

NEW EMBO MEMBER'S REVIEW

Lysosomal cysteine proteases: facts and opportunities

Vito Turk¹, Boris Turk and Dušan Turk

Department of Biochemistry and Molecular Biology, J.Stefan Institute, Ljubljana, Slovenia

¹Corresponding author
e-mail: vito.turk@ijs.si

From their discovery in the first half of the 20th century, lysosomal cysteine proteases have come a long way: from being the enzymes non-selectively degrading proteins in lysosomes to being those responsible for a number of important cellular processes. Some of the features and roles of their structures, specificity, regulation and physiology are discussed.

Keywords: cathepsins/cystatins/lysosomal cysteine proteases/regulation/structure

Heritage of the 20th century

Lysosomal cysteine proteases, generally known as the cathepsins, were discovered in the first half of the 20th century. Cathepsin C (also known as dipeptidyl peptidase I or DPP1), as the first pure enzyme, was found in the 1940s (Gutman and Fruton, 1948). We had to wait until the early 1970s for more enzymes, when cathepsins B and H were identified, followed soon after by cathepsin L (reviewed in Barrett *et al.*, 1998). However, the first amino acid sequences of mammalian cathepsins did not appear until the early 1980s, when the sequences of rat cathepsins B and H were published (Takio *et al.*, 1983). In 1990, the first crystal structure of a lysosomal cysteine protease, human cathepsin B, was determined (Musil *et al.*, 1991), indicating rapid progress in the field. Indeed, the 1990s were the golden era of lysosomal cysteine protease research, with six out of 11 known human enzymes identified. With the completion of human genome sequencing, this number will probably increase, especially since several new mouse cathepsins without apparent human counterparts were discovered recently (Sol-Church *et al.*, 2000). The 1990s also provided more clues on the physiological role of cysteine proteases: several cathepsin knockouts demonstrated that the role of cathepsins is not simply that of scavengers, which was long believed to be their major function (Chapman *et al.*, 1997; Turk *et al.*, 2000). In addition, the first genetic disorder, pycnodysostosis, was found to be linked with a lysosomal cysteine protease, cathepsin K (Gelb *et al.*, 1996).

Structure and specificity

Lysosomal cysteine proteases are optimally active in the slightly acidic, reducing milieu found in lysosomes. They comprise a group of papain-related enzymes, sharing

similar amino acid sequences and folds (Figure 1). Purified cathepsins are generally composed of disulfide-connected heavy and light chains. The enzymes are monomers with an M_r of ~30 kDa, with the exception of the tetrameric cathepsin C. Papain-like enzymes have a two-domain structure with the V-shape active site cleft extending along the two-domain interface. The left (L-)domain is dominated by three α -helices and the right (R-)domain is based on a β -barrel motif (McGrath, 1999). The catalytic Cys25 (papain numbering) is located at the N-terminus of the characteristic α -helix. It has an unusually low pK_a value and forms an ion pair with His159 positioned in the β -barrel domain on the opposite site of the active site cleft (Brocklehurst, 1994). Substrate binds in an extended conformation along the active site cleft. The only well defined substrate binding sites are S2, S1 and S1', with S2 and S1' being the major specificity determinants (Figure 2; Turk *et al.*, 1998).

Most of the enzymes are endopeptidases (Table I). Cathepsin C is an aminodipeptidase (Barrett *et al.*, 1998), cathepsin X a carboxy-mono or -dipeptidase (Klemenčič *et al.*, 2000), cathepsin B is a carboxydipeptidase and cathepsin H is an aminopeptidase. Cathepsins B and H also exhibit endopeptidase activity (Barrett *et al.*, 1998). In exopeptidases, access to the substrate binding sites is restricted by additional structural features: loops in cathepsins B (Musil *et al.*, 1991) and X (Klemenčič *et al.*, 2000), or propeptide parts in the aminopeptidases cathepsin H (Gunčar *et al.*, 1998) and C (D.Turk, unpublished data).

Regulation of activity

The activity of lysosomal cysteine proteases can be regulated in a number of ways, the most important being zymogen activation and inhibition by endogenous protein inhibitors.

