



Published in final edited form as:

*Biol Chem.* 2012 March ; 393(3): 177–186. doi:10.1515/hsz-2011-0219.

## Biochemical characterization and structural modeling of human cathepsin E variant 2 in comparison to the wild-type protein

Vida Puizdar<sup>a,1</sup>, Tajana Zajc<sup>a,1</sup>, Eva Žerovnik<sup>1</sup>, Miha Renko<sup>1</sup>, Ursula Pieper<sup>2</sup>, Narayanan Eswar<sup>2</sup>, Andrej Šali<sup>2</sup>, Iztok Dolenc<sup>1,\*</sup>, and Vito Turk<sup>1,\*</sup>

<sup>1</sup>Department of Biochemistry and Molecular and Structural Biology, Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia

<sup>2</sup>Department of Bioengineering and Therapeutic Sciences, California Institute for Quantitative Biosciences at UCSF, Department of Pharmaceutical Chemistry, University of California, San Francisco, UCSF MC 2552, Byers Hall, 1700 4th Street, Suite 503 B, San Francisco, CA 94158, USA

### Abstract

Cathepsin E splice variant 2 appears in a number of gastric carcinoma. Here, we report detecting this variant in HeLa cells using polyclonal antibodies and biotinylated inhibitor pepstatin A. An overexpression of GFP fusion proteins of cathepsin E and its splice variant within HEK-293T cells was performed to show their localization. Their distribution under a fluorescence microscope showed that they are colocalized. We also expressed variant 1 and variant 2 of cathepsins E, with propeptide and without it, in *Echerichia coli*. After refolding from the inclusion bodies, the enzymatic activity and circular dichroism spectra of the splice variant 2 were compared to those of the wild-type mature active cathepsins E. While full-length cathepsin E variant 1 is activated at acid pH, the splice variant remains inactive. In contrast to the active cathepsin E, the splice variant 2 predominantly assumes  $\beta$ -sheet structure, prone to oligomerization, at least under *in vitro* conditions, as shown by Atomic Force Microscopy as shallow disk-like particles. A comparative structure model of splice variant 2 was computed based on its alignment to the known structure of cathepsin E intermediate (Protein Data Bank code 1TZS), and used to rationalize its conformational properties and loss of activity.

### Keywords

cathepsin E; splice variant; aspartic protease; structural modelling; atomic force microscopy

### Introduction

Aspartic peptidases represent one of the five mechanistic classes of peptidases. Among them, clan AA contains the large A1 family (the pepsin family). The members of this family

\*Corresponding authors addresses: Vito Turk, Department of Biochemistry and Molecular and Structural Biology, Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia. Fax: +386 1 4773984; vito.turk@ijs.si Iztok Dolenc, Department of Biochemistry and Molecular and Structural Biology, Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia. Fax: +386 1 4773984; iztok.dolenc@ijs.si.

<sup>a</sup>Vida Puizdar and Tajana Zajc equally contributed to this work

include cathepsin D and cathepsin E (Rawlings et al., 2008). In contrast to cathepsin D, which is a widely distributed lysosomal enzyme, cathepsin E, unlike cathepsin D, is not a lysosomal enzyme, but has limited cell and tissue distribution (Bennett et al., 1992; Saku et al., 1991). Cathepsin E was discovered in bone marrow (Lapresle and Webb, 1962) and in bovine spleen (Puizdar et al., 1985). It was found to be present in large amounts in polymorph nuclear cells and in smaller amounts in macrophages (Stefanovic et al., 1962), in the cells of immune system, including antigen-presenting cells (Yamamoto et al., 1978), lymphoid tissues, erythrocytes (Yonezawa and Nakamura, 1991) and gastric cells (Samloff et al., 1987). Cathepsin E is localized in the endosome, plasma membrane, Golgi complex and endoplasmic reticulum (Finley and Kornfeld, 1994). For a detailed review see ref. (Zaidi and Kalbacher, 2008).

The sequence of cathepsin E resembles the aspartic proteases in the A1 family. The sequence of preprocathepsin E was determined from its isolated cDNA (Azuma et al., 1992). The known 3D structure of cathepsin E provides some information about activation of an intermediate (Ostermann et al., 2004). Cathepsin E is expressed in the form of an inactive proenzyme; however, its propeptide is not required for protein stability (Tsukuba et al., 2006). Complete activation of procathepsin E occurs by proteolytic removal of the N-terminal 36-amino acid residue propeptide (Yasuda et al., 2005b). The maturation generates an exposed active site cleft, with an Asp-Thr-Gly motif from each of the N- and C-terminal protein lobes (Ostermann et al., 2004). The Asp side chains cleave the peptide bond of a substrate, with the aid of an activated water molecule (James, 2004). Physiological substrates and endogenous protein inhibitors of cathepsin E are not known. The activities of cathepsins E and D are similar and difficult to distinguish from each other. However, beside specific substrates (Abd-Elgaliel and Tung, 2010; Yasuda et al., 2005a), a new approach based on a combination of a cathepsin E antibody with its substrate, can differentiate their activities in antigen-processing organelles (Zaidi et al., 2007).

Cathepsin E has been implicated in various physiological and pathological processes. The enzyme participates in processing of the neurotensin precursors and other bioactive peptides (Kageyama et al., 1995; Lees et al., 1990), as well as functions in antigen processing in dendritic cells (Chain et al., 2005). Inhibition and gene deletion have been reported to compromise processing and presentation of antigens, such as tetanus toxin, myoglobin, and ovalbumin (Burster et al., 2008; Kakehashi et al., 2007). Cathepsin E-deficient mice are also found to spontaneously develop atopic dermatitis-like skin lesions (Tsukuba et al., 2003). The enzyme has a substantial role in host defense against tumor cells through TRAIL-dependent apoptosis without affecting normal cells (Kawakubo et al., 2007). In particular, it was shown that tumor growth arrest through inhibition of angiogenesis is induced by stable expression of cathepsin E (Shin et al., 2007). Recent reviews summarize physiological and pathological processes involving cathepsin E (Yamamoto et al., 2011; Zaidi et al., 2008).

Alternative pre-mRNA processing is a key regulator of gene expression, as it generates numerous transcripts from a single protein-coding gene (Ben-Dov et al., 2008). Alternative splicing generates protein isoforms with different biological properties, differing in protein-protein interactions, subcellular localization, and enzymatic activity (Stamm et al., 2005). It has been recently shown in two gastric adenocarcinoma cell lines that human procathepsin E

gene is subject to alternative splicing (Tatnell et al., 2003). Interestingly, alternatively spliced transcripts have already been observed for various proteases and appear to be preferentially expressed in cancerous cells and tissues. The procathepsin E gene consists of nine exons (Arnold et al., 1997), but its splice variant has a deletion of exon 7, consisting of 142 bp (Tatnell et al., 2003). This deletion dictates a frame-shift as exon 8 and 9 are introduced to an alternative transcript. The missing exon 7 encodes the second Asp-Thr-Gly active site motif, presumably resulting in an inactive protein (Tatnell et al., 2003). Another variant of cathepsin E (variant 3), with an additional 15 bp in its gene, was found in human intestine. The short insert occurs in the intron between the exons 4 and 5. No differences in the localization and retention between this variant and cathepsin E were observed (Finley and Kornfeld, 1994).

Here, the pro- and mature forms of the splice variant 2 of human cathepsin E have been expressed in *Escherichia coli*, refolded and used for biochemical characterization. Cathepsin E activity was confirmed, whereas its splice form was found to be inactive. A circular dichroism (CD) spectrum of the splice variant allowed us to characterize its oligomeric structure in relation to that of the wild type cathepsin E. Finally, a comparative structure model of splice variant 2 was computed based on its alignment to the known structure of cathepsin E intermediate (Protein Data Bank code 1TZS), and used to rationalize its conformational properties and loss of activity.

