

Review

Structure-function relationships in class CA1 cysteine peptidase propeptides[★]

Bernd Wiederanders[✉]

Institute of Biochemistry, Klinikum, Friedrich-Schiller-University Jena, Nonnenplan 2, D-07743 Jena, Germany

Received: 30 May, 2003; revised: 18 July, 2003; accepted: 29 August, 2003

Key words: papain family, cathepsin L-like peptidases, propeptide inhibition, processing, foldase

Regulation of proteolytic enzyme activity is an essential requirement for cells and tissues because proteolysis at a wrong time and location may be lethal. Proteases are synthesized as inactive or less active precursor molecules in order to prevent such inappropriate proteolysis. They are activated by limited intra- or intermolecular proteolysis cleaving off an inhibitory peptide. These regulatory proenzyme regions have attracted much attention during the last decade, since it became obvious that they harbour much more information than just triggering activation.

In this review we summarize the structural background of three functions of clan CA1 cysteine peptidase (papain family) propeptides, namely the selectivity of their inhibitory potency, the participation in correct intracellular targeting and assistance in folding of the mature enzyme. Today, we know more than 500 cysteine peptidases of this family from the plant and animal kingdoms, e.g. papain and the lysosomal cathepsins L and B. As it will be shown, the propeptide functions are determined by certain structural motifs conserved over millions of years of evolution.

Cysteine peptidases of clan CA family C1 (papain family) can be found in the animal and plant kingdoms as well as in some viruses and prokaryotes. Five hundred and fifty nine members of this peptidase family are actually registered in the respective tree of the

[★]Presented at the XXX Winter School of Faculty of Biotechnology, Jagiellonian University, Kościelisko, Poland, 28th February-4th March, 2003.

[✉]To whom correspondence should be addressed: tel.: (49 3641) 938 611; fax: (49 3641) 938 612; e-mail: bwie@mti.uni-jena.de

Abbreviations: BCP, *Bombyx* cysteine protease; BCPI, *Bombyx* cysteine protease inhibitor; DPPI, dipeptidyl peptidase I; ER, endoplasmic reticulum; GlcNAc, *N*-acetylglucosamine; MMPs, matrix metalloproteinases; PPIV, papaya protease IV. The suffix E70p means that this residue is located in the propeptide part of the sequence.

MEROPS protease database (<http://merops.sanger.ac.uk>) (last entry from June 12, 2003). The enzymes share their general architecture but also the micro-arrangement of the three catalytic residues Cys 25, His 159 and Asn 175 (according to the papain numbering). The ionized state of the nucleophilic cysteine residue in the active site is independent of substrate binding making these and other cysteine proteases *a priori* active (Polgar & Halasz, 1982). This catalytic mechanism is basically different from that of serine proteases whose serine residue in the catalytic triad becomes ionized only upon substrate binding.

Eukaryotic papain family peptidases comprise three parts: an N-terminal signal sequence (10–20 amino acids) is followed by the prosequence (between 38 and 250 amino acids), the third part represents the mature enzyme, generally 220–260 amino acids long. The tertiary structure of the enzyme part is characterized by two domains (R and L) of comparable size with the active site cleft in between. The catalytic site is already preformed in the precursor. It is localized at the bottom of the active site cleft and involves the three residues mentioned above. For more details see the recent review by McGrath (1999).

Most CA1 peptidases act as endopeptidases. Some peculiarities and exceptions from this general rule can be explained by structural details of the catalytic domains as the “occluding loop” in cathepsin B which favours the binding of protein C-termini thus enabling its peptidyl dipeptidase activity (Musil *et al.*, 1991), the Cys 331 of cathepsin C which is necessary for tetramerization and thus for dipeptidyl peptidase activity (Horn *et al.*, 2002), or the “mini-chain” of cathepsin H anchoring the positively charged amino group of the substrate N-terminus (Guncar *et al.*, 1998). Reasons of other features, such as the carboxypeptidase activity of cathepsin X, the lack of the activation of procathepsin W or the functions of C-terminal extensions of parasite derived enzymes still remain to be elucidated.

The functions of CA1 peptidases are different in various organisms. Only few examples of viral, prokaryotic and yeast CA1 peptidases are listed in the MEROPS protease database. There is only little knowledge of their physiological functions.

Plant proteinases of this class are mainly used to mobilize storage proteins in seeds. Protein bodies of seeds contain both storage proteins and protease precursors. The latter become activated after germination and start degradation of the stored proteins (Schlereth *et al.*, 2001). Some of these enzymes have medical significance because they are resorbed in the gut as active enzymes and exert an immunogenic potential (Furmona-viciene *et al.*, 2000; Nettis *et al.*, 2001).

Most parasitic cysteine peptidases act extracellularly. They help the parasites to invade tissues and cells, to gain nutrients, to hatch, to enter and to leave cysts, or even to evade the host immune system. For details see the review by Sajid & McKerrow (2002). Primitive organisms depending on phagocytosis use cysteine proteases to digest phagocytosed proteins. The enzymes of these organisms are already packed in lysosomes or acidified lysosome-like structures (Volkel *et al.*, 1996; Krasko *et al.*, 1997; Gotthardt *et al.*, 2002).

Mammalian CA1 cysteine peptidases are considered as primarily lysosomal enzymes. Only cathepsin W seems to be retained in the endoplasmic reticulum (ER) (Wex *et al.*, 2001). Some cathepsins are found in nearly all tissues and cells (cathepsins B, C, H, L, O), thus probably fulfilling housekeeping functions. Others show a restricted organ distribution (cathepsins S, K, V, F, X, W) suggesting specific functions. Recent information about cathepsin functions came from modern genetic approaches as mutational analyses and gene knock out animals. A recent review of the physiological and pathological roles of mammalian and parasitic CA1 peptidases is recommended for interested readers (Lecaille *et al.*, 2002).

PROPEPTIDE INHIBITION

Structural background

The inhibition of CA1 cysteine proteases by their respective propeptide parts was first observed by Fox *et al.* (1992). They observed a strong inhibition of cathepsin B by a 56 amino acids long synthetic peptide corresponding to residues -62 to -7 of rat liver procathepsin B. The inhibition was pH dependent. At pH 6.0, the inhibition was a slow binding one step reaction, whereas the inhibition at pH 4.0 followed the classical scheme. The propeptide was also slowly degraded by the enzyme at pH 4.0. The authors suggested a loose complex between the pro and the mature domain at acidic pH.

Taylor *et al.* (1995a) expressed the pro-regions of two proteases from *Carica papaya*, papain and PPIV (papaya protease IV), as recombinant proteins in *Escherichia coli* and studied the inhibitory activity of the peptides toward papain, caricain, chymopapain and PPIV. They found different K_i values being three orders of magnitudes higher for PPIV than for the others. They discussed this selectivity for the first time on the basis of structural differences. Numerous reports by different groups confirmed later the general fact that the inhibitory propeptide parts are regulatory elements of cysteine protease activity (Volkel *et al.*, 1996; Maubach *et al.*, 1997; Visal *et al.*, 1998; Guay *et al.*, 2000; Billington *et al.*, 2000).

The structural background of this inhibition was elucidated by X-ray structure analyses. Papain was the first mature cysteine protease whose structure was published (Drenth *et al.*, 1968). A 1.65 Å resolution revealed later a two domain fold of papain with the active site located in a groove between the two domains (Kamphuis *et al.*, 1984). This particular feature is highly conserved amongst cysteine proteases of this type in all kingdoms (Musil *et al.*, 1991; McGrath *et al.*, 1995; Roche *et al.*, 1999).

The first structures of cysteine protease precursors were published by Cygler *et al.* (1996), Turk *et al.* (1996) and Coulombe *et al.* (1996). Their studies revealed how the propeptide is attached to the mature enzyme (Coulombe *et al.*, 1996). The most striking result was the elucidation of the mode of propeptide inhibition. The authors clearly showed that the propeptide covers the active site cleft in a non-productive orientation. The S subsite is occupied by the C-terminal residues (e.g. Gly77p, Leu78p and Gln79p in procathepsin L) whereas the S' subsites are mainly occupied by the N-terminal residues. Such an orientation does not allow the hydrolysis of the peptide bond, however, the tightly bound molecule hinders the access of substrate molecules to the active site. This mode of inhibition is found in all class CA1 cysteine peptidases whose zymogen structures have been resolved.

The propeptides contain some characteristic elements which are highly conserved in evolution. Karrer *et al.* (1993) compared the N-terminal amino-acid sequences of 15 cysteine protease zymogens. They found a consensus sequence known as ERFNIN motif present in the $\alpha 2$ helix of a great number of cysteine protease propeptides of numerous species, including *Tetrahymena*. However, in cathepsin B, the $\alpha 2$ helix is much shorter and does not contain the ERFNIN motif. On this basis, the authors defined two subfamilies of cysteine proteases, the cathepsin L like containing the ERFNIN motif, and the cathepsin B like lacking this motif. In cathepsins F (Wang *et al.*, 1998; Nagler *et al.*, 1999) and W (Linnevers *et al.*, 1997; Brown *et al.*, 1998; Wex *et al.*, 1998), the Ile and Asn residues of the ERFNIN motif are replaced by Ala and Gln, defining a third subgroup of cysteine proteases characterized by the ERFNAQ motif, called cathepsin F like subgroup (Wex *et al.*, 1999).

Another highly conserved motif is the GxNxFx₂D heptapeptide motif (GNFD in brief) which can also be found in most of the cys-

teine protease propeptides. The motif is located at the kink of the β -sheet immediately before the chain builds the third helix turning down into the active site cleft. The Asp residue in the GNFD motif seems to be essential for the correct processing of the protease precursors since replacement of this Asp by Asn, Tyr, Met, Val or Glu resulted in non functional papain mutants (Vernet *et al.*, 1995).

Procathepsin F contains a propeptide which is more than twice as long as the propeptides of a closely related ERFNAQ subfamily member, cathepsin W. Besides this distinctive feature it shows a cystatin-like domain which makes the propeptide of cathepsin F unique amongst the CA1 cysteine peptidases (Nagler *et al.*, 1999). It is still unclear whether or not the cystatin-structure element contributes much to the inhibitory potency of the cathepsin F propeptide *in vivo*.

Inhibition type and constants

Human papain like proteases are involved in a variety of pathological processes, such as malignant tumour invasion and chronic destructive processes. Moreover, proteases of this type are also important for the life cycle and infectivity of parasitic pathogens like *Fasciola hepatica*, *Trypanosoma* species and others. Therefore, enzymes of this class represent attractive targets for the development of therapeutic inhibitors. However, since the proteases play important roles in normal protein turnover and protein processing, and due to the broad substrate specificity of cysteine proteases, the development of inhibitors with high selectivity is a great challenge.

The structural details of the interaction between the propeptides of cysteine proteases and their cognate enzymes suggest a selectivity of inhibitory propeptide action. Kinetic constants of this inhibition have been studied by several groups. The selectivity has various reasons and it is by far not as good as might be expected. Nevertheless, the studies were the basis for the development of peptide derived

low M_r inhibitors mentioned below. Table 1 shows details of kinetic experiments in which the propeptides act in *trans*.

