Structure determinants defining the specificity of papain-like cysteine proteases

Anastasiia I. Petushkova, Lyudmila V. Savvateeva, Andrey A. Zamyatnin Jr

Keywords: Papain-like cysteine protease, Cysteine cathepsin, Protease specificity, Binding site, Selective inhibitor, Protein engineering

Abstract

Papain-like cysteine proteases are widely expressed enzymes that mostly regulate protein turnover in the acidic conditions of lysosomes. However, in the last twenty years, these proteases have been evidenced to exert specific functions within different organelles as well as outside the cell. The most studied proteases of this family are human cysteine cathepsins involved both in physiological and pathological processes. The specificity of each protease to its substrates is mostly defined by the structure of the binding cleft. Different patterns of amino acid motif in this area determine the interaction between the protease and the ligands. Moreover, this specificity can be altered under the specific media conditions and in case other proteins are present. Understanding how this network works would allow researchers to design the diagnostic selective probes and therapeutic inhibitors. Moreover, this knowledge might serve as a key for redesigning and de novo engineering of the proteases for a wide range of applications.

Contents

1. Introduction
2. Proteolytic activity of PLCPs
  2.1. Substrates and endogenic inhibitors of PLCPs
  2.2. Role of pH in PLCP activity and specificity
  2.3. Impact of ions on PLCP activity and specificity
  2.4. Macromolecules in regulation of PLCP activity and specificity
3. Structure of cysteine cathepsins
  3.1. The general structure of cysteine cathepsins
  3.2. Structure and specificity of S1 binding site
  3.3. Structure and specificity of S2 binding site
  3.4. Structure and specificity of S3 binding site
  3.5. Structure and specificity of the primed side
4. Examples of rationally designed PLCP inhibitors
5. Plcps engineering for industrial applications
6. Summary and outlook

© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
1. Introduction

Proteases are proteins that cleave peptide bonds thus maintaining the protein catabolism and performing cell signaling. Based on the structure of their catalytic site, proteases are subdivided into eight classes: aspartic, cysteine, glutamic, serine, threonine, metalloproteases, asparagine peptide lysases, and peptidases of mixed catalytic type [1]. Papain-like cysteine proteases (PLCPs) are one of the most thoroughly studied families (family C1, clan CA) among the cysteine peptidases. It comprises endo- and exopeptidases homologous to papain. PLCPs are expressed as preproenzymes and the expressed polypeptide chain consists of a signal peptide, a propeptide, and a catalytic domain. The signal peptide is necessary for transporting a protein into the endomembrane system, where it is cleaved by signal peptidases [2,3]. The propeptide catalyzes folding of the catalytic domain and inhibits the protease catalytic activity [4,5]. The protease is activated in the acidic environment of endosomes and lysosomes via eliminating the propeptide in an autocatalytic manner or by other proteases [3]. The activity of mature PLCPs is tightly regulated by the post-translational modifications and inhibitors such as cystatins and serpins [6,7]. The genes encoding PLCPs have been identified in bacteria, archaea, protozoa, plants, animals, and viruses [1]. Although these proteases are mainly found in acidic conditions of the apoplasm, they can also perform essential functions in the nucleus, cytosol, cell surface, and extracellular space [2,8]. Despite mounting evidence gathered in the last decade the complexity and intricacies of this network are not fully elucidated.

In different organisms, PLCPs play specific roles contributing to processes ranging from cell differentiation to cellular senescence. In plants, PLCPs are involved in stress response, mobilization of storage proteins during seed germination, induction of defense reactions, senescence, and regulated cell death [6,9–11]. Some aspects of PLCP regulation and application are better investigated on plant PLCPs, so we will discuss them below. In parasites, PLCPs contribute to the regulated cell death, host immune response modulation, hydrolisis of ingested host macromolecules, and progression disease [12–16]. The most studied animal PLCPs are human cysteine cathepsins which are ubiquitously expressed being essential for antigen processing, cellular stress signaling, autophagy, senescence, cells differentiation, and prohormones activation [2,17–19]. Apart from their usual roles under normal physiological conditions, human cathepsins are also involved in tumor growth, invasion, angiogenesis, and metastasis [20,21]. In the present review we will focus our attention on structural features of mammalian cysteine cathepsins.

Cysteine cathepsins were reported to play a certain role in muscular dystrophy, rheumatoid arthritis, bone disorders, and Alzheimer's disease [22]. In particular, cathepsin K (CtsK) is excessively expressed in osteoclasts and is responsible for bone turnover [23]. Its elevated activity leads to osteoporosis and arthritis, while its deficiency is associated with learning and memory deficits [24,25]. The excess activity of CtsC promotes various inflammatory diseases, while its down-regulation results in autoimmunity disorders [26,27]. CtsL is involved in cell differentiation and proliferation playing an important role in skin homeostasis and antigen presentation [28–30]. CtsL and K contribute to adipogenesis and liberation of thyroid hormones [31,32]. Their activity in the extracellular space facilitates cardiovascular and pulmonary diseases and cancer [33]. Intracellular CtsB has been shown to inhibit cancer via promoting apoptosis, while along with CtsL, it can protect the cells against anticancer chemotherapies [34,35]. CtsL, S, F, and V degrade the invariant chain (Ii) during the major histocompatability complex (MHC) class II antigen presentation [36]. Since the excessive activity of human cysteine cathepsins is tightly linked to multiple disorders and tumor progression in particular, they represent promising molecular targets for developing novel therapeutic avenues [37].

Presently, drug design is determined by the structure of the ultimate target. All PLCPs belong to one peptidase family sharing a similar structure [1]. However, their remarkable functional diversity may require a set of rather selective inhibitors to be harnessed. In this review, we delve into the structural peculiarities determining the specific roles of PLCPs or allowing for their interchange ability. We also overview how these specificities can be employed for designing selective inhibitors and engineering new proteases.

2. Proteolytic activity of PLCPs

2.1. Substrates and endogenic inhibitors of PLCPs

PLCPs perform multiple functions since they can digest different substrates and often substitute each other. For instance, human cysteine cathepsins digest the extracellular matrix made up of elastins, collagens, and proteoglycans [38]. Some of them are listed in Table 1, but the complete information on discovered PLCP substrates can be received in MEROPS Database [1]. Elastin is mainly cleaved by CtsV, K, S, F, L, and B [39,40]. CtsB degrades type IV collagen, laminin, and fibronectin [41]. CtsL and CtsS can release glycosaminoglycans (GAGs) such as aggrecan from cartilage proteins [42], while CtsK and CtsL participate in bone resorption [43]. Even CtsC lacking the exclusion domain (endopeptidase CtsCΔEx) can cleave azocasein, elastin, and gelatin [44]. The other substrates degraded by cysteine cathepsins are thyroglobulins, and each cathepsin generates distinct fragments [45]. The truncated active form of pro-apoptotic Bid and the anti-apoptotic Bcl-2 family proteins (Bcl-2, Bcl-XL, Mcl-1, and XIAP) can be cleaved by CtsB, K, L, and H, whereas CtsS is involved in degrading pro-apoptotic Bax [46,47]. Cathepsins are also involved in generating peptides which are presented to MHC class-II molecules. Experiments in mice identified CtsL and S as the key enzymes in the final steps of β degradation. However, CtsS has been reported to compensate their loss in macrophages [48–50]. Still, some substrates are sensitive only to one protease; for instance, native type I collagen can be digested only by CtsK [51,52]. CtsL was shown to cleave histone H3 in the nucleus [28,53] and to take part in the processing of the coronavirus [54,55]. CtsX was found to cleave β2 integrin, thus regulating T-cell migration [2,56].

To reveal the distinct specificities of cysteine cathepsins various synthetic peptides are used. Applying the tetra-peptide substrate library, Choe et al. compared CtsL, V, K, S, F, and B, papain, and bromelain. Most of these proteases prefer hydrophobic amino acids at the P2 position, in contrast to bromelain which favors basic amino acids [66]. All cathepsins cleave substrates with Arg and Lys at the P1 position, while at the P3 and P4 positions, substrates can contain various amino acids [66]. CtsB, L, and S specificity was compared using Z-RR-AMC (Z – benzyl chloroformate, AMC – 7-amino-4-methylcoumarin), Z-RR-AMC, Z-FK-AMC, and Z-KK-AMC. While CtsB efficiently cleaved all substrates, CtsL and S had little activity towards Z-RR-AMC and Z-KK-AMC [67]. There are several selective substrates for cysteine cathepsins. For example, the selective CtsB substrate (Ac-Cha-Leu-Glu(Bzl)-Arg-ACC (Ac – acetyl, Cha – cyclohexylalanin, Bzl – benzyl, ACC – 7-amino-4-carbamoylmethylcoumarin)) was digested by CtsB 3500-fold more efficiently than by CtsL [68]. Abz-HPC/GpQ-EDDnp (Abz – 2-amino benzoic acid, EDDnp – N-(2,4-dinitrophenyl)-ethylenediamine) was found to be a highly selective substrate for CtsK, which mainly depends on the specificity for Pro at P2 and P2’ and the ionization state of His at P3 [59]. CRWHTVG/LRWE-Lys(Dnp)-DArg-NH2 (Dnp – 2,4-dinitrophenyl) substrate is selective for CtsS as compared with CtsB, H, L, and X [2,70].
Some of the human cysteine cathepsin endogenous substrates. O- hydroxyproline.