## Results and discussion

Since the first report of the splice variant of cathepsin E in human intestinal carcinoma cells (Finley and Kornfeld, 1994), an RNA message for this enzyme was also found in gastric adenocarcinoma cells (Tatnell et al., 2003). No cathepsin E splice variant protein has been reported so far. When we scrutinized HeLa cells for the presence of mRNA of cathepsin E variant 2 and its endogenous expression, low amounts of the mRNA and the protein were indeed detected (Fig. 1). Polyclonal antibodies used to detect the protein were raised against the peptide from the C-terminal lobe of cathepsin E variant 2, whose sequence bears no apparent relationship to that of normal cathepsin E. The N-terminal part of the molecule of human origin is unchanged and includes a single N-glycosylation, in comparison to rat cathepsin E with two sites (Yasuda et al., 1999; Tsukuba et al., 2006). Enzymatic deglycosylation, using PNGase F shows partial N-glycosylation, probably due to a long retention of the protein in endoplasmic reticulum (Fig. 1). When proteins misfold, they aggregate, thus preventing N-glycosylation. Consequently, they are retained in the ER and later degraded by the proteasome. (Ellgaard et al., 1999).

To study localization of both cathepsin E forms in the cell, fusions of cathepsin E with red fluorescent protein and cathepsin E variant 2 with green fluorescent protein were coexpressed in HEK-293T cells. The distribution of fusion proteins within cells under a fluorescence microscope showed that they are colocalized (Fig. 2).

Various heterologous gene expression systems have been used to express recombinant cathepsin E and its splice variant for biochemical characterization (Hill et al., 1993; Tsukuba et al., 1993; Yamada et al., 1994). Acidification rapidly activated the recombinant

procathepsin E prepared in *E. coli*, resulting in a heterogeneous N-terminus of the mature cathepsin E (Fowler et al., 1995; Hill et al., 1993; Ostermann et al., 2004; Tatnell et al., 1997). To avoid N-terminal micro heterogeneity of the resultant mature enzymes, cathepsin E and cathepsin E variant 2 were also expressed without the propeptide. The expression of a cathepsin E mutant with propeptide deletion in mammalian cells yielded a stable protein that was retained in the endoplasmic reticulum, indicating the importance of the propeptide in folding and localization (Tsukuba et al., 2006; Yasuda et al., 2005b). When recombinant mature enzymes were expressed in *E. coli*, they were in the form of insoluble inclusion bodies. An additional refolding step was required to obtain soluble proteins (Fig. 3). Detailed search for a soluble form of the cathepsin E variant 2, using specific polyclonal antibodies, showed low level of protein solubility (Fig. 3). The ability of the purified recombinant mature cathepsin E to hydrolyze protein substrates with broad specificity, such as hemoglobin, albumin and insulin, was compared with that of human cathepsin E variant 2. We found that cathepsin E hydrolyzes the substrates; in contrast, no activity was detected for cathepsin E variant 2. Altering the reaction conditions to neutral pH did not improve the ability of the splice variant 2 to hydrolyze these substrates; however, variant 2 was able to bind pepstatin A. Biotinylated inhibitor (Fig. 1) or pepstatin A bound to the Sepharose resin (results not shown) and interacted with both variants of cathepsin E from HeLa cell lysate, indicating that the biological function of cathepsin E variant 2 is not completely lost.

Based on the CD spectra, we observed that the mature cathepsin E has assumed a defined tertiary structure and a secondary structure typical of an  $\alpha/\beta$  protein (Fig. 4). In contrast, the protein containing the propeptide did not refold properly. For cathepsin D, both the proenzyme and the two-chain processed enzyme were properly folded *in vitro* (Lah et al., 1984). Interestingly, the propeptides of the papain-like cathepsins, such as cathepsins S and L, were involved in the refolding of the mature enzymes (Wiederanders, 2000). For procathepsin L, the propeptide is in a molten globule state at lower pH values, which could play a role in proenzyme activation (Jerala et al., 1998). In contrast to the wild-type cathepsin E, the spliced variant 2 apparently folds to a defined native state, if it is intact and includes the propeptide. However, it does not possess as intense a CD signal in the near UV range as that of cathepsin E, and its secondary structure composition is predominantly  $\beta$ -sheet, which often leads to the amyloid-like aggregation. Indeed, in the gel filtration of the splice variant 2 (Fig. 5A), a peak of high molecular weight has been observed. The protein peak has been isolated and checked for amyloid-like characteristics. The Atomic Force Microscope image (Fig. 5B) showed shallow disk-like particles of height 3 to 4 nm and diameter of 10 to 20 nm. The benzothiol dye, Thioflavin-T (ThT), was used to determine protein self-assembly (fibrillization). ThT binding was positive. Additionally, partial N-linked glycosylation could be due to poor folding and therefore retention in the endoplasmic reticulum.

To rationalize the observed conformational properties and lack of activity of cathepsin E splice variant 2, we built its comparative structure model. Several crystal structures are available for A1 aspartic proteases, including the structure of an activation intermediate of cathepsin E (PDB code 1TZS). In the following discussion, we assume that the structure of 1TZS is similar to the structure of cathepsin E variant 1, since the sequences are identical,

except for the propeptide in 1TZS. The Cathepsin E variant 1 model shows three topologically distinct regions, including an N-terminal domain (residues 67–200), a C-terminal domain (residues 251–377), and a six-stranded anti-parallel  $\beta$ -sheet between the two terminal domains, comprising three strands that are connecting the N and C-terminal domains, the propeptide, and the C-terminal  $\beta$ -strand following the C-terminal domain (Ostermann et al., 2004). Both the N- and C-terminal domains each contribute an Asp-Thr-Gly active site motif with the catalytic aspartic acid residues. Modeling of cathepsin E variant 2, using the structure 1TZS as a template, resulted in a model covering residues 67–363, with reasonable overall quality criteria values (Fig. 6). However, considering the statistically insignificant sequence similarity between residues 268–363 of cathepsin E variant 2 and 1TZS as well as inconclusive fold assignments by MODELLER, GeneThreader (Lobley et al., 2009) and i-Tasser (Roy et al., 2010), this part of the model should be considered unreliable.

To support or reject the cathepsin E variant 2 model, we prepared mutation P268D of cathepsin E variant 2. No activity was observed. Far UV CD spectra of the variant 2 and its P268D mutant show the peak of negative ellipticity at 216 nm, which is characteristic of  $\beta$ -sheet proteins. When the proregion is removed, the secondary structure composition shifts to that characteristic of a more unfolded state. Therefore, it is reasonable to conclude that the structures of the two isoform sequences are similar to each other and 1TZS from the N-terminus to the frame shift. Beyond the frame shift, the sequences are not related. Nevertheless, fold assignments by MODELLER, GeneThreader and i-Tasser predict a very similar structure for both variants up to residue 300 (Fig. 7A). The remainder of the sequence of variant 2 is predicted to be mostly  $\beta$ -sheet, with a short helix at residues 340–345 in some predictions. These results are consistent with secondary structure predictions by PsiPred (Jones, 1999) (Fig. 7B) and with experimental CD spectra. Both catalytic aspartate residues in cathepsin E variant 1 are found within the Asp-Thr-Gly motifs (residues 96–98 and 281–283), whereas variant 2 only possesses one of these motifs (residues 96–98); the second motif position 282–284 is after the frame shift. The mutations of Pro268 or Glu282 into Asp of the cathepsin E variant 2 have been expressed and refolded, and shown to be inactive. Thus, the structure of the splice variant sequence after the frame shift is likely significantly different from that of the wild-type.

