

# Lysosomal cysteine proteases (cathepsins): promising drug targets

Dušan Turk\* and Gregor Gunčar

Department of Biochemistry and Molecular  
Biology, Jozef Stefan Institute, Jamova 39,  
1000 Ljubljana, Slovenia

Correspondence e-mail: dusan.turk@ijs.si

Dušan Turk is a research associate at Jozef Stefan Institute in Ljubljana, Slovenia, heading the Structural Biology group in the Department of Biochemistry and Molecular Biology. He received his PhD degree at the Technical University, Munich in 1992 in the laboratory of Professor Robert Huber at Max-Planck Institute of Biochemistry in Martinsried, Germany, after completing his BS in chemistry and masters degree in the area of computational chemistry at Ljubljana University (Chemical Institute). His postdoctoral experience was divided between the laboratories of Professors Robert Huber and Vito Turk, Jozef Stefan Institute and he started a macromolecular crystallography laboratory in the latter. His principal interests are in the structural biology of proteases, predominantly cysteine proteases, and their control mechanisms, and in the development of computational methods for macromolecular crystal structure determination assembled in the computer program MAIN. Gregor Gunčar obtained a degree in chemistry in 1995 and his doctorate in 2000, both from the University of Ljubljana, Slovenia. He joined the Structural Biology group at Jozef Stefan Institute after his BSc and worked in the field of the cysteine proteases and their inhibitors, of which he determined several structures. He is now continuing his work in the field and is looking forward to finding a nice postdoctoral position in the near future.

Papain-like lysosomal cysteine proteases are processive and digestive enzymes expressed in organisms from bacteria to humans. Their ubiquity alone makes them potential drug targets, with the assumption that appropriate specificities may be achieved. These enzymes have rather short active-site clefts, comprising three well defined substrate-binding subsites (S2, S1 and S1') and additionally have comparatively broad binding areas (S4, S3, S2', S3'). This geometry distinguishes them from other protease classes, such as serine and aspartic proteases, with six and eight substrate-binding sites, respectively. Exopeptidases (cathepsins B, C, H and X), in contrast to endopeptidases (such as cathepsins L, S, V and F), possess structural features that facilitate binding of N- and C-terminal groups of substrates in the active-site cleft. Other than a clear preference for free chain termini in the case of exopeptidases, the substrate-binding sites exhibit no strict specificities. Instead, their subsite preferences arise more from specific exclusions of substrate type. This presents a challenge for the design of inhibitors to target a specific cathepsin: only the cumulative effect of an assembly of inhibitor fragments can produce the desired result. The small number of papain-like lysosomal cysteine proteases (11 human enzymes are known) and the small number of substrate-binding sites calls for a innovative and empirical approach.

Received 19 July 2002

Accepted 20 November 2002

## 1. Introduction

Proteolytic activity is important for normal functioning of an organism and must be rigorously controlled to avoid potentially dangerous excess protein degradation. Failure in biological control mechanisms of proteolytic activities causes a wide range of diseases, among them cancer, rheumatoid arthritis and osteoarthritis, Alzheimer's disease, multiple sclerosis and muscular dystrophy (reviewed in Turk *et al.*, 2000). For many diseases resulting from excess proteolysis, no inhibitors have yet been identified with the necessary profile for therapeutic use. Thus, research into the physiological roles of proteases and into the discovery of substances to modulate them will remain a priority of both science and the pharmaceutical industry for the foreseeable future. For now, drug design proceeds hand in hand with the discovery of the biological roles of enzymes; when a specific role has been identified by an inhibitor, this compound is already a drug candidate.

About 500–600 proteases have been identified in the human genome (Lopez-Otin & Overall, 2002). Of these, about 60 are lysosomal proteases (Mason, 1995), which include a group of about a dozen papain-like lysosomal cysteine proteases. For historical reasons, intracellular proteases were named

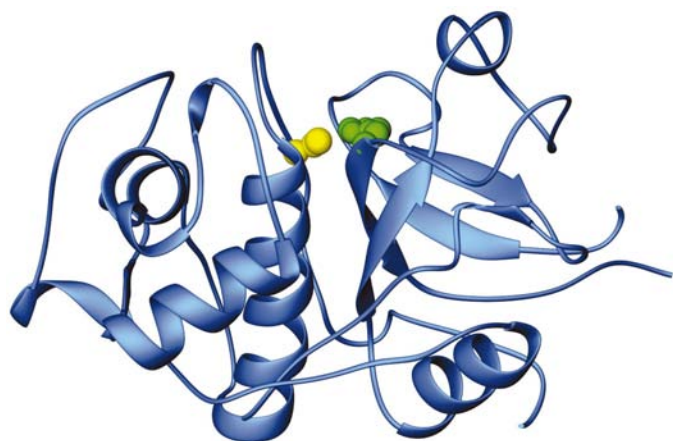
**Table 1**

Primary citations and the PDB codes of cathepsin structures.

The parallel entries indicate structures that were determined simultaneously by several groups.

Cathepsin	PDB code	Citation	Species
B	1huc	Musil <i>et al.</i> (1991)	Human
C	1jqp	Olsen <i>et al.</i> (2001)	Rat
	1k3b	Turk, Janjič <i>et al.</i> (2001)	Human
F	1m6d	Somoza <i>et al.</i> (2002)	Human
H	8pch	Gunčar <i>et al.</i> (1998)	Porcine
L	1icf	Gunčar <i>et al.</i> (1999)	Human
K	1mem	McGrath <i>et al.</i> (1997)	Human
	1atk	Zhao <i>et al.</i> (1997)	Human
S	1glo	Turkenburg <i>et al.</i> (2002)	Human
V	1fh0	Somoza <i>et al.</i> (2000)	Human
X	1ef7	Gunčar <i>et al.</i> (2000)	Human

cathepsins (the root of the word originates from the Greek language, where it means to digest); however, there is no strict rule that links the reactive mechanism and localization of cathepsins with their name. All known lysosomal cysteine proteases are cathepsins, but not all cathepsins are lysosomal or cysteine proteases. Cathepsins D and E are aspartic proteases, whereas cathepsins A and G are serine proteases; cathepsins E and G are not lysosomal proteases. Discovery of legumain (Chen *et al.*, 1997, 1998), also a lysosomal cysteine protease belonging to clan CD, added to the confusion in nomenclature. This review focuses on the group of papain-like cysteine proteases, which are ubiquitous among living organisms (including bacteria, viruses and plants, and lower and higher animals, including parasites). Particular attention is paid to human lysosomal enzymes (cathepsins) and their mammalian homologues. The relatively small size of the group, the uniquely reactive cysteine sulfohydryl group ( $pK_a$  in the range 2.5–3.5; Pinitglang *et al.*, 1997) and their unique reactive mechanism make these enzymes attractive targets for drug design. There are 11 human enzymes currently known



**Figure 1**

Fold of cathepsin L. Cathepsin L (1icf) is shown as a ribbon in its standard orientation, viewed along the two-domain interface with the central  $\alpha$ -helix in a vertical orientation and the active site at the top. The side chains of the catalytic residues Cys25 and His159 are shown as yellow and green atom balls, respectively. This figure and Fig. 2 were prepared using the program RIBBONS (Carson, 1997).

**Table 2**

Primary citations and the PDB codes of proenzyme cathepsin structures.

The parallel entries indicate structures that were determined simultaneously by several groups.

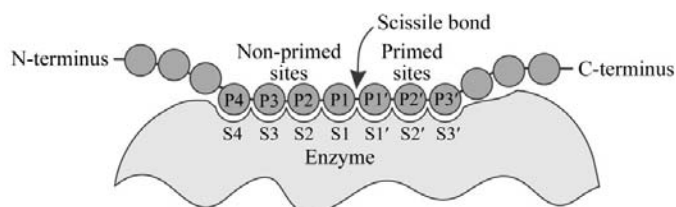
Zymogen	PDB code	Citation	Species
B	1mir	Cyglér <i>et al.</i> (1996)	Rat
	1pbh	Turk <i>et al.</i> (1996)	Human
	2pbh		
	3pbh	Podobnik <i>et al.</i> (1997)	Human
L	1cjl	Coulombe <i>et al.</i> (1996)	Human
K	7pck	Sivaraman <i>et al.</i> (1999)	Human
	1by8	LaLonde <i>et al.</i> (1999)	Porcine
X	1deu	Sivaraman <i>et al.</i> (2000)	Human

(cathepsins B, C, F, H, L, K, O, S, V, X and W; Turk *et al.*, 2000; Turk, Turk *et al.*, 2001) and it is quite likely that the list has already been completed. Human gene data bank searches



