The EMBO Journal Vol.18 No.4 pp.793–803, 1999

Crystal structure of MHC class II-associated p41 Ii fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S

Gregor Gunčar, Galina Pungerčič, Ivica Klemenčič, Vito Turk and Dušan Turk[1]

Department of Biochemistry and Molecular Biology, Jožef Stefan Institute, Jamova 39, SLO-1000 Ljubljana, Slovenia

[1]Corresponding author
e-mail: dusan.turk@ijs.si

The lysosomal cysteine proteases cathepsins S and L play crucial roles in the degradation of the invariant chain during maturation of MHC class II molecules and antigen processing. The p41 form of the invariant chain includes a fragment which specifically inhibits cathepsin L but not S. The crystal structure of the p41 fragment, a homologue of the thyroglobulin type-1 domains, has been determined at 2.0 Å resolution in complex with cathepsin L. The structure of the p41 fragment demonstrates a novel fold, consisting of two subdomains, each stabilized by disulfide bridges. The first subdomain is an α-helix–β-strand arrangement, whereas the second subdomain has a predominantly β-strand arrangement. The wedge shape and three-loop arrangement of the p41 fragment bound to the active site cleft of cathepsin L are reminiscent of the inhibitory edge of cystatins, thus demonstrating the first example of convergent evolution observed in cysteine protease inhibitors. However, the different fold of the p41 fragment results in additional contacts with the top of the R-domain of the enzymes, which defines the specificity-determining S2 and S1′ substrate-binding sites. This enables inhibitors based on the thyroglobulin type-1 domain fold, in contrast to the rather non-selective cystatins, to exhibit specificity for their target enzymes.

Keywords: cathepsin/crystal structure/invariant chain/MHC class II/thyroglobulin type-1 domain

Introduction

The majority of protein antigens must be converted into short peptides before they can trigger an immune response (Mellman *et al*., 1995). Peptides resulting from limited proteolysis are loaded into the binding groove of major histocompatibility complex (MHC) class I and II molecules and presented at the cell surface for recognition by $CD8^+$ and $CD4^+$ T lymphocytes, respectively (Germain and Margulies, 1993). In general, the MHC class I pathway is responsible for processing of intracellular proteins, whereas the class II pathway deals with extracellular proteins, which are transported into cells via endocytosis or phagocytosis (Goldberg and Rock, 1992; Fineschi and Miller, 1997; Chapman, 1998).

The proteolytic environment present in endosomes is responsible for maturation of MHC class II molecules and processing of peptide antigens. MHC class II molecules are liberated from the trimer of class II αβ dimers associated with the trimer of invariant chains (Ii) by gradual degradation of Ii by lysosomal proteases (Roche and Cresswell, 1991; Newcomb and Cresswell, 1993; Cresswell, 1996), of which the cysteine proteases cathepsin L and S play a dominant role (Riese *et al*., 1996, 1998; Deussing *et al*., 1998; Nakagawa *et al*., 1998).

As the degradation of Ii and antigen processing are regulated processes (Chapman, 1998), the proteolytic activity of the enzymes involved must somehow be controlled. It is known that the presence of the p41 invariant chain enhances antigen presentation (Peterson and Miller, 1992). In humans, four variants of Ii, p33, p35, p41 and p43, are products of both alternative splicing and alternative translation initiation (Strubin *et al*., 1986; O'Sullivan *et al*., 1987). The p41 and p43 variants of Ii are produced by splicing in a 64 amino acid fragment (hereafter called the p41 fragment) located before the C-terminal domain and encoded by exon 6B (Strubin *et al*., 1986). The discovery of the Ii p41 fragment tightly bound to cathepsin L (Ogrinc *et al*., 1993) led to the suggestion that the invariant chain may enhance antigen presentation by providing a mechanism to inhibit otherwise destructive cathepsin L activity (Rodriguez and Diment, 1995). This would be similar to the action of cystatin C, a rather non-selective inhibitor of lysosomal cysteine proteases, which regulates the maturation of MHC class II molecules in dendritic cells by inhibiting cathepsin S (Pierre and Mellman, 1998). Interestingly, the cystatins in general do not discriminate well between closely related proteases, and so the mechanism whereby the p41 fragment distinguishes between the closely related cysteine proteases cathepsins L and S requires an explanation.

Previously it was not clear how the p41 fragment could inhibit proteases since it shows no homology to known families of cysteine protease inhibitors (Bevec *et al*., 1996). Its sequence is, however, similar to those of a number of thyroglobulin type-1 domains, originally found in thyroglobulin (Molina *et al*., 1996a), but present in various proteins of different origin and function (Lenarčič and Bevec, 1998). Although the role of these domains has not been revealed, they are thought to be involved in the control of proteolytic degradation of either their cognate or foreign proteins (Molina *et al*., 1996b; Lenarčič and Bevec, 1998). Homologues include insulin-like growth factor-binding proteins (IGFBPs), saxifilin, nidogen, carcinoma-associated antigen GA732 and others. IGFBPs represent a family of proteins that modify the growth-stimulating effects of the IGFs by inhibiting degradation of IGFBP-4 (Fowlkes *et al*., 1997). Saxiphilin, a protein with unknown function, binds the neurotoxin saxitoxin with high affinity (Llewellyn *et al*., 1997). Nidogen, a

150 kDa glycoprotein, may play a central role in the supramolecular organization of basement membranes (Aumailley *et al*., 1993), which control a large number of cellular activities including adhesion, differentiation and apoptosis (Ashkenas *et al*., 1996). Thus, the thyroglobulin type-1 domain may represent a novel protease inhibitor family.

To elucidate the inhibitory mechanism of p41, as well as to clarify its role in cathepsin L inhibition during antigen presentation, we have determined the structure of the complex of these two proteins at 2.0 Å resolution. The structure reveals the features that enable the p41 fragment to act as a protease inhibitor, and to inhibit cathepsin L selectively in the presence of cathepsin S, thereby facilitating antigen processing. The revealed fold of the thyroglobulin type-1 domain enables three-dimensional models of the homology domains to be built using the disulfide bridges and other constraints recognizable from the p41 fragment structure and consistent with their sequence alignment table. It supports the function of the thyroglobulin type-1 domain as a protease inhibitor scaffold upon which can be placed specificity-determining side chains selective for individual cysteine proteases.

Results

Cathepsin L residues are numbered as in the procathepsin L structure (Coulombe *et al*., 1996) and p41 fragment residues as they occur in the p41 form of Ii with the letter 'F' added.

Overall structure

The crystal structure of the p41 fragment in complex with cathepsin L comprises human Ii p41 residues from Leu194F to Ser258F, and the heavy (Ala1–Thr175) and light (Asn179–Val220) chains of human cathepsin L (Figure 1). The majority of the p41 fragment and cathepsin L residues are defined unambiguously by the electron density maps, the exceptions being a few side chains and the termini of the chains. The positioning of the N-terminal Leu194F, C-terminal residues Glu257F–Ser258F, the C-terminal residues of the heavy chain (Phe172–Thr175) and the N-terminal residue Asn179 of the light chain of cathepsin L is not revealed by the electron density maps.

