UDC 577.152.34:579.842.11
ISSN 1330-9862 *original scientific paper*

(FTB-1007)

# Endoproteolytic Pattern Observed During Refolding of a Human Exopeptidase Proenzyme, Procathepsin H, Produced in *Escherichia coli*

*Marko Dolinar*, Andreja Mehle[1], Bojana Mozetič Francky[2], Ana Schweiger and Vito Turk*

[1] current address: Ljubljana Pharmacies, Ul. stare pravde 11, SI-1000 Ljubljana

[2] current address: Biohit d.o.o., Gorazdova 7, SI-1000 Ljubljana

Department of Biochemistry and Molecular Biology, Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia

Received: November 11, 1999
Accepted: January 21, 2000

**Summary**

Cathepsin H is a lysosomal cysteine proteinase with exopeptidase activity. Total mRNA was purified from human endometrium cells and converted to cDNA. The procathepsin H coding region was amplified by specific primers, cloned and expressed by the T7-polymerase controlled expression vector pET3a and the bacterial strain *Escherichia coli* BL21[DE3]pT-Trx. The majority of the recombinant procathepsin H was present in the insoluble form, which comprised 25 % of total bacterial proteins. After solubilisation of inclusion bodies, procathepsin H was refolded by dialysis. In the process of renaturation, recombinant procathepsin H was proteolytically degraded into several distinct fragments which were detected by monoclonal antibodies directed toward the N- or C- terminus of the mature enzyme. The degradation pattern was typical for an endopeptidase and a stable LMW fragment of 14 kDa could be assigned to the C-terminal region of the mature enzyme. In refolding, human procathepsin H thus undergoes autolysis by an endopeptidase mechanism, previously disputed for this enzyme.

*Key words:* cathepsin H, endopeptidase, exopeptidase, *Escherichia coli*

## Introduction

Human lysosomes are complex organelles involved in intracellular protein breakdown. They contain among others a number of related cysteine proteinases, including cathepsin H (*1*), which in contrast to other cysteine cathepsins is an aminopeptidase, cleaving single amino acid residues from the N-terminus of the substrate.

The three-dimensional structure of porcine cathepsin H (*2*) revealed that the enzyme's mini-chain, an octapeptide derived from the pro-part of the enzyme and attached to the C-terminal part of the mature enzyme by a disulphide bond, defines the aminopeptidase function of the enzyme. Endopeptidase activity was detected on synthetic peptide substrates only (*3*).

*In vivo*, cathepsin H is synthesised as a preproenzyme. After removal of the signal peptide and posttranslational modifications it migrates into lysosomes as a glycosylated proenzyme (41 kDa) (*4*). Cathepsin H can be found in lysosomes as a single-chain mature protein of 28 kDa (*5*). It can be isolated from tissues in a single-chain and a two-chain form with varying ratios

* Corresponding author; Tel.: ++386 (0)61 1773 623; Fax: ++386 (0)61 273 594; E-mail: Marko.Dolinar@IJS.SI

which most likely reflect the stringency of isolation procedure used (*6*). Cathepsin H (but not its pro-form) is irreversibly inactivated in neutral and alkaline environment (*7*). Latest investigations on macrophages (*8*) localise the bulk of the cathepsin H into early endosomes rather than lysosomes. It has been recently observed that increased levels of cathepsin H correlate with various types of cancer and that it can be used as a prognostic marker, *e.g.* in metastatic melanoma (*9*).

Because of its many functions, we decided to produce recombinant human procathepsin H in bacterial cells, taking advantage of the established T7 expression system, which we used with success for preparation of several related cysteine proteinases. For isolation, our starting material was the insoluble fraction of bacterial lysates which contained highly enriched, though misfolded, procathepsin H. In the process of renaturation, we unexpectedly observed trimming of the full-size protein. Here, we report on this observation and suggest explanations for endopeptidase activity of an exopeptidase proenzyme, procathepsin H.

