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 DPPI), 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 Mₐ of ~30 kDa, with the exception of the tetrameric cathepsin C. Papain-like enzymes have a two-domain structure with the V-shaped 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 pKᵢ 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 aminopeptidase (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 (Guncar 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. 1. (A) Structure-based amino acid sequence alignment of mature parts of papain and related human lysosomal cysteine proteases was performed by the CLUSTAL_W program as described previously (Turk et al., 2000). The sequences were taken from the SWISS-PROT or GenBank databases. The active site residues Cys25 and His159 are marked with asterisks and numbered. (B) Fold of cathepsin L, a typical human lysosomal cysteine protease (Güncar et al., 1999).
cystatins and kininogens. Stefs 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; Günčar 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

---

**Fig. 2.** A substrate model APRIL bound along the active site of cathepsin B (Turk et al., 1995; ICSB). 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. Papain 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).

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 steins,
Table I. Human lysosomal cysteine proteases: nomenclature and properties

<table>
<thead>
<tr>
<th>Name</th>
<th>Synonyms</th>
<th>Endopeptidase</th>
<th>Exopeptidase</th>
<th>Chromosome localization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cathepsin L</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>9q21</td>
</tr>
<tr>
<td>Cathepsin V</td>
<td>cathepsin L2</td>
<td>+</td>
<td>–</td>
<td>9q21</td>
</tr>
<tr>
<td>Cathepsin S</td>
<td>cathepsin U</td>
<td>+</td>
<td>–</td>
<td>1q21</td>
</tr>
<tr>
<td>Cathepsin K</td>
<td>cathepsin O</td>
<td>+</td>
<td>–</td>
<td>1q21</td>
</tr>
<tr>
<td>Cathepsin W</td>
<td>lymphopain</td>
<td>n.d.</td>
<td>n.d.</td>
<td>11q13</td>
</tr>
<tr>
<td>Cathepsin F</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>11q13</td>
</tr>
<tr>
<td>Cathepsin O</td>
<td>–</td>
<td>n.d.</td>
<td>n.d.</td>
<td>4q31-32</td>
</tr>
<tr>
<td>Cathepsin B</td>
<td>cathepsin B1</td>
<td>+</td>
<td>dipeptidase</td>
<td>8p22-23</td>
</tr>
<tr>
<td>Cathepsin X</td>
<td>cathepsin Z</td>
<td>–</td>
<td>+ monopeptidase also dipeptidase</td>
<td>20q13</td>
</tr>
<tr>
<td>Cathepsin H</td>
<td>cathepsin I</td>
<td>+</td>
<td>–</td>
<td>15q24-25</td>
</tr>
<tr>
<td>Cathepsin C</td>
<td>dipeptidyl peptidase I</td>
<td>–</td>
<td>+ monopeptidase</td>
<td>11q14</td>
</tr>
<tr>
<td></td>
<td>cathepsin J</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


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 Levy, 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–Lefèvre syndrome, characterized in patients by severe, early-onset periodontitis and palmoplantar 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 Gunicar 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**


Gunicar, G., Podobnik, M., Pungerčič, J., Štrukelj, B., Turk, V. and Turk, D.


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