Table 1

| Substrates | Cathepsins | Cleavage site | Ref. |
|------------|------------|---------------|------|
| Elastin    | CtsS       | PGCV [57]     | [57–60] |
| Collagen type I | CtsK | ECGPN [57] | |
| Collagen type II | CtsS | GPGD [61] | [58] |
| Collagen type III | CtsB | PEPF [61] | [59] |
| Collagen type IV | CtsL | FPFH [63] | [57] |
| Collagen type XVIII | CtsL | REEA [64] | [62] |
| Proteoglycan 4 | CtsS | LGPF [59] | [63] |
| Collagen II | CtsB | NFFP [64] | [61] |
| Casein kinase II | CtsL | EPFF [63] | [66] |
| Thyroglobulin | CtsB | REAA [65] | [60] |
| Bid        | CtsB       | SRSF [47]    | |
| CtsH       | SGLG [48]  | AEHI          | |
| CtsK       | SGLG [48]  | AEHI          | |
| CtsL       | SGLG [48]  | AEHI          | |
| CtsS       | SGLG [48]  | AEHI          | |
|            |            | VQAY [65]     | |
|            |            | DQG [47]      | |

Most endogenous inhibitors of PLCPs bind to the binding site cleft. Thus, investigating PLCPs specificity to inhibitors may provide more information about the overall specificity of these proteases. PLCP specificity to different endogenous inhibitors allows for regulating the proteolytic network. Endogenous cysteine cathepsins inhibitors include cystatins (including stefins and kininogens), thyropins, and serpins, which are competitive reversible inhibitors [2]. Those cystatins which were shown to inhibit human cysteine cathepsins are listed in Table 2. The endogenous inhibitors of the cathepsins are characterized by various potency towards distinct proteases. For example, bovine steins A, B, and C suppress the endopeptidases papain, CtsL, and S more efficiently than CtsB [71–73]. Mouse stein A variants inhibit the endopeptidases CtsL and S to a greater extent than exopeptidases CtsB, C, and H [74]. Human cystatin C, stein B, and chicken cystatin significantly reduce the activity of the endopeptidase cruzipain, cysteine protease of Trypanosoma cruzi [75]. CtsX can be inhibited by various cystatin-type protease inhibitors although a modeling study showed that the side chains of His23 and Tyr27 (Protein Data Bank (PDB) ID: 1EF7) collide with the residues Val55 and Ala56 in the QVVG region of stefin B [76]. Bacterial leupeptin is an aldehyde that covalently binds to nucleophilic Ser and Cys at the active site of the serine and cysteine proteases, respectively [77]; however, it more potently inhibits cathepsins in rats rather than in humans [23]. Elastolytic activity of CtsV and K is inhibited by GAGs such as chondroitin sulfate [78]. Cathepsin activity of CtsK is suppressed by the negatively charged polymers, polyglutamates, and oligonucleotides [2,79]. Human CtsK and rat CtsB, H, K, and L can be inhibited by T-kininogen (only found in rats [80]) [81,82]. Multiple PLCPs are negatively affected by reactive oxygen species and reactive nitrogen species [83]. Papain, CtsB and K can be reversibly inhibited by nitric oxide via S-nitrosylation of their catalytic Cys, while hydrogen peroxide has been shown to oxidize triticain-α from Trichinella spirum L., CtsK, and S to either a reversible sulfenic or to an irreversible sulfonic acid [83]. S-Methyl methanethiosulfonate can induce the disulfide bond formation between Cys22 and catalytic Cys25 in tricitain-α (active protease numbering), thus protecting the protease from irreversible oxidation [84].

The variability of PLCP substrates and inhibitors enables modulating the proteolytic network. Some proteases can be inactive, while some other proteases are not. At the same time, certain proteases have limited activity performing various specific functions. For instance, CtsK can possess collagenase activity, while its elastolytic activity is suppressed [78]. The complexity of this network enables maintaining homeostasis. Still, there are other ways of regulating PLCP activities to be discussed below.

2.2. Role of pH in PLCP activity and specificity

Hydrolysis takes place in water solution and the components of the solution can drastically change the effectiveness of the reaction. Such parameters as pH and ion strength influence the charges on amino acid residues and thus impact intra- and intermolecular interactions [90]. For instance, CtsB activation proceeds faster at acidic pH [2,91] because CtsB propeptide binds into the binding groove of the protease more efficiently at pH 4.0 than at pH 6.0 [92]. At pH 5.5, the elastolytic activity of CtsC is lower than that of other cathepsins. In contrast, at pH 7.4, elastin is cleaved by CtsC [75]. Trypanosoma cruzi proteases extract, usually referred as ‘bromelain’, and investigate their biochemical properties [94]. The basic bromelain forms 1 and 2 cleaved pGlu-FL-p-nitroanilide (pGlu – pyroglutamic acid) in the pH range of 4.0–5.4, while anain form 2 cleaved the substrate in the range of 6.5–7.0 [95]. pH 6.0 was optimal for cleavage of Z-RR-AMC and Z-FR-AMC by basic bromelain form 1, while the optimum pH of the activity of its form 2 towards Z-FR-AMC and Z-RR-AMC was 6.0 and close to neutral, respectively [94]. Basic bromelain forms 1, 2, and 3 possess a decreased enzymatic activity under acidic conditions [94]. Both acidic bromelain forms are active at acidic pH as dipeptide [94]. Ph5.5 hinders the digestion of collagen by CtsK and L, in contrast to pH 6.8 [93]. Even very similar proteases may exercise different activities at distinct pH levels. It was demonstrated by Matagne et al. who managed to separate distinct PLCPs from crude Ananas comosus proteases extract, usually referred as ‘bromelain’, and investigate their biochemical properties [94]. The basic bromelain forms 1 and 2 cleaved pGlu-FL-p-nitroanilide (pGlu – pyroglutamic acid) in the pH range of 4.0–5.4, while anain form 2 cleaved the substrate in the range of 6.5–7.0 [95]. pH 6.0 was optimal for cleavage of Z-RR-AMC and Z-FR-AMC by basic bromelain form 1, while the optimum pH of the activity of its form 2 towards Z-FR-AMC and Z-RR-AMC was 6.0 and close to neutral, respectively [94]. Basic bromelain forms 1, 2, and 3 possess a decreased enzymatic activity under acidic conditions [94]. Both acidic bromelain forms are active at acidic pH as low as 4.0, except for the form 2 with Z-RR-AMC, while a significant loss in their activity is observed at pH 7 or above it [94]. The optimum pH for PLCPs from Toxocara canis larvae for digesting Z-FR-AMC and Z-RR-AMC were 5.0 and 6.5, respectively [14]. The reduction in cleavage ratio of Z-FR-AMC to Z-RR-AMC changed from 16:1 at pH 5.0 to about 1:1 at pH 6.5 suggesting that the deprotonation of Glu205 (papain numbering; PDB ID: 1BP4) in S2 increased the affinity to Arg [14,96]. Fasciola hepatica CtsL2 (FhCL2)
had 6-fold greater affinity for Z-PR-NH-AMC at pH 5.5 and 3-fold greater affinity at pH 7.3 than FhCL1 [97]. However, both FhCL2 and FhCL3 were able to digest native collagen at neutral pH, whereas CtsK required acidic conditions [13]. Wheat triticain-coefficientally degraded gluten in the pH range of 3.0–6.5 only [98]. Other plant PLCPs responsive to dehydration protease 21A (RD21A), RD21B, RD21C, RDL2, and CtsB-like protease 3 (CTB3) can be labeled by the activity-based probe MV201 at any pH, AALP was labeled only in a narrow neutral pH range, labeling xylem cysteine protease 1 (XCP1) and XCP2 was at wider range of neutral pH, while senescence-associated gene 12 protease (SAG12), RD19A, triticine protease 1 (XCP1) and XCP2 was at wider range of neutral pH, can be labeled by the activity-based probe MV201 at any pH, AALP other plant PLCPs responsive to dehydration protease 21A.

Table 2

| Cystatin   | Cathepsins | K_i  | Recognition sequence | Ref. |
|------------|------------|------|----------------------|------|
| Cystatin A | CtsB       | 8.2 nM | QIVAG                | [85,86]|
|            | CtsH       | 0.31 nM |                       |      |
|            | CtsL       | 1.3 nM |                       |      |
|            | CtsS       | 0.05 nM |                       |      |
| Cystatin B | CtsB       | 73 nM  |                       |      |
|            | CtsH       | 0.58 nM |                       |      |
|            | CtsL       | 0.23 nM |                       |      |
|            | CtsS       | 0.07 nM |                       |      |
| Cystatin C | CtsB       | 0.25 nM | QIVAG                | [86] |
|            | CtsH       | 0.28 nM |                       |      |
|            | CtsL       | 0.005 nM |                    |      |
|            | CtsS       | 0.008 nM |                       |      |
| Cystatin D | CtsB       | >1000 nM | QVMAA                |      |
|            | CtsH       | 8.5 nM  |                       |      |
|            | CtsL       | 25 nM   |                       |      |
|            | CtsS       | 0.24 nM |                       |      |
| Cystatin SN| CtsB       | 19 nM   | QTVGG                | [87,88]|
| Cystatin M/E| CtsB      | 32 nM   | QLVAG                | [87,88]|
|            | CtsL       | 1.78 nM |                       |      |
|            | CtsS       | 0.47 nM |                       |      |
| Cystatin F | CtsB       | 0.31 nM | QIVKG                | [89] |
|            | CtsL       | 0.28 nM |                       |      |

The presence or absence of particular ions may also significantly affect the interactions between molecules. Ions can shield charges on the amino acid residues in the binding cleft of the protease thus preventing salt bridges from forming, or they can increase the strength of the hydrophobic interactions [108,109]. Unlike CtsC, whose activity was investigated towards GF-p-nitroanilide at pH 5.0 and GF-4MjNA (4-methoxy-β-naphthylamide) at pH 3.6–8.2, CtsCEx did not require halide ions for enzymatic activity [110,111]. However, CtsCEx optimal activity towards Z-LR-AMC at pH 5.6 was observed between 50 and 200 mM NaCl [44]. Specificity constants (k_cat/K_M) of CtsL and V for Abz-ALR;SSQK-EDDnp, Abz-KLR;SSQK-EDDnp, and Abz-ELR;SSQK-EDDnp were lower in the presence of NaCl. The hydrolysis of Abz-KLR;SSQK-EDDnp by CtsV was the most sensitive to NaCl. These results indicate the importance of the electrostatic interaction between the positively charged P3 and CtsL and V at pH 5.5. NaCl also reduced the CtsV activity at pH 5.5 towards elastin but not towards Z-RR-MCA (MCA – 7-methoxycoumarin-4-acetyl) [112]. Activity of all stem bromelain proteases for Z-RR-AMC and Z-PR-AMC at pH 6.25 was completely inhibited by 1 mM Zn^{2+}, typical of PLCPs [113–115]. 10 mM Mg^{2+}, Ca^{2+}, and Mn^{2+} significantly suppressed the activity of basic bromelain forms 1, 2, and 3, ananain forms 1 and 2, and comosain [94]. A 25 % reduction in acidic bromelain 1 was observed in the presence of 10 mM Mn^{2+}, but not of 10 mM Mg^{2+} or Ca^{2+}. 0.1–1 mM Mg^{2+}, 0.1–0.5 mM Mn^{2+}, and 0.1–0.5 mM Ca^{2+} stimulated the activity of ananain form 1. The properties of the ananain form 2 were similar to form 1, however, the increase of the activity was observed at 0.1–5 mM of Mg^{2+} and 0.1 mM of Mn^{2+}, 0.1–1 mM of Mn^{2+} also stimulated comosain [94]. The activity of the cysteine protease zingibain (from Zingiber officinale rhizomes) towards bovine casein at pH 7.0 was stimulated by 2 mM of Ca^{2+} and reduced in the presence of Zn^{2+}, Mn^{2+}, and Mg^{2+} [116]. Milk-clotting activity of procerain B from Calotropis procera was elevated by 0.01–10 mM Mn^{2+} and completely abolished by 0.01–10 mM of Zn^{2+} at pH 7.5 [117]. Caseinolytic activity of microcarp from the Ficus microcarpa latex was suppressed by 42 %, 79 %, 85 %, and 92 % by Ca^{2+}, Mg^{2+}, Mn^{2+}, and Zn^{2+} at 5 mM concentration, respectively, at pH 8.0 [94,118]. The protease from Parabacteroides distasonis (Pd_dinase) was inhibited by excessive Zn^{2+} via a coordination of Cys56, His340, and the side-chain carboxylate oxygens of Asp298 at pH 8.0. As a result, the tip of the flexible loop (residues 290–310) occluded the active site pocket [119].