## Materials and Methods

### *PCR amplification and cloning*

From human uteral endometrium, mRNA was isolated and converted to cDNA by the cDNA Synthesis kit (Amersham International, Little Chalfont, UK). Amplification primers used were:

5′-AGGAAAGAATTCATATGGCCGAACTGTCCGTGAACTCCTTA-3′ and

5′-AATCTAGAAAGCTTGGATCCCGGGACGGCTCACACCAGAGGGA-3′

The 5′ amplification primer was designed so as to include *Eco*RI and *Nde*I restriction sites and a starting Met codon, followed by 24 bases complementary to the coding region for procathepsin H. The 3′ amplification primer contained restriction sites *Hind*III, *Bam*HI and *Sma*I on the flanking region and 22 bases complementary to the procathepsin H cDNA sequence. The obtained PCR product was inserted into pALTER (Stratagene, La Jolla, USA) and the DNA sequence was checked using universal sequencing primers and an internal primer and using the T7 sequencing kit (Pharmacia, Uppsala, Sweden).

Due to the limited choice of restriction sites in the expression vector pET3a (Novagen, Madison, USA) and having an internal *Bam*HI site in procathepsin H coding region, we adapted the cloning scheme accordingly. The vector was first digested by *Bam*HI, the ends were filled with Klenow enzyme and cleaved by *Nde*I, while the procathepsin H coding fragment was cleaved by *Nde*I and *Sma*I and ligated with the vector fragment. The resulting vector (pBMF2) was introduced into *E. coli* DH5α by the procedure of Inoue *et al.* (*10*). The restriction pattern was checked and the plasmid was isolated from the bacterial cells and used for transformation of the *E. coli* strain BL21[DE3] (Novagen) harbouring an additional plasmid pT-Trx (kindly provided by dr. Shunsuke Ishii, RIKEN Tsukuba Life Science Centre) (*11*), coding for *E. coli* thioredoxin.

### *Heterologous gene expression*

An overnight culture of a single *E. coli* BL21[DE3] pT-Trx pBMF2 colony was prepared in LBAC medium (yeast extract 5 g/L, casein hydrolysate 10 g/L, NaCl 10 g/L, ampicillin 200 mg/L, chloramphenicol 25 mg/L). A starter overnight culture (100 mL) was prepared in LBAC, centrifuged (10 min at 20 °C and 6500 g) and the cells resuspended in 40 mL of the same medium. 5 mL of this suspension was used for inoculation of 500 mL LBAC in 2 L shaking flasks. In a preparative experiment, 8 × 500 mL cultures were grown at 200 rpm and 37 °C until $A_{550}$ reached 0.7–1.0. At this stage, the cultures were supplemented with IPTG (isopropyl-β-*-D*-thiogalactopyranoside) to 0.4 mM final concentration. Growth was continued for 3 h post induction.

### *Protein purification*

Cells from 4 L culture were harvested by centrifugation (10 min at 20 °C and 6500 g), resuspended in 40 mL of 10 mM TrisHCl pH=9.0, frozen and thawn (three times) and sonified (3 × 30 s on ice). To the cell suspension 5 mg of lysozyme was added and the lysate was incubated for 25 min at 22 °C. After centrifugation (20 min at 4 °C and 9000 g), supernatants were discarded. To the pellets, 40 mL of 50 mM TrisHCl pH=8.0, 1 % Triton X-100 was added and the suspension was mixed for 20 min on ice. The centrifugation step was repeated (as above) and the pellet was mixed with 40 mL of 50 mM TrisHCl pH=8.0, 1 M guanidine hydrochloride (GdmHCl) as above. Centrifugation was repeated and pellet was dissolved by mixing over night at 4 °C in 45 mL of 0.3 M $Na_2SO_3$, 7 M GdmHCl (pH set to 8 – 8.5).

To 15 mL of solubilized inclusion bodies, 2.5 mL 0.25 mM NTSB (disodium 2-nitro-5-thiosulfobenzoate) was added and mixed in dark for 30 min on ice. The sample was immediately applied to a Sephacryl S-200 column (2.7 cm × 61.5 cm) and eluted with 3 M GdmHCl, 20 mM TrisHCl pH=8.0, 1 mM EDTA, and 0.25 M NaCl. Fractions were collected each 5 min at a flow rate of 39 mL/h.