The asymmetric unit of the crystal contains two complexes. The structures of both crystallographically independent p41 fragments and cathepsin L molecules are similar, the r.m.s. deviation between superimposed equivalent C_α atoms being 0.26 and 0.13 Å, respectively. Superposition of the two complexes in the asymmetric unit shows that the p41 fragment structures are oriented slightly differently relative to their cathepsin L partners, being rotated about an axis positioned along the active site cleft. They are nearly identical within the active site cleft, whereas at the most distant part the equivalent atoms are displaced by >1 Å. The p41 fragment was found to be glycosylated at Asn240F, where positioning of the first attached carbohydrate ring is well defined by the electron density maps, whereas the electron density around other known glycosylation sites on cathepsin L and the p41 fragment did not enable us to position the associated carbohydrate rings.

p41 fragment structure

The p41 fragment structure has a wedge shape best shown in the standard view along the active site cleft of cathepsin L (Figure 1A). The views along (Figure 1A) and across (Figure 1B) the active site cleft show that the p41 fragment structure is composed of two subdomains, each stabilized by disulfide bonds. The first subdomain is characterized by a short α-helix–β-strand arrangement, and the second by three strands forming a short antiparallel β-sheet (Figure 2). At the sharp end of the wedge, there are three interacting turns. The first turn, which is also the broadest,

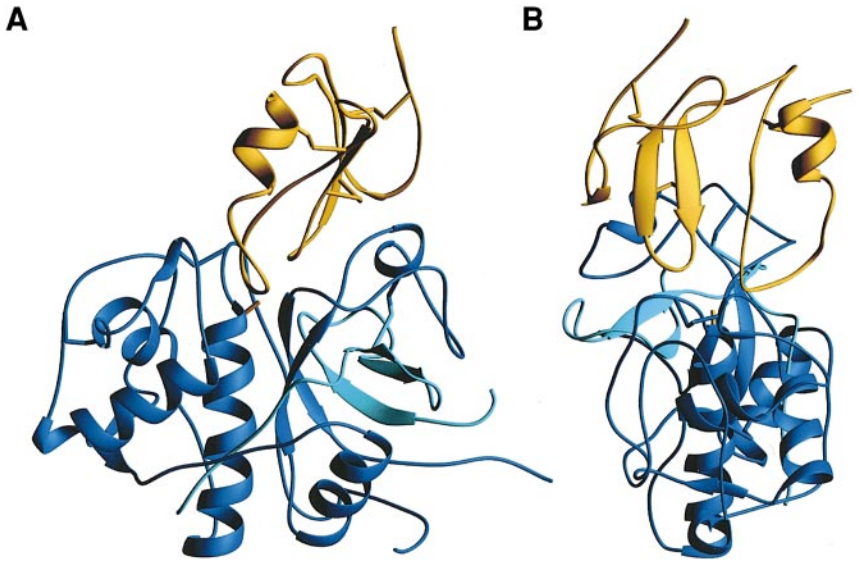

Fig. 1. Ribbon diagram of the cathepsin L–p41 fragment complex. Cathepsin L is shown in blue (heavy chain, dark blue; light chain, light blue) and the p41 fragment in yellow. The catalytic residue Cys25 positioned on the top of the central α-helix and the disulfide bridges are shown with sticks. The three turns of the p41 fragment contact the enzyme surface. The p41 fragment starts with the short N-terminal α-helix. The figure was produced using the program RIBBONS (Carson, 1991). (**A**) Standard view, along the interdomain interface and the active site cleft of cathepsin L, with L- and R-domains on the left and right. (**B**) Side view, perpendicular to the standard view.

belongs to the first subdomain, whereas the other two are tight β-turns and belong to the second subdomain.

The chain of the first subdomain starts with a short N-terminal α-helix (Thr195F–His203F), with a short two residue extension (Ile204F–Pro205F) which continues into a 90° turn at Ala206F (Figure 3). From the C_α atom of Ala206F to the C_α atom of Gly210F the chain runs in a β-strand-like conformation and forms the bottom of the first turn. At Gly210F, the chain makes another 90° turn and then continues in an extended conformation towards the top of the molecule where a disulfide bond between the Cys197F and Cys216F bridges the structure of the first subdomain. The chain continues to the top, where it forms a tight β-turn with Glu218F and Asn219F side chains pointing into the solvent, and then proceeds horizontally to the second subdomain.

A small hydrophobic core forms the inner part of the first subdomain (Figure 3). The core starts with the disulfide bond at the top and continues towards the bottom with hydrophobic residues organized in pairs, each coming from a different side (Val201F–Pro214F, Ile204F–Arg213F). The core is surrounded by a layer of hydrophilic interactions. The Thr195F Oγ1 atom forms a hydrogen bond with the Gly220F carbonyl oxygen at the top, and the Arg213F guanidinium group bridges the extended chain with the end of the helical region forming a hydrogen bond with the carbonyl group of Ile204F in the lower part of the subdomain. At the bottom, the His208F ring δ-atoms are within hydrogen-bonding distance of the carbonyl groups of Ala206F and Ser211F on the left and right sides of the ring, respectively. This probably implies that the ring has alternative orientations. The two hydrophobic

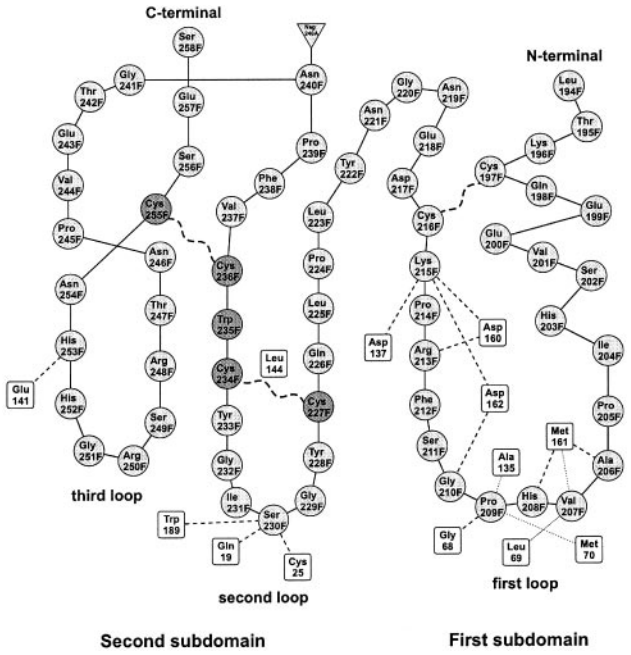

Fig. 2. Schematic representation of the p41 fragment fold and its contacts with cathepsin L. p41 fragment residues are represented as circles, and interacting cathepsin L residues as squares. Dark grey shaded circles represent the C–CWC–C building element. Disulfide bridges are represented as dashed winding lines, hydrogen bonds and electrostatic interactions as dashed lines, and hydrophobic interactions as thin dotted lines. The Asn240F residue was observed in the electron density map to be glycosylated and is represented by a triangle.

side chains of Val207F and Pro209F, placed at the bottom of the first subdomain, point downwards.

The chain comes horizontally from the first into the second subdomain (Figure 3). At Pro224F, the chain runs downwards in a β-strand conformation and, at the bottom with a β-turn (with Ser230F and Ile231F as the bottom residues), turns upwards. The β-turn deviates from standard geometry, probably due to interaction of the Ser230F side chain with the negatively charged Sγ atom of the active site cysteine. After Cys236F, the chain turns right and forms a horizontally located β-turn, with Pro239F and glycosylated Asn240F at the edge. The chain continues to the left and at Pro245F turns downwards, where Ser249F forms a short antiparallel β-sheet hydrogen bond arrangement with Cys234F in the middle strand. A broad turn starts at Arg250F as the lowest residue, and then continues upwards behind the short three-stranded β-sheet, where its position is stabilized with the disulfide bond between Cys236F and Cys255F.

The Trp235F ring is the central feature of the second subdomain (Figure 3). It is almost entirely shielded from solvent; from the front by the side chains of Glu243F–Arg248F involved in electrostatic interactions, from the back by the main chain β-sheet atoms and Gln226F side chain, from the top by the horizontally placed Cys236F and Val237F, from the bottom by Tyr228F and Ile231F, from the left by Tyr233F and from the right by the aromatic ring of Phe212F, a residue from the first subdomain. A well ordered solvent molecule is placed near to Trp235F Nε1, fixed between it and the Gly241F carbonyl. At the back, below the disulfide 236F–255F, is His253F which fills the middle of the last loop and stabilizes its formation with two hydrogen bonds formed to the main chain carbonyl oxygen of Thr247F.

The interface of the two subdomains is not stabilized by main chain–main chain interactions. Most of the interface is a zipper-like formation of side chains (Lys196F, Tyr222F, Gln226F, Phe212F and Tyr228F; the residues are ordered from top to bottom as shown in Figure 3). The Tyr222F ring is placed at the top and in the middle of the two subdomains between the disulfide 197F–216F from the first subdomain and a top turn from the second subdomain. Its hydroxyl group is involved in triangular interactions with the negatively charged Glu200F and the positively charged amino group of Lys196F located within the first turn of the α-helix. The amino group of Lys196F is also involved in a hydrogen bond with the Asn240F main chain carbonyl, thereby stabilizing the positioning of the two subdomains. Below Tyr222F, the Gln Oε1 atom and Lys215F amide proton form another cross-subdomain hydrogen bond linkage. The interface hydrogen bond pattern is completed with the hydrogen bond between the hydroxyl group of Tyr228F and the carbonyl group of Gly210F at the bottom.