It is not clear what role, if any, the enzymatically inactive splice variant plays in cells. Recently, it was demonstrated that cathepsin E specifically induces tumor growth arrest and apoptosis in human prostate carcinoma cell lines without affecting normal cells (Kawakubo et al., 2007). The same group showed that cathepsin E expression in tumor cells induces tumor growth arrest *via* inhibition of angiogenesis and enhanced immune response (Shin et al., 2007). When tumor cells express the functionally inactive splice variant of cathepsin E, probably it substitutes the authentic cathepsin E. The loss of cathepsin E activity and, consequently, an absence of tumor growth arrest can be expected. Another pathological situation due to absence of cathepsin E activity is expected in keratinocyte terminal differentiation, where the cathepsin E activity is functionally linked to the expression of terminal differentiation markers (Kawakubo et al., 2011).

In conclusion, this study characterized cathepsin E and its inactive splice variant, providing a basis for further studies of the role of cathepsin E spliced variant in pathological conditions.

## Materials and methods

The source cDNA clones for cathepsin E (IRAKp961K0951) and cathepsin E splice variant 2 (IMAGp998E045582) were obtained from the Resource Center of the German Human Genome Project (imaGenes, Germany).

### Antibodies

An integrated commercial procedure (GenicBio, Shanghai, China) was used to generate peptides, peptide-carrier conjugates for immunization and antisera. Rabbit polyclonal antibodies were raised against synthetic peptides with an N-terminal Cys to couple the keyhole limpet hemocyanin (KLH). For antibody production against cathepsin E variants 1 and 2, the peptides C-LITGPSDKIKQLQ and C-TLQLGPSGSWGMS, respectively, were used.

### RNA isolation and RT-PCR

Total RNA was isolated from HeLa cells using RNeasy Mini Kit (Qiagen, Germany), according to the manufacturer protocol. Two micrograms of total RNA was reverse-transcribed to cDNA using Omniscript RT Kit (Invitrogen) in a 50- $\mu$ l total reaction volume, followed by polymerase chain reaction (PCR). For exponential amplification, PCR was performed for 30 cycles, followed by visualization of the product by Sybr Safe (Invitrogen) staining and 1 % agarose gel electrophoresis. The specific primers used for cathepsin E variant were described previously (Tatnell et al., 2003).

### Bacterial expression

The cloning and protein expression were carried out as previously described (Hill et al., 1993). Human cathepsin E and cathepsin E variant 2 (splice variant of cathepsin E) were sub-cloned without the N-terminal signal sequence. Two recombinant constructs for cathepsin E and two recombinant constructs for cathepsin E variant 2 were prepared. The longer genes encode the enzymes with the propeptide, while the shorter genes encode mature enzymes without the propeptide. Fragments were amplified from the plasmids by PCR, using Pfu polymerase. The primers used in the reaction are shown in Table 1. In all cloning procedures, an NdeI and XhoI restriction sites were introduced for sub-cloning the PCR products into pET22 and pET28 vectors (Invitrogen, Germany). The nucleotide sequences of newly prepared constructs were verified by DNA sequence analysis (Macrogen, Korea). Expression studies were carried out in the *E. coli* strain BL21-CodonPlus (DE3)-RP cells (Stratagene, USA). Cells were grown at 37 °C in 400 ml of Luria-Bertani medium, containing an appropriate antibiotic. The expression was induced with 1M IPTG. Cells were grown for an additional three hours, then harvested by centrifugation, followed by resuspending the pellets in 50 ml of Wash Buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% Triton X-100). Cells were lysed, by adding lysozyme, to a final protein concentration of 100  $\mu$ g/ml and incubating at 37 °C for 30 min, followed by

sonication on ice. Inclusion bodies were collected by centrifugation at 10,000g for 10 min. The pellet was washed three times with 50 ml of Wash buffer and again collected by centrifugation. The purified inclusion bodies were dissolved in 40 ml of 8 M urea, 20 mM Tris-HCl pH 7.5, 150 mM NaCl at room temperature, and centrifuged at 30,000g for 30 minutes to remove insoluble material. The solubilized sample was reduced before renaturation with DTT at the final concentration of 150 mM. Renaturation was performed in Refolding buffer pH 9.5 (50 mM Tris-HCl pH 9.5, 1m M EDTA and 150 mM NaCl). 5 ml of the cathepsin E variant 2 with concentration 15 mg/ml was diluted in 500 ml of Refolding buffer pH 9.5. 1 mM reduced glutathione and 0.2 mM oxidized glutathione were added and then the solution was dialyzed over night against 20 mM Tris-HCl pH 8.0, containing 1mM EDTA. Finally, the protein sample was concentrated. The renaturation of the cathepsin E variant 2 was performed with a similar protocol. The dilution and dialysis were performed with Refolding buffer pH 8.0 (50 mM Tris-HCl pH 8.0, 1m M EDTA and 150 mM NaCl) and no glutathione was required.

Mutations of cathepsin E variant 2 were performed at positions P268D and E282D. The sets of primers each harbored a single amino acid residue mutation (Table 1). The amplification of the expression vector pET22 including the cathepsin E variant 2 gene was performed. The polymerase chain reaction products were treated with DpnI to digest the parental DNA template. The reaction mixture was directly transformed into *E. coli* for amplification. Sequencing verified the mutation. Expression and renaturation were performed as for the wild-type protein.

### Expression in mammalian cells

Human cathepsin E and cathepsin E variant 2 were sub-cloned into pDsRed-Monomer-N1 and pEGFP-N2, respectively (Table 1). Oligonucleotides were designed to allow using the XhoI and KpnI restriction enzymes for sub-cloning. Preceding the signal sequences of both genes, the Kozak sequence was included. The stop codons were removed to obtain fusions with red and green fluorescent proteins.

The HEK-293T cells (ATCC, USA) were grown in DMEM medium (Invitrogen, Germany) with Glutamax, supplemented with 10% solution of FCS, 500 U/ml penicillin and 500 mg/ml streptomycin to approximately 95% confluence in a 10-cm dish. The culture was washed twice in PBS, detached with trypsin and seeded onto 12-well plates with round 15 mm glass cover slips (Electron Microscopy Sciences, USA) at a density of 100,000 cells/well 24 h before transfection. On the day of co-transfection, 2 mg of each plasmids DNA were combined, diluted into 500  $\mu$ l Optimem medium (Invitrogen, Germany) and supplemented with 2,5  $\mu$ l FuGene HD transfection reagent (Roche, Germany). The following procedure was performed according to manufacturer recommendation. Cells were further grown in 12-well dishes on glass cover slips. On first, second and third days after transfection, cells on glass cover slips were fixed in 4% paraphormaldehyde in phosphate-buffered saline and permeabilized with 0.1% Triton-X100, and then washed 4 times with 0,1 % glycine in 1 x PBS buffer. Afterwards, fixed cells were mounted in the Vectashield mounting medium, which includes the DAPI fluorescent nuclear stain (Vector Labs, Burlingame, CA). The fluorescence of cell samples was followed with a Hamamatsu Orca-

R2 CCD on an inverted Olympus IX-81 microscope. The resulting digitized images were artificially colored.

### Western blot analysis

Proteins were transferred from SDS electrophoresis onto the PVDF membrane, and detected by goat polyclonal anti-cathepsin E antibodies (Abcam, UK). Non-specific binding was blocked with 2% BSA and 0.4% Tween 20 in PBS, pH 7.4. The membrane was incubated with antibodies to cathepsin E (5µg/ml), followed by HRP-conjugated rabbit anti-goat IgG secondary antibody (Jackson ImmunoResearch Lab., USA).