Except for Zn^{2+} inhibiting PLCPs, other ions can either suppress or stimulate PLCP activity depending on the enzyme. Normally the excess concentrations (5 mM or more) of bivalent ions inhibit PLCPs. However, at low concentrations the impact differs. The possible explanation is that in some cases charged residues play important role in substrate binding while ions shield these residues. On the other hand, large ionic strength favors hydrophobic interactions [108,109].

2.4. Macromolecules in regulation of PLCP activity and specificity

The propeptides inhibiting PLCPs can also promote their folding and affect their activity and even specificity. The X-ray structures of the papainzymogen, enzyme kinetic analyses, intact and tryptic digested peptide mass analyses, and circular dichroism data of mature proteases showed that mutations in the propeptide induce the conformation changes in the mature papain altering its activity and specificity [120].

Table 2

| Cathepsins | Recognition sequence | Ref. |
|-----------|----------------------|------|
| CtsB      | QIVAG                | [85,86]|
| CtsH      |                       |      |
| CtsL      |                       |      |
| CtsS      |                       |      |
| CtsB      | 73 nM                |      |
| CtsH      | 0.58 nM              |      |
| CtsL      | 0.23 nM              |      |
| CtsS      | 0.07 nM              |      |
| CtsB      | 0.25 nM              | QIVAG |
| CtsH      | 0.28 nM              |      |
| CtsL      | 0.005 nM             |      |
| CtsS      | 0.008 nM             |      |
| CtsB      | >1000 nM             | QVMAB | [86] |
| CtsH      | 8.5 nM               |      |
| CtsL      | 25 nM                |      |
| CtsS      | 0.24 nM              |      |
| CtsB      | 19 nM                | QTVGG | [87,88]|
| CtsB      | 32 nM                | QLVAG | [87,88]|
| CtsL      | 1.78 nM              |      |
| CtsS      | 0.47 nM              |      |
| CtsB      | 0.31 nM              | QIVKG | [89] |
| CtsL      | 0.28 nM              |      |
Fig. 1. Charge distribution on the surface of different human cysteine cathepsins at pH 4.5 and 7.5 (PDB IDs: 1CSB for CtsB, 1M6D for CtsF, 1CJL for CtsL, 2FQ9 for CtsS, 1EF7 for CtsX) [76,100–103]. The electrostatic potential field was modelled using ProKSim (Protein Kinetics Simulator) software [104,105]. The values of protein atom charges were automatically calculated using the Henderson–Hasselbalch equation according to standard values of the dissociation constant $pK_a$ for amino acid residues [106]. Surface of the protein colored by electrostatic potential value from $-100$ mV (red) to $+100$ mV (blue). The figure was created by PyMOL Molecular Graphics System [107]. The total charges of each binding pocket are shown to the right of each structure. They were calculated as a sum of the charges of the amino acid residues identified for the pockets. The disposition of the pockets indicated on the structure of CtsB and on Fig. 2. N.d. – not determined. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
The inhibitors that abolish or suppress PLCP activity were already discussed above in detail. Meanwhile, there are other molecules that can promote the digestion efficiency. For instance, GAGs were shown to regulate the collagenolytic activity of CtsK [78,121,122]. The collagenolytic activity of CtsK was stimulated by chondroitin and keratan sulfate and inhibited by heparan sulfate or heparin [2]. CtsK-dependent cleavage of triple collagen helix requires the complex with chondroitin sulfate to be formed [121]. Chondroitin sulfate binds and orients several CtsK molecules in ‘beads-on-a-string’ orientation [123]. However, substituting the binding site for chondroitin into the CtsL-like site did not abolish the collagenase activity [2]. The autocatalytic activation of cathepsins was also facilitated by the negatively charged molecules, such as GAGs and dextran sulfate [124,125].

To sum up, the specificity of PLCPs is quite variable and tightly regulated by pH level, ions, inhibitors, and other molecules. Even the protease prodomain is crucial for the activity of the mature enzyme. All these regulation mechanisms are specific and determined by the structure of a particular protease.

3. Structure of cysteine cathepsins

While describing structures of PLCPs, we will discuss the particular amino acid residues of cysteine cathepsin binding clefts. Their numbers are indicated according to the following PDB structures unless indicated otherwise: 1BP4 for papain, 1CSB for CtsB, 1M6D for CtsF, 1ATK for CtsK, 1CJL for CtsL, 2FQ9 for CtsS, 1FH0 for CtsV, 1EF7 for CtsX [76,96–103,126,127].

3.1. The general structure of cysteine cathepsins

The general structure of animal PLCPs is presented on Fig. 2. It is similar to that of papain and consists of two domains: the left N-terminal domain consisting of α-helices (l-domain) and the right C-terminal domain comprising a β-barrel and a short α-helical motif (r-domain) (Fig. 2A) [128]. The domains form a ‘V-shaped’ active site cleft containing a catalytic triad of Cys, His, and Asn (Fig. 2B) [2,76]. The catalytic domain of the pro-cathepsins is preserved in the active conformation, with the prodomain interacting with the proregion binding loop and the substrate-binding site. The corresponding contacting surfaces of the prodomain and the catalytic domain have opposite electrostatic characters [129].

Cysteine cathepsins have distinct structural features as well. Based on sequence analysis, the cathepsins can be divided into two distinct subfamilies, designated as CtsL-like and CtsB-like proteases [130,131]. Proteptides of the CtsL subfamily (CtsL, V, K, S, W, F, and H) contain a propeptide of about 100 residues with two conserved motifs: the ERFNIN and the GNFD ones. The proteases in CtsB subfamily either lack only the GNFD motif (CtsB) or both of them (CtsC, O, and X) [2]. All proteases can also be classified according to the structure of their catalytic cleft. In case of human cysteine cathepsins in endopeptidases CtsF, L, K, S, V, and W, the active-site cleft is extended on both sides from catalytic cysteine [132]. In carboxypeptidases CtsB and X and in aminopeptidases CtsH and C, the binding sites are located only on the non-primed side and on the primed side, respectively. The other side of the cleft is filled with the additional domains [132]. The occluding loop of CtsB and the mini-loop of CtsX contain His residues to dock the C-terminal carboxylic group of a substrate. The CtsH mini-chain and the CtsC exclusion domain carry carboxylic groups for docking the positively charged N-termini of the substrates [133,134]. The mutations of His residues in the occluding loop and deletion of the loop demonstrated their importance for the CtsB endopeptidase activity [2,135,136].

The binding cleft is indicated on a PLCP structure on Fig. 2C. It consists of two parts: the non-primed side for binding of the N-terminal side of a substrate (S1, S2, S3, etc.) and the primed side for binding of the C-terminal side of a substrate (S1′, S2′, S3′, etc.). The amino acids located in these areas define the specificity of a protease in the first place, so we will pay special attention to them in the remaining part of the review.

3.2. Structure and specificity of S1 binding site

The S1 binding site of PLCPs binds the first amino acid from the cleavage site on the N-terminal side. Its structure is a groove in the wall of the left domain [137] formed by Ser21, Cys22, and Gly62 on one side and Gly62, Cys63, and Asn64 on the other side that are bridged by the disulfide bond Cys22-Cys63 (papain numbering) [76]. The residues in S1 of the cathepsins, which were shown to interact with the substrates, are illustrated on Fig. 3 and discussed further in the present subsection. The S1 site of CtsB is mainly composed of Gln23, Gly27, Cys29, Gly74, and Gly198 [138]. CtsB prefers Arg and Lys in the P1 position among all natural residues because the occluding loop harbours the negatively charged Glu122. However, it also well accommodates Tyr, norleucine (Nle), Thr and Gly [139], while some artificial bulky and hydrophobic derivatives of Lys and Cys (Lys(2-Cl-Z), Cys(Bzl), Cys(Me-Bzl) (Me – methyl), Cys(MeOBzl), as well as Nle(OBzl)) are hydrolyzed faster than Arg [68]. The S1 pocket is shallow in the L domain of CtsF and is created by Gly23, Cys63, Met64, and Gly65. Only Met64 side chain participates in forming the pocket, this residue being unique among various PLCPs [102]. CtsF prefers the positively charged amino acid residues at P1 as well [66]. Gly65 at this position is conserved across human cysteine cathepsins, while Gly23 is conserved in the family except for CtsW (Gly is replaced with Asn) [102]. The S1 subsite of CtsV hydrolyzes the substrates with Leu, Gln, Arg, and Lys at P1 because of the aliphatic side chains in S1, while CtsL prefers Arg followed by Lys at P1 [112]. Substrates with the negatively charged and polar non-charged resi-
3.3. Structure and specificity of S2 binding site