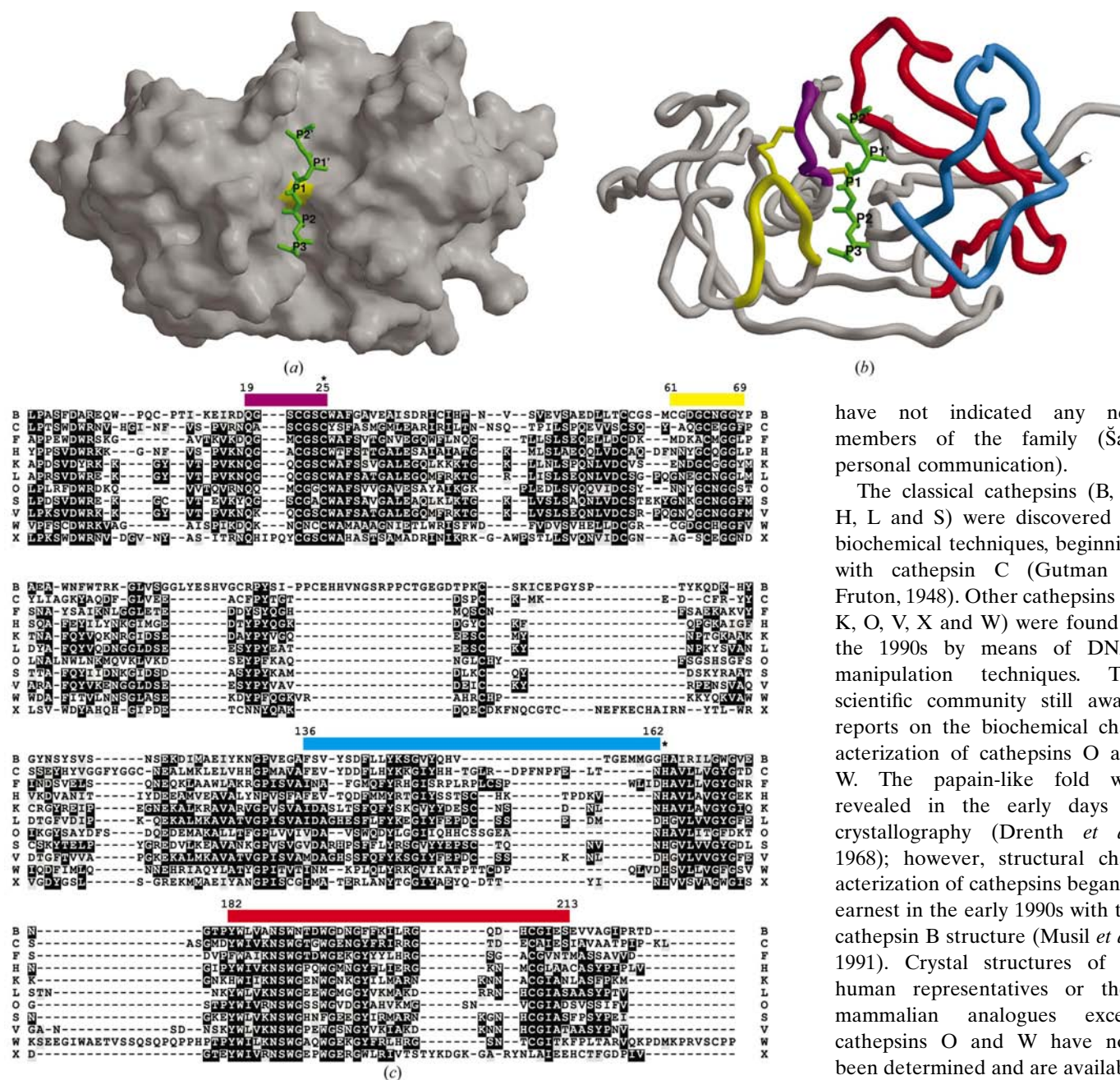
**Figure 2**

Fold of procathepsin L (1cjl). The mature enzyme part of cathepsin L is shown in blue and the propeptide is shown in red.



**Figure 3**

Schechter and Berger's definition of substrate-binding sites (Schechter & Berger, 1967).



**Figure 4**

Substrate-binding sites. (a) A view from the top: a polyalanine substrate model bound in the active-site cleft of cathepsin L. (Modelling of the binding geometry of a substrate is based on information gained from the crystal structures of substrate-analogue inhibitors and their interactions with a papain-like protease active site; see Figs. 5a, 5b and 5c). Substrate residues are shown as green sticks and are denoted using the Schechter and Berger nomenclature. Cathepsin L is shown with a grey surface representation. The surface of the catalytic cysteine side chain is yellow. (b) The same as Fig. 4(a), only that in this case cathepsin L is shown as a chain trace. Most of the chain is grey, whereas the loops building the substrate-binding sites are colour-coded: the L-domain loops (19–25 and 61–69) are purple and yellow and the R-domain loops (136–162 and 182–213) are blue and red. The loops building the substrate-binding sites do not only contain the residues directly contributing to the surface, but also include those that provide the foundation for it. (c) Structure-based amino-acid alignment of sequences of papain-like domains of all known human cathepsins. Structural alignment was made using the program *Modeller* (Šali and Blundell, 1993), then the sequences of cathepsins F, O and W were aligned to the template with the *ClustalW* program (Higgins *et al.*, 1996). The sequences were taken from SWISS-PROT or GENBANK databases and the structures from the PDB. The loops building the substrate-binding sites are marked at the top marked with stripes and using the same colour code as in Fig. 4(b). Figs. 4(a), 4(b), 5, 8 and 9 were prepared with *MAIN* (Turk, 1992) and rendered with *Raster3D* (Merritt & Bacon, 1997). The cathepsin L surface was generated with *GRASP* (Nicholls *et al.*, 1991).

have not indicated any new members of the family (Šali, personal communication).

The classical cathepsins (B, C, H, L and S) were discovered by biochemical techniques, beginning with cathepsin C (Gutman & Fruton, 1948). Other cathepsins (F, K, O, V, X and W) were found in the 1990s by means of DNA-manipulation techniques. The scientific community still awaits reports on the biochemical characterization of cathepsins O and W. The papain-like fold was revealed in the early days of crystallography (Drenth *et al.*, 1968); however, structural characterization of cathepsins began in earnest in the early 1990s with the cathepsin B structure (Musil *et al.*, 1991). Crystal structures of all human representatives or their mammalian analogues except cathepsins O and W have now been determined and are available from the PDB (Table 1). The relevance of cathepsins as potential drug targets is best indicated by the fact that four (K, S, V and recently F) of the nine structures of cathepsins were published by industrial research groups. Structures of cathepsins L (Fujishima *et al.*, 1997) and S (McGrath *et al.*, 1998) have been also reported by industrial groups but are not yet publically available. An additional publication describing the complexes of cathepsin S with synthetic inhibitors is in preparation (Cygler & Rath, personal

communication). So far, structures of four zymogens are known (Table 2). Their structure determination generally followed the structures of the active enzymes. An exception was the procathepsin L structure (Coulombe *et al.*, 1996), which preceded the mature enzyme structure by 3 years.

### 2. Physiological roles and localization

The papain-like lysosomal cysteine proteases have long been believed to be responsible for protein degradation in lysosomes (Kirschke *et al.*, 1995). Analyses of gene knockouts suggested that this function is not exclusively dependent on any single cathepsin (Saftig *et al.*, 1998; Shi *et al.*, 1999; Pham & Ley, 1999; Deussing *et al.*, 1998; Nakagawa *et al.*, 1998, 1999; Roth *et al.*, 2000). However, analyses of gene knockouts and the locations of mutations on genes of lysosomal cysteine proteases responsible for some hereditary diseases revealed several specific biological functions. These functions are a consequence of limited proteolysis of their target substrates and additionally rely on co-localization and timing.