An interesting observation was the existence of proteins with homology to the propeptide regions of cysteine proteases, however, expressed independently from the co-expressed enzyme. The occurrence of such proteins has been reported in activated mouse T lymphocytes and mast cells, they were named CTLA-1 and -2 (Denizot *et al.*, 1989; Delaria *et al.*, 1994), and in the hemolymph of the silkworm *Bombyx mori*, named BCPI α and β (*Bombyx* cysteine protease inhibitor (Yamamoto *et al.*, 1999a; 1999b; Kurata *et al.*, 2001; Yamamoto *et al.*, 2002). Moreover, a search in the SwissProt databank revealed the occurrence of further sequences homologous to CTLA-2 in rat (R/CTLA-2) as well in the *Drosophila* genome (D/CTLA-2). Whereas the kinetic constants of BCPIs and CTLAs have been measured, the rat and the *Drosophila* proteins have not been characterized so far. Yamamoto *et al.* (2002) suggested the origin of this class of inhibitory peptides by partial gene duplication of an ancestor cysteine protease gene thus leading to a new class of endogenous inhibitors without relation to cystatins and other endogenous inhibitors. A biological role of these proteins has not been reported so far.

Selectivity of inhibition

Selectivity of propeptide inhibition has to consider two different meanings: 1st how selective is an effect between two members of the family within a given species (e.g. human cathepsin L *vs* human cathepsin K), and 2nd how selective is the effect in an interspecies comparison (e.g. human cathepsin L *vs* *F. hepatica* cathepsin L). This is important if propeptide derived inhibitors should be taken as leading structures for the development of pharmaceuticals. Numerous experiments have been performed to characterize the type of propeptide inhibition toward different cysteine peptidases, moreover, truncated and

Table 1. Inhibition constants of cysteine protease propeptides toward different proteases, either the cognate or a non-cognate enzyme, using low MW substrates

Propeptide species	Truncated/full length	Target enzyme	pH	Substrate	K_i [nM]	Reference
Hum Cat L	full	Cat L	5.5	Z-LR-AMC	0.12	Guay <i>et al.</i> (2000)
	full	Cat K	5.5	Z-LR-AMC	0.27	Guay <i>et al.</i> (2000)
	full	Cat S	5.5	Z-LR-AMC	65	Guay <i>et al.</i> (2000)
	full	Cat B	5.5	Z-LR-AMC	>600	Guay <i>et al.</i> (2000)
	full	Papain	5.5	Z-LR-AMC	>600	Guay <i>et al.</i> (2000)
Hum Cat K	full	Cat K	5.5	Z-LR-AMC	5.5	Guay <i>et al.</i> (2000)
	full	Cat L	5.5	Z-LR-AMC	3.6	Guay <i>et al.</i> (2000)
	full	Cat S	5.5	Z-LR-AMC	6.3	Guay <i>et al.</i> (2000)
	full	Cat B	5.5	Z-LR-AMC	>600	Guay <i>et al.</i> (2000)
	full	Papain	5.5	Z-LR-AMC	>600	Guay <i>et al.</i> (2000)
Hum Cat S	full	Cat S	5.5	Z-LR-AMC	7.6	Guay <i>et al.</i> (2000)
	full	Cat L	5.5	Z-LR-AMC	0.46	Guay <i>et al.</i> (2000)
	full	Cat K	5.5	Z-LR-AMC	7.0	Guay <i>et al.</i> (2000)
	full	Cat B	5.5	Z-LR-AMC	>600	Guay <i>et al.</i> (2000)
Hum Cat S	full	Cat S	6.5	Z-VVR-AMC	0.27	Maubach <i>et al.</i> (1997)
	full	Rat Cat L	6.5	Z-FR-AMC	0.36	Maubach <i>et al.</i> (1997)
	full	Par. Cat L	6.5	Z-FR-AMC	0.37	Maubach <i>et al.</i> (1997)
Hum Cat S	full	Cat S	6.5	Z-VVR-AMC	0.049	Guo <i>et al.</i> (2000)
	full	Cat L	6.5	Z-FR-AMC	0.079	Guo <i>et al.</i> (2000)
P.c. Cat L*	full	Cat L	6.5	Z-VR-AMC	20.9	Guo <i>et al.</i> (2000)
	full	Cat S	6.5	Z-FR-AMC	>150	Guo <i>et al.</i> (2000)
Hum Cat K	full	Cat K	6.0	Z-FR-AMC	2.61	Billington <i>et al.</i> (2000)
	full	Cat K	4.0	Z-FR-AMC	147	Billington <i>et al.</i> (2000)
	full	Cat L	6.0	Z-FR-AMC	26.1	Billington <i>et al.</i> (2000)
	full	Cat B	6.0	Z-FR-AMC	> 400	Billington <i>et al.</i> (2000)
B.m. CP **	full	BCP	5.25	Z-FR-AMC	0.0059	Kurata <i>et al.</i> (2001)
	full	Cat L	5.5	Z-FR-AMC	0.036	Kurata <i>et al.</i> (2001)
	full	Cat H	6.8	R-AMC	82	Kurata <i>et al.</i> (2001)
	full	Cat B	5.5	Z-RR-AMC	> 4000	Kurata <i>et al.</i> (2001)
	full	Papain	6.0	Z-FR-AMC	> 1000	Kurata <i>et al.</i> (2001)
Synth. Cat B	full	Cat B	6.0	Z-FR-AMC	0.4	Fox <i>et al.</i> (1992)
	full	Cat B	4.7	Z-FR-AMC		Fox <i>et al.</i> (1992)
	full	Cat B	4.0	Z-FR-AMC	64	Fox <i>et al.</i> (1992)
	full	Papain	6.0	Z-FR-AMC	5600	Fox <i>et al.</i> (1992)
	full	Papain	4.0	Z-FR-AMC	2800	Fox <i>et al.</i> (1992)
CTLA-2 β	full	Cat L	5.5	Z-FR-AMC	24	Delaria <i>et al.</i> (1994)
	full	Cat B	5.5	Z-FR-AMC	> 1000	Delaria <i>et al.</i> (1994)
	full	Cat H	6.8	R-AMC	67(IC ₅₀)	Delaria <i>et al.</i> (1994)
	full	Papain	7.4	Z-FR-AMC	25	Delaria <i>et al.</i> (1994)
Fasc. hep. Cat L	full	FC CL1	6.5	Bz-VLK-AMC	1.73	Roche <i>et al.</i> (1999)
	full	FC CL2	6.5	Bz-VLK-AMC	12.0	Roche <i>et al.</i> (1999)
	full	Cat L	5.5	Z-FR-AMC	> 10000	Roche <i>et al.</i> (1999)
	full	Papain	6.0	Z-FR-AMC	> 10000	Roche <i>et al.</i> (1999)
	full	Cat B	5.5	Z-RR-AMC	> 10000	Roche <i>et al.</i> (1999)
HumCat L	full	Cat K	5.5	Bz-FVR-AMC	> 10000	Roche <i>et al.</i> (1999)
	4p-90p	Cat L	5.5	Z-FR-AMC	0.088	Carmona <i>et al.</i> (1996)
	1p-81p	Cat L	5.5	Z-FR-AMC	0.66	Carmona <i>et al.</i> (1996)
	21p-95p	Cat L	5.5	Z-FR-AMC	11.5	Carmona <i>et al.</i> (1996)
	52p-95p	Cat L	5.5	Z-FR-AMC	2900	Carmona <i>et al.</i> (1996)
	4p-90p	Cat S	6.0	Z-FR-AMC	44.6	Carmona <i>et al.</i> (1996)
	21p-95p	Cat S	6.0	Z-FR-AMC	>1000	Carmona <i>et al.</i> (1996)
	4p-90p	Cat B	6.0	Z-FR-AMC	>1000	Carmona <i>et al.</i> (1996)
	21p-95p	Cat B	6.0	Z-FR-AMC	>1000	Carmona <i>et al.</i> (1996)
	4p-90p	papain	6.0	Z-FR-AMC	>1000	Carmona <i>et al.</i> (1996)
CatS/P.c.CatL chimera *	S1-54 / L44-86	Cat S	6.5	Z-VR-AMC	0.1	Guo <i>et al.</i> (2000)
		Cat L	6.5	Z-FR-AMC	1.4	Guo <i>et al.</i> (2000)
		Cat S	6.5	Z-VR-AMC	>150	Guo <i>et al.</i> (2000)
CatS/P.c.CatL chimera*	L1-43 / S55-98	Cat L	6.5	Z-FR-AMC	13.6	Guo <i>et al.</i> (2000)
		Cat S	6.5	Z-VR-AMC	>150	Guo <i>et al.</i> (2000)
Rec BCPI	Full	BCP	5.25	Z-FR-AMC	0.11	Kurata <i>et al.</i> (2001)
	$\Delta 2$	BCP	5.25	Z-FR-AMC	0.38	Kurata <i>et al.</i> (2001)
	$\Delta 6$	BCP	5.25	Z-FR-AMC	2.4	Kurata <i>et al.</i> (2001)
	$\Delta 10$	BCP	5.25	Z-FR-AMC	no inh.	Kurata <i>et al.</i> (2001)

*P.c. *Paramecium caudatum*; **B.m. *Bombyx mori*

chimeric propeptides were involved in these studies in order to determine which part of the propeptide is of importance for selectivity.

The results can be summarized as follows:

- ◆ Mammalian cathepsin propeptides with a complete $\alpha 2$ helix are poor inhibitors of cathepsin B and of papain (Delaria *et al.*, 1994; Carmona *et al.*, 1996; Guay *et al.*, 2000; Billington *et al.*, 2000; Kurata *et al.*, 2001);
- ◆ The cathepsin B propeptide is a poor inhibitor of papain (Fox *et al.*, 1992);
- ◆ The species differences between inhibitory propeptide and enzyme do not follow a general rule (Roche *et al.*, 1999; Guo *et al.*, 2000; Kurata *et al.*, 2001);
- ◆ There is only little selectivity of propeptide inhibition between cathepsins S, L and K (Guay *et al.*, 2000);
- ◆ N-terminal truncation of the propeptide reduces the inhibitory potency more than truncation at the C-terminus (Carmona *et al.*, 1996; Kurata *et al.*, 2001);
- ◆ The N-terminal part of the propeptide shows a greater influence on selectivity than the C-terminal part (Guo *et al.*, 2000).

Some of the findings from different laboratories are not compatible, e.g. the inhibition of cathepsin L by the propeptide of cathepsin S showed a difference in K_i of one order of magnitude (Maubach *et al.*, 1997 and Guay *et al.*, 2000) *vs* (Guo *et al.*, 2000). Most of the propeptides are very hydrophobic. We observed a remarkable adsorption of the propeptides to the surface of assay tubes. This resulted in incorrectly high K_i values (Maubach *et al.*, 1997) which we later corrected on the basis of a more careful analysis (Guo *et al.*, 2000). Some of the deviating data may be caused by such experimental details, but others may also be explained by different substrate and pH conditions as the authors already discussed (Guay *et al.*, 2000).