### *Refolding*

After gel filtration, pooled fractions containing proteins of the expected size (32–35 kDa) were subjected to dialysis with delayed addition of cysteine (Marko Dolinar and Barbara Kahne, unpublished procedure), later referred to as 'standard renaturation procedure'. Recombinant protein (2 mg/mL) was diluted with 2.5 M GdmHCl, 20 mM TrisHCl pH=8.0, 0.015 % Triton X-100 to a final protein concentration of 50 μg/mL and dialysed against 17 volumes of 1.2 M GdmHCl, 20 mM TrisHCl pH=8.0 and 0.015 % Triton X-100 at 4 °C. After 2 h, cysteine was added to the dialysis buffer (to 3 mM final concentration) and dialysis continued for further 22 h. The tubing was then transferred into 17 volumes of 20 mM TrisHCl, 3 mM cysteine and 0.015 % Triton X-100 (no GdmHCl). After 4 h dialysis at 4 °C, the dialysis buffer was replaced by 20 mM TrisHCl and dialysis continued for 24 h. A reversible protease inhibitor MMTS (methyl methane-thiosulphonate) was added after the refolding and the sample was centrifuged for 20 min at 4 °C and 6500 g to remove any precipitate.

Since proteolytic degradation in the course of renaturation was observed in the first two experiments (parallel 1 and 2 in Fig. 3), the procedure was later modified in that the last dialysis step (24 h against 20 mM Tris-HCl) was replaced by 3 × 2 h dialysis against each time 17 volumes of 20 mM TrisHCl pH=8.0 ('modified renaturation procedure' in Fig. 3). Triton X-100 was omitted from all buffers in this modified refolding protocol. Addition of MMTS and centrifugation were performed as above.

Refolded proteins were concentrated by ultrafiltration through YM-10 membranes in a 350 mL Amicon apparatus. Final concentrating was performed on Centriprep-10 cartridges.

*Characterisation*

The size and uniformity of the recombinant protein were assayed by SDS-PAGE followed by immunoblotting using two different monoclonal antibodies, 1D10 and 2E4 (*12*) which recognise specific epitopes in the N- and the C- terminal part of mature cathepsin H, respectively. Proteins in GdmHCl were precipitated on ice by addition of 10 % TCA and washed by 2 × 200 μL ethanol : ether (1:1) before they were dissolved in the reducing gel loading buffer. Bacterial lysates and crude protein samples were analysed on 1 mm thick vertical polyacrylamide gels while renatured protein was resolved on the Phast System apparatus (Pharmacia). For Western blotting, proteins from the Phast System gel were blotted onto a PVDF membrane. They were then incubated with monoclonal mouse anti-cathepsin H antibodies and detected by a secondary goat anti-mouse IgG, conjugated with horse-radish peroxidase after reaction with $H_2O_2$ and diaminobenzidine.

## Results and Discussion

*cDNA amplification and cloning*

After PCR amplification, a single product of the expected size (≈950 bp) was observed on 1 % agarose gel electrophoresis (data not shown). DNA manipulations were complicated by the presence of a unique *Bam*HI site (normally used for insertion into the expression vector pET3) in the 3'-region of the cathepsin H coding DNA. The 3'-terminus was correctly processed since both original PCR primers successfully reamplified the coding region from the expression vector pBMF2.

*DNA sequence*

We found the sequence of the procathepsin H coding region identical to the published sequence (*13,14*) with the exception of 3 nucleotides. The nucleotide 493 (A) in the GCA (Ile) triplet was replaced by a G (silent mutation), nucleotide 526 (G) in the TTG (Leu) codon was replaced by an A (silent mutation) and the T at position 569 of the TAC triplet (Tyr) was replaced by a C, resulting in a Tyr to His mutation (nucleotide numbering according to the Genbank entry HSCATHH, accession code X16832). A His was found at this position also by Ritonja *et al.* (*15*) who determined the complete amino acid sequence of natural human cathepsin H. We thus believe that the CAC (His) triplet at position 569 is naturally occurring and is not an artefact of PCR.

*Expression and purification*

Three hours after addition of IPTG recombinant procathepsin H represented the most abundant protein species as judged from SDS-PAGE. The expression level was estimated to be 25 % of total bacterial proteins (Fig. 1).