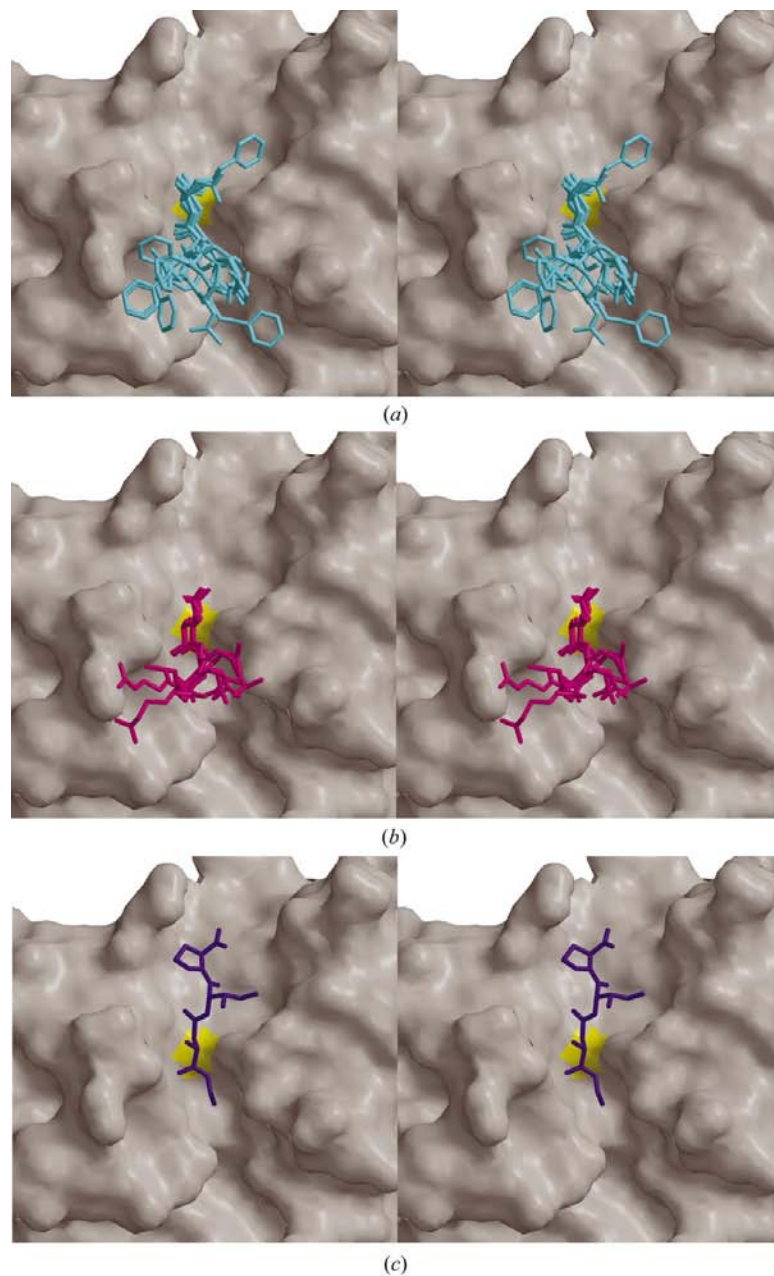
(i) Cathepsin K was found to be crucial in bone remodelling (Chapman *et al.*, 1997; Saftig *et al.*, 1998).

(ii) Cathepsin S is the major processing enzyme of the MHC class II associated invariant chain and is thus essential for the normal functioning of the MHC class II associated antigen processing and presentation (Nakagawa *et al.*, 1998, 1999; Shi *et al.*, 1999). Cathepsins L and F were shown to participate in the same process, primarily in tissues or cells not expressing cathepsin S (Nakagawa *et al.*, 1998; Shi *et al.*, 2000), although the role of the former has probably been taken by cathepsin V in humans (Brömme *et al.*, 1999).

(iii) Cathepsin L-deficient mice developed periodic hair loss and epidermal hyperplasia, indicating that cathepsin L is involved in epidermal homeostasis and regular hair-follicle morphogenesis and cycling (Roth *et al.*, 2000). One-year-old cathepsin L-deficient mice (Stypmann *et al.*, 2002) exhibited histomorphological and functional alterations of the heart, resulting in dilated cardiomyopathy, which is a frequent cause of heart failure.

(iv) Cells derived from cathepsin C-deficient mice fail to activate groups of serine proteases from granules of immune (cytotoxic T lymphocytes, natural killer cells) and inflammatory (neutrophils, mast cells) cells primarily involved in the defence of the organism, demonstrating that cathepsin C is involved in their activation (Pham & Ley, 1999; Wolters *et al.*, 2001). The current list of unprocessed zymogens of proteases in cathepsin C knockout mice contains granzymes A, B and C, cathepsin G, neutrophil elastase and a chymase.

More data about the processes in which the lysosomal papain-like cysteine proteases participate can be found elsewhere (Kirschke *et al.*, 1995; Chapman *et al.*, 1997; Barrett *et al.*, 1998; McGrath, 1999; Turk *et al.*, 2000; Turk, Turk *et al.*, 2001; Brömme & Kaleta, 2002).



**Figure 5**

Low-molecular-weight inhibitor-binding geometry. The inhibitors (shown as sticks) from structures of complexes with papain-like cysteine proteases are superimposed on top of the cathepsin L surface. The catalytic site Cys25 surface is coloured yellow. PDB codes are given in parentheses. Complexes with plant enzymes are also included. (a) Substrate-analogue inhibitors: fluoro- and chloromethylketone-based inhibitors and leupeptin are shown in light blue. Inhibitors are taken from structures of complexes with the following enzymes: cruzipain (1aim, 2aim), papain (1pad, 1pop, 5pad, 6pad), glycyI endopeptidase (1gec) and cathepsin B (1the, 1cte). (b) E-64 and derivative are shown in magenta. Inhibitors are taken from structures of complexes with the following enzymes: actinidin (1aec), caricain (1meg), cathepsin K (1atk) and papain (1pe6, 1ppp). (c) CA030 inhibitor from the complex with cathepsin B (1csb) is shown in blue.

### 3. Pathology

Lysosomal cysteine proteases have been found to be associated with a number of pathologies, including cancer, inflammation, rheumatoid arthritis and osteoarthritis, Alzheimer's disease, multiple sclerosis, muscular dystrophy, pancreatitis, liver disorders, lung disorders, lysosomal disorders, Batten's disease, diabetes and myocardial disorders. In many of these diseases, the lysosomal enzymes have been found in the extracellular and extralysosomal environment in their (zymogenic) 'pro' forms, which are substantially more stable than the mature enzymes (reviewed in Kirschke *et al.*, 1995; Chapman *et al.*, 1997; Barrett *et al.*, 1998; Kos & Lah, 1998; Turk, Turk *et al.*, 2001). Cathepsins also participate in apoptosis, although the exact mechanism is not yet clear (Stoka *et al.*, 2001; Salvesen, 2001; Leist & Jüttelä, 2001; Turk *et al.*, 2002).

Several genetic disorders have been traced to genes of lysosomal cysteine proteases. Pycnodysostosis, an autosomal recessive osteochondrodysplasia characterized in humans by severe bone abnormalities, was found to be associated with the loss-of-function mutation of cathepsin K (Gelb *et al.*, 1996), while loss-of-function mutation in the cathepsin C gene leads to Papillon-Lefevre syndrome, an autosomal recessive

disorder characterized in patients by palmoplantar keratosis and severe early-onset periodontitis (Toomes *et al.*, 1999; Hart *et al.*, 1999; Hart, Hart, Michalec, Zhang, Firatli *et al.*, 2000; Hart, Hart, Michalec, Zhang, Marazita *et al.*, 2000; Allende *et al.*, 2001). These effects are quite likely to be a result of incomplete processing of some as yet unidentified proteases presumably involved in establishing or maintaining the structural organization of the epidermis of the extremities and the integrity of the tissues surrounding the teeth and in the processing of proteins such as keratins (Nuckolls & Slavkin, 1999). In addition, cathepsin C may be involved in chronic airway diseases such as asthma (Wolters *et al.*, 2000).

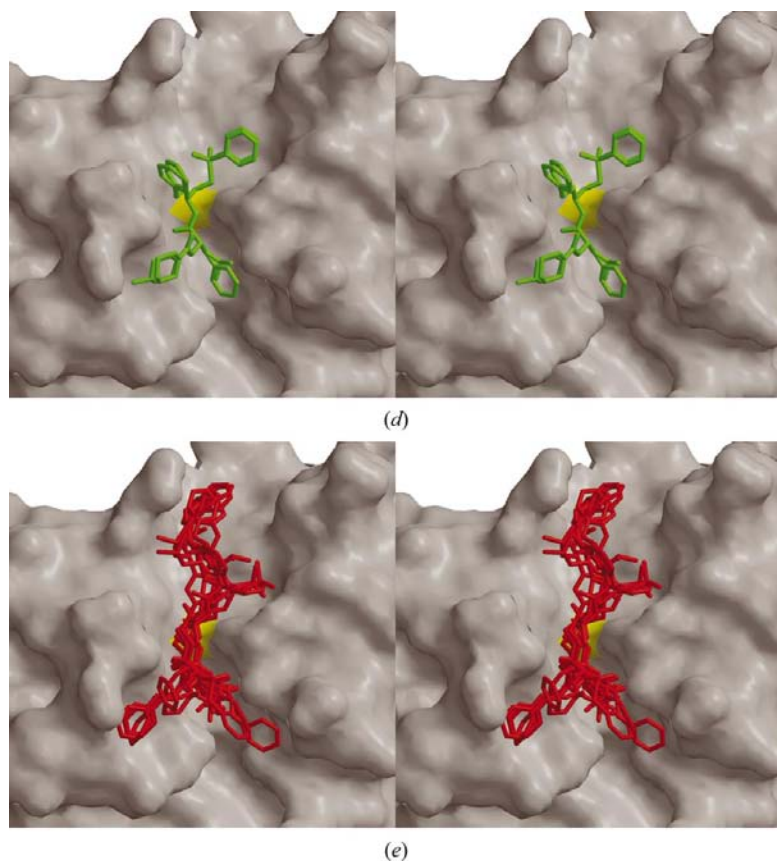
Similarly, down-regulation of natural inhibitors, as demonstrated by a mutation in the gene for stefin B, predisposes affected individuals to a hereditary form of myoclonal epilepsy (Pennacchio *et al.*, 1996; Lalioti *et al.*, 1997).