The S2 subsite of PLCPs is considered the most crucial binding spot [90]. It comprises the residues at positions 67, 68, 86, 133, 157, 160, and 205 (papain numbering), which are quite variable among PLCPs [13]. In contrast to the positively charged P1 amino acid residue, PLCPs mainly prefer bulky hydrophobic side chains (such as Phe) at P2 [66,144]. The residues in S2 of the cathepsins, which were shown to form bonds with the substrates, are illustrated on Fig. 4 and discussed further in the present subsection. The hydrophobic S2 subsite of CtsL is formed by Leu69, Met70, Ala135, Met161, and Gly164 at the bottom [128]. The side chain of Met70 (Pro76 for CtsB) lies across the cleft resulting in a shallower S2 pocket. If we look down the active site cleft of CtsL, orienting the primed subsites being above and unprimed subsites below, Asp160, Met161, Asp162, and Ala214 form a wall on the side of the S2 pocket [128]. Despite Asp160 and Asp162 at the S2 subsite, CtsL prefers aromatic residues, mostly Phe, Trp, Tyr and also Leu. The specificity to aromatic residues distinguishes CtsL from other cysteine cathepsins with exception for CtsV [61,66]. CtsS is specific to hydrophobic Val, Met, and Nle at the P2 position [2,70]. The deep S2 pocket of CtsS is formed by the side chain of Met68 and the methylene groups of Gly133 and Gly160 at its bottom; the side chains of Ph67 and Val157 shape its sides, and the side chain of Phe205 at the distal position closes the pocket [138]. Gly133, Val157, and Phe205 are unique residues present in CtsS [140]. Gly133 of CtsS is substituted by Ala in CtsK, L, and V so that CtsS has the deeper S2 pocket than the other cathepsins. Conversely, Val157 is replaced with Leu in CtsK and V and Met in CtsL. In CtsS, this residue restricts the entrance into the S2 pocket [138]. Phe205 (Leu in CtsK or Ala in CtsV and L) defines the specificity of CtsS to branched hydrophobic P2 residues [137]. The side chain of Phe205 in S2 of CtsS facilitates two conformations that accommodate either Leu or a bulkier side chain of Phe at the P2 site, which is a unique feature for CtsS [140]. Phe205 in the open position is stabilized by interactions with the aromatic ring of Tyr118 and the side chain of Val157 [140]. The Phe205Glu substitution in CtsS leads to a decreased affinity to hydrophobic residues at P2 but 77 times more efficient digestion of the RR dipeptide [90]. The S2 site of CtsV requires hydrophobic residues with a preference for Phe and Leu leading to a less efficient cleavage of peptides containing Tyr, Val, and Ile at the P2 position. Trp appeared to be least favorable among hydrophobic residues at the P2 position.
The S2 pocket of CtsF is also hydrophobic being formed by Leu67, Pro68, Ala133, Ala160, and Met205 [102]. The width of the S2 pocket is determined by Leu67 on one side, and Ile157 and Asp158 on the other side of the pocket [102]. The amino acids at the P2 position influence the specificity to inhibitors. Peptides carrying exopeptidase warheads with hydrophobic residues (Leu, Ile, Phe, nor-Ile, Trp, Tyr, Val) at P2 more efficiently inhibit PLCPs than peptides containing small or hydrophilic residues (Ala, Asp, Glu, Lys, Gln, Gly, His, Pro, Ser, Arg, Thr) at P2 [3].

Some PLCPs possess unique specificity to Pro at P2 resulting in a collagenolytic activity. The S2 subsite of CtsK consists of Tyr67, Met68, Ala134, Leu160, Ala163, and Leu209 [23]. The side chains of Tyr67, Leu160, and Leu209 form a barrier to substrate binding, as in CtsB, limiting the length of the substrate and probably determining the specificity to Pro in the S2 position and collagenolytic activity of CtsK [138]. This preference is unique among the mammalian cysteine cathepsins [145]. CtsK can cleave synthetic substrates containing either Pro or Phe at the P2 position. Nevertheless, it preferentially cleaves type I collagen after Pro20 at P2, while CtsS cleaves after Phe23 at P2 [93]. The mutations Tyr67Leu and Leu209Ala (as in CtsS) abolished the Pro specificity and collagenolytic activity of CtsK [13,146], whereas the reverse mutations Leu69Tyr and Ala214Leu in CtsL introduced Pro specificity [147]. Rat CtsK displays similar preferences, but rat CtsK has weaker peptidase activity than human CtsK [148]. The Ser134-Ala and Val160Leu mutants (the same as in human CtsK) are more active than the WT enzymes (human CtsK numbering) [149]. Simil-ilar to other PLCPs, CtsCAEx prefers the P2 position containing large hydrophobic residues (Phe, Leu, or Pro, but not Val) [44]. It efficiently digests Z-FR-AMC, Z-LR-AMC, and Boc-VLK-AMC. However, it can also digest Z-GPR-AMC containing Pro at P2. Almost no activity was detected in case of Bz-Arg-AMC (Bz – benzoyl) and small hydrophobic residues (Val, Ala, or Gly) or Arg at the P2 position [44].

Some cathepsins can also accommodate polar amino acid residues at the P2 position. The S2 subsite of CtsB comprises Pro76, Ala173, Gly198, Ala200, and Glu245 [138]. CtsB binds bulky hydrophobic amino acids (homoserine(Bzl) (hSer(Bzl)), Glu(Bzl), or Glu(Chx) (Chx – cycloheximide)) at the P2 position, synthetic residues being recognized better than natural amino acids [68]. But unlike most cathepsins, the S2 subsite of CtsB has a negative charge of the Glu245 side chain, which enables the binding of CtsB with the positively charged residues (Arg, Lys, and His) at P2 at neutral pH levels [138,144,150]. Glu245Gln mutation in CtsB limited the digestion of substrates that contain Arg at P2 but had no effect on the P2 Phe substrates [2,151]. However, at acidic pH levels CtsB can digest the substrates with Glu at P2 [99]. CtsX S2 pocket is formed by Gly74, Asp76, Gly154, Leu160, Val181, and His234 [76]. The Gly74 amide proton and carbonyl oxygen are essential for the CtsV and L activity [112]. The residues in S3 of the cathepsins, which were shown to be involved in interactions with the substrates, are illustrated on Fig. 5 and discussed further in the present subsection. The central part of the CtsL S3 subsite contains Gly63 and Asn66, and the carbonyl oxygen of Gly61 [128]. CtsL favors both basic amino acids (Lys and Arg) in the P3 position since it contains the acidic Glu63 at the S3 subsite [112], as well as aliphatic residues [66]. S3 of CtsB favors its specificity to aliphatic side chains (Leu, Nle, hCha, 2-amino-o-nocnnoic acid (2Aoc), hLeu, octahydroindole-2-carboxylic acid (Oic)) and basic amino acids as well Phe-derivatives [68]. CtsK prefers the positively charged residues at P3 [2,69]. The S3 pocket of CtsK is formed by Gly65 and Gly66 (conserved among PLCPs), the side-chains of Asp60, Lys61, Leu67, Asn70, and the carbonyl oxygen of Met59 [102]. Residues 67 and 61 vary widely across the PLCPs. In CtsF, residue 60 is Asp, whereas in other enzymes, it is represented by Gly, Ala, or most commonly, Asn [102].

CtsS contains a three-residue insertion relative to CtsK in the loop between Ser57 and Asn60 in the S3 pocket [140]. In CtsL and V, the insertion is only two amino acid residue long. Therefore, CtsS has a smaller S3 pocket than the other cathepsins [140]. The S3 subsite of CtsS contains Gly65 and Gly66, the side-chains of Lys61 and Phe67, and the backbone atoms of Gly59 and Asn60. Lys61 (Asp in CtsK, Gln in CtsV, and Glu in CtsL) and Gly59 (Glu in CtsK and Gly in CtsV and L) are unique for the S3 pocket of PLCP

3.4. Structure and specificity of S3 binding site

The S3 binding site possesses a lower specificity than the S2 subsite, although amino acids with a side chain in the P3 position are essential for the CtsV and L activity [112]. The residues in S3 of the cathepsins, which were shown to be involved in interactions with the substrates, are illustrated on Fig. 5 and discussed further in the present subsection. The central part of the CtsL S3 subsite contains Gly63 and Asn66, and the carbonyl oxygen of Gly61 [128]. CtsL favors both basic amino acids (Lys and Arg) in the P3 position since it contains the acidic Glu63 at the S3 subsite [112], as well as aliphatic residues [66]. S3 of CtsB favors its specificity to aliphatic side chains (Leu, Nle, hCha, 2-amino-o-nocnnoic acid (2Aoc), hLeu, octahydroindole-2-carboxylic acid (Oic)) and basic amino acids as well Phe-derivatives [68]. CtsK prefers the positively charged residues at P3 [2,69]. The S3 pocket of CtsK is formed by Gly65 and Gly66 (conserved among PLCPs), the side-chains of Asp60, Lys61, Leu67, Asn70, and the carbonyl oxygen of Met59 [102]. Residues 67 and 61 vary widely across the PLCPs. In CtsF, residue 60 is Asp, whereas in other enzymes, it is represented by Gly, Ala, or most commonly, Asn [102].

CtsS contains a three-residue insertion relative to CtsK in the loop between Ser57 and Asn60 in the S3 pocket [140]. In CtsL and V, the insertion is only two amino acid residue long. Therefore, CtsS has a smaller S3 pocket than the other cathepsins [140]. The S3 subsite of CtsS contains Gly65 and Gly66, the side-chains of Lys61 and Phe67, and the backbone atoms of Gly59 and Asn60. Lys61 (Asp in CtsK, Gln in CtsV, and Glu in CtsL) and Gly59 (Glu in CtsK and Gly in CtsV and L) are unique for the S3 pocket of PLCP

The figure shows the aligned structures of CtsB, F, L, S, X, and V (PDB IDs: 1CSB, 1M6D, 1CJL, 2FQ9, 1EF7, 1ATK, and 1FH0, respectively) [76,100–103,126,127]. The figure shows the aligned structures of CtsB, F, L, S, X, and V (PDB IDs: 1CSB, 1M6D, 1CJL, 2FQ9, 1EF7, 1ATK, and 1FH0, respectively) [76,100–103,126,127]. Binding sites are indicated by the same colors as on Fig. 2. Amino acid residues of cts are green, cts – magenta, cts – blue, cts – orange, cts – black, CtsK – dark green, and CtsV – beige. Arrows indicate the binding groove. The PyMOL Molecular Graphics System was used to create the figure [107]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 5. Amino acid residues of the S3 binding pocket of human cysteine cathepsins.
The positively charged side chains of Lys61 and of Phe67 form a specific binding site [138]. CtsS S3 subsite is composed of Gln60 and Arg72, which is a distinct feature of CtsV. Thus, CtsV prefers substrates with a negatively charged Glu in the P3 position, which may interact with the guanidino group of Arg72 in the S3 [112].

Such diversity of amino acids that enables substrate cleavage by different PLCPs indicates a low specificity of this pocket. Similarly to the S2 binding site, the S3 binding site generally contains hydrophobic amino acids. However, some human cysteine cathepsins harbor negatively charged (Asp60 in CtsF and Glu63 in CtsL) or positively charged (Lys61 in CtsS and Arg72 in CtsV) amino acid residue at S3 (Fig. 5) expanding the range of sensitive substrates.