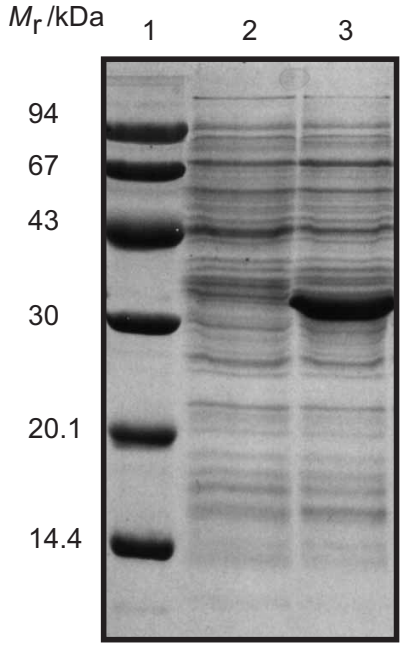


Fig. 1. 15 % PAGE in the presence of SDS and β-mercaptoethanol; samples: 1, LMW markers; 2, total bacterial proteins prior to addition of inducer (IPTG); 3, total bacterial proteins 3 h after addition of inducer

Despite the presence of thioredoxin, expressed from pT-Trx, a great majority of the recombinant procathepsin H was present in the insoluble fraction. Proteins from the solubilized material were separated by gel filtration and the fractions of the major peak were pooled (data not shown). After gel filtration (run in 3 parallels) we obtained 162 mg procathepsin H from 4 L of bacterial culture with a purity > 90 % as estimated from PAGE (Fig. 2).

*Refolding and trimming*

After refolding, conformational properties of the recombinant protein were followed by Trp fluorescence and by circular dichroism (data not shown). Refolded protein could be unfolded by addition of 2 M guanidinium hydrochloride. Secondary structural elements were observed but spectroscopic results could not be quantified because natural non-glycosylated procathepsin H is practically impossible to obtain as control.

Renaturation was first attempted by dialysis and with delayed addition of cysteine (standard procedure in Materials and Methods). In the modified procedure, dialysis steps were shorter. We observed that longer di-

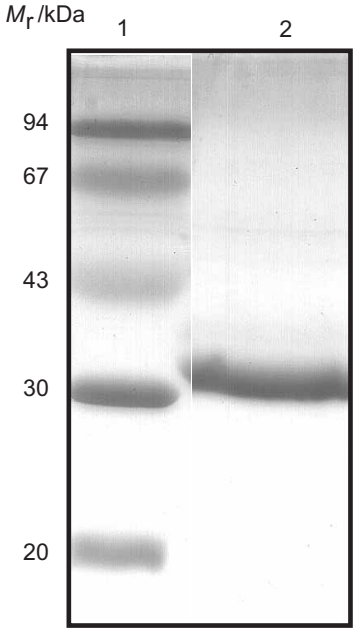


Fig. 2. 15 % PAGE in the presence of SDS and β-mercaptoethano; samples: 1, LMW markers; 2, pooled denatured procathepsin H after gel filtration (approx. 10 μg)