Zymogen activation: crucial step in the regulation of cathepsin activity

Similarly to most other proteases, the cathepsins are synthesized as inactive precursors, and are activated by proteolytic removal of the N-terminal propeptide. *In vitro*, removal of the propeptide can be facilitated either by activation by other proteases such as pepsin or cathepsin D, or by autocatalytic activation at acidic pH. The latter is a bimolecular process in which one of the cathepsin molecules activates the other in a chain reaction manner (Turk *et al.*, 2000). Crystal structures of procathepsins B revealed that the propeptide runs through the active site cleft in the orientation opposite to that of a substrate, blocking access to the already structured active site (Cygler *et al.*, 1996; Turk *et al.*, 1996). The propeptide, which *in vitro* is an inhibitor of the enzyme, can be at least

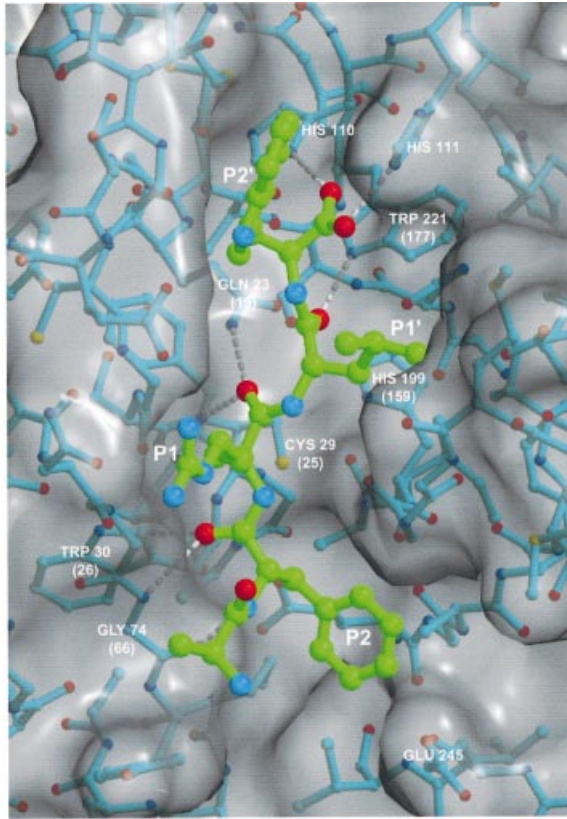


Fig. 2. A substrate model APRLW bound along the active site of cathepsin B (Turk *et al.*, 1995; 1CSB). Bonds of substrate and cathepsin B structure are shown in cyan and green, respectively. Non-hydrogen atoms are shown as small colored spheres: oxygens are red, sulfurs yellow and nitrogens blue, whereas carbons are shown in cyan and green, corresponding to the coloring code of covalent bonds of cathepsin B and substrate models. The GRASP generated surface (Nicholls *et al.*, 1991) of cathepsin B is gray. Crucial residues involved in the binding of substrate main chain atoms are labeled with their residue name and sequence ID. Pappain numbering is shown in parentheses for the contacts between essential enzyme residues and substrate. Hydrogen bonds along the substrate main chain are shown as white broken lines. The figure was prepared with MAIN (Turk, 1992) and rendered with RENDER (Merritt and Bacon, 1997).

partially removed from the active site, explaining the catalytic activity of the precursor. Activation can be facilitated by a drop in pH and/or by glycosaminoglycans (Turk *et al.*, 2000). However, only the endopeptidases can be autoactivated, whereas the true exopeptidases, cathepsins X and C, require endopeptidases, such as cathepsins L and S, for their activation (Nägler *et al.*, 1999; Dahl *et al.*, 2001).

Cystatins, non-selective baits for escaped cathepsins

Once activated, cathepsins have enormous disruptive potential, since their total concentration inside lysosomes can well exceed 1 mM. Their inappropriate action is controlled by cystatins, the endogenous protein inhibitors of lysosomal cysteine proteases, and not by pH (Turk *et al.*, 1993). It seems that they are there to trap proteases that have accidentally escaped from lysosomes, and to help in the defense of the organism against intruders. On the basis of sequence homology, cystatins are divided into stefins,

cystatins and kininogens. Stefins are intracellular inhibitors, whereas cystatins and kininogens are extracellular (Turk and Bode, 1991). A cystatin N-terminal trunk and two conserved hairpin loops are involved in interaction with the conserved features of the active site of the target enzymes. The catalytic cysteine residue is surrounded by the residues of the trunk and the first hairpin loop (Stubbs *et al.*, 1990). The cystatins are not very selective. They inhibit endopeptidases in the picomolar range and exopeptidases in the nanomolar range (Turk *et al.*, 1997).

