Abstract A novel cDNA encoding a cysteine protease of the papain family named cathepsin X was obtained by PCR amplification from a human ovary cDNA library. The cathepsin X cDNA is ubiquitously expressed in human tissues and contains an open reading frame of 912 nucleotides encoding a predicted protein of 303 amino acids. All highly conserved regions in papain-like cysteine proteases including the catalytic residues are present in cathepsin X. The mature part of cathepsin X is 26–32% identical to human cathepsins B, C, H, K, L, O, S and W. The cathepsin X sequence contains several unique features: (i) a very short proregion; (ii) a three amino acid residue insertion in a highly conserved region between the glutamine of the putative oxyanion hole and the active site cysteine; and (iii) a second insertion of 15 amino acid residues that can be aligned with the occluding loop region in cathepsin B.

Published by Elsevier on behalf of the Federation of European Biochemical Societies.

Key words: Cysteine protease; Cathepsin X; cDNA cloning; Expressed sequence tags database

1. Introduction

Cysteine proteases of the papain family include the lysosomal cathepsins B, C, H, L and S. The name 'cathepsin' is derived from the Greek term meaning 'to digest' and was initially introduced as a general term for digestive enzymes with a slightly acidic pH optimum [1]. Later, only lysosomal proteases were referred to as cathepsins and were thought to function primarily in general intracellular protein degradation and turnover within the endosomal/lysosomal system [2]. More recent findings suggest that in addition to protein degradation, individual enzymes can fulfill specific functions such as proenzyme activation [3,4], antigen processing [5–7], hormone maturation [8] and bone resorption [9,10]. Evidence has been accumulating that cysteine proteases can also be secreted and function extracellularly in pathophysiological processes such as muscular dystrophy [11], arthritis [12], tumor invasion and metastasis [13,14], and Alzheimer's disease [15,16]. Among the lysosomal cysteine proteases, cathepsin S has attracted recent interest. Its high elastinolytic activity in lung macrophages at neutral pH raised the possibility that cathepsin S may be involved in macrophage-mediated tissue destruction [17]. Cathepsin S has also been demonstrated to play an important role in MHC class II antigen processing and to specifically degrade the invariant chain Ii generating CLIP peptides [6] and therefore targeted inhibition of this enzyme may be beneficial for modulating MHC class II immune responses. An additional role for cathepsin S in lung biology, specifically in cilial function, has been suggested but functional studies have yet to be performed to test this hypothesis [18]. Taken together, these findings clearly demonstrate that cysteine proteases are important targets for inhibitor development.

Several novel members of this growing family have been cloned recently, e.g. cathepsin O [19], cathepsin K [20–22] and cathepsin W [23]. Among these, a specific in vivo function was demonstrated only for cathepsin K. Cathepsin K was shown to be expressed at high levels in osteoclastomas [22] and is a major enzyme in bone resorption [10]. Mutations in the human cathepsin K gene were linked to pycnodysostosis, an autosomal recessive osteochondrodysplasia [24]. However, cathepsin K is also expressed in a variety of other tissues and recent results showed a high level of cathepsin K expression in human breast carcinoma [25]. Cathepsin O was initially isolated from a breast cancer cDNA library, but then found to be expressed in a variety of different tissues [19]. Cathepsin W displays a more limited tissue distribution, restricted to lymphatic tissues [23]. A detailed characterization of these enzymes will provide further insight in their biological function and potential therapeutical applications.

In this article we report the identification from the EST (expressed sequence tags) database, cloning and tissue distribution of a novel human cysteine protease of the papain family, termed cathepsin X. The primary structure of cathepsin X reveals strong homology to a cDNA clone isolated from bovine heart encoding a short fragment corresponding to the C-terminal region of a putative cysteine protease [26] and to two sequences from *Urechis caupo* and *Onchocerca volvulus* [27,28]. Several unique features in the primary structure of cathepsin X, such as a very short proregion or the position of cysteines, suggest that it may belong to a new subfamily within the papain family of cysteine proteases.

2. Materials and methods

2.1. Materials

Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs. Human cDNA libraries, Northern blot and the Advantage-GC cDNA Polymerase Mix were from Clontech Laboratories, Inc. (Palo Alto, CA, USA). Double-stranded DNA probes were radiolabeled with ^{32}P-dCTP (3000 Ci/mmol) using a commercial random-priming kit (Ready-To-Go Labeling Kit, Amersham Pharmacia Biotech, Uppsala, Sweden). The vector pGEM-T was obtained from Promega (Madison, WI, USA).