### 4. Fold and specificity

The papain-like lysosomal cysteine proteases are monomeric proteins with MW between 22 and 28 kDa. The only exception is cathepsin C, which is a tetrameric molecule with an MW of 200 kDa (Dolenc *et al.*, 1995). They all share the common fold of a papain-like structure. Cathepsin L, as a typical endo-

peptidase, has been chosen as a representative of the family (Fig. 1). A papain-like fold consists of two domains, reminiscent of a closed book with the spine at the front. The domains separate at the top in a V-shaped active-site cleft, in the middle of which the residues Cys25 and His159, one from each domain, form the catalytic site of the enzyme. The most prominent feature of the left (L) domain is central  $\alpha$ -helix of about 30 residues in length; the right (R) domain forms a kind of  $\beta$ -barrel, which includes a shorter  $\alpha$ -helical motif. (The terms left and right domain refer to the standard view shown in Fig. 1.)

Lysosomal cathepsins are encoded as 'pre-proenzymes'. Following cotranslational cleavage of an amino-terminal signal peptide that mediates transport across the endoplasmic reticulum membrane (Erickson, 1989), procathepsins undergo proteolytic processing to the active mature enzyme form in the acidic environment of late endosomes or lysosomes (Nishimura *et al.*, 1988; Kominami *et al.*, 1988). The crystal structures of proenzymes (Table 2) showed that the structure of the mature enzyme is already formed in the zymogen form. Propeptide chain builds a  $\alpha$ -helical domain, which continues along the active-site cleft towards the N-terminus of the mature enzyme in a predominantly extended conformation in the direction opposite to substrate binding, blocking access to the active site. The procathepsin L structure (Coulombe *et al.*, 1996) has been chosen as a representative of the family (Fig. 2). Propeptides are in fact inhibitors of their cognate enzymes, as demonstrated by kinetic data (Guay *et al.*, 2000). Among them, cathepsin K (Sivaraman *et al.*



**Figure 5 (continued)**

Low-molecular-weight inhibitor-binding geometry. (d) Vinylsulfone-based inhibitors taken from structure of complexes with cathepsin K (1mem) and cathepsin V (1fh0) are shown in green. (e) A group of non-covalent cathepsin K inhibitors are shown in red (1ayu, 1ayv, 1ayw, 1au0, 1bgo, 1au2, 1au3, 1au4).

*al.*, 1999; LaLonde *et al.*, 1999) has the longest N-terminal peptide and cathepsin X (Sivaraman *et al.*, 2000) the shortest. The propeptide of cathepsin X is also the only one covalently attached to the reactive-site cysteine *via* a disulfide bond.

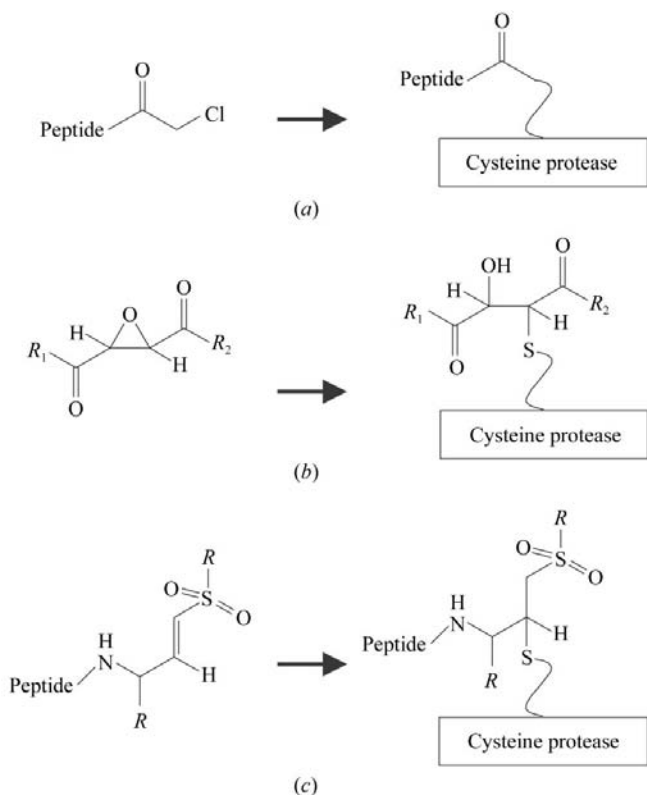
#### 4.1. Substrate-binding sites

When in 1967 Schechter and Berger reported their fundamental work on the substrate-binding sites of papain they had to rely solely on kinetic data (Schechter & Berger, 1967). They studied the dependence of substrate kinetics on the length of a polyaniline chain and discovered the kinetics to be influenced by the polypeptide-chain length up to a length of seven amino acids and concluded that there are seven substrate-binding sites on the papain molecule. Their definition of substrate-enzyme interactions and their nomenclature became the standards (Fig. 3) for the assignment of interaction sites of a polypeptide substrate and a proteolytic enzyme.

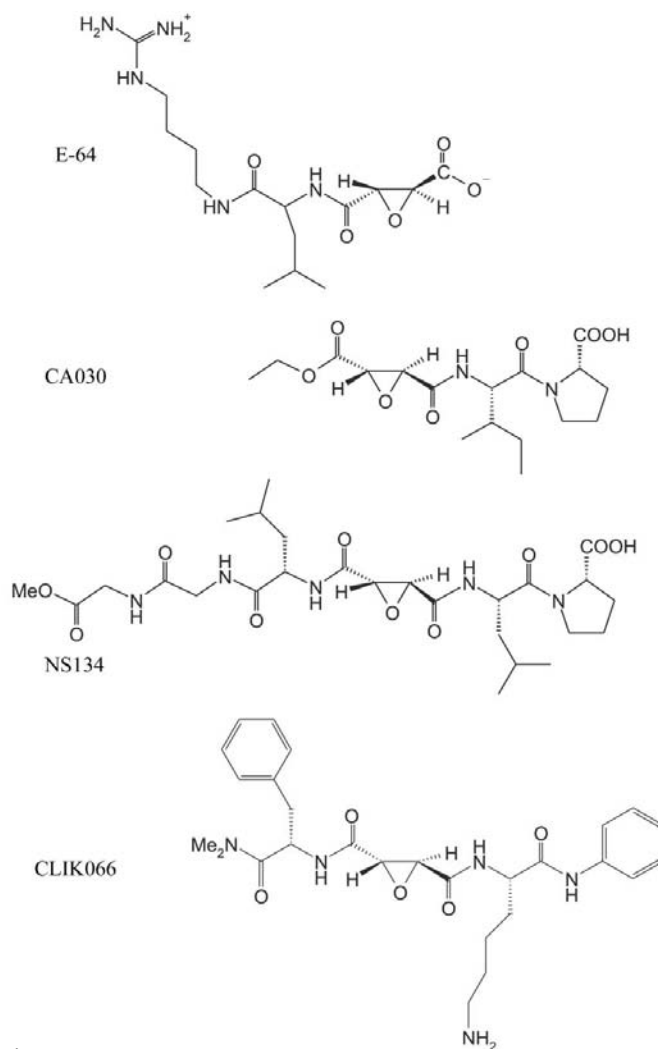
Three decades later, when a sufficient number of protease-inhibitor structures became available, the definition of Schechter and Berger substrate-binding sites on the papain-like enzymes was revisited and redefined (Turk *et al.*, 1998). The base and walls of the substrate-binding sites are formed by four chain segments comprising two shorter loops in the L-domain (19–25, 61–69) and two longer loops in the R-domain (136–162, 182–213; Fig. 4). A third loop from the L-domain might also be named if the disulfide (Cys22–Cys65)

which connects the two L-domain loops at the top is considered to be an additional loop closure.

The superimposed structures of complexes of substrate-analogue inhibitors and cathepsins (Figs. 5*a* and 5*c*) have revealed that substrate residues bind along the active-site cleft in an extended conformation, with the side chains alternately oriented toward the L- and R-domains. Residues P2, P1 and P1' bind into well defined binding sites. Positioning of these residues is governed by interactions which involve both main-chain and side-chain atoms. The S2 binding site is a deep pocket, whereas the S1 and S1' sites provide a binding surface. The positioning of the P3 residue is mediated only by side-chain interactions. For this reason, the binding geometries of the latter are scattered over a broad area and are unique for each substrate. On the prime side of the binding cleft, the P2' residue-binding site appears to be quite well defined. However, current knowledge is based on specific interactions between CA030 and the parts of cathepsin B structure responsible for its carboxydi-peptidase activity (Fig. 5*c*) (Turk *et al.*, 1995). It thus remains possible that interactions within the S2' site of an endopeptidase would be different.



**Figure 6**  
Schemes of three most frequent reactive groups before and after binding to the reactive-site cysteine. (a) Chloromethylketone, (b) epoxysuccinyl, (c) vinylsulfone.