In addition to cystatins, several other inhibitors of lysosomal cysteine proteases have been found. The most important are the thyropins, inhibitors homologous to the thyroglobulin type I domains (Lenarčič and Bevec, 1998). Among these, the only known mammalian representative is the major histocompatibility (MHC) class II-associated p41 invariant chain fragment, which is a selective cathepsin L inhibitor (Bevec *et al.*, 1996; Gunčar *et al.*, 1999).

Physiological roles of lysosomal cysteine proteases

Are lysosomal cysteine proteases redundant?

Based on the results from the cathepsin gene knockouts, one might imagine that the anticipated function of intralysosomal protein degradation is not exclusively dependent on any cathepsin, since in all the cathepsin-deficient mice there have been no reports of defects in protein degradation (Deussing *et al.*, 1998; Nakagawa *et al.*, 1998, 1999; Saftig *et al.*, 1998; Pham and Ley, 1999; Shi *et al.*, 1999; Roth *et al.*, 2000). However, gene knockouts have revealed that lysosomal cathepsins have specific and individual functions, which are very important for the normal functioning of an organism. These specific functions are often associated with the restricted tissue localization of the cathepsins, as demonstrated for cathepsins S, V and K. Although cathepsins B, H, L, F, C, X and O are ubiquitous, this does not preclude them from being involved in some more specialized processes.

Cathepsin K was shown to be crucial for normal bone remodeling (reviewed in Chapman *et al.*, 1997) and cathepsin K-deficient mice developed similar symptoms (Saftig *et al.*, 1998) to those of the patients with pycnodysostosis (Gelb *et al.*, 1996).

The major role of lysosomal cathepsin S, and the lesser role of cathepsin L, in the processing of the MHC class II-associated invariant chain (reviewed in Chapman *et al.*, 1997), which is essential for the normal functioning of the immune system, has also been confirmed (Nakagawa *et al.*, 1998, 1999; Shi *et al.*, 1999). However, it seems that the role played by cathepsin L in the mouse (Nakagawa *et al.*, 1998) has been taken by its close homologue, cathepsin V, in humans (Brömme *et al.*, 1999). The list of cathepsins involved in invariant chain processing and MHC class II antigen presentation does not stop here, and cathepsin F has recently been suggested to be an important player in this process in macrophages (Shi *et al.*, 2000). It is also quite likely that there could be more cathepsins involved in this process, including the very abundant cathepsins B and D (Watts, 2001).

Yet another function of cathepsin L has come to light through gene knockout analysis: mice deficient in

Table I. Human lysosomal cysteine proteases: nomenclature and properties

Name	Synonyms	Endopeptidase	Exopeptidase		Chromosome localization
			Carboxypeptidase	Aminopeptidase	
Cathepsin L	–	+	–	–	9q21
Cathepsin V	cathepsin L2 cathepsin U	+	–	–	9q21
Cathepsin S	–	+	–	–	1q21
Cathepsin K	cathepsin O cathepsin O2 cathepsin X	+	–	–	1q21
Cathepsin W	lymphopain	n.d.	n.d.	n.d.	11q13
Cathepsin F	–	+	–	–	11q13
Cathepsin O	–	n.d.	n.d.	n.d.	4q31-32
Cathepsin B	cathepsin B1	+	+	–	8p22-23
Cathepsin X	cathepsin Z cathepsin P cathepsin Y	–	+	–	20q13
Cathepsin H	cathepsin I	+	–	+	15q24-25
Cathepsin C	dipeptidyl peptidase I cathepsin J	–	–	+	11q14
				monopeptidase dipeptidase	

Data are taken from Barrett *et al.* (1998), Kirschke *et al.* (1995), Chapman *et al.* (1997) and Turk *et al.* (2000). n.d., not determined.

cathepsin L developed periodic hair loss and epidermal hyperplasia, indicating that cathepsin L is essential for epidermal homeostasis and regular hair follicle morphogenesis and cycling (Roth *et al.*, 2000).