Cathepsin L structure

The structure of cathepsin L shares the common fold of papain-like cysteine proteases. The structure contains two domains, left (L-) and right (R-), reminiscent of a closed book with the spine in front. The domains separate on the top in a 'V'-shaped active site cleft, in the middle of which are residues Cys25 and His163, one from each domain, forming the catalytic site of the enzyme (Figure 1).

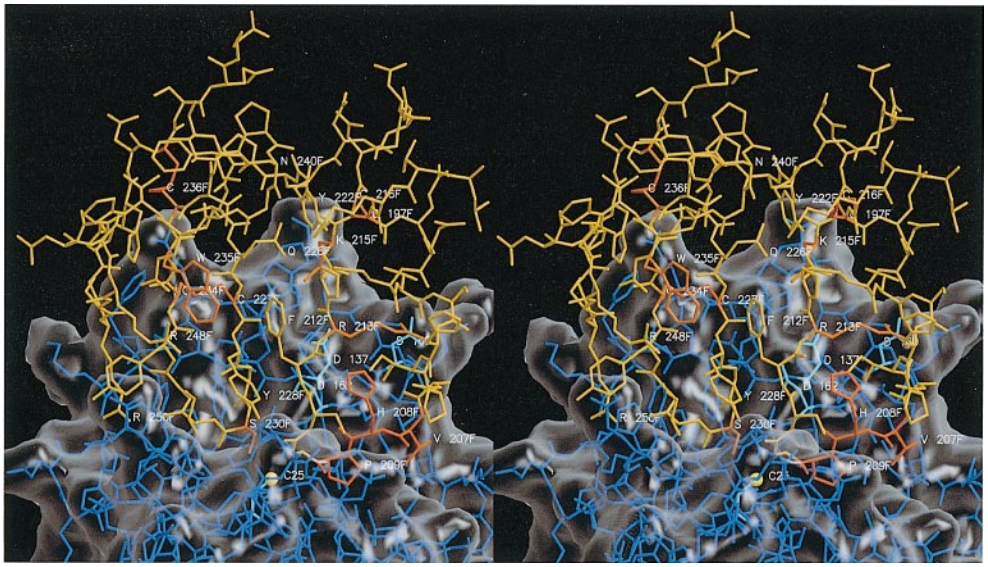

Fig. 3. Side view of the p41 fragment and the interaction region of cathepsin L. Cathepsin L and the p41 fragment are shown with blue and yellow/orange sticks, respectively. The side chains of catalytic cysteine and negatively charged residues involved in electrostatic interactions (Asp137, Glu159, Asp160 and Asp162) are highlighted with light blue. p41 fragment residues of the first loop (Val207F, His208F and Pro209F), side chains of Ser230F, all cysteines, Trp235F and side chains of Arg213F and Lys215F involved in electrostatic interactions with the negatively charged cathepsin L surface are highlighted with orange. The catalytic SG atom is shown as a yellow ball. The single residue letter code and sequence IDs were used to mark the residues. The transparent surface, created with GRASP (Nicholls *et al*., 1991), demarcates cathepsin L from the p41 fragment. MOLSCRIPT (Kraulis, 1991) was used for generation of the sticks model. Both image files were merged and finally rendered with the Raster3D program (Merritt and Bacon, 1997).

The structure is almost identical to that of the mature enzyme domains of procathepsin L (Coulombe *et al*., 1996), with the r.m.s. deviation between equivalent C_α atoms being 0.31 Å. The structures differ at Gln21, the top residue of the short Gln19–Cys25 loop, where the Gln21 C_α atom is shifted by 1.2 Å against the centre of the active site cleft, thus making it slightly narrower. Comparison with the mature enzyme structure (Fujishima *et al*., 1997) could not be performed due to its coordinates not being available.

Interaction between the p41 fragment and cathepsin L

The wedge shape structure of the p41 fragment is anchored into the active site cleft of the cathepsin L molecule by the entire length of its sharp edge (Figure 3). The interaction is mediated through the three turns of the p41 fragment, and via well defined solvent molecules.

The first turn residues are bound to the cathepsin L surface by hydrogen bonds and hydrophobic interactions (Figure 3). Four hydrogen bonds are formed between the CO of Ala206F and NH of Met161, NH of His208F and CO of Met 61, CO of Pro209F and NH of Gly68 and between NH of Gly210F and CO of Asp162. The side chain of Val207F is packed between the side chains of Leu69 and Met161. The hydrophobic ring of Pro209F is located adjacent to the side chains of Met70 above and of Ala135 on the right, and points into the S2 binding site of cathepsin L. The neighbouring structure around the Gly68 CO is not disturbed by this positioning of the proline. The rest of the S2 binding site below Leu69 is filled with two well ordered solvent molecules. The guanidinium group of Arg213F is in the vicinity of Asp160; however, the conformation of the carboxyl group of the aspartic residue cannot be revealed from the electron density, so no rigid salt bridge formation can be observed. Additional electrostatic interactions between His208F and Asp162 are mediated through a layer of solvent molecules. Lys215F lies above the S1′ binding site of cathepsin L. It is involved in electrostatic interactions within the area of negatively charged Asp137, Asp160 and Asp162; however, the positioning of the atoms at the end of its side chain in the molecule could not be resolved in the electron density maps.

The residues of the second interacting loop are tightly packed in the active site, in the vicinity of the catalytic residues. The Ser230F side chain hydroxyl group is involved in interactions with the negatively charged active site sulfur atom as well as with the oxyanion hole Nε2 atom of Gln19. Its carbonyl group forms a hydrogen bond with Trp189 Nε1 (Figures 2 and 3).

The third turn is placed higher above the active site cleft than the other two, and the side chain of Arg250F is hydrogen bonded to the carbonyl groups of Gly232F and Ile231F (Figure 3). The Arg248F guanidinium group interacts with the Gln21 side chain (which is displaced from the position it takes up in the proenzyme structure). At the back of the second subdomain, the disulfide 227F–234F packs against the side chain of Leu144 and thus protects it from exposure to the solvent. The His253F ring forms a salt bridge with the Glu141 carboxylic group. The position of the cathepsin L Phe145 side chain is different from its position in the proenzyme structure, presumably due to interactions with the p41 fragment.

Discussion

Binding mode of the p41 fragment

The fold of the p41 fragment has no features in common with the known propeptide structures of cysteine proteases