Guay *et al.* (2000) discussed the inhibition of cathepsins S, K and L *vs.* papain by the respective propeptides on the basis of homologies of the propeptide sequences between Thr55p

and Leu78p (procathepsin K numbering) which cover the active site cleft. Whereas the cathepsins S, K and L share 10 of the 24 amino acids, papain shares only 4, and papain shows an additional insertion between the residues binding to S1 and S2 (Guay *et al.*, 2000).

From some data one might speculate that propeptides of cysteine peptidases have acquired increasing selectivity and increasing inhibitory potency during evolution:

- ◆ Mammalian propeptides are by far the best inhibitors of their cognate enzymes in comparison to the propeptide mediated inhibition of *Paramecium tetraurelia* or plant enzymes (Taylor *et al.*, 1995; Roche *et al.*, 1999; Guo *et al.*, 2000);
- ◆ The *Paramecium tetraurelia* cathepsin L propeptide inhibits the parental enzyme in a two step reaction (Guo *et al.*, 2000) (see Scheme 2 below and the respective discussion). This is a unique observation in the papain family up to now;
- ◆ The K_i values of the inhibition of plant enzymes by their propeptides have been deduced from steady-state measurements. Pre-steady-state was not recorded at all (Taylor *et al.*, 1995).

Measuring the pre-steady-state inhibition kinetics of plant peptidases by their respective propeptides would be the key experiment in order to support the optimisation hypothesis.

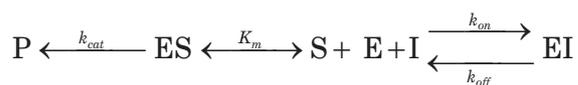
A closer view on the active site of cathepsins L, K and S shows that the two bulky aromatic residues Phe63p and Phe71p of the cathepsin L propeptide do not fit well into the respective binding pocket of cathepsin S due to steric hindrance by Phe146 (cathepsin S). In cathepsin L, Leu144 is in this position which allows a tight contact of the propeptide with the mature enzyme. This may explain the differences in K_i of cathepsin L propeptide inhibition toward cathepsins L and S (Guay *et al.*, 2000; Guo *et al.*, 2000). The propeptide of cathepsin K shows in this position two branched amino acids, Leu63p and Val71p. The hydrophobic interactions of these residues with Phe146 and Leu144 in cathepsins S

and L, respectively, are obviously much more tight than with Gln143 in the cognate cathepsin K. The different kinetic constants for this inhibition detected by two independent groups may be explained by the different conditions: whereas Guay *et al.* (2000) determined the activity at pH 5.5 toward Z-LR-AMC, the activity of cathepsins L and K was measured with Z-FR-AMC at pH 6.0 (Billington *et al.*, 2000).

In cathepsin B, the occluding loop prevents the propeptide of other cathepsins from binding to the active enzyme.

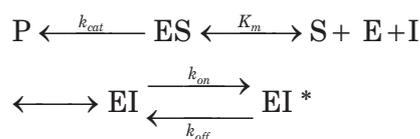
The electrostatic surface potentials of papain, various cathepsins and their propeptides reveal significant differences amongst them (see Fig. 1). Thus, the overall surface charge besides the atomic interactions may in some cases also contribute to the selectivity of propeptide inhibition.

Two different types of inhibition can be observed when propeptides are incubated with various enzymes. The general mechanism by which the propeptides react with their cognate peptidases is tight binding. It is a one step reaction according to Scheme 1.



Scheme 1

Another type of inhibition can also be observed in which the enzyme first forms a loose contact to the propeptide (Scheme 2). The initial contact induces in both the propeptide and the enzyme a conformational shift resulting in k_{off} values of the complexes resembling those of the cognate propeptide/enzyme pairs. This was observed with mutants destabilizing the mini domain between helices $\alpha 1$ and $\alpha 2$ in procathepsin S (Schilling *et al.*, 2001) and also for the inhibition of cathepsin H by the cathepsin L propeptide (own unpublished result).



Scheme 2

Propeptide derived peptidomimetic inhibitors

The development of protease inhibitors is an obvious challenge, either for the treatment of diseases whose progress is caused by uncontrolled protease activity, or for intervention into the life cycle of parasites. Most inhibitors blocking the active site of a protease react covalently. They are directed to the catalytic cysteine residue common within all members of the family – the basis of the clan classification – and act, therefore, with low selectivity. However, the CA1 cysteine peptidase precursors provide examples of non-covalent inhibition with relatively high selectivity as mentioned above, and there are some approaches to use their structural features for the development of effective medicine. Furthermore, modifications stabilizing the peptide moieties against degradation are also necessary, as e.g. introduction of D-amino acids.

Carmona *et al.* (1996) used already truncated forms of recombinant cathepsin L propeptides in order to study the contributions of defined segments to the inhibitory potency. However, truncation led to a dramatic loss of affinity suggesting that extensive molecular contacts are necessary for efficient inhibition. Chowdhury *et al.* (2002) carefully analysed the binding mode of the pentapeptide moiety MNGFQ (residues 75p to 79p of the cathepsin L propeptide) spanning the active site groove of cathepsin L from the S2' to S3 subsites in order to synthesize an optimized non-covalent selective cathepsin L inhibitor. They ended up with a series of compounds, the best of which showed a K_i of 19 nM toward cathepsin L, the K_i values toward

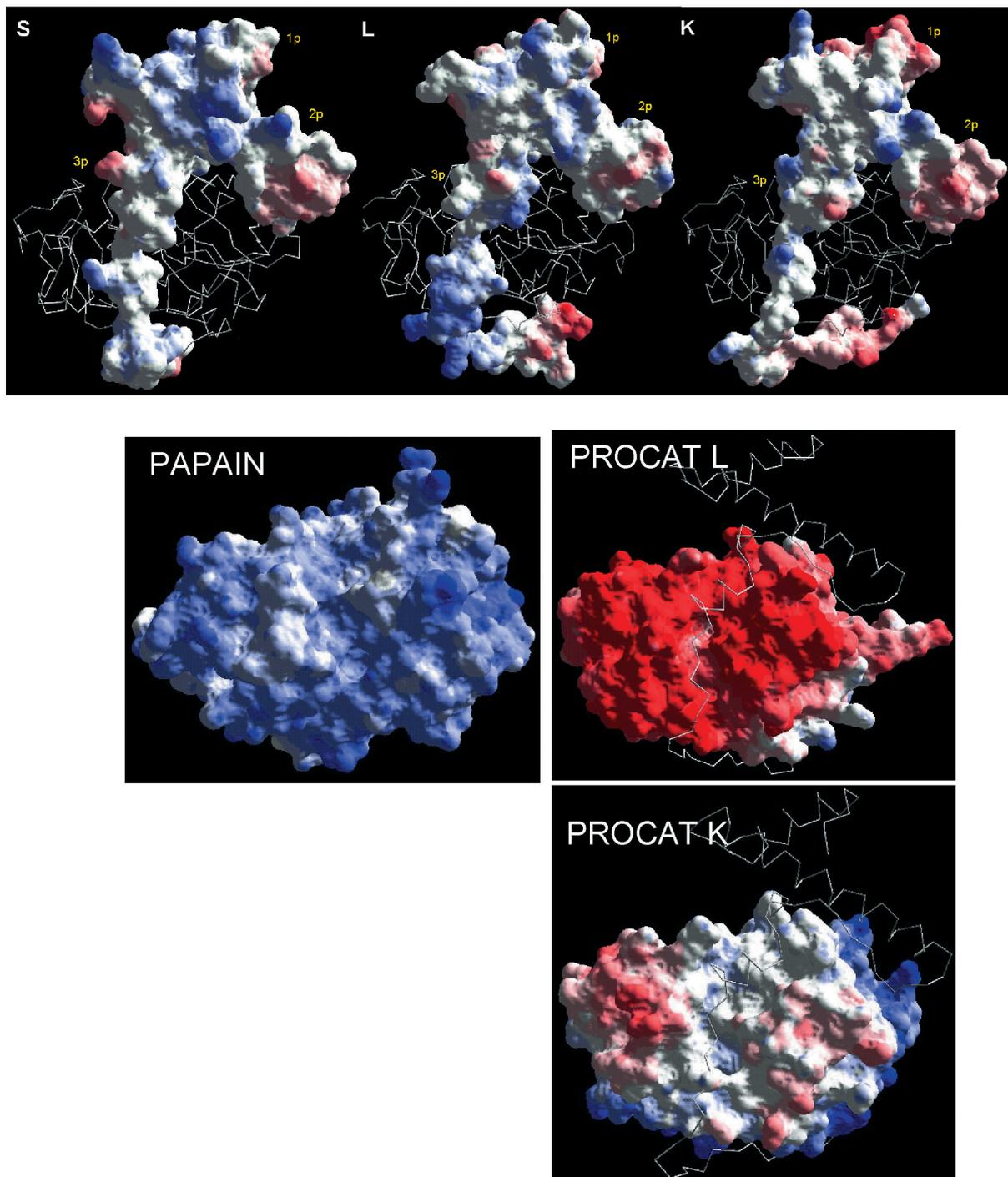


Figure 1. Model of electrostatic potentials on the surfaces of three cysteine protease propeptides and of three mature cysteine proteases.

The models were calculated by Swiss-PdbViewer (Guex & Peitsch, 1997) using atomic partial charges for each molecule. The colour of the electrostatic potential maps ranges from red (-8 kT/mol) *via* white (neutral) to blue ($+8$ kT/mol). **Upper row:** propeptides of cathepsins S (Kaulmann, PhD thesis), K (PDB entry 1BY8), and L (PDB entry 1CS8). The mature enzymes are shown as thin white sequence chains, i.e. the model does not show the contact area with the mature enzyme. **Lower rows:** papain (PDB entry 9PAP), procathepsin L (PDB entry 1CS8) and procathepsin K (PDB entry 1BY8). View on the active-site cavity and on the contact area of the mature enzyme with the propeptide. The proparts of cathepsins K and L are shown as thin white sequence chains.

cathepsins K and B being 5.9 and 4.1 μM , respectively, which is remarkable since the full length propeptide of cathepsin L shows only a 2-fold better selectivity toward cathepsin K.

The propeptide of cathepsin B was also systematically truncated from the N- as well as from the C-terminus in steps of 5 amino acids, and the change of K_i toward cathepsin B was recorded. The stepwise truncation from either the N- or the C-terminus increased the K_i moderately (Chen *et al.*, 1996).

Schaschke *et al.* (1998) in a very sophisticated approach used the Leu-Gly-Gly (44p-46p of cathepsin B propeptide) motif to develop selective *trans*-epoxysuccinyl derivative based cathepsin B inhibitors. This sequence binds in the anti-substrate orientation to the S subsite thus mimicking the binding mode of the propeptide, whereas the Leu-Pro-OH moiety occupies the S' subsite. The inhibitor MeO-GGL-*trans*-epoxysuccinyl-LP-OH showed remarkable selectivity of cathepsin B inhibition: the k_2/K_i ratios (which is the apparent second-order rate constant calculated from $k_{\text{obs}}/[I]$ for $[I] \ll K_i$) of cathepsins B and L *versus* papain inhibition were 103 and even 1262, respectively.