2.2. Identification and cloning of human cathepsin X

A TBLASTN search of the EST database maintained by the National Center for Biotechnology Information (NCBI) was conducted,

*Corresponding author. Fax: (1) (514) 496-5143.

NRCC Publication No. 41436. The nucleotide sequence reported in this paper has been submitted to the GenBank/EMBL data bank under GenBank accession number AF073890.

Published by Elsevier on behalf of the Federation of European Biochemical Societies.
PII: S0014-5793(98)00964-8

using peptide sequences derived from highly conserved regions in papain-like cysteine proteases. A human EST clone (accession number AA283747) was identified initially as a fragment of a novel gene encoding the C-terminus of a putative papain-like cysteine protease. The novel sequence was used in a second TBLASTN search and three additional human clones with high similarity to the query sequence were retrieved. Despite a poor sequence quality, it was possible to align the new sequences using CLUSTALW and to compile a consensus. The alignment was used to design a new query sequence upstream of the first one. The TBLASTN search, coupled with CLUSTALW alignment of the newly identified clones, was repeated five times and yielded a total of 25 human and murine clones that most likely belong to the same gene. The CLUSTALW alignments were combined and a predicted sequence encoding part of a novel putative cysteine protease of the papain family was generated. By using this approach, it was possible to discriminate between the correct reading frame and frameshifts due to sequencing errors. The compiled sequence did not include the 5′ end of the novel cDNA. The full-length cDNA was amplified by PCR from a human ovary cDNA library (Marathon-Ready cDNA) using the Advantage-GC cDNA Polymerase Mix (Clontech), the adaptor primer (5′-CCATCCTAATACGACTCACTATAGGGC-3′) and a gene specific antisense primer (5′-TTAAACGATGGGGTCCCCAAATGTACAG-3′) derived from the human EST clone (accession number AA283747). The obtained 1100-bp fragment was subcloned into the vector pGEM-T and sequenced using an automatic DNA sequencer (ABI PRISM 377, PE Applied Biosystems, Foster City, CA, USA). A part of the 3′ untranslated region was obtained by PCR from the same library, using the following sense primer: 5′-TACTTCCGCCGGGGACAGACCTGCTAC-3′.

2.3. Tissue distribution of cathepsin X

Normalized multiple tissue cDNA panels (Clontech) were screened by PCR for an 830-bp fragment of the cathepsin X cDNA using Advantage-GC cDNA Polymerase Mix (Clontech) and the two gene specific primers. Nested sense and antisense primers (5′-GGCCTCATGAGTACCTGTCCCCAGC-3′ and 5′-GCTCCTCGATGGCAAGGTTGTATCTGG-3′, respectively) were used in a second PCR reaction to verify the initial PCR products. A human multiple tissue Northern blot (Clontech) containing 2 μg poly(A)$^+$ RNA per lane was probed with the 830-bp fragment of the cathepsin X cDNA, labeled with ^{32}P-dCTP by random priming as recommended by the manufacturer. The blot was washed at high stringency (40 min with 2×SSC, 0.05% SDS, 22°C, 40 min with 0.1×SSC, 0.1% SDS, 50°C) and exposed at −80°C to a Kodak X-OMAT AR autoradiography film with 2 intensifying screens. The blot was stripped with boiling SDS solution (0.5%) for 10 min and rehybridized with a β-actin probe (Clontech) as described (data not shown).