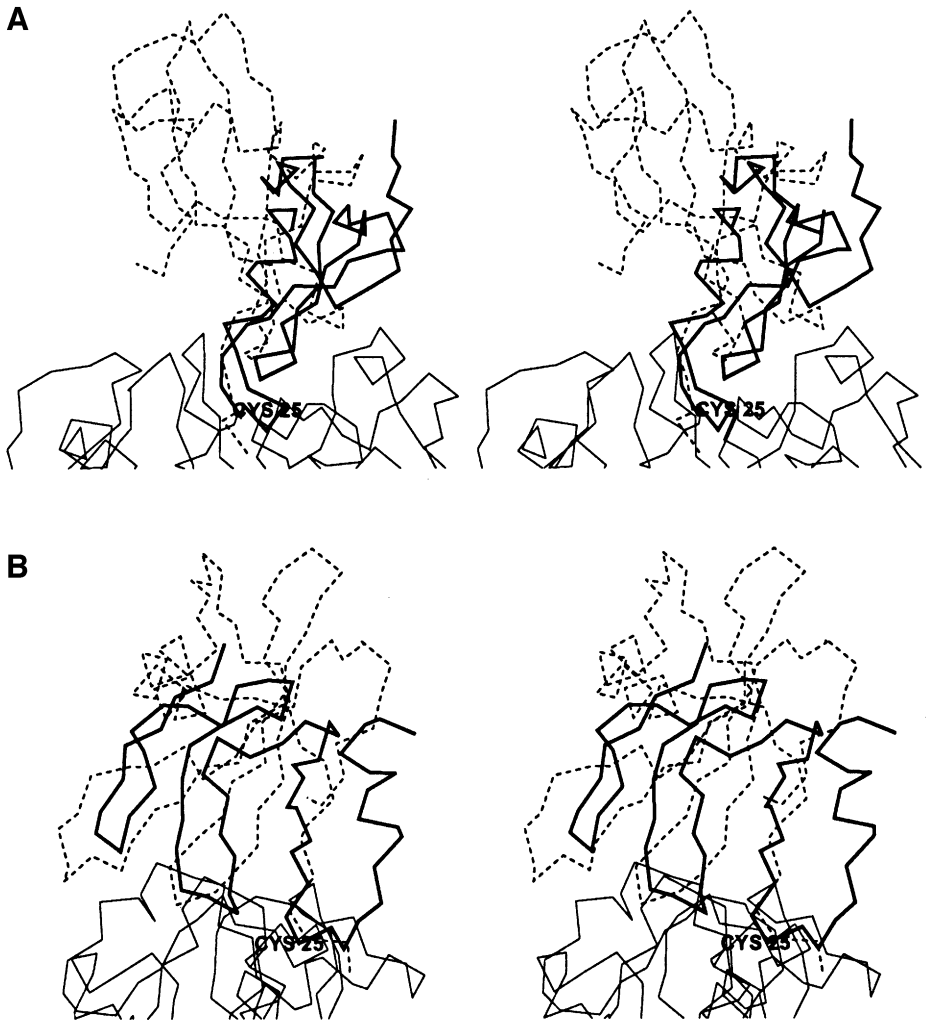

Fig. 4. Comparison of the p41 fragment, in complex with cathepsin L, with stefin B. Superposition of the p41 fragment and stefin B is based on the three-dimensional alignment of papain and cathepsin L structures in the complexes of papain–stefin B (Stubbs *et al*., 1990) and cathepsin L–p41 fragment. The p41 fragment is represented by a thick line, stefin B by a dashed line, and cathepsin L by a thin line. The figure was prepared with the program MAIN (Turk, 1992). (**A**) standard view, (**B**) side view.

(Coulombe *et al*., 1996; Cygler *et al*., 1996; Groves *et al*., 1996; Turk *et al*., 1996; Podobnik *et al*., 1997); however, its structural elements bound into the active site cleft region are similar to those of stefin B (Stubbs *et al*., 1990). The p41 fragment is about two-thirds of the size of stefin B, with a different topology of the parts not involved in binding to a cysteine protease active site. As a similar fold was not found in the database using the DALI program (Holm and Sander, 1996), the p41 fragment structure constitutes a fold of a novel class of cysteine protease inhibitors (a thyroglobulin type-1 domain fold) as well as a new protein fold.

Stefins and cystatins exhibit relatively little specificity in their inhibition of the papain-like family of enzymes, whereas the p41 fragment is capable of discriminating between the different members (Bevec *et al*., 1996; Turk *et al*., 1997). Figure 4 reveals the obvious similarities and differences between the p41 fragment and stefin B, each in an inhibitory complex with an enzyme. Stefin B interacts with the enzyme surface only at the bottom of the active site cleft, whereas the broader p41 fragment also interacts with the higher regions of the cleft, in particular with the R-domain loops contributing to the S2 and S1′ binding sites. These binding sites form the basis of substrate specificity within the family (Turk *et al*., 1998) and, as discussed below, also form the basis on which the p41 fragment can discriminate among them.

The positioning of the residues of the first interacting loop of the p41 fragment from Val207F to Arg213F overlaps with the positioning of the N-terminal trunk residues Met6–Ser12 of stefin B. Both chains run across the S2 binding site in a substrate-like orientation. However, the fragment starts as a short α-helix and, at the bottom, forms a broad loop extending along the non-primed substrate-binding region, whereas stefin interacts with the active site cleft with its N-terminal residues. In contrast to the main chain atoms of stefin Ser8, Pro209F is not capable of forming the short β-antiparallel-like hydrogen-bonding ladder with the Gly68 (66 in papain) typical of a substrate-like mode of binding into the S2 binding site (Turk *et al*., 1998).

The second interacting loop of the p41 fragment and the first stefin B hairpin loop bind into similar positions on the primed site of the active site cleft. Both turns form a hydrogen bond between a main chain carbonyl (Ser230F from the p41 fragment and Val55 from stefin B) and

Trp189 (177 in papain); however, the hydrogen bond is formed by the Ser230F carbonyl and not by Ile231F which is positionally equivalent to the stefin B Val55 carbonyl. Furthermore, the loops have quite different conformations. The p41 fragment loop approaches the catalytic Cys25 and forms a hydrogen bond between the Ser230F hydroxyl and the negatively charged Sγ atom, whereas the stefin loop is positioned away from the catalytic cysteine. Stefins and cystatins are capable of interacting with carboxymethylated enzymes, whereas binding of the p41 fragment appears from the structure to require an unperturbed catalytic site.

The third interacting loop of the p41 fragment resembles, both in its conformation and positioning, the second hairpin loop of stefin, although, in contrast to stefin where the loop lies along the active site cleft, the third loop of the p41 fragment lies across the cleft, thus making extensive contacts with the underlying R-domain residues.

The thyroglobulin type-1 domain and cystatin/stefin family of cysteine protease inhibitors are based on different folds, which seem to have converged in the way they interact with the bottom of the active site cleft of papain-like cysteine proteases. They are both wedge-shaped structures, which fill the active site cleft with three short binding segments; three hairpin loops of the p41 fragment versus the N-terminal and two hairpin loops of stefins. The similarity is reminiscent of the canonical conformation of inhibitors of serine proteases (Bode and Huber, 1992); however, the similarity is in topology and not in conformation and the pattern of hydrogen bonds of the binding region. It is thus intriguing to suggest that the congruent binding of cystatin/stefin and thyroglobulin type-1 domain-based inhibitors has revealed the canonical topology of inhibitors of papain-like cysteine proteases.

Selective inhibition of cathepsins with the p41 fragment

As already pointed out, the structural basis of specificity is to be sought among the contacts on the top of the cathepsin L R-domain, where the lower and upper loops (Trp189–Ser213 and Asp137–Asp162, respectively) cover the top of the R-domain β-barrel. In particular, the upper loop forms most of the contacts with the p41 fragment.

The p41 fragment inhibits cathepsin L (K_i = 1.7 pM) and cruzipain (K_i = 58 pM); papain (K_i = 1.4 nM) and cathepsin H (K_i = 5.3nM) are only weakly inhibited, and cathepsins B and S are not inhibited (Bevec *et al*., 1996). It was reported (Fineschi and Miller, 1997) that the p41 fragment also inhibits cathepsins U and K, but no kinetic data were provided. Among these, cathepsins B, H, L, S and U are of physiological relevance for MHC class II maturation and antigen processing. Cathepsins B, H, L and S are common lysosomal enzymes (Turk *et al*., 1997). Cathepsin U (V,L2), whose sequence is closely related to that of cathepsin L, is a tissue-specific protein expressed in thymus (D.Brömme, personal communication). Cathepsin K is a tissue-specific protease expressed in bone osteoclasts (Drake *et al*., 1996), whereas papain and cruzipain are proteases of plant and parasite origin, respectively. Further discussion is thus confined to the first group of enzymes.

Stefins and cystatins inhibit exopeptidases cathepsins B and H, although with lower affinity than the related endopeptidases (Turk *et al*., 1997). This indicates that the mini-chain of cathepsin H and occluding loop of cathepsin B, which partially occupy the active site cleft, are rather flexible and can be displaced from the active site cleft by an approaching inhibitor (Illy *et al*., 1997; Podobnik *et al*., 1997; Gunčar *et al*., 1998). The fact that cathepsin H, but not cathepsin B, is inhibited by the p41 fragment suggests that a feature on the cathepsin B surface other than the occluding loop prevents binding of the p41 fragment. A good candidate is the region between His190 and Gly198, where the cathepsin B chain follows a unique path. It forms part of the S1′ binding site surface, in contrast to the other related enzymes where it forms part of the S2 binding site surface. The region is positioned much higher and would thus offer steric hindrance to the binding of the p41 fragment, whereas the binding of the stefin chains, which do not lean towards the R-domain, is not hindered (Figure 5).