3.5. Structure and specificity of the primed side

The abovementioned S1-S3 binding sites are referred to the non-primed side of the active cleft. The C-terminal part of the substrate binds to the so-called primed side of the binding cleft. It is less thoroughly investigated than the other side of the binding cleft but there are still some noteworthy features [2]. The residues in the primed side of the cathepsins able to interact with the substrates are illustrated on Fig. 6 and discussed further in the present subsection. S1’ of CtsL possesses Glu141 and Glu192 that make this side negatively charged [112]. CtsL most efficiently cleaves the substrates with His, Asn, Phe, and Ser at P1’ and less efficiently digests ligands with Arg, Gln, Gly, and Val. Glu at P1’ was the least favorable substrate, while the substrate with Pro at P1’ was resistant to hydrolysis [112]. The S2’ position of CtsL is hydrophobic and is formed by highly conserved Trp189 and Trp193 [129]. CtsL seems to lack a clear preference for residues at P2’, but Ser and Glu are the most and least preferable residues, respectively [112]. The primed side of CtsL also contains Asn18, Gly20, Gln21, and Leu144 [128].

In CtsB, S1’ and S1’ are negatively charged because of Glu122 and Glu194 side chains [138]. However, the occluding loop (105–125) of CtsB exposes the positively charged His110 and His111 in the S3’ subsite [138]. It limits the length and orientation of a substrate underlying CtsB exopeptidase activity [138]. CtsX contains Trp202 in its S1’ binding pocket [76]. The three residues in the mini-loop of CtsX (Ile24, Pro25, and Gln26) form the wall of the S2’ binding site, which is occupied by the imidazole ring of His23. In turn, His23 can interact with the C-terminal carboxylic group of a substrate [76]. CtsX can be inhibited by CA074, an epoxysuccinyl-based inhibitor with IP dipeptide which binds to S1’ and S2’; however, the His23 ring is rotated into a position equivalent to a position occupied by His110 in CtsB to create an open S2’ [76].

The proximal S1’ of CtsS consists of His159 and Trp177 at its bottom and Ala136, Arg137, and Asn158 in the walls of the pocket [140]. Arg137 is an unusual residue as papain, cruzain, and CtsL, K, and B have Ala, Ser, Gly, Ser, and Tyr at the corresponding position, respectively [137,140]. The S1’ distal subsite is defined by the main chain atoms of His138 and Pro139 and the side chain of Phe142 [140]. Phe142 (Gln in CtsK and V and Leu in CtsL) can form a hydrophobic interaction with the substrate [138], and CtsS was shown to prefer hydrophobic branched side chains at the P1’ [2,70].

In CtsV the preferable amino acid residue at S1’ was Val, followed by Asn and Ser [112], CtsV also hydrolyzes substrates with Arg or Glu at P1’ with low but similar efficiency. The peptides with Pro at P1’ are resistant to hydrolysis. At S2’ of CtsV, Val and Ile are the most preferable amino acids [112].

The primed side of most PLCPs is formed by two Trp residues which may form the stacking bonds with large aromatic amino acids of the substrate. However, the S1’ binding sites of PLCPs generally contain charged amino acids enabling the interactions with basic (CtsB and L), acidic (CtsS), or both types of the charged amino acid residues (CtsV). The S2’ binding site in endopeptidases is hydrophobic, while carboxypeptidases CtsB and X possess positively charged His residues which interact with the C-terminal carboxylic group of a substrate (Fig. 6). Different structural features of the binding cleft of PLCPs allow for determining the specific substrates even for similar proteases. These differences facilitate the development of selective inhibitors or probes for PLCPs that can be further used in therapy and diagnostics.

4. Examples of rationally designed PLCP inhibitors

The activity of proteases is commonly modulated through their inhibition. PLCP inhibitors mainly interact with the S1, S2, S3, and S1’ pockets of the protease [102]. Thus, knowing the structure and specificity (Table 3) of each binding site, we can develop an effective inhibitor as those discussed in the present subsection and listed in Table 4. E-64 is a non-selective inhibitor of all the cysteine cathepsins (except for CtsC) and the calpain [2,155]. The inhibitor was modified to obtain a specific CtsB inhibitor via mimicking a carboxy-terminal dipeptide by replacing the terminal guanidinobutyramine of E-64 with Pro and terminating the chain with a free carboxylic group [152,153]. Modifications of the other terminus of the epoxysuccinyls to various ester and amide derivatives resulted in new inhibitors named CA030 and CA074 [152,153]. KI of CA030 were 4.38, 42,900, and 40,000 nM, while KI of CA074 were 1.94, 75,000, and 233,000 nM for CtsB, H, and L, respectively [152]. The crystal structure (illustrated on Fig. 7) confirmed the key role of residues His110 and His111 in the occluding loop of CtsB in C-terminal carboxylic group recognition at S2’ binding site [101].
Amino acid residues at the positions P3 to P3' of ligands interacting with the binding clefts of the listed proteases. The table is based on the sources cited in the present review. Hydrophobic amino acid residues are shown in gray, aromatic ones are shown in purple, polar uncharged amino acids are shown in green, positively charged amino acids are shown in blue, and negatively charged amino acids are shown in red. The graphs in the first column visualize types of amino acids favored in indicated positions [154].

| PLCP | P3 | P2 | P1 | P1' | P2' | P3' | Ref. |
|------|----|----|----|----|----|----|------|
| CtsL | A/L/ M | V/L/I | Q  | G/N | S  | K  | [61,66,112,128] |
|      | K/R | F/Y/W | K/R | F/S/N/Q/R/H | H  |    |      |
| CtsV | G/A | V/L/I | L  | V  | V/I | K  | [66,112] |
|      | P  | F/Y/W | Q  | S/N/R/E | S  |    |      |
|      | K/H |    | K/R |    |    |    |      |
|      | E  |    |    |    |    |    |      |
| CtsK | K/R/H | L/I | G  | G  | F/P | F  | [23,66,69,93,145] |
|      | F  | K/R | F  | S/P | S/N/Q/K/R |    |      |
|      | P  |    |    |    |    |    |      |
|      | R  |    |    |    |    |    |      |
| CtsS | I  | V/L/M | G  | A/V/L/I | R  | M/I | [61,66,70,140] |
|      | T/P | F/W | T/N | I  | E  | F/W/Q |      |
|      | R  |    | K/R |    |    | K    |      |
| CtsB | L  | A/V/L/M | G/A/L | A/L/I | A/ | N.d. | [61,66,153,67,68,99,101,138,139,150,152] |
|      | S  | Y  | L  | F  | V  |      |      |
|      | R  | F/Y | S/T/N | F  | P  |      |      |
|      | K/R/H | K/R | K/R |    |    |      |      |
| CtsX | G  | F  | W  | I  | W  | N.d. | [76] |
|      | F  |    | R  |    |    |      |      |
|      | R  |    |    |    |    |      |      |
CtsX can also be inhibited by CA074 in case the His23 ring is rotated into a position equivalent to a position occupied by His110 in CtsB to create an open S2\(_0\) (Fig. 7) [76].

Experimental measurements with a series of vinyl sulfone inhibitors with different P2 substituents revealed that the inhibitors containing P2-Leu and P2-Phe are the most effective for CtsS [156]. They inhibited the protease with similar \(K_I\) of 13 and 18 nM, respectively. The vinyl sulfone inhibitors with P2-Leu and P2-Phe showed less effective inhibition of the other cysteine cathepsins: respective \(K_I\) values were 110 nM and 36 \(\mu\)M for CtsK and 120 nM and 110 nM for CtsL [156]. The crystal structures of two covalent inhibitors, vinyl sulfone inhibitor (4-morpholinecarbonyl-Leu-hPhe-\(\Psi\)(CH = CH-SO\(_2\)-Ph), Ph – phenyl) with P2-Leu and the aldehyde inhibitor (4-morpholinecarbonyl-Phe-hPhe-\(\Psi\)(CH = CH-SO\(_2\)-Ph) with P2-Phe, in a complex with CtsS were obtained and described [140]. According to structural data, the good packing of either P2-Leu or P2-Phe in the S2 pocket of CtsS is accomplished through the conformational adjustment of Phe211 (PDB ID: 1NPZ) [140]. CtsK contains a more rigid Leu209, while Ala is found at this position in CtsL and V, which is illustrated on Fig. 8. Another distinctive structural feature is Lys64 in the S3 subsite of CtsS. In the investigated complexes the partially negatively charged ether oxygen of the P3-morpholine ring is located nearby the NH\(_3^+\) group of Lys64 (PDB ID: 1NPZ) [140]. The positively charged Lys64 is substituted with an acidic residue in CtsK and CtsL and with a neutral but polar residue in CtsV (Fig. 8) [140].

N-Desmethyl thalassospiramide C covalent inhibitor was developed as a cysteine cathepsin inhibitor and its binding modes in different cathepsins have been described [157]. Biochemical research showed that the inhibitor is selective for CtsK and L (respective IC\(_{50}\) is 3 nM and 1 nM) and suppresses CtsB and S less effective (respective IC\(_{50}\) is 65 nM and 46 nM) [157]. The complex of the inhibitor with CtsK was crystallized and its detailed structural analysis by X-ray crystallography was conducted (PDB ID: 6HGY) [157]. The macrocycle of the inhibitor is anchored in the S1\(_0\) site of the enzyme and interacts with Asn161 and Gln19 (Fig. 9). The S1\(_0\) pocket of CtsL is similar to CtsK and there is no steric interference with the bound conformation of the inhibitor [158]. By contrast, Ser138Arg substitution in CtsS may hinder binding of the C11–C12 region of the inhibitor [159]. CtsS has an alternative fold in the S1\(_0\) site, which is unique in the cathepsin family [160]. This con-

---

### Table 4

| Inhibitors                                      | Cathepsins | \(K_I\)  | Ref.    |
|------------------------------------------------|------------|---------|---------|
| CA030                                          | CtsB       | 4.38 nM | [152]   |
| CA074                                          | CtsB       | 1.94 nM |         |
| Z-RK-AOMK                                      | CtsB       | 130 nM  | at pH 7.2|
| 4-morpholinecarbonyl-Leu-hPhe-\(\Psi\)(CH = CH-SO\(_2\)-Ph) | CtsS       | 13 nM   | [156]   |
| 4-morpholinecarbonyl-Phe-hPhe-\(\Psi\)(CH = CH-SO\(_2\)-Ph) | CtsS       | 18 nM   |         |
| \(\mu\)HLFRSAA\(_\mu\) (R\(_{\mu}\))             | CtsL       | 0.52 \(\mu\)M | [161]   |
| \(\mu\)HLFRSA\(_{\mu}\) (R\(_{\mu}\))            | CtsL       | 1.11 \(\mu\)M | [161]   |
| \(N\)-Desmethyl thalassospiramide C            | CtsK       | IC\(_{50}\) = 3 nM | [157]   |
|                                                | CtsL       | IC\(_{50}\) = 1 nM | [157]   |

---

Fig. 7. The crystal structure of the complex of CA030 (brown) bound to CtsB (gray surface; PDB ID: 1CSB) [100,101]. In PyMOL, the crystal structure of CtsX (PDB ID: 1EF7) was aligned to the structure of CtsB to compare the localization of the residues involved in interactions with the inhibitor [76,107]. The residues of CtsB are green and the residue of CtsX is black. His23 is illustrated in ‘closed’ position. Dashed line indicates possible non-covalent bond. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 8. Crystal structure of the complex of the vinyl sulfone inhibitor (brown) bound to CtsS (gray surface; PDB ID: 1NPZ) [100,140]. In PyMOL, the crystal structures of CtsK, L, and V (PDB IDs: 1ATK, 1CJL, and 1FH0, respectively) were aligned to the structure of CtsS to compare the localization of the residues involved in interactions with the inhibitor [103,107,126,127]. The residues of CtsS are orange, the residues of CtsK are dark green, the residues of CtsL are blue, and the residues of CtsV are beige. Dashed lines indicate possible non-covalent bonds. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
formational differences may account for the lower potency in CtsB and S (Fig. 9) [157].