### Interaction with the biotinylated pepstatin A and pepstatin A-Sepharose

Pepstatin was labeled with biotin using amine-reactive biotinylation reagent NHS-PEG4-Biotin (Thermo Fisher Scientific, Germany) according to manufacturer's instructions. 20 mM pepstatin in 50% ethanol was mixed with 2 mg of activated biotin. The cell lysate of three 25 cm<sup>2</sup> tissue culture flasks (TPP, Switzerland) of confluent HeLa cells were prepared using non-denaturing extraction buffer (20 mM Tris HCl, 137 mM NaCl, 10% glycerol, 1% Nonidet P-40, 2 mM EDTA pH 8). The extract was incubated with biotinylated pepstatin A for two hours at 4°C and then Streptavidin Agarose from the same producer was used to remove unbound proteins. The experiment was performed using pepstatin A-Sepharose affinity chromatography, as well, under identical conditions as described for purification of cathepsin D (Smith and Turk, 1974). SDS-PAGE and Western blot were performed to analyze binding of cathepsin E and its splice variant.

### Immunoprecipitation

The cell lysate of three 25 cm<sup>2</sup> tissue culture flasks (TPP, Switzerland) of confluent HeLa cells was prepared using non-denaturing extraction buffer (20 mM Tris HCl, 137 mM NaCl, 10% glycerol, 1% Nonidet P-40, 2 mM EDTA pH 8). Half of the sample was mixed with the specific rabbit antibodies against cathepsin E variant1 and another half with the specific rabbit antibodies against cathepsin E variant 2. Samples were incubated overnight at 4°C and applied on protein A Sepharose (GE Healthcare). Nonspecifically bound proteins were removed at pH 8.5 (150 mM glycine/NaOH buffer), and proteins were eluted at pH 3.5 (20 mM glycine/HCl buffer). Eluted sample was immediately neutralized to pH 8.0 using 1M Tris/HCL buffer. Both samples were analyzed with SDS-PAGE and Western blot using polyclonal mouse antibodies against cathepsin E.

### Enzyme Assay

Cathepsin E activity was assayed according to the protocols published for the KYS-1 substrate (MOCAc-Gly-Ser-Pro-Ala-Phe-Leu-Ala-Lys(Dnp)-D-Arg-NH<sub>2</sub>) (Yasuda et al., 2005a), FITC-hemoglobin, albumin, insulin, and substance P, as well as in 100 mM sodium phosphate buffer pH 7.0.

### Deglycosylation of cathepsin E variant 2

The cell lysate of HeLa cells was prepared under denaturation conditions using RIPA buffer. 500 µg of sample was deglycosylated using PNGase F (New England Biolabs, UK) or Endo



H (New England Biolabs, UK) under reducing conditions according to the manufacturer's instructions. All reactions were performed in duplicate. Deglycosylated lysates and media were characterized by SDS-PAGE analysis and Western blotting.

### CD spectra measurements

CD spectra were recorded using an Aviv (Lakewood, NJ, USA) model 60 DS CD spectropolarimeter. A 10 mm mini-rectangular cell (near UV CD), 1 mm rectangular cell and 0.2 mm demountable rectangular cell (both for far UV CD) were used. The corresponding bandwidths were set at 0.5 nm (for the near UV) and 1.0 nm (for the far UV). The temperature was 20° C. Data in the far UV were collected every 1 nm, using a scan rate of 0.25 nm/s and data in the near UV every 0.5 nm at a scan rate of 0.16 nm/s. The usual protein concentration was A280 = 1.8 for the near UV and A280 = 0.3 for the far UV region. As no extinction coefficient are known for cathepsin E and its splice variant, the ellipticity was recalculated to  $\text{deg.cm}^2.\text{dmol}^{-1}$  using estimated extinction coefficients.

### Structural Analysis

Comparative protein structure modeling was carried out using the automated comparative protein structure modeling pipeline MODPIPE (Pieper et al., 2011), which relies primarily on various modules of MODELLER (Sali and Blundell, 1993) for its functionality. Briefly, template selection was performed using a variety of sequence-sequence, sequence-profile and profile-profile methods, using an E-value threshold of 1.0, and the resulting alignments were then used to compute comparative models using MODELLER. The resulting model was assessed to be reliable by the normalized DOPE (-0.37), GA341 (1.0), and MPQS (1.3) scores (ref1). More details are available at [http://salilab.org/modbase/search?dataset\\_name=MW-cath\\_E](http://salilab.org/modbase/search?dataset_name=MW-cath_E).

### Acknowledgments

This work was supported by the Slovenian Research Agency research project J1-2307 (Vito Turk).

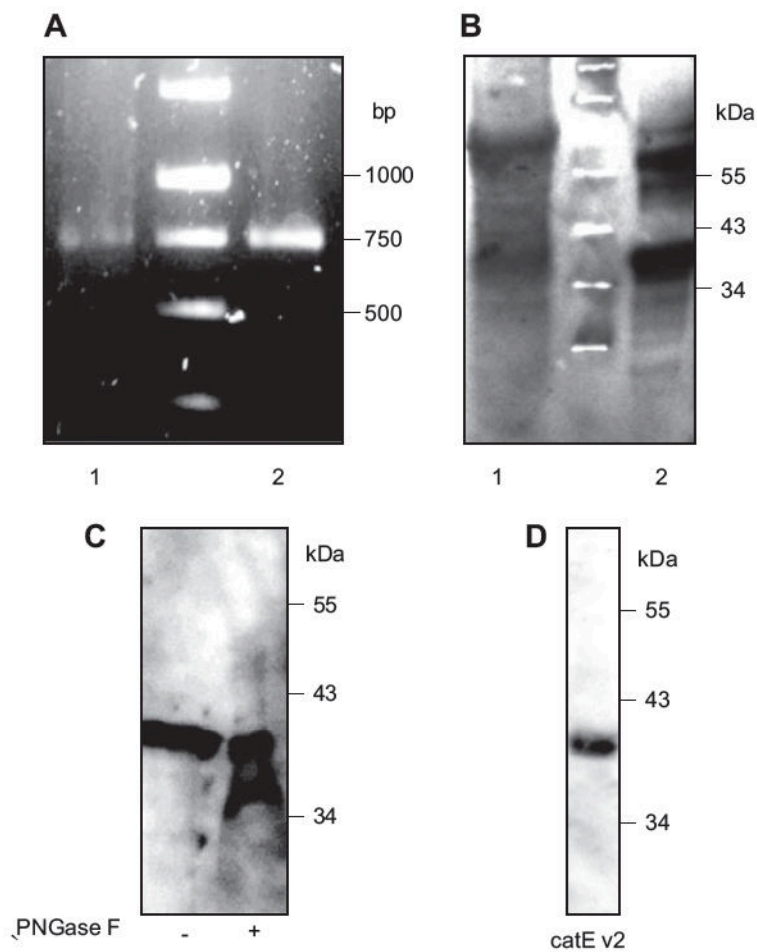
### References

- Abd-Elgaliel WR, Tung CH. Selective detection of Cathepsin E proteolytic activity. *Biochimica et Biophysica Acta (BBA)-General Subjects*. 2010; 1800:1002–1008.
- Arnold D, Keilholz W, Schild H, Dumrese T, Stevanovi S, Rammensee HG. Evolutionary conserved cathepsin E substrate specificity as defined by N-terminal and C-terminal sequencing of peptide pools. *Biol Chem*. 1997; 378:883–891. [PubMed: 9377485]
- Azuma T, Liu W, Vander Laan DJ, Bowcock AM, Taggart RT. Human gastric cathepsin E gene: Multiple transcripts result from alternative polyadenylation of the primary transcripts of a single gene locus at 1q31–q32. *J Biol Chem*. 1992; 267:1609–1614. [PubMed: 1370478]
- Ben-Dov C, Hartmann B, Lundgren J, Valcarcel J. Genome-wide analysis of alternative pre-mRNA splicing. *J Biol Chem*. 2008; 283:1229–1233. [PubMed: 18024428]
- Bennett K, Levine T, Ellis JS, Peanasky RJ, Samloff IM, Kay J, Chain BM. Antigen processing for presentation by class II major histocompatibility complex requires cleavage by cathepsin E. *Eur J Immunol*. 1992; 22:1519–1524. [PubMed: 1601038]
- Burster T, Reich M, Zaidi N, Voelter W, Boehm BO, Kalbacher H. Cathepsin E regulates the presentation of tetanus toxin C-fragment in PMA activated primary human B cells. *Biochem Biophys Res Commun*. 2008; 377:1299–1303. [PubMed: 18996084]