3. Results and discussion

3.1. Human cathepsin X cDNA

A novel cDNA encoding a papain-like cysteine protease that we have named cathepsin X was obtained by PCR am-

```
ccgaggccgccggggcgggatccagagcgggagccggcgcgggatctgggactcggagcgggatccggagcgggacccaggagccggcgc          90

ggggccatggcgaggcgcgggccagggtggcggccgcttctgctgctggtgctgctggcgggcgcggcgcagggcggcctctacttccgc         180
       M  A  R  R  G  P  G  W  R  P  L  L  L  L  V  L  L  A  G  A  A  Q  G ▲G  L  Y  F  R    (5p)
                                                                          1p pro
cggggacagacctgctaccggcctctgcgggggacgggctggctccgctggggcgcagcacatacccccggcctcatgagtacctgtcc         270
 R  G  Q  T  C  Y  R  P  L  R  G  D  G  L  A  P  L  G  R  S  T  Y  P  R  P  H  E  Y  L  S   (35p)

ccagcggatctgcccaagagctgggactggcgcaatgtggatggtgtcaactatgccagcatcacccggaaccagcacatcccccaatac         360
 P  A  D ▲L  P  K  S  W  D  W  R  N  V  D  G  V  N  Y  A  S  I  T  R  N  Q  H  I  P  Q  Y   (27)
     38p 1 mature
tgcggctcctgctgggcccacgccagcaccagcgctatggcggatcggatcaacatcaagaggaagggagcgtggccctccaccctcctg         450
 C  G  S  C  W  A  H  A  S  T  S  A  M  A  D  R  I  N  I  K  R  K  G  A  W  P  S  T  L  L   (57)

tccgtgcagaacgtcatcgactgcggtaaccgtggctcctgtgaaggggtaatgacctgtccgtgtgggactacgcccaccagcacggc         540
 S  V  Q  N  V  I  D  C  G  N  R  G  S  C  E  G  G  N  D  L  S  V  W  D  Y  A  H  Q  H  G   (87)

atccctgacgagacctgcaacaactaccaggccaaggaccaggagtgtgacaagtttaaccaatgtgggacatgcaatgaattcaaagag         630
 I  P  D  E  T  C  N  N  Y  Q  A  K  D  Q  E  C  D  K  F  N  Q  C  G  T  C  N  E  F  K  E  (117)

tgccacgccatccggaactacaccctctggagggtgggagactacggctccctctctgggagggagaagatgatggcagaaatctacgca         720
 C  H  A  I  R  N  Y  T  L  W  R  V  G  D  Y  G  S  L  S  G  R  E  K  M  M  A  E  I  Y  A  (147)

aatggtcccatcagctgtggaataatggcaacagaaagactggcaaactacaccggaggcatctatgccgaataccaggacaccacatat         810
 N  G  P  I  S  C  G  I  M  A  T  E  R  L  A  N  Y  T  G  G  I  Y  A  E  Y  Q  D  T  T  Y  (177)

ataaaccatgtcgtttccgtggctgggtggggcatcagtgatgggactgagtactggattgtccggaattcatggggtgaaccatggggc         900
 I  N  H  V  V  S  V  A  G  W  G  I  S  D  G  T  E  Y  W  I  V  R  N  S  W  G  E  P  W  G  (207)

gagagaggctggctgaggatcgtgaccagcacctataaggatgggaagggcgccagatacaaccttgccatcgaggagcactgtacattt         990
 E  R  G  W  L  R  I  V  T  S  T  Y  K  D  G  K  G  A  R  Y  N  L  A  I  E  E  H  C  T  F  (237)

ggggaccccatcgtttaaggccatgtcactagaagcgcagtttaagaaaaggcatggtgacccatgaccagaggggatcctatggttatg        1080
 G  D  P  I  V  -                                                                          (242)

tgtgccaggctggctggcaggaactggggtggctatcaatattggatggcgaggacagcgtggtactggctgcgagtgttcctgagagtt        1170

gaaagtgggatgacttatgacacttgcacagcatggctctgcctcacaatgatgcagtcagccacctggtgaagaagtgacctgcgacac        1260

aggaaacgatgggacctcagtcttcttcagcagaggacttgatattttg                                            1309
```

Fig. 1. Nucleotide and predicted amino acid sequence of the human cathepsin X cDNA. Potential cleavage sites between the signal peptide and the proregion and between the proregion and the mature form are indicated (▲). The RGD motif is underlined, the two potential glycosylation sites are double underlined. Numbers in parentheses correspond to the protein sequence. The proregion is numbered from 1p (N-terminus) to 38p (C-terminus) and the mature enzyme starts at residue number 1.