Structural differences between cysteine cathepsins were used for the design of novel P1 thioamide peptide inhibitors selective to CtsL [161]. The obtained peptides are non-covalent reversible mixed-type inhibitors. While the primed positions of these peptides possess Ala at the P1', P2', and P3' positions, sequences for the non-primed side were designed based on substrate profiling study by Choe et al [66]. At the P1 position, similar to other human cathepsins CtsS prefers basic amino acids, but unlike CtsK, S, and B, CtsL is selective to the substrates with aromatic residues (Phe, Trp, Tyr) at P2 [61,66]. CtsV also favors Trp and Tyr, so Phe was chosen at P2 for CtsL. At P3, CtsL prefers basic and aliphatic residues, but since basic residue is preferred at P1, an aliphatic Leu (Leu is the most preferred aliphatic according to [66]) was chosen for P3 to avoid multiple cleavage sites. At P4, CtsL shows a preference for His. The P1 position was chosen for thioamid [162,163]. Obtained peptides μHLFR³AAAµ (R₁₃A) (µ - 7-methoxycoumarin-4-yl-alanine) and μHLFK⁵AAAµ (K₁₃A) were completely resistant to proteolysis by CtsV, K, and S, but cleaved by CtsL and B (μHLFK⁵AAAµ or μHLFR³-AAAµ) [161]. Shorter thiopeptide (μHLFR³µ; R₁₃A) was left intact by all investigated cathepsins: CtsL, V, K, S, and B. The thioamide peptides R₁₃A and R₃₈A were very good inhibitors of CtsL, with respective Kₐ values of 0.60 ± 0.15 µM and 0.52 ± 0.12 µM. The R₃₈A peptide was slightly worse with a Kₐ value of 1.11 ± 0.22 µM. The R₃₈A peptide exhibited a 26-fold increase in Kₐ for CtsV (26.22 ± 8.42 µM). No significant decrease in activities of CtsB, S, and K was observed in the presence of up to 50 µM of R₃₈A peptide [161]. Molecular docking with all five cathepsins showed that P1 thioamide bond N–H of the R₃₈A and R₁₃A peptides can interact with catalytic His163 exclusively in CtsL. Hydrogen bond prevents His163 from efficient deprotonation Cys25 thus inhibiting the proteolytic activity [161]. Finally, the R₁₃A peptide was shown to be able to inhibit CtsL in human liver carcinoma HepG2 lysates (IC₅₀ = 19 µM) [161].

A specific pH-dependent inhibitor was developed for CtsB [99]. At pH 7.2, CtsB prefers the basic residues Arg and Lys, and also Nle and Tyr at P1. At pH 7.2 CtsB possesses Glu245, so it favors the basic residues Arg, Lys, and His, as well as Trp at P2 [139]. At pH 4.6, CtsB prefers noncharged Thr and Gly residues, and the basic residue Arg at P1, while at P2, CtsB favors the substrates containing either the acidic Glu residue or hydrophobic Val. Glu245 can interact with the Glu residue and has no preference to the basic residues as at neutral pH. 7.2 Glu residues are uncharged [99]. CtsB had the highest ratio of pH 7.2/pH 4.6 activities towards Z-RK-AMC and the highest ratio of pH 4.6/pH 7.2 activities towards Z-EK-AMC among the investigated dipeptide-AMC substrates (Z-RK-AMC, Z-KR-AMC, Z-KR-AMC, Z-RR-AMC, Z-EK-AMC, and Z-ER-AMC). No CtsL and V activity was detected for Z-RK-AMC at pH 7.2 or Z-EK-AMC at pH 4.6. AMC group was replaced with the acryloylketone (AOMK) warfarehead to obtain the Z-RK-AOMK and Z-EK-AOMK peptide inhibitors [119,164]. Z-RK-AOMK displayed selective inhibition of CtsB at pH 7.2 compared to pH 4.6 with Kₐ values of 130 nM at pH 7.2 and 15,000 nM at pH 4.6. Z-EK-AOMK displayed less effective inhibition of CtsB at both pH levels with Kₐ values of 2,300 nM and 7,900 nM at pH 7.2 and pH 4.6, respectively. At pH 7.2, Z-RK-AOMK inhibited CtsB with IC₅₀ of 20 nM, which was more potent than CtsV, CtsS, and CtsC inhibition with respective IC₅₀ of 440 nM, 2,200 nM, and 850 nM. CtsK and H were minimally inhibited by Z-RK-AOMK at 16 µM. Computational modeling of Z-RK-AOMK binding to CtsB showed that the P2 Arg residue of Z-RK-AOMK has a strong polar interaction with the carbonylate of Glu245 in the S2 subsite of the enzyme. In the S1 binding pocket Lys residue of the inhibitor interacts with Glu122 and Asn72 of the enzyme [139]. Z-RK-AOMK was shown to inhibit CtsB at 1 µM concentration in human neuroblastoma cells SHSY5Y [99]. There are quite a few different types of PLCP inhibitors both covalent and non-covalent [165]. Their structures located in the protease binding cleft of the proteases might provide important information for improving their selectivity. Structural analysis of the complexes and sequence alignment of different PLCPs might be employed to design selective inhibitors for other proteases.

5. Plcps engineering for industrial applications

The information about the structural features of proteases provides a basis for ligand design as well as templates for engineering of the novel enzymes. The examples of enzyme redesign include an improved thermostability, tolerance to organic solvents, protein–protein interactions, altered substrate promiscuity/specificity, enhanced enzymatic activity, and inversion of enantioselectivity [166]. The obtained engineered enzymes can be of specific use in food, pharmaceuticals, biofuels and scientific applications [167]. Multiple plant PLCPs, especially papain, bromelain, and ficin, are often implemented in distinct industries [168], such as brewing,
meat softening, milk-clotting, cancer treatment, digestion and viral disorders [169–171]. Papain is one of the most commonly used proteases in the food industry [172,173]. Papain, as well as bromelain and ficin, are used to tenderize meat improving the sensory evaluation rating [174]. The application of PLCPs in the bakery industry may reduce risks of gluten intolerance since PLCPs digest allergenic gliadins [175]. In the textile industry PLCPs can be used to improve tactile behavior and wettability of wool and silk thus increasing the dye uptake [176,177]. Papain and bromelain as the ingredients in the cosmetic products digest the protein of dead cells in the upper layer of the skin alleviating wrinkles, acne, and dry skin [178]. Bromelain can also reduce bruising and swelling of skin after cosmetic injection treatments [179,180]. Papain is used in different formulations for chemomechanical caries removal [181,182] while bromelain and papain gel is used for deproteinization before orthodontic bracket attachment [183]. Though intact PLCPs are applied in numbers of industries, the need for new proteases with appealing physicochemical properties are emerging and protease reengineering may be the key. The importance of research in this area is underlined by the fact that the Nobel Prize in Chemistry was awarded to Frances H. Arnold in 2018 for developing the method for a directed evolution of enzymes [184]. However, there is still no published data on PLCP redesign harnessing the directed evolution. Directed evolution is used to simulate the process of natural evolution with the gene encoding the enzyme of interest being used to generate a library of mutants. After that, the enzyme variants with required properties are selected by an appropriate high-throughput screening method [167,185,186]. Typically, the cycles of mutagenesis and selection are repeated for several times. Most mutations in directed evolution occur outside the catalytic cleft and affect proteolytic activity only indirectly. Moreover, after introducing several mutations, improvements in activity often reach a plateau due to epistasis or stability-threshold effects [187]. In contrast, rational design is based on mutating specific residues in the protein structure [188,189]. Rational design is based on profound knowledge on the protein structure, function, and catalytic site. Alignment of multiple sequences and computational methods such as molecular docking and molecular dynamics simulation are used to analyze the enzyme structure [190]. The redesign is often based on a template enzyme with the innate activity for the target reaction [166]. As the libraries generated by rational design are generally smaller than the libraries obtained from a directed evolution, analyzing the obtained mutants does not require extensive screening [167]. However, chemical versatility that is available for a protein engineer is limited to a generally narrow range of reaction types within a superfamily [166]. Rational design was used to stimulate the catalytic activity of ervatamisin C (Erv-C), a PLCP from the plant *Ervatamia coronaria* [168]. Contrasted with Erv-A and Erv-B, Erv-C is highly stable in a wide range of pH (2–12), at a high temperature (70 °C) and a high concentration of chemical denaturants. Still, it possesses a significantly lower catalytic activity than Erv-A. The structures of Erv-A and Erv-C were compared to identify the mutations in the catalytic cleft that had to be introduced into Erv-C to enhance its proteolytic activity. Two single mutants (Ser32Thr and Ala67Tyr) and one double mutant of Erv-C (PDB ID: 2PNS) displayed ~8 times higher catalytic efficiency than WT with no effects on the thermal stability [168,191]. The obtained protease can be used as a biocatalyst with a potential for industrial applications. In order to identify the amino acid residues determining PLCPs’ specificity, there were several mutants designed. The papain with the Ser205Glu mutation in the S1 pocket digested Z-RR-AMC 100-fold more efficiently than WT [14]. However, mutations of PLCPs are mostly introduced into the most studied S2 subsite. The papain Ile86Leu mutant preferred branched hydrophobic residues at the P2 positions, while the Ile86Phe mutant preferred aromatic amino acids and Ile86Ala mutant demonstrated a decreased catalytic activity against the majority of the substrates (PDB ID: 3TNX) [192]. The gelatinolytic and caseinolytic activities of the mutants were reduced as well [120]. The Phe205Glu CtsS mutant had a lower affinity to the P2 hydrophobic residues and 77-fold increase in its activity towards the dipeptide RR. Adding one more mutation Gly133Ala (as in CtsB and L) promoted the digestion of RR [90]. The Glu245Gln CtsB mutant with the mutated S2 site did not digest substrates with P2 Arg in contrast to WT proteases and possessed unaltered activity towards the substrates with P2 Phe [2,151]. Similarly, Met205 in CtsF and Ala214 in CtsL can be replaced to Glu (as in CtsB) thus introducing specificity to positively charged residues at the P2 position of the substrates. Several mutants were created to study the collagenolytic activity of CtsK. The double CtsK mutant (Tyr67Leu and Leu209Ala as in CtsL) lacked Pro specificity and collagenolytic activity [13,146]. CtsL with the reverse mutations (Leu67Tyr and Ala214Leu) gained the ability to bind P2 Pro [147], but was not able to cleave collagen the way CtsK did [2,193]. However, the triple CtsL mutant (Leu67Tyr, Gly164Ala, and Ala214Leu) cleaved collagen similarly to CtsK [93]. It would be also interesting to evaluate the specificity at P2 to Pro and collagenolytic activity of the relative mutants of other cathepsins: in particular, triple mutant of CtsS (Phe67Tyr, Gly160Ala, and Phe205Leu), double mutant of CtsF (Leu67Tyr and Met205Leu; CtsF already possesses Ala160), and single mutant of CtsB (Glu245Leu; CtsB already possesses Tyr75 and Ala200). Rational design is being successfully employed to devise novel proteases with the desired specificity and improved stability. However, this method is based on the mutagenesis of the particular residues and requires identifying these residues. That is why it is so crucial to investigate structural features of the proteases and their impact in the specificity and activity of the enzyme.