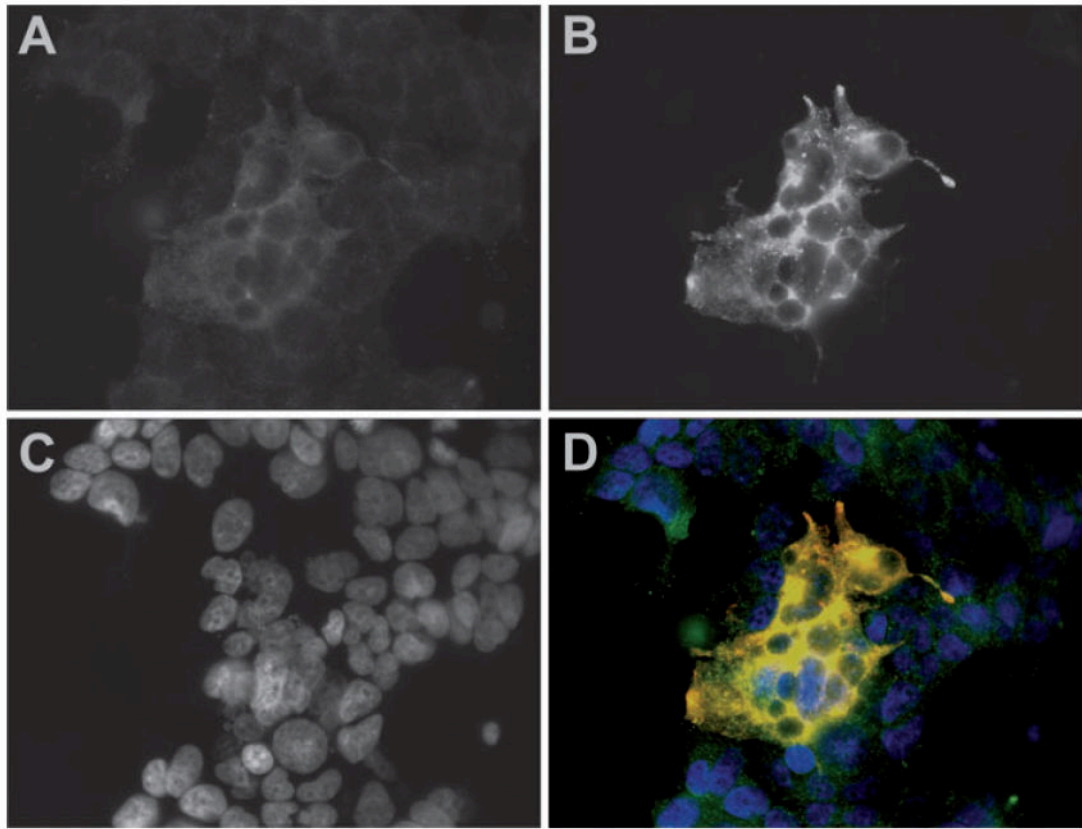
- Chain BM, Free P, Medd P, Swetman C, Tabor AB, Terrazzini N. The expression and function of cathepsin E in dendritic cells. *J Immunol.* 2005; 174:1791–1800. [PubMed: 15699105]
- Ellgaard L, Molinari M, Helenius A. Setting the Standards: Quality Control in the Secretory Pathway. *Science.* 1999; 286:1882–1888. [PubMed: 10583943]
- Finley EM, Kornfeld S. Subcellular localization and targeting of cathepsin E. *J Biol Chem.* 1994; 269:31259–31266. [PubMed: 7983070]
- Fowler SD, Kay J, Dunn BM, Tatnell PJ. Monomeric Human Cathepsin-E. *FEBS Lett.* 1995; 366:72–74. [PubMed: 7789521]
- Hill J, Montgomery DS, Kay J. Human cathepsin E produced in *E. coli*. *FEBS Lett.* 1993; 326:101–104. [PubMed: 8325357]
- James, MNG. Catalytic pathway of aspartic peptidases. In: Barrett, AJ.; Rawlings, NDF.; WJ, editors. *Handbook of Proteolytic Enzymes.* Elsevier; Amsterdam: 2004. p. 12-19.
- Jerala R, Zerovnik E, Kidric J, Turk V. pH-induced conformational transitions of the propeptide of human cathepsin L. A role for a molten globule state in zymogen activation. *J Biol Chem.* 1998; 273:11498–11504. [PubMed: 9565563]
- Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol.* 1999; 292:195–202. [PubMed: 10493868]
- Kageyama T, Ichinose M, Yonezawa S. Processing of the precursors to neurotensin and other bioactive peptides by cathepsin E. *J Biol Chem.* 1995; 270:19135–19140. [PubMed: 7642580]
- Kakehashi H, Nishioku T, Tsukuba T, Kadowaki T, Nakamura S, Yamamoto K. Differential regulation of the nature and functions of dendritic cells and macrophages by cathepsin E. *J Immunol.* 2007; 179:5728–5737. [PubMed: 17947645]
- Kawakubo T, Okamoto K, Iwata JI, Shin M, Okamoto Y, Yasukochi A, Nakayama KI, Kadowaki T, Tsukuba T, Yamamoto K. Cathepsin E prevents tumor growth and metastasis by catalyzing the proteolytic release of soluble TRAIL from tumor cell surface. *Cancer Res.* 2007; 67:10869–10878. [PubMed: 18006832]
- Kawakubo T, Yasukochi A, Okamoto K, Okamoto Y, Nakamura S, Yamamoto K. The role of cathepsin E in terminal differentiation of keratinocytes. *Biological Chemistry.* 2011; 392:571–585. [PubMed: 21521076]
- Lah T, Drobnic-Kosorok M, Turk V, Pain RH. Conformation, structure and activation of bovine cathepsin D. Unfolding and refolding studies. *Biochem J.* 1984; 218:601–608. [PubMed: 6712634]
- Lapresle C, Webb T. The purification and properties of a proteolytic enzyme, rabbit cathepsin E, and further studies on rabbit cathepsin D. *The Biochemical Journal.* 1962; 84:455–462. [PubMed: 14462361]
- Lees WE, Kalinka S, Meech J, Capper SJ, Cook ND, Kay J. Generation of human endothelin by cathepsin E. *FEBS Lett.* 1990; 273:99–102. [PubMed: 2226872]
- Lobley A, Sadowski MI, Jones DT. pGenTHREADER and pDomTHREADER: New methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics.* 2009; 25:1761–1767. [PubMed: 19429599]
- Ostermann N, Gerhartz B, Worpenberg S, Trappe J, Eder J. Crystal structure of an activation intermediate of cathepsin E. *J Mol Biol.* 2004; 342:889–899. [PubMed: 15342244]
- Pieper U, Webb BM, Barkan DT, Schneidman-Duhovny D, Schlessinger A, Braberg H, Yang Z, Meng EC, Pettersen EF, Huang CC, Datta RS, Sampathkumar P, Madhusudhan MS, Sjölander K, Ferrin TE, Burley SK, Sali A. ModBase, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res.* 2011; 39:D465–D474. [PubMed: 21097780]
- Puizdar V, Lapresle C, Turk V. Evidence for the presence of large amounts of cathepsin E in rat spleen. *FEBS Lett.* 1985; 185:236–238. [PubMed: 3996601]
- Rawlings ND, Morton FR, Kok CY, Kong J, Barrett AJ. MEROPS: The peptidase database. *Nucleic Acids Res.* 2008; 36:D320–D325. [PubMed: 17991683]
- Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols.* 2010; 5:725–738.
- Saku T, Sakai H, Shibata Y, Kato Y, Yamamoto K. An immunocytochemical study on distinct intracellular localization of cathepsin E and cathepsin D in human gastric cells and various rat cells. *Journal of Biochemistry.* 1991; 110:956–964. [PubMed: 1794985]

- Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993; 234:779–815. [PubMed: 8254673]
- Samloff IM, Taggart RT, Shiraishi T. Slow moving proteinase. Isolation, characterization, and immunohistochemical localization in gastric mucosa. *Gastroenterology.* 1987; 93:77–84. [PubMed: 3556306]
- Shin M, Kadowaki T, Iwata JI, Kawakubo T, Yamaguchi N, Takii R, Tsukuba T, Yamamoto K. Association of cathepsin E with tumor growth arrest through angiogenesis inhibition and enhanced immune responses. *Biol Chem.* 2007; 388:1173–1181. [PubMed: 17976010]
- Smith R, Turk V. Cathepsin D: rapid isolation by affinity chromatography on haemoglobin agarose resin. *Eur J Biochem.* 1974; 48:245–254. [PubMed: 4448170]
- Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj TA, Soreq H. Function of alternative splicing. *Gene.* 2005; 344:1–20. [PubMed: 15656968]
- Stefanovic J, Webb T, Lapresle C. Etude des cathepsines D et e dans des preparations de polynucleaires, de macrophages et de lymphocytes de lapin. *Ann Ins Pasteur.* 1962; 103:276–284.
- Tatnell PJ, Cook M, Kay J. An alternatively spliced variant of cathepsin E in human gastric adenocarcinoma cells. *Biochim Biophys Acta Gene Struct Expression.* 2003; 1625:203–206.
- Tatnell PJ, Lees WE, Kay J. Cloning, expression and characterisation of murine procathepsin E. *FEBS Lett.* 1997; 408:62–66. [PubMed: 9180269]
- Tsukuba T, Hori H, Azuma T, Takahashi T, Taggart RT, Akamine A, Ezaki M, Nakanishi H, Sakai H, Yamamoto K. Isolation and characterization of recombinant human cathepsin E expressed in Chinese hamster ovary cells. *J Biol Chem.* 1993; 268:7276–7282. [PubMed: 8463263]
- Tsukuba T, Ikeda S, Okamoto K, Yasuda Y, Sakai E, Kadowaki T, Sakai H, Yamamoto K. Characterization of rat cathepsin E and mutants with changed active-site residues and lacking propeptides and N-glycosylation, expressed in human embryonic kidney 293T cells. *FEBS J.* 2006; 273:219–229. [PubMed: 16367762]
- Tsukuba T, Okamoto K, Okamoto Y, Yanagawa M, Kohmura K, Yasuda Y, Uchi H, Nakahara T, Furue M, Nakayama K, Kadowaki T, Yamamoto K, Nakayama KI. Association of Cathepsin E Deficiency with Development of Atopic Dermatitis. *Journal of Biochemistry.* 2003; 134:893–902. [PubMed: 14769879]
- Wiederanders B. The function of propeptide domains of cysteine proteinases. *Adv Exp Med Biol.* 2000; 477:261–270. [PubMed: 10849753]
- Yamada M, Azuma T, Matsuba T, Iida H, Suzuki H, Yamamoto K, Kohli Y, Hori H. Secretion of human intracellular aspartic proteinase cathepsin E expressed in the methylotrophic yeast, *Pichia pastoris* and characterization of produced recombinant cathepsin E. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology.* 1994; 1206:279–285.
- Yamamoto K, Katsuda N, Kato K. Affinity purification and properties of cathepsin-E-like acid proteinase from rat spleen. *Eur J Biochem.* 1978; 92:499–508. [PubMed: 33048]
- Yamamoto K, Kawakubo T, Yasukochi A, Tsukuba T. Emerging roles of cathepsin E in host defense mechanisms. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics.* 2011 In Press, Corrected Proof.
- Yasuda Y, Ikeda S, Sakai H, Tsukuba T, Okamoto K, Nishishita K, Akamine A, Kato Y, Yamamoto K. Role of N-glycosylation in cathepsin E. *Eur J Biochem.* 1999; 266:383–391. [PubMed: 10561578]
- Yasuda Y, Kohmura K, Kadowaki T, Tsukuba T, Yamamoto K. A new selective substrate for cathepsin E based on the cleavage site sequence of alpha 2-macroglobulin. *Biol Chem.* 2005a; 386:299–305. [PubMed: 15843176]
- Yasuda Y, Tsukuba T, Okamoto K, Kadowaki T, Yamamoto K. The role of the cathepsin E propeptide in correct folding, maturation and sorting to the endosome. *Journal of Biochemistry.* 2005b; 138:621–630. [PubMed: 16272574]
- Yonezawa S, Nakamura K. Species-specific distribution of cathepsin E in mammalian blood cells. *Biochimica et Biophysica Acta - General Subjects.* 1991; 1073:155–160.
- Zaidi N, Hermann C, Herrmann T, Kalbacher H. Emerging functional roles of cathepsin E. *Biochem Biophys Res Commun.* 2008; 377:327–330. [PubMed: 18938134]

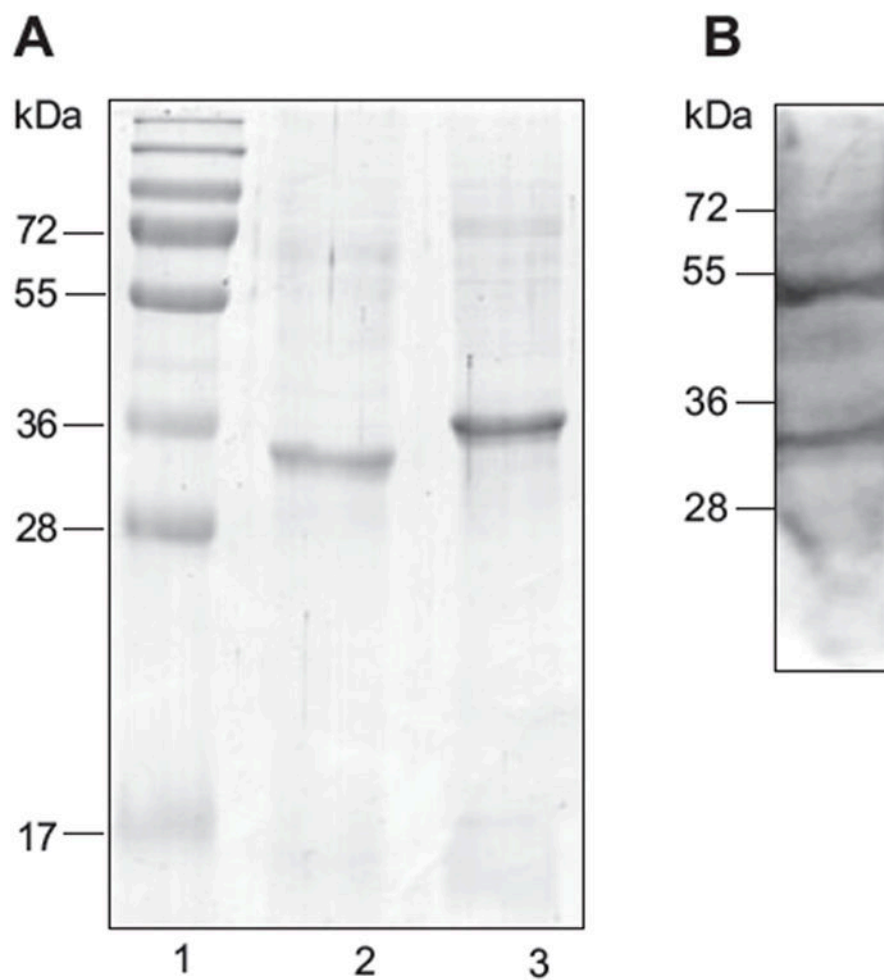
- Zaidi N, Herrmann T, Baechle D, Schleicher S, Gogel J, Driessen C, Voelter W, Kalbacher H. A new approach for distinguishing cathepsin E and D activity in antigen-processing organelles. *FEBS J.* 2007; 274:3138–3149. [PubMed: 17521331]
- Zaidi N, Kalbacher H. Cathepsin E: A mini review. *Biochem Biophys Res Commun.* 2008; 367:517–522. [PubMed: 18178150]