Truncation of recombinant *Bombyx mori* cysteine protease inhibitor (BCPI) resulted in a loss of inhibitory activity, but here, the C-terminus seemed to be responsible for efficient inhibition, because removal of the 10 C-terminal amino acids completely abolished the inhibition toward the cognate BCP (Kurata *et al.*, 2001).

The parasite cysteine proteases cruzipain and congopain are also targets for the development of selective inhibitors since the parasites cause severe human and animal diseases in tropical regions. A systematic study of 23 overlapping 15-mer peptides and peptide amides, respectively, mimicking the active site-covering sequences of the propeptides resulted in four competitive inhibitors of congopain, cruzipain and recombinant cruzain. The peptides did not significantly inhibit cathepsins B and L (K_i above 100 μM) and

thus showed a satisfying degree of selectivity. However, the best K_i values were around 5 μM and, therefore, were too high to make these peptides efficient inhibitors (Lalmanach *et al.*, 1998). Nevertheless, the pentapeptide YHN-GA was present in all four peptides and constituted an essential element for selectivity and inhibition of parasite proteases and, therefore, was discussed as a promising leading structure for the development of anti-parasitic therapeutic drugs.

It must be mentioned that the development of substrate derived low M_r inhibitors resulted in very promising substances. They represent a much greater group which, however, will not be reviewed here.

PROCESSING

Here, we summarize the current knowledge about the mechanistic and general aspects of the cysteine protease precursor activation processes. The processing of protease precursors into active enzymes includes two steps, the activation of the zymogen and one or more limited proteolytic cleavages within the polypeptide backbone as well as at the N- and C-termini, respectively (Mach *et al.*, 1994, Menard *et al.*, 1998). If this is catalysed by the molecule itself, it is called intramolecular or unimolecular. However, if it is catalysed by a second molecule it should be named intermolecular or bimolecular (Rowan *et al.*, 1992). This clarity of nomenclature seems to be necessary because the term intramolecular processing is sometimes incorrectly used when a purified recombinant protease precursor is self-processing just by a shift from neutral to acidic pH.

Some authors discriminate two steps in the intramolecular activation of a protease precursor molecule: the first makes the scissile bond accessible, mostly by a conformational shift, and the second leads to the – at least transient – removal of the inhibitory propeptide (Menard *et al.*, 1998). If a bimolecular

process is considered, the scissile bond(s) is/are always accessible due to the high flexibility of the C-terminal loop of the propeptide.

The *in vivo* conditions for protease precursor activation are generally very complex involving not only the pH but also other proteases, their endogenous inhibitors, macromolecular components like glycosaminoglycans (Ishidoh & Kominami, 1995; Kihara *et al.*, 2002) and clusters of negative charges (Mason & Massey, 1992) influencing the maturation process. Moreover, the liberation of the propeptides or their parts from the precursors generates additional inhibitors with moderate affinity for the active enzymes (see above). The *in vivo* action of such peptides is not very likely but cannot be completely excluded. Most of the split products may be rapidly degraded within the acidic environment of the vacuoles where they arise. The propeptides remain stable until pH values around 5.0, but may change their native conformation below this critical pH (Maubach *et al.*, 1997). However, the presence of cathepsin L propeptide could still be demonstrated after 2 h incubation with active cathepsin L at pH 5.1 (Menard *et al.*, 1998). This leads to the suggestion that cysteine protease propeptides may not only play a role in the regulation of the protease activity itself but may also reduce the generation of active enzyme.

Lysosomal cysteine protease precursor activation was first intensively studied by Nishimura *et al.* (1987a and 1987b). They identified the lysosomes not only as the site of cathepsin action but also of their activation. They were able to prove that correctly trimmed carbohydrate chains and an acidic pH are necessary prerequisites for processing of procathepsins to mature enzymes (Nishimura *et al.*, 1988a; 1988b; Oda *et al.*, 1991). Using group specific inhibitors, they suggested the involvement of at least one pepstatin sensitive protease, probably cathepsin D, in the activation of procathepsins B, H and L (Nishimura *et al.*, 1988c; Nishimura *et al.*, 1989; Kawabata *et al.*, 1993), a result also confirmed by us (Wiederanders & Kirschke, 1989). Later

on, the same group showed that a leupeptin sensitive protease may also be involved in the processing of cathepsin precursors (Nishimura *et al.*, 1995). They confirmed a suggestion made earlier by Salminen & Gottesman (1990). To make the confusion complete, matrix metalloproteinases (MMPs) seem to be involved, too, since Hara *et al.* (1988) were able to show that metal chelators such as *o*-phenantroline inhibit also the processing of cathepsin L, B and H precursors in macrophages.

Further progress in understanding the order of the processing steps came from experiments with recombinant proenzymes. Procathepsins expressed in yeast systems can be activated *in vitro* by proteases with different substrate specificities. This indicates that the exposition of scissile bond(s) rather than the amino-acid sequence of the loop itself are determinants of the cleavage (Menard *et al.*, 1998). As a consequence, the primary *in vitro* processing products did not always show the terminal sequences of *in vivo* processed mature enzymes. *In vitro* processing of propapain resulted in enzymes with N-terminal extensions of 5 and 3 amino acids (Vernet *et al.*, 1991). Only 10% of the *in vitro* processed recombinant precursors had the N-terminus of wild type papain (Vernet *et al.*, 1990). Rowan *et al.* (1992) studied the processing of a rat procathepsin B mutant (C29S; S115A) by five different proteases *in vitro*. Compared to the wild-type, all processed enzymes revealed N-terminal extensions whose lengths varied between 4 and 32 residues. Interestingly, trimming of the extensions was observed after incubation of the primary processing products with the exopeptidases cathepsin H or dipeptidyl peptidase I (DPPPI). The trimming stopped before the Leu1-Pro2 bond of mature cathepsin B thus resulting in the N-terminal sequence typical for all lysosomal cathepsins, X-Pro.

***In vitro* processing**

In vitro activation studies of CA1 peptidase precursors have mainly concentrated on two

major points. The first question is whether the precursors are able to autoactivate or this process needs the assistance of another active protease. The second is whether the *in vitro* processing products are identical to those which can be isolated from tissues or cells. The latter question is of certain interest for large scale production of recombinant active enzymes.

Recombinant propapain was successfully expressed and activated as one of the first cysteine protease precursors (Vernet *et al.*, 1990). The processing reaction seems to be a unimolecular process (Vernet *et al.*, 1991). In contrast to propapain, a recombinant precursor of papaya protease IV (PPIV) was incapable of autocatalytic processing due to very stringent substrate specificity although the enzyme activity of the expressed PPIV precursor has been documented (Baker *et al.*, 1996).

Autocatalytic maturation of recombinant procathepsin L has been reported in which intra- and intermolecular reactions might be involved (McDonald & Emerick; 1995; Nomura & Fujisawa, 1997; Menard *et al.*, 1998). Detailed studies of the concentration dependence of activation progress revealed participation of both bimolecular as well as unimolecular components in the activation (Menard *et al.*, 1998).

In an early report, human recombinant procathepsin B was suggested to be activated at acidic conditions by unimolecular cleavage because the activation was concentration-independent (Mach *et al.*, 1994), whereas Rozman *et al.* (1999) suggested an intermolecular rather than intramolecular activation process. Recent experiments using the binding hierarchy of cystatin C to different processing products resulted for the first time in a direct experimental proof of unimolecular procathepsin B activation (Quraishi & Storer, 2001).

Recombinant procathepsin K was auto-activated at pH 4.0 resulting in an enzyme with an N-terminal dipeptide extension (Bossard *et*

al., 1996; McQueney *et al.*, 1997), whereas pepsin treatment of recombinant procathepsin K at this pH resulted in the "correct" sequence of mature enzyme (Linnevers *et al.*, 1997). Recombinant procathepsin S expressed in a yeast system can be activated autocatalytically as well as by addition of subtilisin BPN' with high yield (Bromme *et al.*, 1993). The pH seems to be the most important parameter for processing, because at pH 8 cathepsin S is still active, but autocatalytic processing of procathepsin S cannot be observed.

An interesting exception is procathepsin W. Procathepsin W was expressed in mammalian cells transfected with the respective cDNA (Wex *et al.*, 1998). However, up to now neither a report of successful procathepsin W activation *in vitro* nor any report of active enzyme have been published (Wex *et al.*, 2001).

Cathepsin X is also unique within the group of lysosomal cysteine proteases. It shows the shortest propeptide of all cathepsins (Santamaria *et al.*, 1998; Sivaraman *et al.*, 2000). The propeptide is covalently attached to the mature enzyme by a disulfide bridge between the active site Cys31 and Cys10p of the propeptide, thus making an autoproteolytic activation impossible. Therefore, *in vitro* maturation of procathepsin X needs assistance by another protease as well as reduction of the S-S bridge.

***In vivo* processing**

In vivo processing experiments have been performed to find out where the activation process occurs and which processing enzymes might be involved.

It has long been discussed that procathepsin B might be activated autocatalytically *in vivo* (Mach *et al.*, 1994; Rozman *et al.*, 1999) or by the complete mixture of cathepsins within the lysosomes. In a recent paper, Ishidoh *et al.* (1999) identified four candidates as the processing proteases: cathepsins H, S, C and

even procathepsin L. Activation of procathepsin B *in vivo* can also occur in the extracellular matrix, e.g. by tissue plasminogen activator (tPA) (Dalet-Fumeron *et al.*, 1996) or active extracellular cathepsin D (van der Strappen *et al.*, 1996).

Procathepsin K is activated from its precursor within the lysosomal compartment of osteoclasts as expected (Rieman *et al.*, 2001). In contrast to other lysosomal cathepsins whose secretion has been reported to occur as zymogens (e.g. in malignant tumours), some cell types secrete active cathepsin K (Dodds *et al.*, 2001; Hou *et al.*, 2002). Furthermore, in osteoarthritis, autocatalytic activation of secreted procathepsin K in damaged cartilage samples has been observed (Konttinen *et al.*, 2002). Surprisingly, a mutation of the C-terminal Met329 of procathepsin K to Ala resulted in a complete inhibition of precursor processing (Claveau & Riendeau, 2001).

Menard *et al.* (1998) demonstrated in their studies that more than one peptide bond in the flexible region between the cathepsin L propeptide and the mature enzyme might be recognized for processing and are cleaved. These alternative cleavage sites of procathepsin L are indeed used *in vivo* resulting in cathepsin L species with three different N-termini (Ishidoh *et al.*, 1998). The authors confirmed in a later experiment the sequence IPTKV as the major N-terminal sequence of fully processed cathepsin L (Ishidoh *et al.*, 1999). They identified the lysosomes as the processing compartment of procathepsin L in NIH 3T3 cells, whereas procathepsins B and D were processed in late endosomes.

Procathepsin S maturation can only occur within the lysosomes or in other acidic vacuoles because the only potential glycosylation site of this enzyme precursor is located in the propeptide part (Shi *et al.*, 1994; Wiederanders *et al.*, 1992).