**Figure 7**  
Alignment of epoxysuccinyl derivatives.

The location of the substrate-binding sites beyond S3 and S2' is not constrained by main-chain interactions. Each substrate residue docks on the surface of an enzyme in its own way (Fig. 5a). In particular, for the non-primed binding sites, there is evidence that a common S4 binding site and also an S3' site do not exist. Therefore, it was suggested that the substrate residue-binding regions beyond S2 and S2' should not be called sites but areas (Turk *et al.*, 1998). The papain-like proteases thus represent a special class of proteolytic enzymes with the smallest number of substrate-binding sites, as opposed to chymotrypsin-like serine proteases which have six (Bode & Huber, 1992) and aspartic proteases which have eight binding sites (Wlodawer & Gustchina, 2000).

#### 4.2. Binding of low-molecular-weight inhibitors

The rather short binding area seems to facilitate covalent interactions with low-molecular-weight inhibitors. Covalent interactions, however, impose hard constraints on the binding geometry. It thus took some time to design inhibitors that bind into the primed as well as non-primed side of the active-site cleft.

The structures of the first complexes of substrate-analogue inhibitors, based on the chloromethyl reactive group, with papain clarified the substrate binding in the non-primed binding sites in the 1970s (Drenth *et al.*, 1976). Other structures followed later (Fig. 5a). At about the same time, a natural cysteine protease inhibitor named E-64 was discovered (Aoyagi & Umezawa, 1975; Hanada *et al.*, 1978). E-64 utilizes an epoxysuccinyl group to covalently interact with the reactive-site cysteine (Fig. 6). Structures of E-64 (Varughese *et al.*, 1992) and its analogues (Yamamoto *et al.*, 1991) revealed that they bind into the non-primed region of the active site, but in the direction of propeptide binding and opposite to substrate binding (Fig. 5b).

The crystal structure of CA030 in complex with human cathepsin B (Turk *et al.*, 1995) showed that E-64 derivatives can also bind into the primed binding side in the direction of a substrate binding. Switching of the binding side was made possible by the specific interactions. The carboxylic group of the C-terminal residue of CA030 mimics the C-terminus of a substrate and docks against the occluding loop residues His110 and His111 (Fig. 5c). Alignment of the E-64 and CA030 binding geometries showed (Fig. 7; Turk *et al.*, 1995) that the epoxysuccinyl group possesses internal symmetry with two carboxylic heads, mimicking a polypeptide C-terminus to which amino-acid residues can be attached. The synthesis of double-head inhibitors followed (Schaschke *et al.*, 1997, 2000; Katunuma *et al.*, 1999). The binding geometry of the double-head inhibitor design has recently been confirmed by the crystal structures of cathepsin L and cathepsin B–inhibitor complexes (Tsuge *et al.*, 1999; Stern *et al.*, unpublished results).

The S1' binding site can also be reached with inhibitor constructs using the vinylsulfone reactive group (Fig. 6c; McGrath *et al.*, 1997; Somoza *et al.*, 2000, 2002) and exceptionally even by a long side chain of a P1-mimicking residue of a chloromethyl inhibitor (Figs. 5a and 5d; Jia *et al.*, 1995).

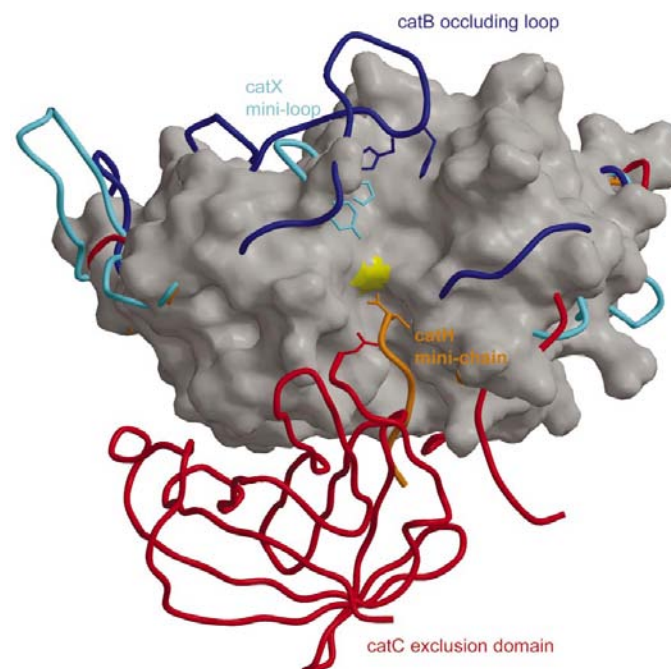
The covalent interaction with the reactive-site cysteine is not mandatory as shown by a series of 'Smith–Kline' compounds (Fig. 5e), which utilize various constructs to tightly block the reactive site, but are not engaged in covalent interactions (Thompson *et al.*, 1997).

Additional information regarding inhibitors and their chemistry can be found elsewhere (Shaw, 1990; Otto & Schirmeister, 1997; Brömme & Kaleta, 2002).

#### 4.3. Exopeptidases

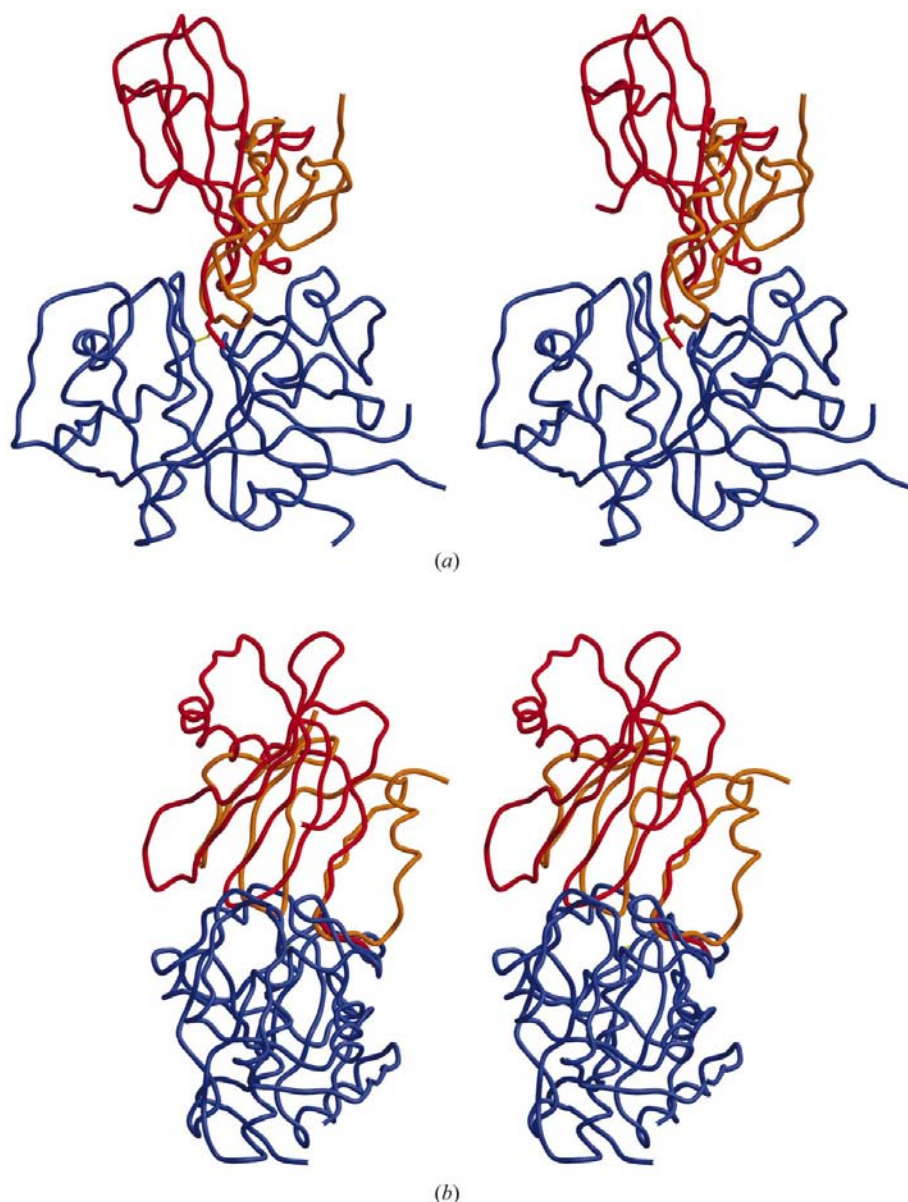
Whereas in endopeptidases (cathepsins F, L, K, O, S and V) the active-site cleft extends along the whole length of the two-domain interface, the exopeptidases (cathepsin B, C, H and X) possess additional features that reduce the number of substrate-binding sites (Fig. 8). The role of these features is dual: they prevent the binding of longer peptidyl substrates and they dock with charged N or C chain termini of substrates by utilizing selective electrostatic interactions.