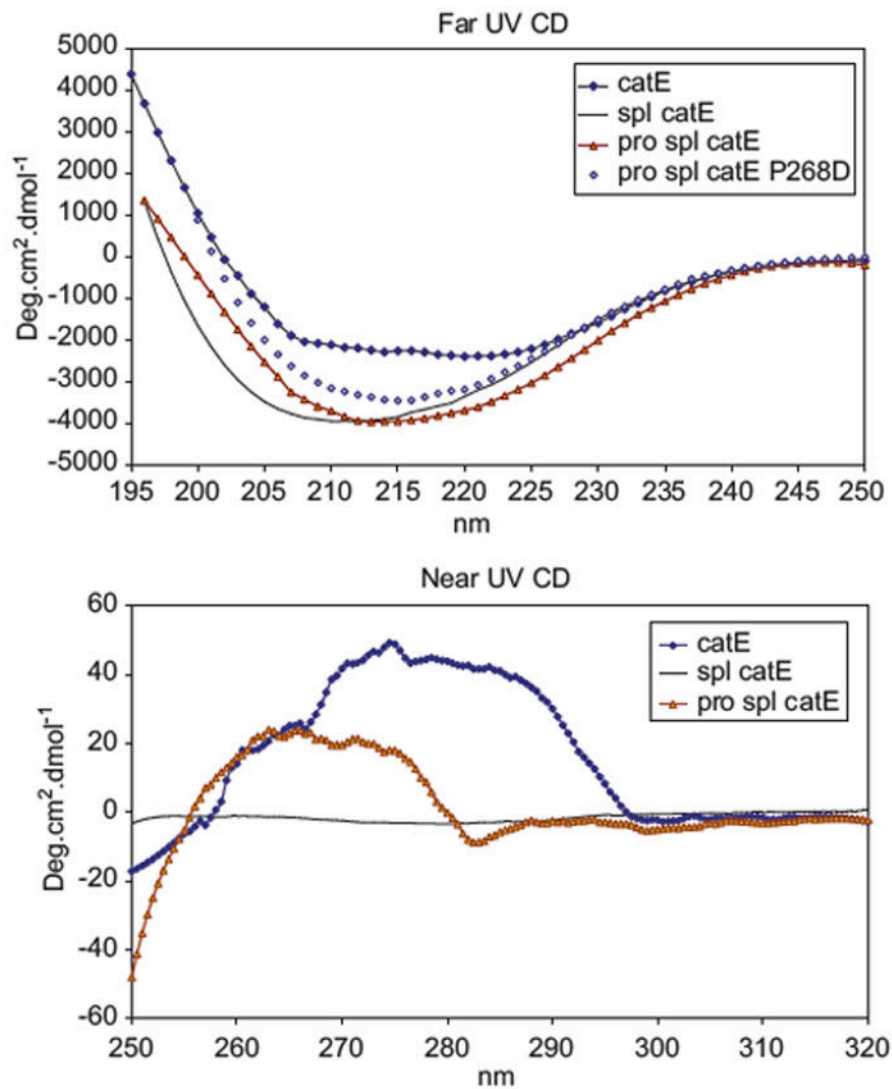
**Fig. 1.** The comparison of endogenously expressed cathepsin E variant 2 in HeLa cells (lines 1) and heterologously expressed protein in HEK-293 cells (lines 2) performing (A) RT-PCR and (B) Western blot. (C) Isolation of both cathepsin E variants using biotinylated pepstatin A, line 1, cathepsin E; line 2, cathepsin E variant 2.



**Fig. 2.** Transient co-expression of human cathepsin E and human cathepsin E variant 2 in HEK-293T cells. Cathepsin E is fused to red fluorescent protein and cathepsin E variant 2 is fused to green fluorescent protein. (A) Green channel; (B) red channel; (C) blue channel; (D) superposition of all three channels.

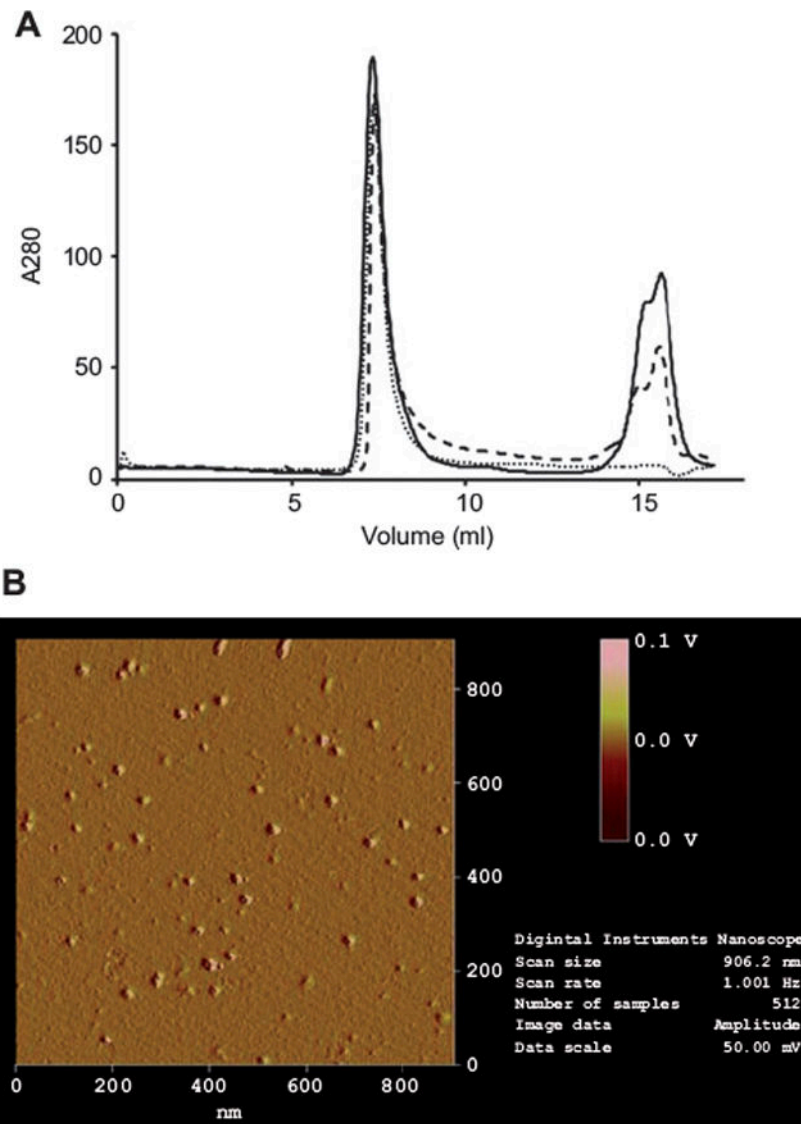


**Fig. 3.** Expression in *E. coli*. (A) SDS-PAGE of expressed and folded recombinant proteins. (1) protein standards; (2) mature cathepsin E variant 2; (3) mature cathepsin E. (B) Western blot of soluble mature cathepsin E variant 2 in cell lysate.