### 6. Summary and outlook

PLCPs can perform a plethora of functions in different organisms inside and outside the cells cleaving a wide range of substrates. Some of them are highly specific for certain PLCPs, while the others can be digested by several proteases. Moreover, PLCP-mediated proteolysis is orchestrated by various parameters such as pH level, ions, GAGs, inhibitors, and cognateodomains. This flexible regulation is involved in maintaining cell homeostasis and any shift or disruption of this network may lead to various disorders, such as cancer. Thus, it is utterly important to reveal the features defining the specificity of PLCPs. Since the catalytic cleavage takes place inside the binding cleft, the amino acid residues that constitute the cleft are the key determinants of PLCPs specificity.

Most PLCPs prefer basic amino acid residues at the P1 position which can interact with the negatively charged amino acid of the S1 position and with the oxyanion hole. CtsS is not an exception from this rule but unlike the other cysteine cathepsins, it possesses Arg in its S1 subsite. This can be useful for development of the selective inhibitors even though the phenyl ring of the vinyl sulfone inhibitor creates a hydrophobic bond with the methylene of the Arg [140]. This might result from the specific local pKa of the residue; however, more evidence is needed. S2 binding pocket of PLCPs is generally hydrophobic favoring the interactions with substrates and inhibitors with a hydrophobic P2 residue. CtsK possesses a unique specificity to P2 Pro, which determines its ability to cleave the intact collagen triple helices. The artificial endoprotease CtsSΔEx can also accommodate P2 Pro [44]. Carboxypeptidases CtsB and CtsX are also distinct from other cathepsins since they possess charged amino acids at the S2 position. This differ-
ence determined the choice of positively charged amino acid residues at the P1 and P2 positions in the selective probe for CtsB Z-RR-AMC [67]. The specificity of the S3 binding pocket is very low. Most PLCPs can efficiently digest substrates with hydrophobic P3 but some of them expose charged amino acids as well. Some human cysteine cathepsins contain negatively charged (CtsF and L) or positively charged (CtsS and V) amino acids that can also be used in drug design. The primed side of PLCPs is formed by highly conservative Trp which may be involved in stacking interactions. However, CtsB, L, S, and V can accommodate hydrophilic amino acid residues at the S1’ binding site. The S2’ binding site in endopeptidases is hydrophobic while carboxypeptidases CtsB and X contain the positively charged His residues interacting with the C-terminal carboxylic group of the substrate. The existing data has already revealed several distinct structural features in the primed side of PLCPs. However, there are much less studies on this side of binding cleft than on the non-primed side [2]. Thus, the artificial substrates, probes, and inhibitors are mainly designed to interact with the non-primed side of the binding cleft. Still, designing elongated molecules may increase strength of binding.

Currently, only a few cysteine proteases were successfully redesigned. Still, enzyme engineering is employed mainly for increasing their stability. Most changes introduced into the structure of redesigned proteases are located outside the active site. This can be accounted for the overall focus on ensuring stability and high catalytic efficiency rather than specificity since proteases are mostly used in industry. The data on PLCPs structural features discussed in this review may prove useful for the protease engineering. Scrutinizing more PLCP structures could also provide valuable data that can further serve as a basis for the de novo synthesis of proteases.

CRediT authorship contribution statement

Anastasiia I. Petushkova: Writing – original draft, Visualization. Lyudmila V. Savvateeva: Writing – review & editing, Funding acquisition. Andrey A. Zamyatnin: Conceptualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by Russian Science Foundation (№ 22-25-00648).

References

[1] Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptides in the PANTHER database. Nucl Acids Res 2018;46:D624–32. https://doi.org/10.1093/nar/gky1114.
[2] Turk V, Stoka V, Vasileva G, Renkonen M, Sain T, Turk B, et al. Cysteine cathepsins: from structure, function and regulation to new frontiers. Biochim Biophys Acta 2012;1824:68–88. https://doi.org/10.1016/j.bbadis.2011.10.002.
[3] Richau KH, Kaschani F, Verdoes M, Pansuriya TC, Niessen S, Stüber K, et al. Lysosomal cathepsins and their regulation in aging and neurodegeneration. Ageing Res Rev 2016;32:22–37. https://doi.org/10.1016/j.arr.2016.04.010.
[4] Mischkowitz M, Ryan KM. Lysosomal proteins in cell death and autophagy. FEBS J 2015;282:1858–70. https://doi.org/10.1111/febs.13253.
[5] Turk B, Turk D, Turk V. Protease signalling: the cutting edge. EMBO J 2012;31:1630–43. https://doi.org/10.1038/emboj.2012.42.
[6] Olson OC, Joyce JA. Cysteine cathepsin proteases: Regulators of cancer progression and therapeutic response. Nat Rev Cancer 2015;15:712–29. https://doi.org/10.1038/nrc4027.
[7] Mohamed MM, Sloane BF. Cysteine cathepsins: multienzyme functions in cancer. Nat Rev Cancer 2006;6:764–75. https://doi.org/10.1038/nrc1940.
[8] Reiser J, Adair B, Reinhecht T. Specialized roles for cysteine cathepsins in health and disease. J Clin Invest 2010;120:3421–31. https://doi.org/10.1172/JCI43016.
[9] Lecaille F, Chazeirat T, Bojariski KK, Renault J, Saidi A, Prasov N, et al. Rat cathepsin K: Enzymatic specificity and regulation of its collagenolytic activity. Biochim Biophys Acta 2020;1866. https://doi.org/10.1016/j.bbadis.2020.11.002.
[10] Roy S, Das Chakraborty S, Biswas S. Not all pycnodysostosis-related mutants of human cathepsin K are inactive – crystal structure and biochemical studies of an active mutant 1249T. FEBS J 2018;285:4265–80. https://doi.org/10.1111/febs.14655.
[11] Dauth S, Sîrbulescu RF, Jordans S, Rehders M, Avena L, Oswald J, et al. Papain-like cysteine cathepsines as hubs in plant immunity. New Phytol 2016;212:983–1000. https://doi.org/10.1111/nph.14117.
[12] Souza DP, Fretas CDT, Pereira DA, Nogueira FC, Silva FDA, Salas CE, et al. Lysosomal cathepsins play a defensive role against hemibiotrophic and necrotrophic phytopathogens. Planta 2011;234:183–93. https://doi.org/10.1007/s00425-011-1392-1.
[13] Rehmi FHH, O’Rourke DJ, De Geyter E, O’Sullivan MD, Gilding EK, Durek T, et al. Papain-like cysteine cathepsines prepare plant cyclic peptide precursors for cyclization. Proc Natl Acad Sci USA 2019;116:7831–6. https://doi.org/10.1073/pnas.1901807116.
[14] Misas-Villamil JC, van der Hoorn RAL, Doehlemann G. Papain-like cysteine cathepsines as hubs in plant immunity. New Phytol 2016;212:902–7. https://doi.org/10.1111/nph.14117.
[15] Ettari R, Previti S, Tamborini L, Cullia G, Grasso S, Zappalà M. The Inhibition of Cysteine Proteases Rhodesian and TbcCatB: A Valuable Approach to Treat Human African Trypanosomiasis. Mini Rev Med Chem 2016;16:1274–91. https://doi.org/10.2174/138920303566610091125243.
[16] Ettari R, Previti S, Di Chio C, Zappalà M. Falcipain-2 and Falcipain-3 Inhibitors as Promising Antimalarial Agents. Curr Med Chem 2021;28:3010–31. https://doi.org/10.2174/0929867376666200701151316.
[17] Stoka V, Turk V, Sain T. Cysteine cathepsine protease from Toxocara canis infective larvae. Mol Biochem Parasitol 1998;92:275–89. https://doi.org/10.1016/S0166-5851(97)00245-4.
[18] Ettari R, Previti S, Tamborini L, Cullia G, Grasso S, Zappalà M. The Inhibition of Cysteine Proteases Rhodesian and TbcCatB: A Valuable Approach to Treat Human African Trypanosomiasis. Mini Rev Med Chem 2016;16:1274–91. https://doi.org/10.2174/138920303566610091125243.
[19] Loukas A, Selzer PM, Mazzeis RM. Characterisation of Tc-cpl-1, a cathepsin L-like cysteine protease from Toxocara canis infective larvae. Mol Biochem Parasitol 1998;92:275–89. https://doi.org/10.1016/S0166-5851(97)00245-4.
[20] Rehm FBH, Jackson MA, De Geyter E, Yap K, Gilding EK, Durek T, et al. Papain-like cysteine cathepsines prepare plant cyclic peptide precursors for cyclization. Proc Natl Acad Sci USA 2019;116:7831–6. https://doi.org/10.1073/pnas.1901807116.
[21] Donnelly S, Dalton JP, Robinson MW. How pathogen-derived cysteine proteases modulate host immune responses. Adv Exp Med Biol 2011;712:192–207. https://doi.org/10.1007/978-1-4419-8414-2_12.
[22] Stack C, Dalton JP, Robinson MW. The phylogeny, structure and function of trematode cysteine proteases, with particular emphasis on the Fasciola hepatica cathepsin L family. Adv Exp Med Biol 2011;712:116–35. https://doi.org/10.1007/978-1-4419-8414-2_8.
[23] Dana D, Pathak SK. A Review of Small Molecule Inhibitors and Functional Probes of Human Cathepsin L Molecules. Molecules 2020;25:698. https://doi.org/10.3390/molecules25030698.
A. Petushkova, L.V. Savvateeva and A.A. Zamyatnin Jr

Computational and Structural Biotechnology Journal 20 (2022) 6552–6569

Obermair J, Repnik U, Jevnikar Z, Turk B, Kreft M, Kos J. Cysteine protease cathepsin X modulates immune response via activation of beta2 integrins. Immunology 2008;124:76–88. https://doi.org/10.1111/j.1365-2567.2007.02740.x.