There are only few reports of procathepsin H activation. The processing was reported to

take place in the lysosomes (Nishimura *et al.*, 1987b) and to depend on a pepstatin sensitive aspartic protease(s) (Nishimura & Kato, 1988).

The dipeptidyl-peptidase cathepsin C is unique amongst the lysosomal cathepsins due to its tetrameric structure which is formed after processing. The activation and cleavage of the propeptide part is in some way different from that of other cathepsins. The maturation is initiated by the cleavage of an N-terminal "residual propeptide" (Cigic *et al.*, 2000) resulting in the accumulation of a 36 kDa precursor with the N-terminal sequence N(134)SKQE. Further processing proceeds with the generation of an "activation peptide", and the heavy and light chains. Whereas the activation peptide is removed the residual propeptide remains non-covalently attached to the heavy and light chains forming a heterotrimer (Cigic *et al.*, 1998). The residual propeptide seems to be essential for oligomerization and stabilization of the final tetrameric structure (Santilman *et al.*, 2002). Cathepsins L and S were identified as the processing proteases of procathepsin C (Dahl *et al.*, 2001). Tetramerization of mature cathepsin C polypeptides leads to occlusion of the endopeptidase-like active site cleft and provides the structural basis of dipeptidyl peptidase activity (Horn *et al.*, 2002). Modifying Cys331, the authors identified this amino acid as essential for tetramerization. During processing of the cathepsin C precursor, this special residue becomes exposed and tetramerization can start. The study shows that also in cathepsin C a part of the propeptide has to be removed first before the activity is exerted which is typical for this unique cathepsin.

It can be concluded from all of these studies that maturation and activation of cathepsin precursors are multistep processes. They may be performed by various proteases and they depend strongly on the cellular or extracellular environment.

FOLDING ASSISTANCE

It has been frequently stated that cysteine proteases need their propeptide parts for proper folding, however, experimental verifications are rare. The first strong hint for propeptide assisted folding of mature papain-like cysteine proteases was the very low specific activity of *in vitro* refolded recombinant procathepsin L with N-terminal truncations. The loss of the specific activity was proportional to the extent of propeptide truncation (Smith & Gottesman, 1989). This result was later confirmed by other groups which also showed that not any but the complete cognate propeptide was necessary for the synthesis of correctly folded procathepsin L (Tao *et al.*, 1994; Ogino *et al.*, 1999). This function of the propeptide cannot be replaced even under optimized folding conditions (Tobbell *et al.*, 2002). Another set of hints came from mutations of the propeptide affecting its structure and, as an obvious consequence of the destabilized structure, also its function to support folding (Yamamoto *et al.*, 1999; Hou *et al.*, 1999; Kreuzsch *et al.*, 2000). The questions arising from these results were whether folding assistance by the propeptide is a general phenomenon valid for all members of the CA1 peptidase family, and whether the propeptide has to be bound covalently to the mature enzyme in order to exert the foldase effect.

The answer to the first question came from experiments with cathepsin B (Mehtani *et al.*, 1998) and from a general observation: Cathepsin B was synthesized as active enzyme from a truncated mRNA lacking exons 2 and 3. Exon 3 codes for a part of the cathepsin B propeptide. A Δ 51 splice variant of cathepsin B is expressed by some malignant tumours, and the resulting active enzyme is correctly folded.

Cathepsins X and B show shorter propeptides than all other mammalian cathepsins, but they show also small extra loops in the mature parts which may facilitate the folding.

Such small loops are considered to be beneficial for folding (Cunningham *et al.*, 1999).

These results showed that not all cathepsins need their propeptide parts for correct folding.

The answer to the second question came from experiments with exogenously added propeptides to refolding assays of denatured proteases similar to those reported earlier for other protease classes (Ohta *et al.*, 1991; Baker *et al.*, 1992; Ogino *et al.*, 1999). Experiments like that with cysteine peptidases were lacking until recently. The propeptides of cysteine proteases fold correctly when expressed as recombinant proteins. They seem to retain their secondary structure over a wide pH range from 6.5 to 3.0 (Jerala *et al.*, 1998). Several experiments have shown that addition of recombinant propeptides can efficiently catalyze the refolding of denatured mature cysteine proteinases (Pietschmann *et al.*, 2002; Yamamoto *et al.*, 2002; Capetta *et al.*, 2002). With respect to cathepsin S, this effect was strongly dependent on the three-dimensional structure of the propeptide. Mutants affecting the aromatic Trp core of the propeptide showed a much weaker foldase effect than the wild type propeptide (Pietschmann *et al.*, 2002). A mutation in the conserved GNFD motif (N70pI/F72pI) of the *F. hepatica* cathepsin L propeptide also diminished the foldase function of the propeptide (Capetta *et al.*, 2002). Thus, a strong correlation exists between the structural integrity of the propeptide, its inhibitory potency and its ability to catalyze correct folding of the mature enzyme (see Fig. 2).

The recently described nucleation-condensation mechanism of protein folding is a combination of the framework model and the hydrophobic collapse model. It implies the possibility of shifting to either of the models depending on the stability of secondary and tertiary structures (Daggett & Fersht, 2003). This fits very well with the phenomenon seen in some peptidases where N-terminal folded structures can provide a scaffold for further

folding and thus facilitate the process *in vivo* (Frydman, 2001). In cathepsin L like cysteine proteases, a mini-domain in the propeptide represents this nucleation centre and is suggested to be the structural correlate of its foldase function (Schilling *et al.*, 2001).

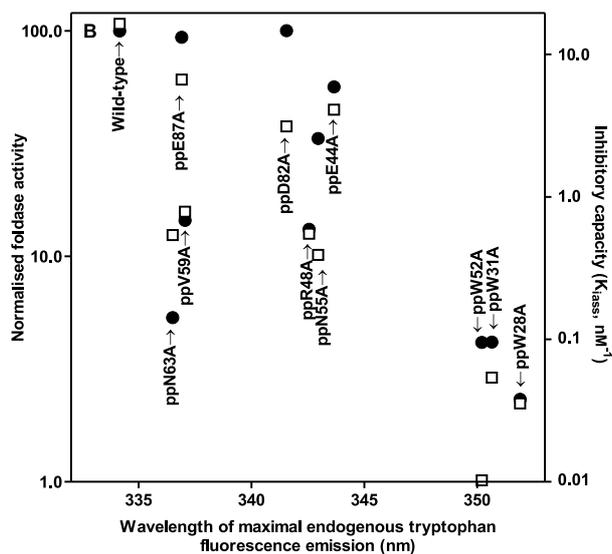


Figure 2. Effect of mutagenesis of the highly conserved prodomain residues on various structural and functional properties of human cathepsin S propeptide.

Renaturation rate (left ordinate, ●), compactness of tertiary structure (abscissa) and inhibitory function (right ordinate, □). Taken from Pietschmann *et al.* (2002).

The name foldase was introduced by Inouye's group for specific folding assistance (Zhu *et al.*, 1989). It describes a real enzyme-like catalysis as the name suggests. It is noteworthy that acceleration of protease folding can be achieved in *cis* and in *trans*, i.e. the propeptide acts either covalently bound to the protein *in statu nascendi* or it is added as a recombinant protein to renaturation assays. The first approach to study the pH-dependent denaturation/renaturation cycle of mature cathepsin B was performed by NMR-spectroscopy (Song *et al.*, 2000).

ENZYME TARGETING

Mannose-6-phosphate dependent sorting mechanism

Mannose-6-phosphate signals for delivery of enzymes to the lysosomes or acidified vesicles can be found in the propeptide parts as well as in the mature catalytic domains. Here, we refer only to the sorting signals located in the propeptide parts.

The transfer of phospho-GlcNAc to mannose residues of N-linked oligosaccharides of lysosomal proteins is catalyzed by the UDP-GlcNAc: lysosomal enzyme *N*-acetylglucosamine-1-phosphotransferase. The removal of terminal GlcNAc results in the generation of mannose-6-phosphate which then attaches the lysosomal proteins to the respective receptors for further transport to the lysosomes. The selectivity of the mannose phosphorylation of lysosomal enzymes has its structural basis in at least two Lys residues on the enzymes' outer surface which are about 34 Å apart. Introduction of putative phosphotransferase recognition sequences into a secretory non-glycosylated protein, pepsinogen, resulted in its proper glycosylation (Baranski *et al.*, 1990).

In cathepsin L, this motif is located in the propeptide. Mutation of only one of the two residues in procathepsin L prevented the molecule phosphorylation (Cuozzo *et al.*, 1995; 1998). However, the recognition motifs for the phosphotransferase are not located in the propeptide part in all cathepsins, as has been shown for cathepsin B (Cuozzo *et al.*, 1998; Lukong *et al.*, 1999). In cathepsin S, the only potential glycosylation site is located in the propeptide 11 amino acids upstream from the N-terminus of the mature enzyme (Wiederanders *et al.*, 1992). In rat hepatocytes, procathepsin B was glycosylated only at one of the two potential glycosylation sites, namely that within the propeptide (Tanaka *et al.*, 2000). These exam-

ples seem to be rather the exceptions, as most other lysosomal proteins are glycosylated in both the pro- and mature regions. In summary, although the correct fold of the propeptide is a prerequisite for proper glycosylation in the ER, the recognition motifs for glycosylation and the glycosylation sites *per se* are not exclusively restricted to the propeptide part of cathepsins.

Mannose-6-phosphate independent sorting mechanisms

The existence of mannose-6-phosphate independent trafficking of lysosomal proteins has long been suggested by various lines of evidence.

- ◆ Site directed mutagenesis experiments in which the essential Asn residues of potential glycosylation sites in lysosomal proteins have been replaced by Gly or Gln. The mass of such non glycosylated enzymes was secreted, however, a small fraction remained intracellular (Nissler *et al.*, 1998; Tanaka *et al.*, 2000).
- ◆ Lysosomal membrane attachment of non-glycosylated procathepsin S has been demonstrated (Nissler *et al.*, 1998).
- ◆ Lymphocytes of patients with I-cell disease show normal cellular levels of lysosomal enzymes despite severely reduced phosphotransferase activity and consequently a lack of mannose phosphorylation (Glickman & Kornfeld, 1993).
- ◆ Some parasite cathepsins are not glycosylated due to mutations of the essential Asn residues (Sajid & McKerrow, 2002); *Dictyostelium discoideum* does not express detectable amounts of mannose-6-phosphate receptors (Cardelli *et al.*, 1986); and murine cell lines defective in mannose-6-phosphate receptor are also known (Gabel *et al.*, 1983). Nevertheless, the lysosomal cathepsins take their correct way into these organisms/cells suggesting that a kind of “primitive targeting” independent of mannose-6-phosphate and its receptor has

evolved early in evolution and is maintained until today.