Cathepsins S and L are the most studied proteases involved in MHC class II maturation. They are both endopeptidases with ~60% identical residues. Cathepsin S is also closely related to other endopeptidases of the papain-like enzyme family, and its crystal structure contains no pronounced features which would discriminate its chain trace from that of the related enzymes (McGrath *et al*., 1998). The repulsive interactions between p41 fragment and cathepsin S are, therefore, to be sought among the side chains. As the coordinates of the crystal structure of the latter are not available, we used a model, generated on the basis of homologous structures by the program Modeller (Šali and Blundell, 1993). A putative complex of cathepsin S with p41 fragment was built by superimposing the model of cathepsin S on the cathepsin L of the complex and checked for possible unfavourable interactions with the p41 fragment.

We have found three selective interaction regions which can enable the p41 fragment to discriminate between cathepsins L and S. The first region involves interactions around cathepsin L Gly139. The equivalent residue in cathepsin S is the positively charged Arg141. The model (Figure 5) indicates a probable clash of the Arg141 side chain (Arg137 with the nomenclature of McGrath *et al*., 1998) with the Lys215F side chain. In addition, there are unfavourable electrostatic interactions. Of the four negatively charged cathepsin L residues (Asp137, Glu159, Asp160 and Asp162), only Asp137 is present in cathepsin S. In the observed complex, the two positively charged p41 fragment residues Arg213F and Lys215F interact with the four negatively charged residues, whereas the equivalent part of the cathepsin S surface offers one negatively and one positively charged residue.

The second region involves an electrostatically favourable contact between the positive charge of His253F from the p41 fragment and the negative charge of Glu141 from cathepsin L. This interaction is missing in the putative complex of p41 fragment with cathepsin S, where a proline residue occupies the position of the glutamate.

The third region involves interactions around cathepsin L Leu144. This residue is well packed, within a hydrophobic area beneath the Cys227F–Cys234F disulfide. In the model of cathepsin S, a larger phenylalanine residue occupies this position.

A similar interaction analysis was performed with a

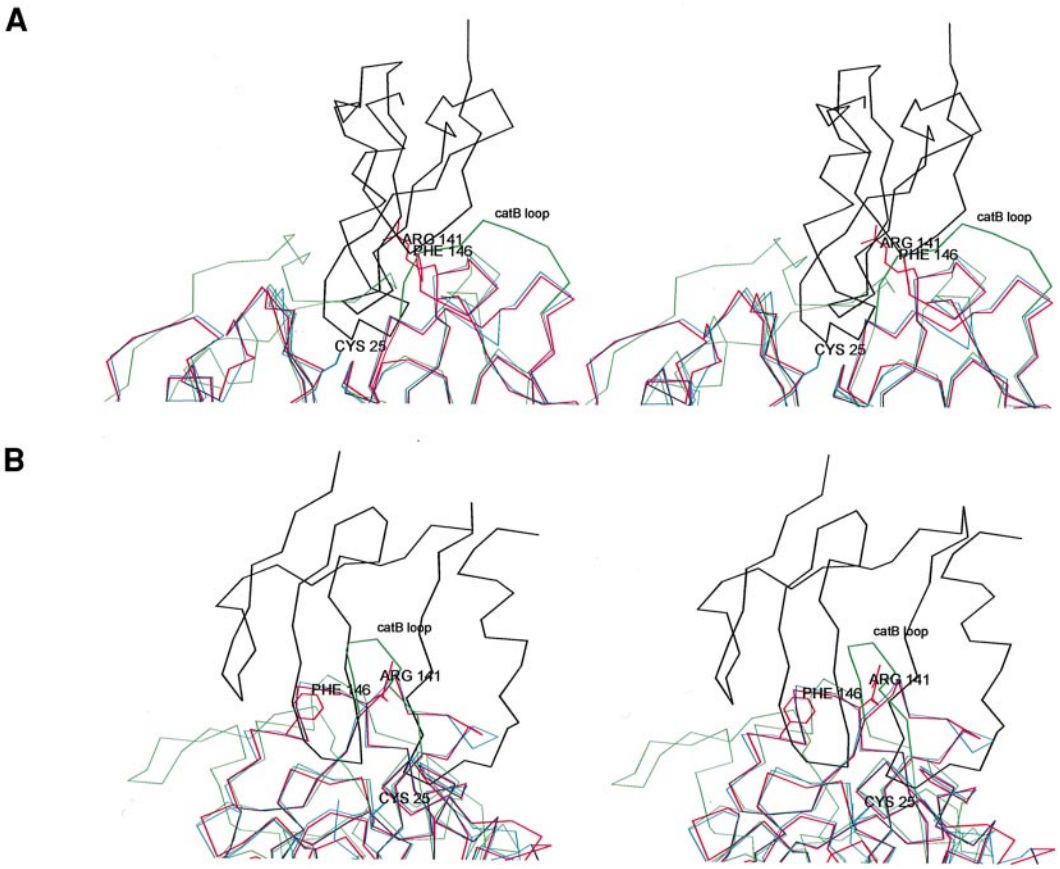

Fig. 5. Stereo C_α plots of the potential complexes of the p41 fragment with cathepsin B and cathepsin S. Complexes were generated by superposition of equivalent C_α atoms from the cathepsin B structure (Musil *et al*., 1991) and cathepsin S modelled onto the cathepsin L part of the p41 fragment complex. The underlying cathepsin L structure with its catalytic Cys25 marked is shown in blue, the p41 fragment in black, and cathepsins B and S in green and magenta, respectively. The cathepsin B chain between residues His190 and Gly198 is shown with a thick green line. Cathepsin S residues Arg141 and Phe146 are marked in magenta. The figure was prepared with the program MAIN (Turk, 1992).

three-dimensional model of cathepsin U (V,L2), generated from the cathepsin L structure using Modeller software (Šali and Blundell, 1993). The modelling study suggests that as a result of the weaker electrostatic interactions within the first and second selective interaction region [cathepsin L residues Glu159, Asp160 and Glu141 versus Lys, Asn, Ser in cathepsin U (V,L2)] and an unfavourable interaction in the third region [cathepsin L Leu144 versus Gln in cathepsin U (V,L2)], p41 fragment would bind to cathepsin U (V,L2) with a lower affinity than to cathepsin L.

The ability of the p41 fragment to bind to various cathepsins is based on electrostatic as well as hydrophobic interactions. The potential clashes, as revealed by the models of putative complexes with cathepsins B and S, would be expected to exert severe effects and to prevent binding.

Thyroglobulin type-1 domain homologues

The sequence homology of thyroglobulin type-1 domains from different proteins suggests that they share a common fold (Figure 6). Most of the conserved residues can be assigned to preserve the structurally characteristic features of the molecules: the six cysteine residues form the three disulfide bonds, all glycine and most proline residues are located at the turns, Phe212F, Tyr222F and Gln226F take part in the interface between the two subdomains, and the CWCV sequence forms the core of the second subdomain. It is involved in two disulfide bridges, Cys227F–Cys234F and Cys236F–Cys255F, which support the short three β-strand arrangements and the attachment of the C-terminal strand. An additional conserved feature is the salt bridge between Arg248F and Glu243F in front of the CWCV tryptophan ring. The salt bridge is inverted in some sequences (EQUI 3, ECST_ACTEQ3, IBP1_HUMAN, G732_HUMAN and G731_HUMAN in Figure 6).

The CWCV sequence motif of thyroglobulin type-1 domains also occurs in other proteins. In the fibrin-binding finger domain of tissue-type plasminogen activator (Downing *et al*., 1992), the CWCV sequence also fixes three strands in a β-sheet formation, although the connectivity of the three strands is opposite to that in p41 fragment. The closing turns between the strands of the finger domain are at the positions where the p41 fragment strands are open and interact with each other non-covalently, and vice versa (Figure 7). This suggests that the CWCV motif itself may appear as a quite general structural element.