Lysosomal cathepsins are also extremely important processing enzymes. Cathepsin C was thus revealed to be one of the major processing machineries known so far. Activation of a number of serine granule proteases, granzymes A and B, cathepsin G, neutrophil elastase and chymase was impaired in cathepsin C-deficient mice (Pham and Ley, 1999). In addition, lysosomal cathepsins are involved in prohormone processing (Tepel *et al.*, 2000) and were suggested to be involved in endostatin generation during angiogenesis (Felbor *et al.*, 2000).

If the system fails...

Disturbance of the normal balance of enzymatic activity may lead to pathological conditions, and lysosomal cysteine proteases are no exception. They have thus been associated with a number of pathological events, such as rheumatoid arthritis and osteoarthritis, cancer, neurological disorders, osteoporosis and lysosomal storage diseases (reviewed in Kirschke *et al.*, 1995; Chapman *et al.*, 1997; Barrett *et al.*, 1998). Cathepsins also participate in apoptosis, although the mechanisms are not clear (Turk *et al.*, 2000; Leist and Jäättelä, 2001; Salvesen, 2001). There is also genetic evidence that abnormalities in cathepsin gene expression may lead to various pathologies: pycnodysostosis, characterized in humans by severe bone abnormalities, was found to be associated with the loss-of-function mutation of cathepsin K (Gelb *et al.*, 1996), while Papillon-Lefevre syndrom, characterized in patients by severe, early-onset periodontitis and palmo-plantar keratosis, is a consequence of loss-of-function mutation in the cathepsin C gene (Toomes *et al.*, 1999). Similarly, problems can be caused by down-regulation of the inhibitors, as demonstrated by the mutation in the gene for stefin B, which was found to be responsible for the

hereditary form of myoclonal epilepsy (Pennacchio *et al.*, 1996).

Acknowledgements

This paper is dedicated to Professor Dušan Hadži on the occasion of his 80th birthday. We are grateful to Gregor Gunčar for help in figure preparation, Miranda Thomas and David Pim for critical reading of the manuscript, and to the Ministry of Schools, Science and Sports of the Republic of Slovenia for financial support.

References

- Barrett,A.J., Rawlings,N.D. and Woessner,J.F.,Jr (eds) (1998) *Handbook of Proteolytic Enzymes*. Academic Press, London, UK.
- Bevec,T., Stoka,V., Pungerčič,G., Dolenc,I. and Turk,V. (1996) Major histocompatibility complex class II-associated p41 invariant chain fragment is a strong inhibitor of lysosomal cathepsin L. *J. Exp. Med.*, **183**, 1331–1338.
- Brocklehurst,K. (1994) A sound basis for pH-dependent kinetic studies on enzymes. *Protein Eng.*, **7**, 291–299.
- Brömme,D., Li,Z., Barnes,M. and Mehler,E. (1999) Human cathepsin V functional expression, tissue distribution, electrostatic surface potential, enzymatic characterization and chromosomal localization. *Biochemistry*, **38**, 2377–2385.
- Chapman,H.A., Riese,R.J. and Shi,G.-P. (1997) Emerging roles for cysteine proteases in human biology. *Annu. Rev. Physiol.*, **59**, 63–88.
- Cyglar,M., Sivaraman,J., Grochulski,P., Coulombe,R., Storer,A.C. and Mort,J.S. (1996) Structure of rat procathepsin B: model for inhibition of cysteine protease activity by proregion. *Structure*, **4**, 405–416.
- Dahl,S.W., Halkier,T., Lauritzen,C., Dolenc,I., Pedersen,J., Turk,V. and Turk,B. (2001) Human recombinant pro-dipeptidyl peptidase I (cathepsin C) can be activated by cathepsins L and S but not by autocatalytic processing. *Biochemistry*, **40**, 1671–1678.
- Deussing,J., Roth,W., Saftig,P., Peters,C., Ploegh,H.L. and Villadangos,J.A. (1998) Cathepsins B and D are dispensable for major histocompatibility complex class II-mediated antigen presentation. *Proc. Natl Acad. Sci. USA*, **95**, 4516–4521.
- Felbor,U., Dreier,L., Bryant,R.A., Ploegh,H.L., Olsen,B.R. and Mothes,W. (2000) Secreted cathepsin L generates endostatin from collagen XVIII. *EMBO J.*, **19**, 1187–1194.
- Gelb,B.D., Shi,G.-P., Chapman,H.A. and Desnick,R.J. (1996) Pycnodysostosis is caused by a deficiency of cathepsin K. *Science*, **273**, 1236–1238.
- Gunčar,G., Podobnik,M., Pungerčar,J., Štrukelj,B., Turk,V. and Turk,D.