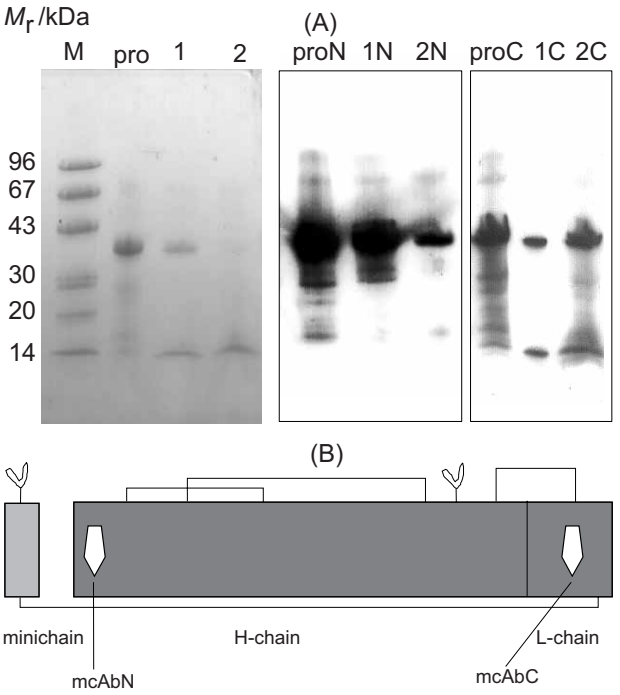


Fig. 3. (A) Coomassie brilliant blue stained SDS-PAGE and corresponding immunoblots obtained by two different monoclonal antibodies; samples: M, molecular size markers; pro, refolded procathepsin H – modified renaturation procedure (2 μg); 1, refolded procathepsin H – standard renaturation procedure, first parallel (2 μg); 2, refolded procathepsin H – standard renaturation procedure, second parallel (2 μg) (see Materials and Methods for details). ProN, 1N, 2N; same samples as above, detected by mcAb, specific for the N-terminal epitope (mcAbN on panel B). ProC, 1C, 2C; same samples as above, detected by mcAb, specific for the C-terminal epitope (mcAbC on panel B). Protein amounts were estimated from $A_{280}$ values

B) Schematic representation of mature cathepsin H molecule with the positions of mini-chain, H- and L- chains, glycosylation sites, disulphide bridges and locations of specific epitopes recognised by each of the antibodies

alysis times (standard procedure) resulted in increased degradation of the 35 kDa protein – the calculated relative molecular mass of procathepsin H is 35.179 Da (Fig. 3).

In refolding, pH was maintained alkaline (pH=8.0). Since mature cathepsin H is believed to be unstable in alkaline conditions (*1*) this should prevent any processing or digestion of procathepsin H molecules in course of renaturation by refolded and eventually processed (pro)cathepsin H.

From the crystal structure of porcine cathepsin H, Gunčar *et al.* (*2*) concluded that the carbohydrate residues attached to the body of the mature enzyme are involved in the positioning of the mini-chain into the active-site cleft. In our experiment, we prepared a non-glycosylated proenzyme form since we believed that glycosylation might not be essential for correct folding of the bulk of the proenzyme. Two other related human glycoproteins, procathepsins B and L, were previously expressed in bacteria, purified and refolded under similar conditions in our laboratory (*16*, *17*). If glycosylation indeed is important for obtaining an active exopeptidase, it obviously might not be essential for correct folding.

Mature cathepsin H is mainly known as an aminopeptidase (*18*), although it can also act as an endopeptidase on short synthetic substrates (*3*). Our observation of endopeptidase activity on procathepsin H molecules thus seems to be in contrast to the above conclusions. However, the endopeptidase activity we observed could only be attributed to proenzyme molecules and not to mature enzyme.

Gunčar *et al.* (*2*) suggested that in the process of procathepsin H activation, there would be an intermediate with an endopeptidase activity. We have observed that in the process of renaturation of reduced and unfolded recombinant procathepsin H, this molecule indeed undergoes proteolysis by an endopeptidase mechanism. We explain this behaviour by inaccurate positioning of the propeptide (because of improper folding or lack of carbohydrate environment, which would direct its positioning).

It could be speculated that procathepsin H was processed by a yet unidentified bacterial endopeptidase. We do not believe this was the case since procathepsin H was recovered from insoluble inclusion bodies, which normally do not contain proteases. We used the same expression system and a very similar purification scheme for related proenzymes, including procathepsins B (*16*), L (*17*) and S (*19*); none of these was proteolytically digested like procathepsin H.

## Conclusion

Our attempt to produce milligram amounts of human procathepsin H in a bacterial system failed because in the course of renaturation, the proenzyme was autolytically processed in an endopeptidolytic manner. This

suggests that either the carbohydrate moiety is essential for the correct positioning of the mini-chain in mature enzyme or that during renaturation an intermediate with endopeptidase activity is formed *in vitro*. We would thus in our future experiments have to switch the expression system to yeast which enables glycosylation, or seek a bacterial system which would lead to a soluble and thus folded proenzyme. The study has nevertheless shown an interesting, yet unobserved feature of cathepsin H proenzyme, namely that it can efficiently act as an endopeptidase while after maturation it is predominately an exopeptidase.

*Acknowledgement*

Presented research was supported by the Slovenian Ministry of science and technology.