Some of the residues at the bottom of the first two turns and involved in contacts with cathepsin L, in particular Pro209F and Gly210F from the first turn and Gly229F and Ser230F from the second turn, are conserved. These turns do not appear to be involved in selectivity, but their structure contributes to the binding strength. An exception

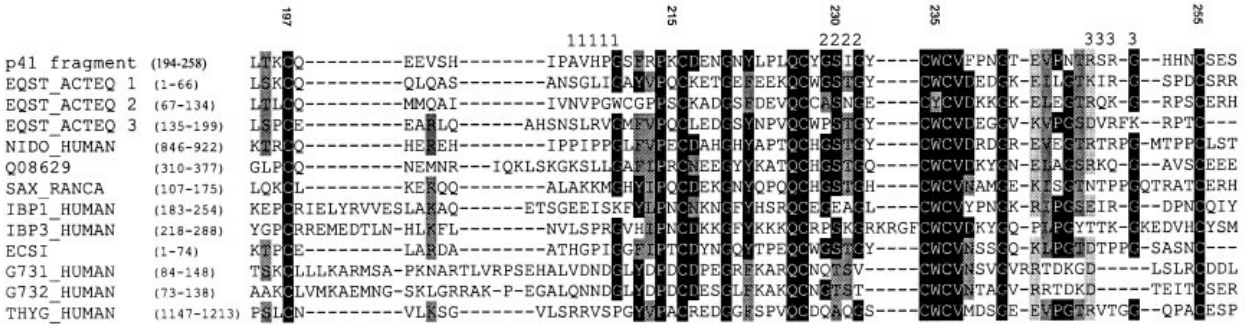

Fig. 6. Sequence alignment of the p41 fragment homologous domains of proteins denoted with their SwissProt codes. Identical residues are shaded in black and similar residues in grey. The three p41 fragment interaction loops are marked and of some the p41 fragment residues are denoted on the top of the alignment with their sequential numbers. p41, p41 fragment (Strubin *et al*., 1986; Bevec *et al*., 1996); EQST_ACTEQ, equistatin (Lenarčič *et al*., 1997); NIDO_HUMAN, nidogen precursor (Nagayoshi *et al*., 1989); Q08629, testican precursor (Alliel *et al*., 1993); SAX_RANCA, saxiphilin precursor (Morabito and Moczydlowski, 1994); IBP, insulin-like growth factor-binding protein precursor (Brinkman *et al*., 1988; Cubbage *et al*., 1990); ECSI, cysteine protease inhibitor of the egg of chum salmon (Yamashita and Konagaya, 1996); G731_HUMAN, pancreatic carcinoma marker protein GA733-1 precursor (Linnenbach *et al*., 1989); G732_HUMAN, major gastrointestinal tumour-associated protein GA733-2 precursor (Strnad *et al*., 1989); THYG_HUMAN, thyroglobulin precursor (Malthiery and Lissitzky, 1987).

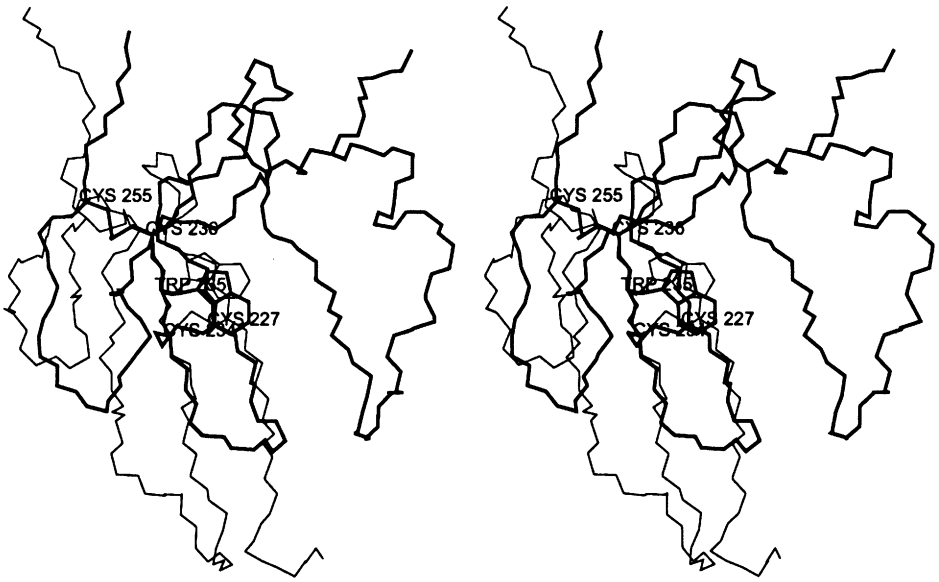

Fig. 7. Structural role of the CWC sequence motif. The CWC sequence forms the core of the interface between the subdomains of the p41 fragment (thick lines) and of the superimposed fibrin-binding finger domain of tissue-type plasminogen activator (thin lines) (Downing *et al*., 1992). The figure was prepared with the program MAIN (Turk, 1992).

is the IBP1_HUMAN sequence, where Ser230F is occupied by a glutamic acid, which is incapable of forming a favourable interaction with the catalytic cysteine of a papain-like cysteine protease. The other three, IGFBPs -3, -5 and -6, inhibit IGFBP-4 degradation (Fowlkes *et al*., 1997).

Arg213F, Lys215F and His253F, involved in electrostatic interactions with cathepsin L negatively charged residues Asp160, Asp137 and Glu141, are not conserved at all. The Arg213F and Lys215F positions appear within a conserved region (Figure 6), whereas His253F belongs to the less conserved C-terminal part. Smaller and neutral residues at positions 213F and 215F may result in a decrease in selectivity. The 215F position shows considerable variability, suggesting that the residue at this position may be crucial for selectivity against particular proteases.

The sequence alignment suggests that the equistatin domain 2 sequence forms an additional disulfide bridge, which covalently links the first and second subdomain at residues Ser211F and Tyr228F, closely positioned in the p41 fragment structure (Figures 2 and 3). It is to be expected that the geometry of the first and the second binding loop of equistatin domain 2 will differ from that observed in p41 fragment, thus also suggesting its different behaviour.

The largest gaps of aligned sequences in Figure 6 are between the first (197F) and the second (216F) cysteine within the first subdomain region. Simple modelling suggested that the short α-helix at the N-terminal part of the p41 fragment structure (Thr195F–His203F), which tightens and then continues in an extended conformation, could be extended down to the first loop for two or more turns and so accommodate additional residues without seriously disrupting the rest of the three-dimensional structure of the first subdomain.

Conclusions

The p41 fragment structure revealed the until now unknown fold of the thyroglobulin type-1 domain and showed unambiguously that the p41 fragment and related cysteine protease inhibitors belong to a new class. The structure suggests that the p41 fragment is capable of discriminating between two very similar cysteine endoproteases, cathepsin L and S, in contrast to rather non-selective cystatins and stefins utilizing electrostatic interactions of Arg213F, Lys215F and His253F and some favourable and unfavourable packing interactions with surface residues on the top of the cathepsin L and S loop embracing the S2–S1′ substrate-binding sites.

It is known that the p41 form of Ii has the ability to slow down degradation of the p31 and p41 forms of invariant chain associated with MHC class II (Fineschi *et al*., 1995). Showing that the p41 fragment inhibits cathepsin L quite specifically led to the question of whether the fragment inhibits cathepsin L when it is still associated with MHC class II molecules or whether it is activated during degradation of Ii (Bevec *et al*., 1996). The structure presented here enables us to formulate this question more precisely and even indicates the answer. The p41 fragment binding surface can bind cathepsin L when it is exposed to solvent and not when it is buried. It seems unlikely that the C-terminal part of Ii containing the whole p41 fragment domain would float around in the nanomeric complex of the three invariant chains and three pairs of MHC class II αβ chain dimers. It is more likely that this part of Ii will be attached to the molecule, probably preserving the 3-fold symmetry of Ii and exposing only a part of its surface area to solvent. The p41 fragment is glycosylated at Asn240F and Asn254F. It is to be expected that both these sites are exposed to solvent. The three binding loops are, however, placed on the side of the p41 fragment structure opposite to the Asn240F glycosylation site. Similarly, the Asn254F site is placed on a side and is not involved in binding to the cathepsin L surface. These findings suggest that the cathepsin L-binding surface of the p41 fragment is at least partially inaccessible to solvent and that degradation of Ii is needed to enable the p41 fragment to bind cathepsin L. This seems consistent with the finding that cathepsin L is responsible for degradation of Ii in cortical thymic epithelial cells (Nakagawa *et al*., 1998).