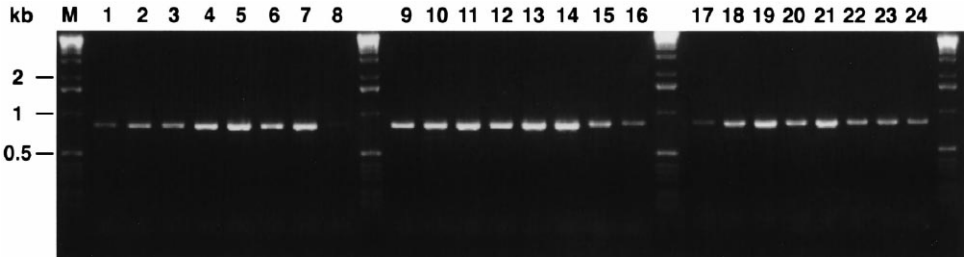

Fig. 2. Multiple tissue cDNA analysis of human cathepsin X. Normalized human multiple tissue cDNA panels were screened by PCR for an 830-bp fragment of the cathepsin X cDNA. After 30 cycles, samples were electrophoresed on a 2% agarose/EtBr gel: brain (1), heart (2), kidney (3), liver (4), lung (5), pancreas (6), placenta (7), skeletal muscle (8), colon (9), ovary (10), peripheral blood leukocyte (11), prostate (12), small intestine (13), spleen (14), testis (15), thymus (16), fetal brain (17), fetal heart (18), fetal kidney (19), fetal liver (20), fetal lung (21), fetal skeletal muscle (22), fetal spleen (23), fetal thymus (24).

plification from a human ovary cDNA library. The human cathepsin X cDNA contains an open reading frame of 912 nucleotides encoding a predicted protein of 303 amino acids (Fig. 1). A typical Kozak consensus [29] was found to surround the initiating ATG codon. The stop codon is followed by an extended non-translated 3′ end. No polyadenylation site was detected within the first 300 nucleotides 3′ of the stop codon. The deduced amino acid sequence contains a signal peptide of 23 residues, using the '(−3,−1)-rule' to identify the cleavage site [30] and a proenzyme of 280 amino acid residues. By analogy with other cysteine proteases, the putative processing site to the mature form of the enzyme has been assigned to the Asp^{38p}-Leu^{1} bond as shown in Fig. 1. Sequences derived from six independent PCR reactions were identical, except for two positions. In position 13p of the protein, a proline to serine mutation was found, attributed to a C/T mutation in nucleotide position 202. In addition, in position 68 an arginine or an alanine was observed, attributed to a CG/GC mutation in nucleotide positions 481–482.

Human cathepsin X contains two potential *N*-glycosylation sites at positions 123 and 163 (Fig. 1). The presence of a signal sequence and potential glycosylation sites indicates that the enzyme may be targeted to the endosomal/lysosomal compartment via the mannose 6-phosphate receptor pathway. Interestingly, the proregion of cathepsin X contains an integrin binding motif Arg-Gly-Asp (RGD). RGD motifs in cell-adhesive proteins such as fibronectin have been shown to serve as ligand recognition sites for their cell-surface receptors, the integrins. RGD sequences are also found in several serine proteases such as pro-protein convertases, factor X and thrombin as well as in cysteine proteases of the caspase family, but are not present in other human cathepsins of the papain family. Mutations within the RGD site in pro-protein convertase (PC1) prevented autoprocessing of PC1 and altered its intracellular trafficking [31]. The RGD site in thrombin has been implicated in tumor-platelet adhesion and tumor metastasis [32,33]. The function of the RGD motif in other proteases is not known.

3.2. Tissue distribution

Cathepsin X expression was detected by PCR in various normalized human cDNA libraries (Fig. 2) and found to be expressed at medium to high levels in placenta, lung, liver, kidney, pancreas, colon, ovary, peripheral blood leukocyte, prostate, small intestine and spleen. Lower expression levels were found in heart, brain, skeletal muscle, testis and thymus. Since PCR based methods are considered to be highly sensitive but less quantitative, cathepsin X was also detected by Northern blot analysis (Fig. 3). A good correlation between the amount of PCR product and the level of expression in the corresponding Northern blot was observed within the 8 tissues analyzed by both methods (Figs. 2 and 3). The approximate size of the cathepsin X mRNA is 1.5 kb, as determined from the Northern blot. Ubiquitous tissue distribution of cathepsin X observed in this study does not exclude the possibility of a specialized function. Cathepsin K, for example, was also shown to be expressed in a variety of different tissues [22] and found to play a more specific role in osteoclastomas [10,24].