The apparent structural background of mannose-6-phosphate independent sorting is obviously a conserved 9 amino acid long peptide motif which was identified to be implicated in the alternative trafficking process (Huete-Perez *et al.*, 1999). It is located close to the N-terminal tail of the propeptides. The motif is highly conserved as exemplified by enzymes from plant, parasite, crustacean and human systems (see Fig. 3). It has to be mentioned, however, that some cathepsins lack this cryptic and ancient sorting motif, e.g. cathepsins B, F, and X, and also baculovirus cathepsin.

GICP2*	31	..WKAGIPKRF..
GICP3*	31	..WKAGIPKRF..
GICP1*	31	..WKAGMPKRF..
human cat-L	31	..WKAMHNRLY..
murine cat-L	31	..WKSTHRRLY..
human cat-S	31	..WKITYGKQY..
human cat-K	29	..WKTHRKQY..
human cat-H	35	..FKSWMSKHR..
human cat-V	31	..WKATHRRLY..
D. mel.**cat-L	32	..FKLEHRKNY..
H. am.***cat-L	23	..FKGKFRKY..
cruzain	41	..FKQKHGRVY..

Figure 3. Homologous nonapeptide sequences in class CA1 peptidase precursors of various species.

This motif is suggested to mediate “primitive” mannose-6-phosphate independent targeting to lysosomes and lysosome-like vesicles. Numbering starts from Met 1, data are taken from the MEROPS database. **Giardia lamblia* cysteine proteases 1, 2 and 3; ***Drosophila melanogaster* cathepsin L; ****Homerus americanus* cathepsin L

A receptor recognizing the motif has not been identified so far, although a 43 kD integral lysosomal membrane protein was described binding mouse procathepsin L in a pH dependent manner (McIntyre & Erickson, 1993). The binding of procathepsin L to this putative “receptor” could be blocked by a synthetic peptide comprising the first 24 N-terminal amino acids of procathepsin L (counted without the signal sequence) con-

taining the respective nonapeptide motif (McIntyre *et al.*, 1994).

REFERENCES

- Baker D, Silen JL, Agard DA. (1992) Protease pro region required for folding is a potent inhibitor of the mature enzyme. *Proteins.*; **12**: 339–44.
- Baker KC, Taylor MA, Cummings NJ, Tunon MA, Worboys KA, Connerton IF. (1996) Autocatalytic processing of pro-papaya proteinase IV is prevented by crowding of the active-site cleft. *Protein Eng.*; **9**: 525–9.
- Baranski TJ, Faust PL, Kornfeld S. (1990) Generation of a lysosomal enzyme targeting signal in the secretory protein pepsinogen. *Cell.*; **63**: 281–91.
- Billington CJ, Mason P, Magny MC, Mort JS. (2000) The slow-binding inhibition of cathepsin K by its propeptide. *Biochem Biophys Res Commun.*; **276**: 924–9.
- Bossard MJ, Tomaszek TA, Thompson SK, Amegadzie BY, Hanning CR, Jones C, Kurdyla JT, McNulty DE, Drake FH, Gowen M, Levy MA. (1996) Proteolytic activity of human osteoclast cathepsin K. Expression, purification, activation, and substrate identification. *J Biol Chem.*; **271**: 12517–24.
- Bromme D, Bonneau PR, Lachance P, Wiederanders B, Kirschke H, Peters C, Thomas DY, Storer AC, Vernet T. (1993) Functional expression of human cathepsin S in *Saccharomyces cerevisiae*. Purification and characterization of the recombinant enzyme. *J Biol Chem.*; **268**: 4832–8.
- Brown J, Matutes E, Singleton A, Price C, Molgaard H, Buttle D, Enver T. (1998) Lymphopain, a cytotoxic T and natural killer cell-associated cysteine proteinase. *Leukemia.*; **12**: 1771–81.
- Cappetta M, Roth I, Diaz A, Tort J, Roche L. (2002) Role of the prosegment of *Fasciola hepatica* cathepsin L1 in folding of the catalytic domain. *Biol Chem.*; **383**: 1215–21.
- Cardelli JA, Golubeski GS, Dimond RL. (1986) Lysosomal enzymes in *Dictyostelium discoideum* are transported to lysosomes at distinctly different rates. *J Cell Biol.*; **102**: 1264–70.
- Carmona E, Dufour E, Plouffe C, Takebe S, Mason P, Mort JS, Menard R. (1996) Potency and selectivity of the cathepsin L propeptide as an inhibitor of cysteine proteases. *Biochemistry.*; **35**: 8149–57.
- Chen Y, Plouffe C, Menard R, Storer AC. (1996) Delineating functionally important regions and residues in the cathepsin B propeptide for inhibitory activity. *FEBS Lett.*; **393**: 24–6.
- Chowdhury SF, Sivaraman J, Wang J, Devanathan G, Lachance P, Qi H, Menard R, Lefebvre J, Konishi Y, Cygler M, Sulea T, Purisima EO. (2002) Design of noncovalent inhibitors of human cathepsin L. From the 96-residue proregion to optimized tripeptides. *J Med Chem.*; **45**: 5321–9.
- Cigic B, Krizaj I, Kralj B, Turk V, Pain RH. (1998) Stoichiometry and heterogeneity of the pro-region chain in tetrameric human cathepsin C. *Biochim Biophys Acta.*; **1382**: 143–50.
- Cigic B, Dahl SW, Pain RH. (2000) The residual pro-part of cathepsin C fulfills the criteria required for an intramolecular chaperone in folding and stabilizing the human proenzyme. *Biochemistry.*; **39**: 12382–90.
- Claveau D, Riendeau D. (2001) Mutations of the C-terminal end of cathepsin K affect proenzyme secretion and intracellular maturation. *Biochem Biophys Res Commun.*; **281**: 551–7.
- Coulombe R, Grochulski P, Sivaraman J, Menard R, Mort JS, Cygler M. (1996) Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.*; **15**: 5492–503.
- Cunningham EL, Jaswal SS, Sohl JL, Agard DA. (1999) Kinetic stability as a mechanism for protease longevity. *Proc Natl Acad Sci U S A.*; **96**: 11008–14.

- Cuozzo JW, Tao K, Wu QL, Young W, Sahagian GG. (1995) Lysine-based structure in the proregion of procathepsin L is the recognition site for mannose phosphorylation. *J Biol Chem.*; **270**: 15611–9.
- Cuozzo JW, Tao K, Cygler M, Mort JS, Sahagian GG. (1998) Lysine-based structure responsible for selective mannose phosphorylation of cathepsin D and cathepsin L defines a common structural motif for lysosomal enzyme targeting. *J Biol Chem.*; **273**: 21067–76.
- Cygler M, Sivaraman J, Grochulski P, Coulombe R, Storer AC, Mort JS. (1996) Structure of rat procathepsin B: model for inhibition of cysteine protease activity by the proregion. *Structure.*; **4**: 405–16.
- Daggett V, Fersht AR. (2003) Is there a unifying mechanism for protein folding? *Trends Biochem Sci.*; **28**: 18–25.
- Dahl SW, Halkier T, Lauritzen C, Dolenc I, Pedersen J, Turk V, Turk B. (2001) Human recombinant pro-dipeptidyl peptidase I (cathepsin C) can be activated by cathepsins L and S but not by autocatalytic processing. *Biochemistry.*; **40**: 1671–8.
- Dalet-Fumeron V, Boudjennah L, Pagano M. (1996) Competition between plasminogen and procathepsin B as a probe to demonstrate the in vitro activation of procathepsin B by the tissue plasminogen activator. *Arch Biochem Biophys.*; **335**: 351–7.
- Delaria K, Fiorentino L, Wallace L, Tamburini P, Brownell E, Muller D. (1994) Inhibition of cathepsin L-like cysteine proteases by cytotoxic T-lymphocyte antigen-2 beta. *J Biol Chem.*; **269**: 25172–7.
- Denizot F, Brunet JF, Roustan P, Harper K, Suzan M, Luciani MF, Mattei MG, Golstein P. (1989) Novel structures CTLA-2 alpha and CTLA-2 beta expressed in mouse activated T cells and mast cells and homologous to cysteine proteinase proregions. *Eur J Immunol.*; **19**: 631–5.
- Dodds RA, James IE, Rieman D, Ahern R, Hwang SM, Connor JR, Thompson SD, Veber DF, Drake FH, Holmes S, Lark MW, Gowen M. (2001) Human osteoclast cathepsin K is processed intracellularly prior to attachment and bone resorption. *J Bone Miner Res.*; **16**: 478–86.
- Drenth J, Jansoniu Jn, Koekoek R, Swen HM, Wolthers BG. (1968) Structure of papain. *Nature.*; **218**: 929–31.
- Fox T, Demiguel E, Mort JS, Storer AC. (1992) Potent slow-binding inhibition of cathepsin-B by its propeptide. *Biochemistry.*; **31**: 12571–6.
- Frydman J. (2001) Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu Rev Biochem.*; **70**: 603–47.
- Furmonaviciene R, Sewell HF, Shakib F. (2000) Comparative molecular modelling identifies a common putative IgE epitope on cysteine protease allergens of diverse sources. *Clin Exp Allergy.*; **30**: 1307–13.
- Gabel CA, Goldberg DE, Kornfeld S. (1983) Identification and characterization of cells deficient in the mannose 6-phosphate receptor: evidence for an alternate pathway for lysosomal enzyme targeting. *Proc Natl Acad Sci U S A.*; **80**: 775–9.
- Glickman JN, Kornfeld S. (1993) Mannose 6-phosphate-independent targeting of lysosomal enzymes in I-cell disease B lymphoblasts. *J Cell Biol.*; **123**: 99–108.
- Gotthardt D, Warnatz HJ, Henschel O, Bruckert F, Schleicher M, Soldati T. (2002) High-resolution dissection of phagosome maturation reveals distinct membrane trafficking phases. *Mol Biol Cell.*; **13**: 3508–20.
- Guay J, Falguyret JP, Ducret A, Percival MD, Mancini JA. (2000) Potency and selectivity of inhibition of cathepsin K, L and S by their respective propeptides. *Eur J Biochem.*; **267**: 6311–8.
- Guex N, Peitsch MC. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis.*; **18**: 2714–23.
- Guncar G, Podobnik M, Pungercar J, Strukelj B, Turk V, Turk D. (1998) Crystal structure of porcine cathepsin H determined at 2.1 Å resolution: location of the mini-chain C-terminal carboxyl group defines cathepsin H