Does regulation of cathepsin L activity by the p41 fragment point to a more general mechanism of proteolysis control used by molecules which are transported into cell compartments, where the local proteolytic milieu is utilized for their maturation and function? The p41 fragment structure has shown that the thyroglobulin type-1 domain structure is a small unit stabilized by multiple disulfide bonds, with a fold that, in contrast to cystatins and stefins, can adopt selective inhibitory properties. These entities of mostly unknown function have been found to be a part of much larger proteins, often in multiple repeats. The p41 fragment, chum salmon eggs inhibitor (ECSI) and equistatin (EQST_ACTEQ) are already known inhibitors of cysteine proteases (Molina *et al*., 1996b; Lenarčič and Bevec, 1998). It is suggestive that other thyroglobulin type-1 domains also have the potential to exhibit inhibitory activity and that proteolytic degradation of some of their cognate proteins may be required to expose these domains to their target enzymes.

Materials and methods

The complex of cathepsin L and p41 fragment was isolated from human kidney (Ogrinc *et al*., 1993). The purified protein complex was concentrated in a spin concentrator (Centricon; Amicon) to a concentration of 10 mg/ml. Crystals of the complex were grown by the hanging drop vapour diffusion method. The reservoir contained 1 ml of 0.2 M Na acetate trihydrate, 30% w/v PEG 8000 and 0.1 M MES, adjusted to pH 6.1. The drop was composed of 2 μl of reservoir solution and 2 μl of the complex (10 mg/ml) in 20 mM Na acetate and 1 mM EDTA, pH 5.0.

Diffraction data were collected from a single crystal using Cu Kα radiation from a Rigaku Ru200 rotating anode X-ray generator and recorded on an 18 cm MAR Research image plate detector. Autoindexing and scaling were done using HKL programs (Otwinowski and Minor, 1996). The crystal diffracted beyond 2.0 Å resolution and belonged to the primitive monoclinic space group $P2_1$ with cell dimensions a = 62.6 Å, b = 80.6 Å, c = 64.2 Å and a β-angle of 96.8°. The asymmetric unit contained two molecules. Orientation and translation of the two molecules were determined using the molecular replacement method implemented in the AMoRe program (Navaza, 1998). The crystal structure of procathepsin L (Coulombe *et al*., 1996) was used as the search model. Two solutions with a correlation factor of 0.532 and an R-value of 0.385 were found using data in the 15–3.5 Å resolution range.

In the subsequent structure determination, the program MAIN (Turk, 1992) was used for density modifications, model building and refinement. The first electron density map was calculated with the cathepsin L part of the molecule (Coulombe *et al*., 1996). The non-interpreted regions of the electron density were masked using solvent flattening statistical density evaluations, then filled with atoms, superimposed and averaged. The density of the solvent region was flipped with the factor 0.4. The inhibitor model was built into averaged density maps. Masks were expanded as the model building was in progress. After completing the model using data to 3.0 Å resolution, resolution of the data was expanded gradually to the final range (10.0–2.0 Å) in subsequent cycles that involved model rebuilding, positional and B-value refinement, density averaging and solvent generation. Standard parameters for protein (Engh and Huber, 1991) and carbohydrate parameters, as provided by X-PLOR, were used for geometry regularization. Kicked omit maps were used throughout the structure determination process to reveal ambiguous parts of the structure. Structure factors for the kicked omit maps were calculated from randomly displaced atoms, up to 0.3 Å along each

Table I. Crystallographic data and refinement statistics

Data collection	
Space group	$P2_1$
Cell parameters	a = 62.6 Å, b = 80.6 Å, c = 64.2 Å α = γ = 90°, β = 96.8
Molecules in asymmetric unit	2
Limiting resolution (Å)	2.0
Measured reflections	132 945
Independent reflections	42 072
R_{sym}	0.11 (99–2.0 Å)
Completeness	0.97 (99–2.0 Å)
Final refinement parameters	
No. of scattering protein atoms	4257
No. of solvent molecules	668
Resolution range in refinement (Å)	10.0–2.0
Reflections used in refinement (1σ cut off)	41 514
R-factor (1σ cut off)	0.186
R_{free} (8% of reflections) (1σ cut off)	0.213
Geometry of the final model	
R.m.s. deviation of bond distances (Å)	0.011
R.m.s. deviation of bond angles (°)	1.38
B-factor values	
Overall (Å^2)	32.9
Protein (average bond deviation) (Å^2)	29.3 (2.54)
Solvent (Å^2)	56.2

coordinate. Water molecules were generated using an automatic procedure in MAIN and then corrected manually. After the crystallographic *R*-value dropped below 0.25 at 2.0 Å resolution, the temperature factor refinement was applied.

The final refinement included all reflections in the resolution range 10–2.0 Å, with the crystallographic *R*-value being 0.186. The geometry of the final model was inspected with MAIN and Procheck (Laskowski *et al*., 1993). All residues lie in the allowed regions of the Ramachandran plot, 0.876 of residues in most favoured regions and 0.124 in additional allowed regions. Other structural parameters for the structure refined at this resolution show no significant deviations from the expected values (the overall *G*-factor is 0.2, which is better than expected for 2.0 Å resolution). The crystallographic data and refinement statistics are summarized in Table I. The coordinates have been deposited in the Brookhaven Protein Data Bank with accession code 1icf and will be on hold for 1 year.

Acknowledgements

The authors gratefully acknowledge Roger H.Pain and Guy Salvesen for critical reading of the manuscript. We also thank Tadeja Bevec, Boris Turk, Marjetka Podobnik and Nataša Kopitar-Jerala for fruitful discussions and help. The financial support of the Slovenian Ministry of Science and Technology is acknowledged.

References

Alliel,P.M., Perin,J.P., Jolles,P. and Bonnet,F.J. (1993) Testican, a multidomain testicular proteoglycan resembling modulators of cell social behaviour. *Eur. J. Biochem.*, **214**, 347–350.

Ashkenas,J., Muschler,J. and Bissel,M.J. (1996) The extracellular matrix in epithelial biology, shared molecules and common themes in distinct phyla. *Dev. Biol.*, **180**, 433–444.

Aumailley,M., Battaglia,C., Mayer,U., Reinhardt,D., Nischt,R., Timpl,R. and Fox,J.W. (1993) Nidogen mediates the formation of ternary complexes of basement membrane components. *Kidney Int.*, **43**, 7–12.

Bevec,T., Stoka,V., Pungerčič,G., Dolenc,I. and Turk,V. (1996) Major histocompatibility complex class II-associated p41 invariant chain fragment is a strong inhibitor of lysosomal cathepsin L. *J. Exp. Med.*, **183**, 1331–1338.

Bode,W. and Huber,R. (1992) Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.*, **204**, 433–451.

Brinkman,A., Groffen,C., Kortleve,D.J., Geurts van Kessel,A. and Drop,S.L. (1988) Isolation and characterization of a cDNA encoding the low molecular weight insulin-like growth factor binding protein (IBP-1). *EMBO J.*, **7**, 2417–2423.

Carson,M. (1991) Ribbons 2.0. *J. Appl. Crystallogr.*, **24**, 958–961.

Chapman,H.A. (1998) Endosomal proteolysis and MHC class II function. *Curr. Opin. Immunol.*, **10**, 93–102.

Coulombe,R., Grochulski,P., Sivaraman,J., Menard,R., Mort,J.S. and Cygler,M. (1996) Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.*, **15**, 5492–5503.

Cresswell,P. (1996) Invariant chain structure and MHC class II function. *Cell*, **84**, 505–507.

Cubbage,M.L., Suwanichkul,A. and Powell,D.R. (1990) Insulin-like growth factor binding protein-3. Organization of the human chromosomal gene and demonstration of promoter activity. *J. Biol. Chem.*, **265**, 12642–12649.