- (1998) Crystal structure of porcine cathepsin H determined at 2.1 Å resolution: location of the mini-chain C-terminal carboxyl group defines cathepsin H aminopeptidase function. *Structure*, **6**, 51–61.
- Gunčar,G., Pungerčič,G., Klemenčič,I., Turk,V. and Turk,D. (1999) Crystal structure of MHC class II-associated p41 Ii fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. *EMBO J.*, **18**, 793–803.
- Gutman,H.R. and Fruton,J.S. (1948) On the proteolytic enzymes of animal tissues VIII. An intracellular enzyme related to chymotrypsin. *J. Biol. Chem.*, **174**, 851–858.
- Kirschke,H., Barrett,A.J. and Rawlings,N.D. (1995) Lysosomal cysteine proteases. In Shterline,P. (ed.), *Protein Profile*. Vol. 2. Academic Press, London, UK, pp. 1587–1643.
- Klemenčič,I., Karaoglanovic-Carmona,A., Juliano,M.A., Juliano,L., Gunčar,G., Turk,D., Križaj,I., Turk,V. and Turk,B. (2000) Biochemical characterization of human cathepsin X revealed that the enzyme is an exopeptidase, acting as carboxy-monopeptidase or carboxy-dipeptidase. *Eur. J. Biochem.*, **267**, 5404–5412.
- Leist,M. and Jäättelä,M. (2001) Triggering of apoptosis by cathepsins. *Cell Death Diff.*, **8**, 324–326.
- Lenarčič,B. and Bevec,T. (1998) Thyropins—new structurally related proteinase inhibitors. *Biol. Chem.*, **379**, 105–111.
- McGrath,M. (1999) The lysosomal cysteine proteases. *Annu. Rev. Biophys. Biomol. Struct.*, **28**, 181–204.
- Merritt,E.A. and Bacon,D.J. (1997) Raster3D: photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.
- Musil,D. et al. (1991) The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *EMBO J.*, **10**, 2321–2330.
- Nägler,D., Tam,W., Storer,A.C., Krupa,J.C., Mort,J.S. and Menard,R. (1999) Human cathepsin X: a cysteine protease with unique carboxypeptidase activity. *Biochemistry*, **38**, 4868–4874.
- Nakagawa,T. et al. (1998) Cathepsin L: critical role in Ii degradation and CD4 T cell selection in the thymus. *Science*, **280**, 450–453.
- Nakagawa,T.Y. et al. (1999) Impaired invariant chain degradation and antigen presentation and diminished collagen-induced arthritis in cathepsin S null mice. *Immunity*, **10**, 207–217.
- Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–376.
- Pennacchio,L.A. et al. (1996) Mutations in the gene encoding cystatin B in progressive myoclonus epilepsy (EPM1). *Science*, **271**, 1731–1734.
- Pham,C.T. and Ley,T.J. (1999) Dipeptidyl peptidase I is required for the processing and activation of granzymes A and B *in vivo*. *Proc. Natl Acad. Sci. USA*, **96**, 8627–8632.
- Roth,V. et al. (2000) Cathepsin L deficiency as molecular defect of furless: hyperproliferation of keratinocytes and perturbation of hair follicle cycling. *FASEB J.*, **14**, 2075–2086.
- Saftig,P., Hunzicker,E., Wehmeyer,O., Jones,S., Boyde,A., Rommerskirch,W., Moritz,J.D., Schu,P. and von Figura,K. (1998) Impaired osteoclastic bone resorption leads to osteopetrosis in cathepsin K-deficient mice. *Proc. Natl Acad. Sci. USA*, **95**, 13453–13458.
- Salvesen,G.S. (2001) A lysosomal protease enters the death scene. *J. Clin. Invest.*, **107**, 21–22.