3.3. Primary structure similarities and differences to other cysteine proteases

The highly conserved regions in papain-like cysteine proteases including the catalytic residues (Cys^{31}, His^{180}) are present in cathepsin X. Based on amino acid sequence alignment, the mature part of cathepsin X displays 26–32% identity when compared to human cathepsins B, C, K, L, S, W and papain. This clearly demonstrates that cathepsin X belongs to the papain family of cysteine proteases (Fig. 4). BLASTP and

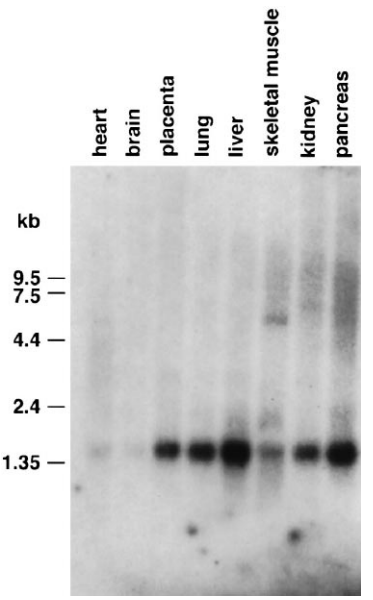

Fig. 3. Northern blot analysis of human cathepsin X in human tissues. A nitrocellulose blot was hybridized with a ^{32}P-labeled cDNA probe of human cathepsin X.

BLASTN homology searches against protein and nucleotide sequence databases revealed that cathepsin X shares a high degree of similarity with a cDNA clone isolated from bovine heart encoding the C-terminal region of a putative cysteine protease (SWISS-PROT accession number: CATX_BOVIN) [26]. This short fragment (73 amino acid residues) is 80% identical to human cathepsin X and since this enzyme has been termed cathepsin X, we used the same name for the human homolog. Two additional sequences with 65–70% identity to cathepsin X were identified in nucleotide databases, one from the nematode *Onchocerca volvulus* (GenBank accession number U71150) and the other from the echiuran *Urechis caupo* (GenBank accession number U30877). The *O. volvulus* enzyme apparently is involved in molting of third stage larvae [28]. These sequences most likely represent variants of the same enzyme in different species.