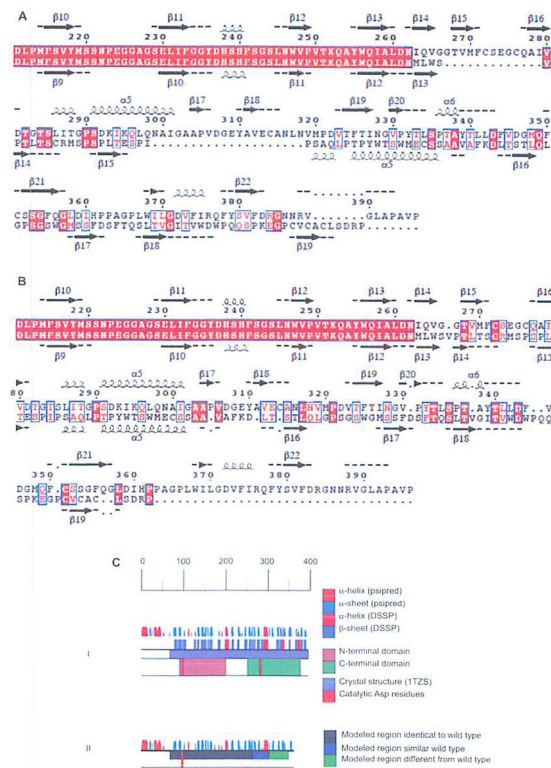


**Fig. 4.** CD spectra of cathepsin E (without proregion) and cathepsin E splice variant. The splice variant with proregion and its P286D mutant were also prepared. Near UV CD (informative about tertiary structure) and far UV CD (informative about secondary structure composition) spectra were recorded.

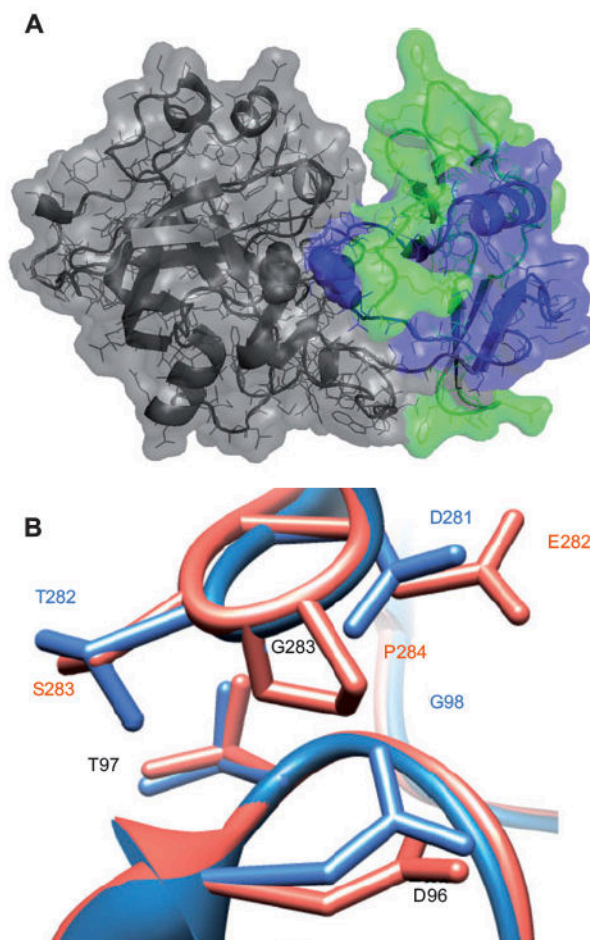




**Fig. 5.** Protein aggregation. (A) Gel filtration of cathepsin E splice variant on Superdex HR 75 column on ÄKTA FPLC. The sample was prepared by refolding from inclusion bodies. The refolded sample before incubation is marked by the solid line. The sample incubated for 2 days at 4° C is marked by the dashed line and the sample incubated 2 days at room temperature is marked by the dotted line. (B) Atomic force microscopy of peak 1 from the gel filtration. The aggregates can be seen. The scale of the view is 800 nm; the particle size is 3 – 4 nm in height and 10 – 20 nm in cross-section.



**Fig. 6.** Alignment of cathepsin E and cathepsin E variant 2. (A) Manual alignment. (B) The ModPipe alignment. (C) Feature Sketch of the two isoforms of Cathepsin E: (I) cathepsin E: Top row: secondary structure prediction (PsiPred); second row: secondary structure assignment (DSSP); third row: sequence region covered by 1TZS; Third row: N-terminal and C-terminal domains, catalytic residues marked in red; (II) cathepsin E splice variant 2: Top row: secondary structure prediction (PsiPred); second row: sequence region covered by 1TZS at 100% sequence identity.



**Fig. 7.** The model of cathepsin E splice variant 2. (A) Part of the molecule, colored in grey, is predicted to be identical to cathepsin E. The part of the sequence indicated in the green/blue color is predicted to be different from cathepsin E. The blue region (residues 263–302) is predicted to have a similar fold as variant 1, (one  $\alpha$ -helix and 2  $\beta$ -strands), while the green region of the model (residues 303–348) is predicted by PsiPred to be predominantly  $\beta$ . The modeled structure for the green region is predicted unreliable. (B) Superposition of the active site residues of the cathepsin E splice variant 2 model (red) with 1TFZ (blue). The Asp-Thr-Gly motifs in 1TFZ and the equivalent residues in cathepsin E splice variant 2 are shown. Black labels indicate identical residue types in both structures.

Table 1

Oligonucleotides for cloning and mutagenesis.

Organism	Insert	Orientation	Nucleotide sequence <sup>a,b</sup>
Mammal	catEsv+catE	5'	GAC ATC TCG AGC GCC ACC ATG AAA ACG CTC CTT CTT TT
Cells	catEsv <sup>c</sup>	3'	GTA GGT ACC CAG GTC TGT CAG ACA GGC AG
	catE <sup>c</sup>	3'	GTA GGT ACC TCG GGG ACT GCT GGG GCC AGT
<i>E. coli</i>	pro-catEsv	5'	CAT CAT ATG GCC CAA GGA TCC CTT CAC
	mat-catEsv	5'	CAT CAT ATG ATC CAG TTC ACC GAG
	catEsv	3'	GTA CTC GAG CTA TCA AGG TCT GTC AGA
	pro-catE	5'	CAT CAT ATG GCC CAA GGA TCC CTT CAC
	mat-catE	5'	CAT CAT ATG ATC CAG TTC ACC GAG
	catE	3'	GTA CTC GAG TCA GGG GAC TGC TGG GGC C
	P268D <sup>d</sup>	5'	GAA TAT GCT GTG GAG TGT GGA TAC CTT AAC GTC ATG CCG G
	P268D <sup>d</sup>	3'	CCG GCA TGA CGT TAA GGT ATC CAC ACT CCA CAG CAT ATT C
	E282D <sup>d</sup>	5'	CTT CAC CAT TAA CGG AIT CCC CTA TAC CCT CAG
	E282D <sup>d</sup>	3'	CTG AGG GTA TAG GGG AAT CCG TTA ATG GTG AAG

<sup>a</sup>Underlined are restriction sites introduced for cloning.

<sup>b</sup>Boldface indicates nucleotides for inserting a desired mutation.

<sup>c</sup>The 3' primers used for mammalian cells for catE (cathepsin E) and catEsv (cathepsin E splice variant 2) have been designed for vectors pDsRed-N1 and pEGFP-N2, respectively.

<sup>d</sup>Mutations of mature cathepsin E variant 2 (mat-catE).