- Shi,G.-P., Villadangos,J.A., Dranoff,G., Small,C., Gu,L., Haley,K.J., Riese,R., Ploegh,H.L. and Chapman,H.A. (1999) Cathepsin S is required for normal MHC class II peptide loading and germinal center development. *Immunity*, **10**, 197–206.
- Shi,G.-P., Bryant,R.A., Riese,R., Verhelst,S., Driessen,C., Li,Z., Brömme,D., Ploegh,H.L. and Chapman,H.A. (2000) Role for cathepsin F in invariant chain processing and major histocompatibility complex class II peptide loading by macrophages. *J. Exp. Med.*, **191**, 1177–1186.
- Sol-Church,K., Frenck,J., Bertenshaw,G. and Mason,R.W. (2000) Characterization of mouse cathepsin R, a new member of a family of placentally expressed cysteine proteases. *Biochim. Biophys. Acta*, **1492**, 488–492.
- Stubbs,M.T., Laber,B., Bode,W., Huber,R., Jerala,R., Lenarčič,B. and Turk,V. (1990) The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *EMBO J.*, **9**, 1939–1947.
- Takio,K., Towatari,T., Katunuma,N., Teller,D.C. and Titani,K. (1983) Homology of amino acid sequences of liver cathepsins B and H with that of papain. *Proc. Natl Acad. Sci. USA*, **80**, 3660–3670.
- Tepel,C., Brömme,D., Herzog,V. and Brix,C. (2000) Cathepsin K in thyroid epithelial cells: sequence, localization and possible function in extracellular proteolysis of thyroglobulin. *J. Cell Sci.*, **113**, 4487–4498.
- Toomes,C. et al. (1999) Loss-of-function mutations in the cathepsin C gene result in periodontal disease and palmoplantar keratosis. *Nature Genet.*, **23**, 421–424.
- Turk,B., Dolenc,I., Turk,V. and Bieth,J.G. (1993) Kinetics of the pH-induced inactivation of human cathepsin L. *Biochemistry*, **32**, 375–380.
- Turk,B., Turk,V. and Turk,D. (1997) Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. *Biol. Chem.*, **378**, 141–150.
- Turk,B., Turk,D. and Turk,V. (2000) Lysosomal cysteine proteases: more than scavengers. *Biochim. Biophys. Acta*, **1477**, 98–111.
- Turk,D. (1992) Weiterentwicklung eines Programms für Molekülgraphik und Elektronendichte-Manipulation und seine Anwendung auf verschiedene Protein-Strukturaufklärungen. PhD Thesis, Technische Universität, München, Germany.
- Turk,D., Podobnik,M., Popovic,T., Katunuma,N., Bode,W., Huber,R. and Turk,V. (1995) Crystal structure of cathepsin B inhibited with CA030 at 2 Å resolution: a basis for the design of specific epoxysuccinyl inhibitors. *Biochemistry*, **34**, 4791–4797.
- Turk,D., Podobnik,M., Kuhelj,R., Dolinar,M. and Turk,V. (1996) Crystal structures of human procathepsin B at 3.2 and 3.3 Å resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide. *FEBS Lett.*, **384**, 211–214.
- Turk,D., Gunčar,G., Podobnik,M. and Turk,B. (1998) Revised definition of substrate binding sites of papain-like cysteine proteases. *Biol. Chem.*, **379**, 137–147.
- Turk,V. and Bode,W. (1991) The cystatins: protein inhibitors of cysteine proteinases. *FEBS Lett.*, **285**, 213–219.
- Watts,C. (2001) Antigen processing in the endocytic compartment. *Curr. Opin. Immunol.*, **13**, 26–31.

Received April 23, 2001; revised July 2, 2001;
accepted July 4, 2001