- aminopeptidase function. *Structure.*; **6**: 51–61.
- Guo YL, Kurz U, Schultz JE, Lim CC, Wiederanders B, Schilling K. (2000) The alpha1/2 helical backbone of the prodomains defines the intrinsic inhibitory specificity in the cathepsin L-like cysteine protease subfamily. *FEBS Lett.*; **469**: 203–7.
- Hara K, Kominami E, Katunuma N. (1988) Effect of proteinase inhibitors on intracellular processing of cathepsin B, H and L in rat macrophages. *FEBS Lett.*; **231**: 229–31.
- Horn M, Baudys M, Voburka Z, Klueh I, Vondrasek J, Mares M. (2002) Free-thiol Cys331 exposed during activation process is critical for native tetramer structure of cathepsin C (dipeptidyl peptidase I). *Protein Sci.*; **11**: 933–43.
- Hou WS, Bromme D, Zhao Y, Mehler E, Dushey C, Weinstein H, Miranda CS, Fraga C, Greig F, Carey J, Rimoin DL, Desnick RJ, Gelb BD. (1999) Characterization of novel cathepsin K mutations in the pro and mature polypeptide regions causing pycnodysostosis. *J Clin Invest.*; **103**: 731–8.
- Hou WS, Li W, Keyszer G, Weber E, Levy R, Klein MJ, Gravalles EM, Goldring SR, Bromme D. (2002) Comparison of cathepsins K and S expression within the rheumatoid and osteoarthritic synovium. *Arthritis Rheum.*; **46**: 663–74.
- Huete-Perez JA, Engel JC, Brinen LS, Mottram JC, McKerrow JH. (1999) Protease trafficking in two primitive eukaryotes is mediated by a prodomain protein motif. *J Biol Chem.*; **274**: 16249–56.
- Ishidoh K, Kominami E. (1995) Procathepsin L degrades extracellular matrix proteins in the presence of glycosaminoglycans *in vitro*. *Biochem Biophys Res Commun.*; **217**: 624–31.
- Ishidoh K, Saido TC, Kawashima S, Hirose M, Watanabe S, Sato N, Kominami E. (1998) Multiple processing of procathepsin L to cathepsin L *in vivo*. *Biochem Biophys Res Commun.*; **252**: 202–7.
- Ishidoh K, Takeda-Ezaki M, Watanabe S, Sato N, Aihara M, Imagawa K, Kikuchi M, Kominami E. (1999) Analysis of where and which types of proteinases participate in lysosomal proteinase processing using bafilomycin A1 and *Helicobacter pylori* Vac A toxin. *J Biochem (Tokyo).*; **125**: 770–9.
- Jerala R, Zerovnik E, Kidric J, Turk V. (1998) pH-Induced conformational transitions of the propeptide of human cathepsin L – A role for a molten globule state in zymogen activation. *J Biol Chem.*; **273**: 11498–504.
- Kamphuis IG, Kalk KH, Swarte MBA, Drenth J. (1984) Structure of papain refined at 1.65 Å resolution. *J Mol Biol.*; **179**: 233–56.
- Karrer KM, Pfeifer SL, DiTomas ME. (1993) Two distinct gene subfamilies within the family of cysteine protease genes. *Proc Natl Acad Sci U S A.*; **90**: 3063–7.
- Kawabata T, Nishimura Y, Higaki M, Kato K. (1993) Purification and processing of rat liver procathepsin B. *J Biochem (Tokyo).*; **113**: 389–94.
- Kihara M, Kakegawa H, Matano Y, Murata E, Tsuge H, Kido H, Katunuma N. (2002) Chondroitin sulfate proteoglycan is a potent enhancer in the processing of procathepsin L. *Biol Chem.*; **383**: 1925–9.
- Konttinen YT, Mandelin J, Li TF, Salo J, Lassus J, Liljestrom M, Hukkanen M, Takagi M, Virtanen I, Santavirta S. (2002) Acidic cysteine endoproteinase cathepsin K in the degeneration of the superficial articular hyaline cartilage in osteoarthritis. *Arthritis Rheum.*; **46**: 953–60.
- Krasko A, Gamulin V, Seack J, Steffen R, Schroder HC, Muller WE. (1997) Cathepsin, a major protease of the marine sponge *Geodia cydonium*: purification of the enzyme and molecular cloning of cDNA. *Mol Mar Biol Biotechnol.*; **6**: 296–307.
- Kreusch S, Fehn M, Maubach G, Nissler K, Rommerskirch W, Schilling K, Weber E, Wenz I, Wiederanders B. (2000) An evolutionarily conserved tripartite tryptophan motif stabilizes the prodomains of cathepsin L-like cysteine proteases. *Eur J Biochem.*; **267**: 2965–72.

- Kurata M, Yamamoto Y, Watabe S, Makino Y, Ogawa K, Takahashi SY. (2001) *Bombyx* cysteine proteinase inhibitor. (BCPI) homologous to propeptide regions of cysteine proteinases is a strong, selective inhibitor of cathepsin L-like cysteine proteinases. *J Biochem (Tokyo)*; **130**: 857–63.
- Lalmanach G, Lecaille F, Chagas JR, Authie E, Scharfstein J, Juliano MA, Gauthier F. (1998) Inhibition of trypanosomal cysteine proteinases by their propeptides. *J Biol Chem*; **273**: 25112–6.
- Lecaille F, Kaleta J, Bromme D. (2002) Human and parasitic papain-like cysteine proteases: their role in physiology and pathology and recent developments in inhibitor design. *Chem Rev*; **102**: 4459–88.
- Linke M, Jordans S, Mach L, Herzog V, Brix K. (2002) Thyroid stimulating hormone upregulates secretion of cathepsin B from thyroid epithelial cells. *Biol Chem*; **383**: 773–84.
- Linnevers C, Smeekens SP, Bromme D. (1997a) Human cathepsin W, a putative cysteine protease predominantly expressed in CD8+ T-lymphocytes. *FEBS Lett*; **405**: 253–9.
- Linnevers CJ, McGrath ME, Armstrong R, Mistry FR, Barnes MG, Klaus JL, Palmer JT, Katz BA, Bromme D. (1997b) Expression of human cathepsin K in *Pichia pastoris* and preliminary crystallographic studies of an inhibitor complex. *Protein Sci*; **6**: 919–21.
- Lukong KE, Elsliger MA, Mort JS, Potier M, Pshezhetsky AV. (1999) Identification of UDP-N-acetylglucosamine-phosphotransferase-binding sites on the lysosomal proteases, cathepsins A, B, and D. *Biochemistry*; **38**: 73–80.
- Mach L, Mort JS, Glossl J. (1994) Maturation of human procathepsin B. Proenzyme activation and proteolytic processing of the precursor to the mature proteinase, *in vitro*, are primarily unimolecular processes. *J Biol Chem*; **269**: 13030–5.
- Mason RW, Massey SD. (1992) Surface activation of pro-cathepsin L. *Biochem Biophys Res Commun*; **189**: 1659–66.
- Maubach G, Schilling K, Rommerskirch W, Wenz I, Schultz JE, Weber E, Wiederanders B. (1997) The inhibition of cathepsin S by its propeptide – specificity and mechanism of action. *Eur J Biochem*; **250**: 745–50.
- McDonald JK, Emerick JM. (1995) Purification and characterization of procathepsin L, a self-processing zymogen of guinea pig spermatozoa that acts on a cathepsin D assay substrate. *Arch Biochem Biophys*; **323**: 409–22.
- McGrath ME. (1999) The lysosomal cysteine proteases. *Annu Rev Biophys Biomol Struct*; **28**: 181–204.
- McGrath ME, Eakin AE, Engel JC, Mckerrow JH, Craik CS, Fletterick RJ. (1995) The crystal-structure of Cruzain – a therapeutic target for Chagas-disease. *J Mol Biol*; **247**: 251–9.
- McIntyre GF, Erickson AH. (1993) The lysosomal proenzyme receptor that binds procathepsin L to microsomal membranes at pH 5 is a 43-kDa integral membrane protein. *Proc Natl Acad Sci U S A*; **90**: 10588–92.
- McIntyre GF, Godbold GD, Erickson AH. (1994) The pH-dependent membrane association of procathepsin L is mediated by a 9-residue sequence within the propeptide. *J Biol Chem*; **269**: 567–72.
- McQueney MS, Amegadzie BY, D'Alessio K, Hanning CR, McLaughlin MM, McNulty D, Carr SA, Ijames C, Kurdyla J, Jones CS. (1997) Autocatalytic activation of human cathepsin K. *J Biol Chem*; **272**: 13955–60.
- Mehtani S, Gong Q, Panella J, Subbiah S, Peffley DM, Frankfater A. (1998) *In vivo* expression of an alternatively spliced human tumor message that encodes a truncated form of cathepsin B. Subcellular distribution of the truncated enzyme in COS cells. *J Biol Chem*; **273**: 13236–44.
- Menard R, Carmona E, Takebe S, Dufour E, Plouffe C, Mason P, Mort JS. (1998) Autocatalytic processing of recombinant human procathepsin L. Contribution of both intermolecular and unimolecular events in

- the processing of procathepsin L in vitro. *J Biol Chem.*; **273**: 4478–84.
- Musil D, Zucic D, Turk D, Engh RA, Mayr I, Huber R, Popovic T, Turk V, Towatari T, Katunuma N, Bode W. (1991) The refined 2.15-Å X-ray crystal-structure of human liver cathepsin-B – the structural basis for its specificity. *EMBO J.*; **10**: 2321–30.
- Nagler DK, Sulea T, Menard R. (1999a) Full-length cDNA of human cathepsin F predicts the presence of a cystatin domain at the N-terminus of the cysteine protease zymogen. *Biochem Biophys Res Commun.*; **257**: 313–8.
- Nagler DK, Zhang R, Tam W, Sulea T, Purisima EO, Menard R. (1999b) Human cathepsin X: A cysteine protease with unique carboxypeptidase activity. *Biochemistry.*; **38**: 12648–54.
- Nettis E, Napoli G, Ferrannini A, Tursi A. (2001) IgE-mediated allergy to bromelain. *Allergy.*; **56**: 257–8.
- Nishimura Y, Kato K. (1987a) Intracellular transport and processing of lysosomal cathepsin B. *Biochem Biophys Res Commun.*; **148**: 254–9.
- Nishimura Y, Kato K. (1987b) Intracellular transport and processing of lysosomal cathepsin H. *Biochem Biophys Res Commun.*; **148**: 329–34.
- Nishimura Y, Kato K. (1988) Identification of latent procathepsin H in microsomal lumen: characterization of proteolytic processing and enzyme activation. *Arch Biochem Biophys.*; **260**: 712–8.
- Nishimura Y, Amano J, Sato H, Tsuji H, Kato K. (1988a) Biosynthesis of lysosomal cathepsins B and H in cultured rat hepatocytes. *Arch Biochem Biophys.*; **262**: 159–70.
- Nishimura Y, Furuno K, Kato K. (1988b) Biosynthesis and processing of lysosomal cathepsin L in primary cultures of rat hepatocytes. *Arch Biochem Biophys.*; **263**: 107–16.
- Nishimura Y, Kawabata T, Kato K. (1988c) Identification of latent procathepsins B and L in microsomal lumen: characterization of enzymatic activation and proteolytic processing in vitro. *Arch Biochem Biophys.*; **261**: 64–71.
- Nishimura Y, Kawabata T, Furuno K, Kato K. (1989) Evidence that aspartic proteinase is involved in the proteolytic processing event of procathepsin L in lysosomes. *Arch Biochem Biophys.*; **271**: 400–6.
- Nishimura Y, Kato K, Furuno K, Himeno M. (1995) Inhibitory effect of leupeptin on the intracellular maturation of lysosomal cathepsin L in primary cultures of rat hepatocytes. *Biol Pharm Bull.*; **18**: 945–50.
- Nissler K, Kreusch S, Rommerskirch W, Strubel W, Weber E, Wiederanders B. (1998) Sorting of non-glycosylated human procathepsin S in mammalian cells. *Biol Chem.*; **379**: 219–24.
- Nomura T, Fujisawa Y. (1997) Processing properties of recombinant human procathepsin L. *Biochem Biophys Res Commun.*; **230**: 143–6.
- Oda K, Nishimura Y, Ikehara Y, Kato K. (1991) Bafilomycin A1 inhibits the targeting of lysosomal acid hydrolases in cultured hepatocytes. *Biochem Biophys Res Commun.*; **178**: 369–77.
- Ogino T, Kaji T, Kawabata M, Satoh K, Tomoo K, Ishida T, Yamazaki H, Ishidoh K, Kominami E. (1999) Function of the propeptide region in recombinant expression of active procathepsin L in *Escherichia coli*. *J Biochem (Tokyo).*; **126**: 78–83.
- Ohta Y, Hojo H, Aimoto S, Kobayashi T, Zhu X, Jordan F, Inouye M. (1991) Pro-peptide as an intramolecular chaperone: renaturation of denatured subtilisin E with a synthetic pro-peptide [corrected]. *Mol Microbiol.*; **5**: 1507–10.
- Pietschmann S, Fehn M, Kaulmann G, Wenz I, Wiederanders B, Schilling K. (2002) Foldase function of the cathepsin S proregion is strictly based upon its domain structure. *Biol Chem.*; **383**: 1453–8.
- Polgar L, Halasz P. (1982) Current problems in mechanistic studies of serine and cysteine proteinases. *Biochem J.*; **207**: 1–10.