The primary structure of cathepsin X contains several unique features that clearly distinguish it from other human cysteine proteases. The predicted proregion is very short (38 amino acid residues) compared to all known cathepsins. Proregions of papain-like cysteine proteases vary in length and range from 62 residues in cathepsin B to around 100 residues in cathepsin L-like enzymes and 206 residues in cathepsin C. In addition, the proregion of cathepsin X does not share a significant degree of similarity with other cathepsins and in particular, it does not contain the so-called ERFNIN motif, present in cathepsin L-like cysteine proteases [34]. The proregion of cathepsin X also contains a cysteine residue (Cys^{10p}). A cysteine residue in the proregion of cathepsin B has been shown to contribute significantly to the inhibition of cathepsin B by its propeptide [35]. Furthermore, a cysteine residue in the proregion of cathepsin H plays an important role in defining the aminopeptidase activity of the enzyme [36]. Similarly, the cysteine residue in the proregion of cathepsin X could play a special role in regulation and/or catalytic activity. In addition, cathepsin X contains 11 cysteine residues in the mature part (including the putative active site cysteine Cys^{31}) which may form up to five disulfide bonds. Based on a previously published alignment [37] and structural data, the first two disulfides (a and b in Fig. 4) are strictly conserved in all cysteine proteases of the papain family and can be predicted for cathepsin X with a high level of confidence. The remaining seven cysteines in cathepsin X may form up to three additional disulfide bonds. By analogy with cathepsin B, one of these could involve two of the three cysteines within the 15 amino acid residue insertion (c in Fig. 4).

```
                                    a  *                                        b         a
cat_X   LPKSWDWRNVDGVN-YASITRNQHIPQYCGSCWAHASTSAMADRINIKRKGAWPSTLLSVQNVIDCG---NRGS-C    71
cat_B   LPASFDAREQWPQCPTIKEIRDQGS---CGSCWAFGAVEAISDRICIHTN-AHVSVEVSAEDLLTCC-GSMCGDGC    71
cat_C   LPTSWDWRNVHGIN-FVSPVRNQAS---CGSCYSFASMGMLEARIRILTN-NSQTPILSPQEVVSCS---QYAQGC    68
cat_K   APDSVDYRKKG----YVTPVKNQGQ---CGSCWAFSSVGALEGQLKKKTG---KLLNLSPQNLVDCV---SENDGC    63
cat_L   APRSVDWREKG----YVTPVKNQGQ---CGSCWAFSATGALEGQMFRKTG---RLISLSEQNLVDCS-GPQGNEGC    65
cat_S   LPDSVDWREKG----CVTEVKYQGS---CGACWAFSAVGALEAQLKLKTG---KLVTLSAQNLVDCSTEKYGNKGC    66
cat_W   VPFSCDWRKVAG---AISPIKDQKN---CNCCWAMAAAGNIKTLWRISFW---DFVDVSVQELLDCG---RCGDGC    64
papain  IPEYVDWRQKG----AVTPVKNQGS---CGSCWAFSAVVTIEGIIKIRTG---NLNEYSEQELLDCD---RRSYGC    63

                                  ?          c       ?  c        b
cat_X   EGGNDLSVWDYAHQH-GIPDE-----TCNNYQAKD-QECDKFN--QCGTCNEF---KECHAI-----------RNY   124
cat_B   NGGYPAEAWNFWTRK-GLVSGGLYESHVGCRPYSI-PPCEHHVNGSRPPCTGEGDTPKCSKI--CEPGYSPTYKQD   143
cat_C   EGGFPYLIAGKYAQD-FGLVE------EACFPYTG------------------TDSPCKMKEDC-----FRYYSS   113
cat_K   GGGYMTNAFQYVQKNRGIDSE-------DAYPYVGQE------------------ESCMY---------NPTGKA   104
cat_L   NGGLMDYAFQYVQDNGGLDSE-------ESYPYEATE------------------ESCKY----------NPKYS   105
cat_S   NGGFMTTAFQYIIDNKGIDSD-------ASYPYKAMD------------------QKCQY---------DSKYRA   107
cat_W   HGGFVWDAFITVLNNSGLASE-------KDYPFQGKVR-----------------AHRCHP----------KKYQK   106
papain  NGGYPWSALQLVAQY-GIHYR-------NTYPYEGVQ------------------RYCRS---------REKGPY   103

                                  ?                                   *
cat_X   TLWRVGDYGSLSGREKMMAEIYANGPISCGIMA-TERLANYTGGIY------AEYQDTTYINHVVSVAGWGIS---   190
cat_B   KHYGYNSYSVSNSEKDIMAEIYKNGPVEGAFSV-YSDFLLYKSGVY-----QHVTGEMMG-GHAIRILGWGVE---   209
cat_C   EYHYVGGFYGGCNEALMKLELVHHGPMAVAFEV-YDDFLHYKKGIYHHTGLRDPFNPFELTNHAVLLVGYGTD---   185
cat_K   AKCRGYREIPEGNEKALKRAVARVGPVSVAIDASLTSFQFYSKGVY---YDESCNSDN--LNHAVLAVGYGIQ---   172
cat_L   VANDTGFVDIPKQEKALMKAVATVGPISVAIDAGHESFLFYKEGIY---FEPDCSSED--MDHGVLVVGYGFESTE   176
cat_S   ATCSKYTELPYGREDVLKEAVANKGPVSVGVDARHPSFFLYRSGVY---YEPSCTQN---VNHGVLVVGYGDL---   174
cat_W   VAWIQDFIMLQNNEHRIAQYLATYGPITVTINM--KPLQLYRKGVIK-ATPTTCDPQL--VDHSVLLVGFGSVKSE   177
papain  AAKTDGVRQVQPYNEGALLYSIANQPVSVVLEAAGKDFQLYRGGIF----VGPCGNK---VDHAVAAVGYGP----   168

                                                                      ?
cat_X   -----------------DGTEYWIVRNSWGEPWGERGWLRIVTSTYKDGKGARYNLAIEEHCTFGDPIV          242
cat_B   -----------------NGTPYWLVANSWNTDWGDNGFFKILR----GQDHCGIESEVVAGIPRTDQYWEKI      260
cat_C   ---------------SASGMDYWIVKNSWGTGWGENGYFRIRR----GTDECAIESIAVAATPIPKL          233
cat_K   -----------------KGNKHWIIKNSWGENWGNKGYILMARN---KNNACGIANLAS--FPKM            215
cat_L   S----------------DNNKYWLVKNSWGEEWGMGGYVKMAKD---RRNHCGIASAAS--YPTV            220
cat_S   -----------------NGKEYWLVKNSWGHNFGEEGYIRMARN---KGNHCGIASFPS--YPEI            217
cat_W   EGIWAETVSSQSQPQPPHPTPYWILKNSWGAQWGEKGYFRLHRG----SNTCGITKFPLTARVQKPDMKPRVSCPP   249
papain  -------------------NYILIKNSWGTGWGENGYIRIKRGTGNSYGVCGLYTSSF--YPVKN            212
```

Fig. 4. Multiple amino acid sequence alignment of human cathepsin X with human cathepsins B, C, K, L, S, W and papain. Sequence information is from SWISS-PROT (accession numbers CATB_HUMAN, CATC_HUMAN, CATK_HUMAN, CATL_HUMAN, CATS_HUMAN, PAPA_CARPA) and from [23]. A previously published alignment was used as a template [37] and cathepsins K, W and X were aligned using CLUSTALW. Active site residues are marked with *; predicted disulfide connectivities are indicated by letters a–c; remaining cysteines are labeled with question marks.