Carboxydipeptidase cathepsin B (Musil *et al.*, 1991) has an insertion of about 20 residues, termed the occluding loop, which blocks the active-site cleft on the primed binding side beyond S2' and provides two histidine residues, His110 and His111, that bind to the charged main-chain carboxylic group of the C-terminal residue of a substrate. Cathepsin B also exhibits an endopeptidase activity that is made possible by the flexible occluding loop, which can be displaced from the



**Figure 8**

Features of exopeptidases. Chain traces of cathepsins H (8pch), C (1k3b), B (1huc) and X (1ef7), coloured orange, red, dark blue and light blue, respectively, are shown superimposed on the cathepsin L structure viewed from the top. The surface of cathepsin L is shown in grey; a yellow colour denotes the surface of the catalytic residue Cys25. Structural elements facilitating the exopeptidase activity are labelled. Residues that play a crucial role in exopeptidase specificity are shown in stick representation.



**Figure 9**

Binding of protein inhibitors. Stefin B superimposed on cathepsin L-p41 complex in views (a) across and (b) along the active-site cleft of cathepsin L. (a) is shown in approximately the standard view (Fig. 1), whereas (b) is generated by an  $90^\circ$  rotation about the vertical axis. Superposition of the p41 fragment and stefin B is based on the three-dimensional alignment of papain and cathepsin L structures in the papain-stefin B and cathepsin L-p41 fragment complexes. Chain traces of the p41 fragment, stefin B and cathepsin L are shown in orange, red and blue, respectively.

active-site cleft (Illy *et al.*, 1997; Podobnik *et al.*, 1997; Nagler *et al.*, 1997).

Cathepsin X is primarily a carboxymonopeptidase (Nagler *et al.*, 1999), which can also act as a carboxydipeptidase (Klemenčič *et al.*, 2000). The crystal structure showed that a histidine residue, His23, positioned within a short loop termed a mini-loop (Nagler *et al.*, 1999), is the anchor for the carboxylic group of the C-terminal substrate residue (Gunčar *et al.*, 2000). In the free-enzyme structure, the histidine ring occupies the position which is the S2' substrate-binding site in related cathepsins. This structure thus corresponds to the

carboxymonopeptidase mode of cathepsin X. A simple modelling study (manual rotation about the side-chain bonds) suggested that the histidine ring can adapt a position equivalent to His110 of cathepsin B. This cathepsin B-like position would thus correspond to the carboxydipeptidase mode.

Cathepsin H is an aminomonopeptidase. The crystal structure of the porcine enzyme (Gunčar *et al.*, 1998) revealed that an eight-residue segment of the propeptide, called the mini-chain, binds in the active-site cleft of the enzyme in the direction of a bound substrate. The negatively charged carboxylic group of its C-terminal residue, Thr83P, attracts the positively charged N-terminus of a substrate and thereby facilitates the aminopeptidase activity of cathepsin H. Thr83P mimics a substrate P2 residue by occupying the position that is the S2 binding site in related enzymes. The mini-chain is additionally fastened to the enzyme surface by a four-residue insertion (Lys155A-Asp155D) and a carbohydrate chain attached to Asn112. The positioning of the cathepsin H mini-chain closely resembles the positioning of the C-terminus of a distant homologue, bleomycin hydrolase (Joshua-Tor *et al.*, 1995).

Cathepsin C (also termed dipeptidyl peptidase I or DPPI) is an aminodipeptidase. The four independent active sites of cathepsin C are located on the external surface of the tetrahedral molecule. In contrast, oligomeric proteolytic machineries such as 20S proteasome (Lowe *et al.*, 1995; Groll *et al.*, 1997), bleomycin hydrolase (Joshua-Tor *et al.*, 1995), trypsin (Pereira *et al.*, 1998) and tricorn protease (Brandstetter *et al.*, 2001) have their active sites on the inside

surface. Proteasomes are barrel-like structures composed of four rings of  $\alpha$ - and  $\beta$ -subunits, which cleave unfolded proteins captured in the central cavity into short peptides. Trypsins are flat tetramers with a central pore in which the active sites reside. The pore restricts the size of accessible substrates and inhibitors. Similarly, the active sites of bleomycin hydrolase and tricorn protease are also located within the hexameric structure. The exposed active sites make cathepsin C a unique oligomeric protease capable of the hydrolysis of protein substrates in their native state regardless of their size. Its design, supported by the oligomeric structure, confines the



activity of the enzyme to an aminodipeptidase and thereby makes it suitable for use in many different environments, where cathepsin C can selectively activate a group of chymotrypsin-like proteases and presumably also other proteins.

The active site of cathepsin C is blocked beyond the S2 binding site by the massive body of the exclusion domain (Turk, Janjič *et al.*, 2001; Olsen *et al.*, 2001). An exposed  $\beta$ -hairpin, the first N-terminal residues of the exclusion domain and the carbohydrate ring attached to Asn5 block undesired access, while Asp1 with its carboxylic group side chain controls entry into the S2 binding pocket by fixing the N-terminal amino group of the substrate. Asp1 simultaneously prevents the positively charged side chains of arginine and lysine residues from binding in the S2 binding pocket. An additional special feature of cathepsin C is the dependence of its activity on chloride ions. One was located at the bottom of the very long S2 binding pocket.

Interestingly, structural comparison and similar interactions within the active-site cleft (Turk, Janjič *et al.*, 2001) suggested that the exclusion domain of cathepsin C was adapted from a metalloprotease inhibitor (Baumann *et al.*, 1995). The N-terminus of the exclusion domain only blocks access to a portion of the active-site cleft, whereas the N-terminus of the metalloprotease inhibitor binds along the primed binding sites and interacts with the reactive-site zinc ion.

#### 4.4. Specificity

Papain-like cathepsins are rather non-specific enzymes with no clear substrate-recognition site. This does not imply that specific inhibitors cannot be designed. It only suggests that specificity is not an issue involving a single binding site, but is rather a cumulative contribution of all interactions. This suggests that inhibitor constructs interacting with regions on both sides of reactive site can be advantageous compared with those which bind to only one side. The specificity of exopeptidases is, however, more a matter of exclusive interactions of free chain termini than side-chain recognition. The design of exopeptidase inhibitors therefore seems easier, as such inhibitors can rely on covalent interactions with the reactive site and electrostatic interactions with negatively charged carboxylic groups or positively charged histidines in the cases of aminopeptidases and carboxypeptidases, respectively.

#### 5. Hints from interactions with protein inhibitors

Stefins and cystatins are rather non-specific endogenous inhibitors of cysteine proteases. They are only able to discriminate between endo- and exopeptidases. Whereas the inhibition of endopeptidases is rapid and tight, almost being pseudo-irreversible, with  $K_i$  values in the picomolar range, the inhibition of exopeptidases is much weaker with  $K_i$  values in the millimolar to nanomolar range (reviewed in Turk & Bode, 1991; Turk *et al.*, 2000). Similar to stefins, an inhibitory fragment of the p41 form of MHC class II associated invariant chain (termed the p41 fragment) inhibits endopeptidase

cathepsin L ( $K_i = 1.7 \text{ pM}$ ) and exopeptidase cathepsin H ( $K_i = 5.3 \text{ nM}$ ); however, it does not inhibit endopeptidase cathepsin S and exopeptidase cathepsin B (Bevec *et al.*, 1996). How can this be explained on a structural basis?

Two crystal structures provided insight into the interactions between a papain-like cysteine protease and its protein inhibitor: those of the complexes of papain–stefin B (Stubbs *et al.*, 1990) and cathepsin L–p41 fragment (Fig. 9; Gunčar *et al.*, 1999).

The wedge shape and the three-loop arrangement of the p41 fragment bound to the active-site cleft of cathepsin L is reminiscent of the inhibitory edge of cystatins and thus demonstrate the first observed example of convergent evolution in the cysteine protease inhibitors. The interactions within the active-site cleft are non-specific. They are either hydrophobic or mediated *via* solvent molecules or involve hydrogen-bond interactions with groups which are conserved throughout the family of papain-like enzymes. This suggests that the p41 fragment, like the stefins, displaces the characteristic structural elements of exopeptidases from the active site, as revealed by the stefin A–cathepsin H complex (Jenko, unpublished results). However, the different fold of the p41 fragment results in additional contacts with the highly variable regions of the loops at the top of the R-domain, which build the surface of the S2 and S1' substrate-binding sites. This enables the p41 fragment to form specific interactions with its target enzymes and simultaneously prevents the approach of cathepsins S and B (Gunčar *et al.*, 1999).

#### 6. Prospectives

Currently, no drug targeted towards a lysosomal cysteine protease is in use; however, many are in development. The knowledge and experience gathered in the field suggest that there are enough leads to drive the research. The protein–inhibitor complexes, however, suggest to designers of low-molecular-weight inhibitors that there are still unexplored areas on the surface of the enzymes. In particular, it may be worthwhile to explore them in the case of endopeptidases, which have fewer constraints within the active-site cleft than the exopeptidases.

The drug-design process is challenged also from another point of view. Are the mature human enzymes really the most appropriate targets for potential drugs? Labelled inhibitors targeting a papain-like cysteine protease from *Trypanosoma cruzi* indicate that such inhibitors may interact already with their proenzyme form within the Golgi complex (Engel *et al.*, 1998).