- Quraishi O, Nagler DK, Fox T, Sivaraman J, Cygler M, Mort JS, Storer AC. (1999) The occluding loop in cathepsin B defines the pH dependence of inhibition by its propeptide. *Biochemistry*; **38**: 5017–23.
- Rieman DJ, McClung HA, Dodds RA, Hwang SM, Holmes MW, James IE, Drake FH, Gowen M. (2001) Biosynthesis and processing of cathepsin K in cultured human osteoclasts. *Bone*; **28**: 282–9.
- Roche L, Tort J, Dalton JP. (1999) The propeptide of *Fasciola hepatica* cathepsin L is a potent and selective inhibitor of the mature enzyme. *Mol Biochem Parasitol*; **98**: 271–7.
- Rowan AD, Mason P, Mach L, Mort JS. (1992) Rat procathepsin B. Proteolytic processing to the mature form *in vitro*. *J Biol Chem*; **267**: 15993–9.
- Rozman J, Stojan J, Kuhelj R, Turk V, Turk B. (1999) Autocatalytic processing of recombinant human procathepsin B is a bimolecular process. *FEBS Lett*; **459**: 358–62.
- Sajid M, McKerrow JH. (2002) Cysteine proteases of parasitic organisms. *Mol Biochem Parasitol*; **120**: 1–21.
- Salminen A, Gottesman MM. (1990) Inhibitor studies indicate that active cathepsin L is probably essential to its own processing in cultured fibroblasts. *Biochem J*; **272**: 39–44.
- Santamaria I, Velasco G, Pendas AM, Fueyo A, Lopez-Otin C. (1998) Cathepsin Z, a novel human cysteine proteinase with a short propeptide domain and a unique chromosomal location. *J Biol Chem*; **273**: 16816–23.
- Santilman V, Jadot M, Mainferme F. (2002) Importance of the propeptide in the biosynthetic maturation of rat cathepsin C. *Eur J Cell Biol*; **81**: 654–63.
- Schaschke N, Assfalg-Machleidt I, Machleidt W, Moroder L. (1998) Substrate/propeptide-derived endo-epoxysuccinyl peptides as highly potent and selective cathepsin B inhibitors. *FEBS Lett*; **421**: 80–2.
- Schilling K, Pietschmann S, Fehn M, Wenz I, Wiederanders B. (2001) Folding incompetence of cathepsin L-like cysteine proteases may be compensated by the highly conserved, domain-building N-terminal extension of the proregion. *Biol Chem*; **382**: 859–65.
- Schlereth A, Standhardt D, Mock HP, Muntz K. (2001) Stored cysteine proteinases start globulin mobilization in protein bodies of embryonic axes and cotyledons during vetch (*Vicia sativa* L.) seed germination. *Planta*; **212**: 718–27.
- Sivaraman J, Nagler DK, Zhang R, Menard R, Cygler M. (2000) Crystal structure of human procathepsin X: a cysteine protease with the proregion covalently linked to the active site cysteine. *J Mol Biol*; **295**: 939–51.
- Smith SM, Gottesman MM. (1989) Activity and deletion analysis of recombinant human cathepsin L expressed in *Escherichia coli*. *J Biol Chem*; **264**: 20487–95.
- Song J, Xu P, Xiang H, Su Z, Storer AC, Ni F. (2000) The active-site residue Cys-29 is responsible for the neutral-pH inactivation and the refolding barrier of human cathepsin B. *FEBS Lett*; **475**: 157–62.
- Tanaka Y, Tanaka R, Kawabata T, Noguchi Y, Himeno M. (2000) Lysosomal cysteine protease, cathepsin B, is targeted to lysosomes by the mannose 6-phosphate-independent pathway in rat hepatocytes: site-specific phosphorylation in oligosaccharides of the proregion. *J Biochem (Tokyo)*; **128**: 39–48.
- Tao K, Stearns NA, Dong J, Wu QL, Sahagian GG. (1994) The proregion of cathepsin L is required for proper folding, stability, and ER exit. *Arch Biochem Biophys*; **311**: 19–27.
- Taylor MAJ, Baker KC, Briggs GS, Connerton IF, Cummings NJ, Pratt KA, Revell DF, Freedman RB, Goodenough PW. (1995a) Recombinant pro-regions from papain and papaya proteinase IV are selective high-affinity inhibitors of the mature papaya enzymes. *Protein Eng*; **8**: 59–62.
- Taylor MAJ, Briggs GS, Baker KC, Cummings NJ, Pratt KA, Freedman RB, Goodenough PW. (1995b) Expression of the pro-regions of

- papain and papaya proteinase-IV in *Escherichia coli* and their inhibition of mature cysteine proteinases. *Biochem Soc Trans.*; **23**: S80–S80.
- Tobbell DA, Middleton BJ, Raines S, Needham MR, Taylor IW, Beveridge JY, Abbott WM. (2002) Identification of *in vitro* folding conditions for procathepsin S and cathepsin S using fractional factorial screens. *Protein Expr Purif.*; **24**: 242–54.
- Turk D, Podobnik M, Kuhelj R, Dolinar M, Turk V. (1996) Crystal structures of human procathepsin B at 3.2 and 3.3 Ångstroms resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide. *FEBS Lett.*; **384**: 211–4.
- van der Stappen JW, Williams AC, Maciewicz RA, Paraskeva C. (1996) Activation of cathepsin B, secreted by a colorectal cancer cell line requires low pH and is mediated by cathepsin D. *Int J Cancer.*; **67**: 547–4.
- Vernet T, Tessier DC, Richardson C, Laliberte F, Khouri HE, Bell AW, Storer AC, Thomas DY. (1990) Secretion of functional papain precursor from insect cells. Requirement for N-glycosylation of the pro-region. *J Biol Chem.*; **265**: 16661–6.
- Vernet T, Khouri HE, Laflamme P, Tessier DC, Musil R, Gour-Salin BJ, Storer AC, Thomas DY. (1991) Processing of the papain precursor. Purification of the zymogen and characterization of its mechanism of processing. *J Biol Chem.*; **266**: 21451–7.
- Vernet T, Berti PJ, de Montigny C, Musil R, Tessier DC, Menard R, Magny MC, Storer AC, Thomas DY. (1995) Processing of the papain precursor. The ionization state of a conserved amino acid motif within the Pro region participates in the regulation of intramolecular processing. *J Biol Chem.*; **270**: 10838–46.
- Visal S, Taylor MA, Michaud D. (1998) The proregion of papaya proteinase IV inhibits Colorado potato beetle digestive cysteine proteinases. *FEBS Lett.*; **434**: 401–5.
- Volkel H, Kurz U, Linder J, Klumpp S, Gnau V, Jung G, Schultz JE. (1996) Cathepsin L is an intracellular and extracellular protease in *Paramecium tetraurelia*. Purification, cloning, sequencing and specific inhibition by its expressed propeptide. *Eur J Biochem.*; **238**: 198–206.
- Wang B, Shi GP, Yao PM, Li Z, Chapman HA, Bromme D. (1998) Human cathepsin F. Molecular cloning, functional expression, tissue localization, and enzymatic characterization. *J Biol Chem.*; **273**: 32000–8.
- Wex T, Levy B, Smeekens SP, Ansorge S, Desnick RJ, Bromme D. (1998) Genomic structure, chromosomal localization, and expression of human cathepsin W. *Biochem Biophys Res Commun.*; **248**: 255–61.
- Wex T, Levy B, Wex H, Bromme D. (1999) Human cathepsins F and W: A new subgroup of cathepsins. *Biochem Biophys Res Commun.*; **259**: 401–7.
- Wex T, Buhling F, Wex H, Gunther D, Malferteiner P, Weber E, Bromme D. (2001) Human cathepsin W, a cysteine protease predominantly expressed in NK cells, is mainly localized in the endoplasmic reticulum. *J Immunol.*; **167**: 2172–8.
- Wiederanders B, Kirschke H. (1989) The processing of a cathepsin L precursor *in vitro*. *Arch Biochem Biophys.*; **272**: 516–21.
- Wiederanders B, Bromme D, Kirschke H, von Figura K, Schmidt B, Peters C. (1992) Phylogenetic conservation of cysteine proteinases. Cloning and expression of a cDNA coding for human cathepsin S. *J Biol Chem.*; **267**: 13708–13.
- Yamamoto Y, Watabe S, Kageyama T, Takahashi SY. (1999a) A novel inhibitor protein for *Bombyx* cysteine proteinase is homologous to propeptide regions of cysteine proteinases. *FEBS Lett.*; **448**: 257–60.
- Yamamoto Y, Watabe S, Kageyama T, Takahashi SY. (1999b) Proregion of *Bombyx mori* cysteine proteinase functions as an intramolecular chaperone to promote proper folding of the mature enzyme. *Arch Insect Biochem Physiol.*; **42**: 167–78.
- Yamamoto Y, Watabe S, Kageyama T, Takahashi SY. (1999c) Purification and characterization of *Bombyx* cysteine proteinase specific inhib-

- itors from the hemolymph of *Bombyx mori*. *Arch Insect Biochem Physiol.*; **42**: 119–29.
- Yamamoto Y, Kurata M, Watabe S, Murakami R, Takahashi SY. (2002) Novel cysteine proteinase inhibitors homologous to the proregions of cysteine proteinases. *Curr Protein Pept Sci.*; **3**: 231–8.
- Zhu XL, Ohta Y, Jordan F, Inouye M. (1989) Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. *Nature.*; **339**: 483–84.