Another distinctive feature of cathepsin X is a three amino acid residue insertion before the consensus CGXC (Cys^{28}-Gly^{29}-Ser^{30}-Cys^{31}, Fig. 4). Such an insertion in a highly conserved region between the glutamine (Gln^{22}) of the putative oxyanion hole and the active site cysteine is highly unusual for a papain-like cysteine protease. A second relatively large insertion of 15 amino acid residues can be aligned with the occluding loop region in cathepsin B [38]. The occluding loop in cathepsin B covers part of the S′ subsites of the enzyme and was shown to be essential for the dipeptidyl carboxypeptidase activity [39] and to interfere with the endopeptidase activity [40]. Two histidine residues (His^{110} and His^{111}) in this loop provide positively charged anchors for the C-terminal carboxylate group of substrates. The insertion in cathepsin X, however, does not contain histidines but other ionizable groups are present and it remains to be determined if any special function is associated with this insertion.

In conclusion, human cathepsin X is a new member of the papain family of cysteine proteases. Unlike recently discovered members of this family, which are closely related to cathepsin L, cathepsin X possesses a number of unique and interesting features, which might confer special properties to the enzyme. Preliminary data show that recombinant procathepsin X does not readily autoprocess at acidic pH. However, the proenzyme is able to hydrolyze small synthetic substrates such as Cbz-Phe-Arg-MCA (unpublished data). The precise biological function of human cathepsin X is unclear. Its ubiquitous tissue distribution and evolutionary conservation may indicate an essential physiological role. Detailed biochemical characterization as well as in vivo studies will provide clues to the precise physiological role of cathepsin X and to its value as a target for therapeutic intervention.

Acknowledgements: We thank Dr. Andrew C. Storer for valuable discussions and Manon Lalumière and Viktoria Lytvyn for expert technical assistance.