The Slovenian Ministry of Education, Science and Sport and the ICGEB are gratefully acknowledged for their financial support.

#### References

- Allende, L. M., Garcia-Perez, M. A., Moreno, A., Corell, A., Corasol, M., Martinez-Canut, P. & Arnaiz-Villena, A. (2001). *Hum. Mutat.* **17**, 152–153.

- Aoyagi, T. & Umezawa, H. (1975). *Proteases in Biological Control*, edited by D. B. Rifkin & E. Shaw, pp. 429–454. Cold Spring Harbor: Cold Spring Harbour Laboratory Press.
- Barrett, A. J., Rawlings, N. D. & Woessner, J. F. Jr (1998). *Handbook of Proteolytic Enzymes*, edited by A. J. Barrett, N. D. Rawlings & J. F. Woessner Jr. London: Academic Press.
- Baumann, U., Bauer, M., Letoffe, S., Delepelaire, P. & Wandersman, C. (1995). *J. Mol. Biol.* **248**, 653–661.
- Bevec, T., Stoka, V., Pungerčič, G., Dolenc, I. & Turk, V. (1996). *J. Exp. Med.* **183**, 1331–1338.
- Bode, W. & Huber, R. (1992). *Eur. J. Biochem.* **204**, 433–451.
- Brandstetter, H., Kim, J. S., Groll, M. & Huber, R. (2001). *Nature (London)*, **414**, 466–470.
- Brömme, D. & Kaleta, J. (2002). *Curr. Pharm. Des.* **8**, 1639–1658.
- Brömme, D., Li, Z., Barnes, M. & Mehler, E. (1999). *Biochemistry*, **38**, 2377–2385.
- Carson, M. (1997). *Methods Enzymol.* **277**, 493–505.
- Chapman, H. A., Riese, J. P. & Shi, G.-P. (1997). *Annu. Rev. Physiol.* **59**, 63–88.
- Chen, J. M., Dando, P. M., Rawlings, N. D., Brown, M. A., Young, N. E., Stevens, R. A., Hewitt, E., Watts, C. & Barrett, A. J. (1997). *J. Biol. Chem.* **272**, 8090–8098.
- Chen, J. M., Rawlings, N. D., Stevens, R. A. & Barrett, A. J. (1998). *FEBS Lett.* **441**, 361–365.
- Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J. S. & Cygler, M. (1996). *EMBO J.* **15**, 5492–5503.
- Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A. C. & Mort, J. S. (1996). *Structure*, **4**, 405–416.
- Deussing, J., Roth, W., Saftig, P., Peters, C., Ploegh, H. L. & Villadangos, J. A. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 4516–4521.
- Dolenc, I., Turk, B., Pungerčič, G., Ritonja, A. & Turk, V. (1995). *J. Biol. Chem.* **270**, 21626–21631.
- Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M. & Wolthers, B. G. (1968). *Nature (London)*, **218**, 929–932.
- Drenth, J., Kalk, K. H. & Swen, H. M. (1976). *Biochemistry*, **15**, 3731–3738.
- Engel, J. C., Doyle, P. S., Palmer, J., Hsieh, I., Bainton, D. F. & McKerrow, J. H. (1998). *J. Cell Sci.* **111**, 597–606.
- Erickson, A. H. (1989). *J. Cell. Biochem.* **40**, 31–41.
- Fujishima, A., Imau, Y., Nomura, T., Fujisawa, Y., Yamamoto, Y. & Sugawara, T. (1997). *FEBS Lett.* **407**, 47–50.
- Gelb, B. D., Shi, G.-P., Chapman, H. A. & Desnick, R. J. (1996). *Science*, **273**, 1236–1238.
- Groll, M., Ditzel, L., Lowe, J., Stock, D., Bochtler, M., Bartunik, H. D. & Huber, R. (1997). *Nature (London)*, **386**, 463–471.
- Guay, J., Falgoutret, J. P., Ducret, A., Percival, M. D. & Mancini, J. A. (2000). *Eur. J. Biochem.* **267**, 6311–6318.
- Gunčar, G., Klemenčič, I., Turk, B., Turk, V., Karaoglanovic Carmona, A., Juliano, L. & Turk, D. (2000). *Structure*, **8**, 305–313.
- Gunčar, G., Podobnik, M., Pungerčar, J., Strukelj, B., Turk, V. & Turk, D. (1998). *Structure*, **6**, 51–61.
- Gunčar, G., Pungerčič, G., Klemenčič, I., Turk, V. & Turk, D. (1999). *EMBO J.* **18**, 793–803.
- Gutman, H. R. & Fruton, J. (1948). *J. Biol. Chem.* **174**, 851–858.
- Hanada, K., Tamai, M., Yamagishi, M., Ohmura, S., Sawada, J. & Tanaka, I. (1978). *Biol. Chem.* **42**, 523–528.
- Hart, T. C., Hart, P. S., Bowden, D. W., Michalec, M. D., Callison, S. A., Walker, S. J., Zhang, Y. & Firatli, E. (1999). *J. Med. Genet.* **36**, 881–887.
- Hart, T. C., Hart, P. S., Michalec, M. D., Zhang, Y., Firatli, E., Van Dyke, T. E., Stabholz, A., Zlorogorski, A., Shapira, L. & Soskolne, W. A. (2000). *J. Med. Genet.* **37**, 88–94.
- Hart, T. C., Hart, P. S., Michalec, M. D., Zhang, Y., Marazita, M. L., Cooper, M., Yassin, O. M., Nusier, M. & Walker, S. (2000). *J. Med. Genet.* **37**, 95–101.
- Higgins, D. G., Thompson, J. D. & Gibson, T. J. (1996). *Methods Enzymol.* **266**, 383–402.
- Illy, C., Quraishi, O., Wang, J., Purisima, E., Vernet, T. & Mort, J. S. (1997). *J. Biol. Chem.* **272**, 1197–1202.
- Jia, Z., Hasnain, S., Hiramata, T., Lee, X., Mort, J. S., To, R. & Huber, C. P. (1995). *J. Biol. Chem.* **270**, 5527–5533.
- Joshua-Tor, L., Xu, H. E., Johnston, S. A. & Rees, D. C. (1995). *Science*, **269**, 945–950.
- Katunuma, N., Murata, E., Kakegawa, H., Matsui, A., Tsuzuki, H., Tsuge, H., Turk, D., Turk, V., Fukushima, M., Tada, Y. & Asao, T. (1999). *FEBS Lett.* **458**, 6–10.
- Kirschke, H., Barrett, A. J. & Rawlings, N. D. (1995). *Protein Profiles*, Vol. 2, edited by P. Sheterline, pp. 1587–1643. London: Academic Press.
- Klemenčič, I., Carmona, A. K., Cezari, M. H., Juliano, M. A., Juliano, L., Gunčar, G., Turk, D., Križaj, I., Turk, V. & Turk, B. (2000). *Eur. J. Biochem.* **267**, 5404–5412.
- Kominami, E., Tsukahara, T., Hara, K. & Katunuma, N. (1988). *FEBS Lett.* **231**, 225–228.
- Kos, J. & Lah, T. T. (1998). *Oncol. Rep.* **5**, 1349–1361.
- Lalioti, M. D., Scott, H. S., Buresi, C., Rossier, C., Bottani, A., Morris, M. A., Malafosse, A. & Antonarakis, S. E. (1997). *Nature (London)*, **386**, 847–851.
- LaLonde, J. M., Zhao, B., Janson, C. A., D'Alessio, K. J., McQueney, M. S., Orsini, M. J., Debouck, C. M. & Smith, W. W. (1999). *Biochemistry*, **38**, 862–869.
- Leist, M. & Jäättelä, M. (2001). *Cell Death Diff.* **8**, 324–326.
- Lopez-Otin, C. & Overall, C. M. (2002). *Nat. Rev. Mol. Cell. Biol.* **7**, 509–519.
- Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W. & Huber, R. (1995). *Science*, **268**, 533–539.
- McGrath, M. E. (1999). *Annu. Rev. Biophys. Biomol. Struct.* **28**, 181–204.
- McGrath, M. E., Klaus, J. L., Barnes, M. G. & Brömme, D. (1997). *Nature Struct. Biol.* **4**, 105–109.
- McGrath, M. E., Palmer, J. T., Brömme, D. & Somoza, J. R. (1998). *Protein Sci.* **7**, 1294–1302.
- Mason, R. W. (1995). *Subcellular Biochemistry*, Vol. 27, edited by J. B. Lloyd & R. W. Mason, pp. 159–190. Dordrecht: Kluwer.
- Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
- Musil, D., Zucić, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovič, T., Turk, V., Towatari, T., Katunuma, N. & Bode, W. (1991). *EMBO J.* **10**, 2321–2330.
- Nagler, D. K., Storer, A. C., Portaro, F. C., Carmona, E., Juliano, L. & Menard, R. (1997). *Biochemistry*, **36**, 12608–12615.
- Nagler, D., Tam, W., Storer, A. C., Krupa, J. C., Mort, J. S. & Menard, R. (1999). *Biochemistry*, **38**, 4868–4874.
- Nakagawa, T. Y., Brissette, W. H., Lira, P. D., Griffiths, R. J., Petrushova, N., Stock, J., McNeish, J. D., Eastman, S. E., Howard, E. D., Clarke, S. R. M., Rosloniec, E. F., Elliott, E. A. & Rudensky, A. Y. (1999). *Immunity*, **10**, 207–217.
- Nakagawa, T., Roth, W., Wong, P., Nelson, A., Farr, A., Deussing, J., Villadangos, J. A., Ploegh, H., Peters, C. & Rudensky, A. Y. (1998). *Science*, **280**, 450–453.
- Nicholls, A., Sharp, K. A. & Honig, B. (1991). *Proteins*, **11**, 281–376.
- Nishimura, Y., Kawabata, T. & Kato, K. (1988). *Arch. Biochem. Biophys.* **261**, 64–71.
- Nuckolls, G. H. & Slavkin, H. C. (1999). *Nature Genet.* **23**, 378–380.
- Olsen, J. G., Kadziola, A., Lauritzen, C., Pedersen, J., Larsen, S. & Dahl, S. W. (2001). *FEBS Lett.* **506**, 201–206.
- Otto, H. H. & Schirmeister, T. (1997). *Chem. Rev.* **97**, 133–172.
- Pinitglang, S., Watts, A. B., Patel, M., Reid, J. D., Noble, M. A., Gul, S., Bokth, A., Naeem, A., Patel, H., Thomas, E. W., Sreedharan, S. K., Verma, C. & Brocklehurst, K. (1997). *Biochemistry*, **36**, 9968–9982.
- Pennacchio, L. A., Lehesjoki, A. E., Stone, N. E., Willour, V. L., Virtaneva, K., Miao, J., D'Amato, E., Ramirez, L., Faham, M., Koskiniemi, M., Warrington, J. A., Norio, R., de la Chapelle, A., Cox, D. R. & Myers, R. M. (1996). *Science*, **271**, 1731–1734.
- Pereira, P. J., Bergner, A., Macedo-Ribeiro, S., Huber, R., Matschner, G., Fritz, H., Sommerhoff, C. P. & Bode, W. (1998). *Nature (London)*, **392**, 306–311.