Vidmar R, Vizovichek M. Turk B, Tur L, Fönovč M. Protease cleavage fingerprinting by label-free in-gel degradomics reveals pH-dependent specificity switch of legumain. EMBO J 2017;36(24):2455–65. https://doi.org/10.1525/faseb.j.2013.04.0439.

Kafeshan W, Brömme D, Buttle DJ, Croucher IJ, Hollander AP. Human thyroid cathepsin K cleaves native type I and II collagens at the N-terminal end of the triple helix. Biochim Biophys Acta 1991;1031:71–9. https://doi.org/10.1016/0005-2728(91)90091-4.

Nosaoka AY, Kanaori K, Teno N, Togame H, Inaoka T, Takai M, et al. Conformational studies on the specific cleavage site of Type I collagen (alpha-1) fragment (157–192) by cathepsins K and L by proton NMR spectroscopy. Biochim Biophys Acta 1999;1457:433–43. https://doi.org/10.1016/S0006-8993(98)00237-2.

Sassi ML, Eriksen H, Risteli L, Niemi S, Mansell J, Gowen M, et al. Immunohistochemical characterization of assay for carboxylterminal telopeptide of human type I collagen: loss of antigenicity by treatment with cathepsin K. Bone 2006;36:267–73. https://doi.org/10.1016/j.bone.2006.07.005.

Biniossek ML, Nägler DK, Becker-Paul C, Schilling O. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. Proteome Res 2011;10:5363–73. https://doi.org/10.1021/pr200621z.

Felbor U, Dreier L, Bryant RA, Ploegh HL, Olsen BR, Mothes W. Secreted cathepsin L generates endostatin from collagen XVIII. EMBO J 2000;19:5761–9. https://doi.org/10.1002/1092-028X(20001117)19:21<5761::AID-EMBOJ10>3.0.CO;2-M.

Regmi SC, Samsom ML, Heynen ML, Jay GD, Sullivan BD, Srinivasan S, et al. Degradation of proteoglycan 4/lubricin by cathepsin S: Potential mechanism for diminished ocular surface lubrication in Sjögren’s syndrome. Exp Eye Res 2016;141:1–8. https://doi.org/10.1016/j.exer.2016.01.017.

Fosang AJ, Neame PJ, Last K, Harding JM, Murphy G, Hamilton JA. The intracellular toll-like receptors. Front Immunol 2013;4:299. https://doi.org/10.3389/fimmu.2013.00299.

Johne L, Schmutz K, Simonsen H, Schröder M, Schubert M, et al. Induction of CPP32 activity by multiple papain-like lysosomal cathepsins. J Biol Chem 2006;281:12824–32. https://doi.org/10.1074/jbc.M513301200.

Dunn AD, Crucifield HD. Trypsin-like cysteine protease (cathepsin X) modulates immune response via activation of beta2 integrins. J Immunol 2008;180:4691–7. https://doi.org/10.4049/jimmunol.180.8.4691.

Johne L, Ni J, Wu Y. Tumor tissues. Biochem J 1992;282:273–8. https://doi.org/10.1042/bj3310727.

Manoury B. Proteases: essential actors in processing antigens and intracellular toll-like receptors. Front Immunol 2013;4:299. https://doi.org/10.3389/fimmu.2013.00299.

Drake FH, Dodds RA, James IE, Connor JR, Debouck C, Richardson S, et al. Cathepsin K, but not cathepsins B, L, or S, is abundantly expressed in human osteoclasts. J Biol Chem 1996;271:12511–6. https://doi.org/10.1016/S0021-9927(96)00271-0.

Lecaile F, Brömme D, Lalmanach G. Biochemical properties and regulation of cathepsin K activity. Biochemistry 2008;90:208–26. https://doi.org/10.1016/j.biochem.2008.01.002.

Adams-Cioaba MA, Krupa JC, Xu C, Mort JS, Min J. Structural basis for the recognition and cleavage of histone H3 by cathepsin L. Nat Commun 2012;3:3179. https://doi.org/10.1038/ncomms4179.

Huang I-C, Hou W-S, Li Z, Büttner FH, Bartnik E, Brömme D, et al. Cleavage site specificity of recombinant human cathepsin K and development of selective and potent catalytic antibodies for the enzyme. EMBO J 2000;19:1187–94. https://doi.org/10.1093/emboj/19.6.1187.

Regmi SC, Samsom ML, Heynen ML, Jay GD, Sullivan BD, Srinivasan S, et al. Degradation of proteoglycan 4/lubricin by cathepsin S: Potential mechanism for diminished ocular surface lubrication in Sjögren’s syndrome. Exp Eye Res 2016;141:1–8. https://doi.org/10.1016/j.exer.2016.01.017.

Fonovic M, Turk B. Cysteine cathepsins and extracellular matrix degradation. Biochim Biophys Acta 2007;1768:1417–31. https://doi.org/10.1016/j.bbph.2007.04.009.

Huang I-C, Hou W-S, Li Z, Büttner FH, Bartnik E, Brömme D, et al. Cleavage site specificity of recombinant human cathepsin K and development of selective and potent catalytic antibodies for the enzyme. EMBO J 2000;19:1187–94. https://doi.org/10.1093/emboj/19.6.1187.

Huang I-C, Bosch BJ, Li F, Li W, Lee KH, Chirnan S, et al. SARS coronavirus, but not human coronavirus NL63, utilizes cathepsin L to infect ACE2-expressing cells. J Biol Chem 2006;281:3198–203. https://doi.org/10.1016/j.bjoc.2005.12.014.

Zhao M-M, Yang W-L, Yang F-Y, Zhang L, Huang W-J, Hou W, et al. Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development. Signal Transduct Target Ther 2021;6:134. https://doi.org/10.1038/s41392-021-00558-8.
[178] Trevisol TC, Henriques RO, Souza AJA, Furigo A. An overview of the use of proteolytic enzymes as exfoliating agents. J Cosmet Dermatol 2022;21:3300–7. https://doi.org/10.1111/jocd.14673.

[179] Emer J, Levy L. Complications of minimally invasive cosmetic procedures: Prevention and management. J Cutan Aesthet Surg 2012;5:121. https://doi.org/10.4103/0974-2077.99451.

[180] Arshad ZIM, Anis A, Yusof F, Jaswir I, Ahmad K, Loke SP. Bromelain: an overview of industrial application and purification strategies. Appl Microbiol Biotechnol 2014;98:7283–97. https://doi.org/10.1007/s00253-014-5889-y.

[181] Venkataraghavan K, Kush A, Lakshminarayana C, Diwakar L, Ravikumar P, Patil S, et al. Chemomechanical Caries Removal: A Review & Study of an Indigenously Developed Agent (Carie Care (TM) Gel) In Children. J Int Oral Heal JIOH 2013;5:84–90.

[182] Balakireva AV, Kuznetsova NV, Petushkova AL, Savvateeva LV, Zamiatnin Jr AA. Trends and Prospects of Plant Proteases in Therapeutics. Curr Med Chem 2019;26:465–86. https://doi.org/10.2174/0929867326666171123204401.

[183] Pithon MM, Campos MS, da S. Coqueiro R. Effect of bromelain and papain gel on enamel deproteinisation before orthodontic bracket bonding. Aust Orthod J 2016;32:23–30.

[184] Arnold FH. Innovation by Evolution: Bringing New Chemistry to Life (Nobel Lecture). Angew Chem Int Ed Engl 2019;58:14420–6. https://doi.org/10.1002/anie.201907729.

[185] Zeng W, Guo L, Xu S, Chen J, Zhou J. High-Throughput Screening Technology in Industrial Biotechnology. Trends Biotechnol 2020;38:888–906. https://doi.org/10.1016/j.tibte.2020.01.001.

[186] Madhavan A, Arun KB, Binod P, Sirohi R, Tarafdar A, Reshmy R, et al. Design of novel enzyme biocatalysts for industrial bioprocess: Harnessing the power of protein engineering, high throughput screening and synthetic biology. Bioresour Technol 2021;325.; https://doi.org/10.1016/j.biortech.2020.124617.

[187] Goldsmith M, Aggarwal N, Ashani Y, Jabran H, Greisen PJ, Ovchinnikov S, et al. Overcoming an optimization plateau in the directed evolution of highly efficient nerve agent bioscavengers. Protein Eng Des Sel 2017;30:333–45. https://doi.org/10.1093/protein/gxz003.

[188] Dadwal A, Sharma S, Satyanarayana T. Progress in Ameliorating Beneficial Characteristics of Microbial Cellulases by Genetic Engineering Approaches for Cellulose Saccharification. Front Microbiol 2020;11:1387. https://doi.org/10.3389/fmicb.2020.01387.

[189] Ali M, Isbqi HM, Husain Q. Enzyme engineering: Reshaping the biocatalytic functions. Biotechnol Bioeng 2020;117:1877–94. https://doi.org/10.1002/bit.27329.

[190] Ghosh R, Dattagupta JK, Biswas S. A thermostable cysteine protease precursor from a tropical plant contains an unusual C-terminal propeptide: cDNA cloning, sequence comparison and molecular modeling studies. Biochem Biophys Res Commun 2007;362:965–70. https://doi.org/10.1016/j.bbrc.2007.03.100.

[191] Roy S, Choudhury D, Aich P, Dattagupta JK, Biswas S. The structure of a thermostable mutant of pro-papain reveals its activation mechanism. Acta Crystallogr D Biol Crystallogr 2012;68:1591–603. https://doi.org/10.1107/0907444912019602.

[192] Cherney MM, Lecaille F, Kienitz M, Nallaseth FS, Li Z, James MMG, et al. Structure-activity analysis of cathepsin K/chondroitin 4-sulfate interactions. J Biol Chem 2011;286:8988–98. https://doi.org/10.1074/jbc.M110.126706.