References

[1] Willstätter, R. and Bamann, E. (1929) Hoppe-Seylers Z. Physiol. Chemie 180, 127–143.
[2] Kirschke, H. and Barrett, A.J. (1987) in: Lysosomes: Their Role in Protein Breakdown (Glaumann, H. and Ballard, F.J., Eds.) pp. 193–238, Academic Press, London.
[3] Samarel, A.M., Ferguson, A.G., Decker, R.S. and Lesch, M. (1989) Am. J. Physiol. 257, 1069–1079.
[4] Goretzki, L., Schmitt, M., Mann, K., Calvete, J., Chucholowski, N., Kramer, M., Gunzler, W.A., Janicke, F. and Graeff, H. (1992) FEBS Lett. 297, 112–118.
[5] Guagliardi, L.E., Koppelman, B., Blum, J.S., Marks, M.S., Cresswell, P. and Brodsky, F.M. (1990) Nature 343, 133–139.
[6] Riese, R.J., Wolf, P.R., Brömme, D., Natkin, L.R., Villandagos, J.A., Ploegh, H.L. and Chapman, H.A. (1996) Immunity 4, 357–366.
[7] Nakagawa, T., Roth, W., Wong, P., Nelson, A., Farr, A., Deussing, J., Villandagos, J.A., Ploegh, H., Peters, C. and Rudensky, A.Y. (1998) Science 280, 450–453.
[8] Wang, P.H., Do, Y.S., Macaulay, L., Shinagawa, T., Anderson, P.W., Baxter, J.D. and Hsueh, W.A. (1991) J. Biol. Chem. 266, 12633–12638.
[9] Delaissé, J.M., Eeckhout, Y. and Vaes, G. (1984) Biochem. Biophys. Res. Commun. 125, 441–447.
[10] Inui, T., Ishibashi, O., Inaoka, T., Origane, Y., Kumegawa, M., Kokubo, T. and Yamamura, T. (1997) J. Biol. Chem. 272, 8109–8112.
[11] Katunuma, N. and Kominami, E. (1987) Rev. Physiol. Biochem. Pharmacol. 108, 1–20.
[12] Mort, J.S., Recklies, A.D. and Poole, R.A. (1984) Arthr. Rheum. 27, 509–515.
[13] Sloane, B.F., Dunn, J.R. and Honn, K.V. (1981) Science 212, 1151–1153.
[14] Sloane, B.F. and Honn, K.V. (1984) Cancer Metastasis Rev. 3, 249–263.
[15] Cataldo, A.M. and Nixon, R.A. (1990) Proc. Natl. Acad. Sci. USA 87, 3861–3865.
[16] Munger, J.S., Haass, C., Lemere, C.A., Shi, G.P., Wong, W.S.F., Teplow, D.B., Selkoe, D.J. and Chapman, H.A. (1995) Biochem. J. 311, 299–305.
[17] Shi, G.-P., Munger, J.S., Meara, J.P., Rich, D.H. and Chapman, H.A. (1992) J. Biol. Chem. 267, 7258–7262.
[18] Chapman, H.A., Riese, R.J. and Shi, G.-P. (1997) Annu. Rev. Physiol. 59, 63–88.
[19] Velasco, G., Ferrando, A.A., Puente, X.S., Sánches, L/M. and López-Otín, C. (1994) J. Biol. Chem. 269, 27136–27142.
[20] Shi, G.-P., Chapman, H.A., Bhairi, S.M., DeLeeuw, C., Reddy, V.Y. and Weiss, S.J. (1995) FEBS Lett. 357, 129–134.
[21] Inaoka, T., Bilbe, G., Ishibashi, O., Tezuka, K., Kumegawa, M. and Kokubo, T. (1995) Biochem. Biophys. Res. Commun. 206, 89–96.
[22] Brömme, D. and Okamoto, K. (1995) Biol. Chem. Hoppe-Seyler 376, 379–384.
[23] Linnevers, C., Smeekens, S.P. and Brömme, D. (1997) FEBS Lett. 405, 253–259.
[24] Gelb, B.D., Shi, G.-P., Chapman, H.A. and Desnick, R.J. (1996) Science 273, 1236–1238.
[25] Littlewood-Evans, A.J., Bilbe, G., Bowler, W.B., Farley, D., Wlodarski, B., Kokubo, T., Inaoka, T., Sloane, J., Evans, D.B. and Gallagher, J.A. (1997) Cancer Res. 57, 5386–5390.
[26] Gay, N.J. and Walker, J.E. (1985) Biochem. J. 225, 707–712.
[27] Rosenthal, E. (1993) Dev. Genet. 14, 485–491.
[28] Lustigman, S., McKerrow, J.H., Shah, K., Lui, J., Huima, T., Hough, M. and Brotman, B. (1996) J. Biol. Chem. 271, 30181–30189.
[29] Kozak, M. (1986) Cell 44, 283–292.
[30] von Heijne, G. (1986) Nucleic Acids Res. 14, 4683–4690.
[31] Lusson, J., Benjannet, S., Hamelin, J., Savaria, D., Chrétien, M. and Seidah, N.G. (1997) J. Biol. Chem. 326, 737–744.
[32] Nierodzik, M.L., Plotkin, A., Kajumo, F. and Karpatkin, S. (1991) J. Clin. Invest. 87, 229–236.
[33] Bar-Shavit, R., Ginzburg, Y., Maoz, M., Vlodavsky, I. and Peretz, T. (1995) Isr. J. Med. Sci. 31, 86–94.
[34] Karrer, K.M., Pfeiffer, S.L. and DiThomas, M.E. (1993) Proc. Natl. Acad. Sci. USA 90, 3063–3067.
[35] Chen, Y., Plouffe, C., Ménard, R. and Storer, A.C. (1996) FEBS Lett. 393, 24–26.
[36] Gunčar, G., Podobnik, M., Pungerčar, J., Štrukelj, B., Turk, V. and Turk, D. (1998) Structure 6, 51–61.
[37] Berti, P.J. and Storer, A.C. (1995) J. Mol. Biol. 246, 273–283.
[38] Musil, D., Zucic, D., Turk, D., Engh, R.A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N. and Bode, W. (1991) EMBO J. 10, 2321–2330.
[39] Illy, C., Quraishi, O., Wang, J., Purisima, E., Vernet, T. and Mort, J. (1997) J. Biol. Chem. 272, 1197–1202.
[40] Nägler, D.K., Storer, A.C., Portaro, F.C.V., Carmona, E., Juliano, L. and Ménard, R. (1997) Biochemistry 36, 12608–12615.