## References

1. H. Kirschke: *Handbook of Proteolytic Enzymes*, A. J. Barrett, N. D. Rawlings, J. F. Woessner (Eds.), Academic Press, London (1998) pp. 628–631.
2. G. Gunčar, M. Podobnik, J. Pungerčar, B. Štrukelj, V. Turk, D. Turk, *Structure, 6* (1998) 51–61.
3. H. Koga, N. Mori, H. Yamada, Y. Nishimura, K. Tokuda, K. Kato, T. Imoto, *Chem. Pharm. Bull. 40* (1992) 965–970.
4. Y. Nishimura, K. Kato, *Biochem. Biophys. Res. Comm. 146* (1987) 159–164.
5. Y. Nishimura, K. Kato, *Biochem. Biophys. Res. Comm. 148* (1987) 329–334.
6. T. Popović, J. Brzin, A. Ritonja, B. Svetic, V. Turk, *J. Chromatogr. 615* (1993) 243–249.
7. A. Barrett, H. Kirschke, *Methods Enzymol. 80* (1981) 535–561.
8. V. Claus, A. Jahraus, T. Tjelle, T. Berg, H. Kirschke, H. Faulstich, G. Griffiths, *J. Biol. Chem. 273* (1998) 9842–9851.
9. J. Kos, B. Štabuc, A. Schweiger, M. Krašovec, N. Cimerman, N. Kopitar-Jerala, I. Vrhovec, *Clin. Cancer Res. 3* (1997) 1815–1822.
10. H. Inoue, H. Nojima, H. Okayama, *Gene, 96* (1990) 23–28.
11. T. Yasukawa, C. Kanei-Ishii, T. Maekawa, J. Fujimoto, T. Yamamoto, S. Ishii, *J. Biol. Chem. 270* (1995) 25328–25331.
12. A. Schweiger, B. Štabuc, T. Popović, V. Turk, J. Kos, *J. Immunol. Methods 201* (1997) 165–172.
13. R. Fuchs, W. Machleidt, H.G. Gassen, *Biol. Chem. Hoppe Seyler, 369* (1988) 469–475.
14. R. Fuchs, H. G. Gassen, *Nucl. Acids Res. 17* (1989) 9471.
15. A. Ritonja, T. Popović, M. Kotinik, W. Machleidt, V. Turk, *FEBS Lett. 228* (1988) 341–345.
16. R. Kuhelj, M. Dolinar, J. Pungerčar, V. Turk, *Eur. J. Biochem. 229* (1995) 533–539.
17. M. Dolinar, D. Barlič Maganja, V. Turk, *Biol. Chem. Hoppe-Seyler, 376* (1995) 385–388.
18. T. Takahashi, A. H. Dehdarani, J. Tang, *J. Biol. Chem. 263* (1988) 10952–10957.
19. G. Kopitar, M. Dolinar, B. Štrukelj, J. Pungerčar, V. Turk, *Eur. J. Biochem. 236* (1996) 558–562.

# Endoproteolitički način razgradnje opažen tijekom renaturacije proenzima ljudske egzopeptidaze, prokatepsina H, proizvedenog u *Escherichia coli*

**Sažetak**

Katepsin H je lizozomska cisteinska proteinaza s egzopeptidaznom aktivnosti. Iz ljudskih stanica endometrija pročišćena je ukupna mRNA i prevedena u cDNA. Kodirajuća regija prokatepsina H bila je proširena specifičnim klicama, klonirana i eksprimirana pomoću vektora ekspresije pET3, a kontrolirana s T7 polimerazom i bakterijskim sojem *Escherichia coli* BL21[DE3]pT-Trx. Glavnina rekombiniranog prokatepsina H bila je prisutna u netopljivom obliku a činila je 25 % od ukupnih proteina stanice. Nakon otapanja inkluzijskih tijela prokatepsin H dijalizom je promijenio prostornu strukturu. Renaturacijom se rekombinantni prokatepsin H proteolitički razgradio u nekoliko definiranih fragmenata identificiranih monoklonskim antitijelima usmjerenim prema N- ili C-terminalnom kraju stvorenog enzima. Način razgradnje bio je tipičan za neku endopeptidazu, a stabilni fragment LMW od 14 kDa mogao se pripisati C-terminalnom dijelu stvorenog enzima. Tijekom prostorne strukturne promjene ljudski se prokatepsin H autolizira endopeptidaznim procesom koji je prije bio sporan za taj enzim.