- Pham, C. T. & Ley, T. J. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 8627–8632.
- Podobnik, M., Kuhelj, R., Turk, V. & Turk, D. (1997). *J. Mol. Biol.* **271**, 774–788.
- Roth, V., Deussing, J., Botchkarev, V. A., Pauly Evers, M., Saftig, P., Hafner, A., Schmidt, P., Schmahl, W., Scherer, J., Anton-Lamprecht, I., von Figura, K. & Peters, C. (2000). *FASEB J.* **14**, 2075–2086.
- Saftig, P., Hunzicker, E., Wehmeyer, O., Jones, S., Boyde, A., Rommerskirch, W., Moritz, J. D., Schu, P. & von Figura, K. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 13453–13458.
- Šali, A. & Blundell, T. L. (1993). *J. Mol. Biol.* **234**, 779–815.
- Salvesen, G. S. (2001). *J. Clin. Invest.* **107**, 21–22.
- Schaschke, N., Assfalg-Machleidt, I., Lassleben, T., Sommerhoff, C. P., Moroder, L. & Machleidt, W. (2000). *FEBS Lett.* **482**, 91–96.
- Schaschke, N., Assfalg-Machleidt, I., Machleidt, W., Turk, D. & Moroder, L. (1997). *Bioorg. Med. Chem.* **5**, 1789–1797.
- Schechter, I. & Berger, A. (1967). *Biochem. Biophys. Res. Commun.* **27**, 157–162.
- Shaw, E. (1990). *Adv. Enzymol.* **63**, 271–297.
- Shi, G.-P., Bryant, R. A., Riese, R., Verhelst, S., Driessen, C., Li, Z., Brömme, D., Ploegh, H. L. & Chapman, H. A. (2000). *J. Exp. Med.* **191**, 1177–1186.
- Shi, G.-P., Villadangos, J. A., Dranoff, G., Small, C., Gu, L., Haley, K. J., Riese, R., Ploegh, H. L. & Chapman, H. A. (1999). *Immunity*, **10**, 197–206.
- Sivaraman, J., Lalumiere, M., Menard, R. & Cygler, M. (1999). *J. Protein Sci.* **8**, 283–290.
- Sivaraman, J., Nagler, D. K., Zhang, R., Menard, R. & Cygler, M. (2000). *J. Mol. Biol.* **295**, 939–951.
- Somoza, J. R., Palmer, J. T. & Ho, D. H. (2002). *J. Mol. Biol.* **322**, 559–568.
- Somoza, J. R., Zhan, H., Bowman, K. K., Yu, L., Mortara, K. D., Palmer, J. T., Clark, J. M. & Mcgrath, M. E. (2000). *Biochemistry*, **39**, 12543.
- Stoka, V., Turk, B., Schendel, S. L., Kim, T. H., Cirman, T., Snipas, S. J., Ellerby, L. M., Bredesen, D., Freeze, H., Abrahamson, M., Bromme, D., Krajewski, S., Reed, J. C., Yin, X. M., Turk, V. & Salvesen, G. S. (2001). *J. Biol. Chem.* **276**, 3149–3157.
- Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarčič, B. & Turk, V. (1990). *EMBO J.* **9**, 1939–1947.
- Stypmann, J., Glaser, K., Roth, W., Tobin, D. J., Petermann, I., Matthias, R., Monnig, G., Haverkamp, W., Breithardt, G., Schmahl, W., Peters, C. & Reinheckel, T. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 6234–6239.
- Thompson, S. K. *et al.* (1997). *Proc. Natl Acad. Sci. USA*, **94**, 14249–14254.
- Toomes, C. *et al.* (1999). *Nature Genet.* **23**, 421–424.
- Tsuge, H., Nishimura, T., Tada, Y., Asao, T., Turk, D., Turk, V. & Katunuma, N. (1999). *Biochem. Biophys. Res. Commun.* **266**, 411–416.
- Turk, B., Stoka, V., Rozman-Pungerčar, J., Cirman, T., Droga-Mazovec, G., Orešič, K. & Turk, V. (2002). *Biol. Chem.* **383**, 1035–1044.
- Turk, B., Turk, D. & Turk, V. (2000). *Biochim. Biophys. Acta*, **1477**, 98–111.
- Turk, D. (1992). PhD thesis, Technische Universität, München, Germany.
- Turk, D., Gunčar, G., Podobnik, M. & Turk, B. (1998). *Biol. Chem.* **379**, 137–147.
- Turk, D., Janjič, V., Stern, I., Podobnik, M., Lamba, D., Dahl, S. W., Lauritzen, C., Pedersen J., Turk, V. & Turk, B. (2001). *EMBO J.* **20**, 6570–6582.
- Turk, D., Podobnik, M., Kuhelj, R., Dolinar, M. & Turk, V. (1996). *FEBS Lett.* **384**, 211–214.
- Turk, D., Podobnik, M., Popovic, T., Katunuma, N., Bode, W., Huber, R. & Turk, V. (1995). *Biochemistry*, **34**, 4791–4797.
- Turk, V. & Bode, W. (1991). *FEBS Lett.* **285**, 213–219.
- Turk, V., Turk, B. & Turk, D. (2001). *EMBO J.* **20**, 4629–4633.
- Turkenburg, J. P., Lamers, M. B. A. C., Brzozowski, A. M., Wright, L. M., Hubbard, R. E., Sturt, S. L. & Williams, D. H. (2002). *Acta Cryst. D* **58**, 451–455.
- Varughese, K. I., Su, Y., Cromwell, D., Hasnain, S. & Xuong, N. H. (1992). *Biochemistry*, **31**, 5172–5176.
- Wlodawer, A. & Gustchina, A. (2000). *Biochim. Biophys. Acta*, **1477**, 16–34.
- Wolters, P. J., Laig-Webster, M. & Caughey, G. H. (2000). *Am. J. Respir. Cell Mol. Biol.* **22**, 183–190.
- Wolters, P. J., Pham, C. T. N., Muilenburg, D. J., Ley, T. J. & Caughey, G. H. (2001). *J. Biol. Chem.* **276**, 18551–18556.
- Yamamoto, D., Matsumoto, K., Ohishi, H., Ishida, T., Inoue, M., Kitamura, K. & Mizuno, H. (1991). *J. Biol. Chem.* **266**, 14771–14777.
- Zhao, B., Janson, C. A., Amegadzie, B., D'Alessio, K., Griffin, C., Hanning, C. R., Jones, C., Kurdyla, J., McQueney, M., Qiu, X., Smith, W. W. & Abdel-Meguid, S. S. (1997). *Nature Struct. Biol.* **4**, 109–111.