Cygler,M., Sivaraman,J., Grochulski,P., Coulombe,R., Storer,A.C. and Mort,J.S. (1996) Structure of rat procathepsin B: model for inhibition of cysteine protease activity by the proregion. *Structure*, **4**, 405–416.

Deussing,J., Roth,W., Saftig,P., Peters,C., Ploegh,H.L. and Villadangos,J.A. (1998) Cathepsins B and D are dispensable for major histocompatibility complex class II-mediated antigen presentation. *Proc. Natl Acad. Sci. USA*, **95**, 4516–4521.

Downing,A.K., Driscoll,P.C., Harvey,T.S., Dudgeon,T.J., Smith,B.O., Baron,M. and Campbell,I.D. (1992) Solution structure of the fibrin binding finger domain of tissue-type plasminogen activator determined by 1H nuclear magnetic resonance. *J. Mol. Biol.*, **225**, 821–833.

Drake,F.H. *et al*. (1996) Cathepsin K, but not cathepsins B, L, or S, is abundantly expressed in human osteoclasts. *J. Biol. Chem.*, **271**, 12511–12516

Engh,R.A. and Huber,R. (1991) Accurate bond and angle parameters for X ray protein structure refinement. *Acta Crystallogr.*, **A47**, 392–400.

Fowlkes,J.L., Thrailkill,K.M., George-Nascimento,C., Rosenberg,C.K., Serra,D.M. (1997) Heparin-binding, highly basic regions within the thyroglobulin type-1 repeat of insulin-like growth factor (IGF)-binding proteins (IGFBPs) -3, -5, and -6 inhibit IGFBP-4 degradation. *Endocrinology*, **138**, 2280–2285.

Fineschi,B. and Miller,J. (1997) Endosomal proteases and antigen processing. *Trends Biochem. Sci.*, **22**, 377–382.

Fineschi,B., Arneson,L.S., Naujokas,M.F. and Miller,J. (1995) Proteolysis of major histocompatibility complex class II-associated invariant chain is regulated by the alternatively spliced gene product, p41. *Proc. Natl Acad. Sci. USA*, **92**, 10257–10261.

Fujishima,A., Imai,Y., Nomura,T., Fujisawa,Y., Yamamoto,Y. and Sugawara,T. (1997) The crystal structure of human cathepsin L complexed with E-64. *FEBS Lett.*, **407**, 47–50.

Germain,R.N. and Margulies,D.H. (1993) The biochemistry and cell biology of antigen processing and presentation. *Annu. Rev. Immunol.*, **11**, 403–450.

Goldberg,A.L. and Rock,K.L. (1992) Proteolysis, proteasomes and antigen presentation. *Nature*, **357**, 375–379.

Groves,M.R., Taylor,M.A., Scott,M., Cummings,N.J., Pickersgill,R.W. and Jenkins,J.A. (1996) The prosequence of procaricain forms an alpha-helical domain that prevents access to the substrate-binding cleft. *Structure*, **4**, 1193–1203.

Gunčar,G., Podobnik,M., Pungercar,J., Štrukelj,B., Turk,V. and Turk,D. (1998) Crystal structure of porcine cathepsin H determined at 2.1 Å resolution: location of the mini-chain C-terminal carboxyl group defines cathepsin H aminopeptidase function. *Structure*, **6**, 51–61.

Holm,L. and Sander,C. (1996) Alignment of three-dimensional protein structures: network server for database searching. *Methods Enzymol.*, **266**, 653–662.

Illy,C., Quraishi,O., Wang,J., Purisima,E., Vernet,T. and Mort,J.S. (1997) Role of the occluding loop in cathepsin B activity. *J. Biol. Chem.*, **272**, 1197–1202.

Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Lenarčič,B. and Bevec,T. (1998) Thyropins—new structurally related proteinase inhibitors. *Biol. Chem.*, **379**, 105–111.

Lenarčič,B., Ritonja,A., Štrukelj,B., Turk,B. and Turk,V. (1997) Equistatin, a new inhibitor of cysteine proteinases from *Actinia equina*. *J. Biol. Chem.*, **272**, 13899–13903.

Llewellyn,L.E., Bell,P.M. and Moczydlowski,E.G. (1997) Phylogenetic survey of the soluble saxitoxin-binding activity in pursuit of the function and molecular evolution of saxiphilin, a relative of transferrin. *Proc. R. Soc. Lond. B*, **264**, 891–902

Linnenbach,A.J., Wojcierowski,J., Wu,S.A., Pyrc,J.J., Ross,A.H., Dietzschold,B., Speicher,D. and Koprowski,H. (1989) Sequence investigation of the major gastrointestinal tumor-associated antigen gene family, GA733. *Proc. Natl Acad. Sci. USA*, **86**, 27–31.

Malthiery,Y. and Lissitzky,S. (1987) Primary structure of human thyroglobulin deduced from the sequence of its 8448-base complementary DNA. *Eur. J. Biochem.*, **165**, 314–317.

McGrath,M.E., Palmer,J.T., Bromme,D. and Somoza,J.R. (1998) Crystal structure of human cathepsin S. *Protein Sci.*, **7**, 1294–1302.

Mellman,I., Pierre,P. and Amigorena, S. (1995) Lonely MHC molecules seeking immunogenic peptides for meaningful relationships. *Curr. Opin. Cell Biol.*, **7**, 564–572.

Merritt,E.A. and Bacon,D.J. (1997) Raster3D: photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.

Molina,F., Bouanani,M., Pau,B. and Garnier,C. (1996a) Characterization of type-1 repeat from thyroglobulin, a cysteine-rich module found in proteins from different families. *Eur. J. Biochem.*, **240**, 125–133

Molina,F., Bouanani,M., Pau,B. and Garnier,C. (1996b) The type-1 repeats of thyroglobulin regulate thyroglobulin degradation and T3, T4 release in thyrocytes. *FEBS Lett.*, **391**, 229–231

Morabito,M.A. and Moczydlowski,E. (1994) Molecular cloning of bullfrog saxiphilin: a unique relative of the transferrin family that binds saxitoxin. *Proc. Natl Acad. Sci. USA*, **91**, 2478–2482.

Musil,D. *et al*. (1991) The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *EMBO J.*, **10**, 2321–2330.

Nagayoshi,T. *et al*. (1989) Human nidogen: complete amino acid sequence and structural domains deduced from cDNAs, and evidence for polymorphism of the gene. *DNA*, **8**, 581–594.

Nakagawa,T. *et al*. (1998) Cathepsin L: critical role in Ii degradation and CD4 T cell selection in the thymus. *Science*, **280**, 450–453.

Navaza,J. (1998) AMoRe: an automated package for molecular replacement. *Acta Crystallogr.*, **A50**, 157–163.

Newcomb,J.R. and Cresswell,P. (1993) Characterization of endogenous peptides bound to purified HLA-DR molecules and their absence from invariant chain-associated alpha beta dimers. *J. Immunol.*, **150**, 499–507.

Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–376.

Ogrinc,T., Dolenc,I., Ritonja,A. and Turk,V. (1993) Purification of the complex of cathepsin L and the MHC class II-associated invariant chain fragment from human kidney. *FEBS Lett.*, **336**, 555–559.

O'Sullivan,D.M., Noonan,D. and Quaranta,V. (1987) Four Ia invariant chain forms derive from a single gene by alternate splicing and alternate initiation of transcription/translation. *J. Exp. Med.*, **166**, 444–460.

Otwinowski,Z. and Minor,W. (1996) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.

Peterson,M. and Miller,J. (1992) Antigen presentation enchanced by the alternatively spliced invariat chain gene product p41. *Nature*, **357**, 596–598

Pierre,P. and Mellman,I. (1998) Developmental regulation of invariant chain proteolysis controls MHC class II trafficking in mouse dendritic cells. *Cell*, **93**, 1135–1145.

Podobnik,M., Kuhelj,R., Turk,V. and Turk,D. (1997) Crystal structure of the wild-type human procathepsin B at 2.5 Å resolution reveals the native active site of a papain-like cysteine protease zymogen. *J. Mol. Biol.*, **271**, 774–788

Riese,R.J., Wolf,P.R., Bromme,D., Natkin,L.R., Villadangos,J.A., Ploegh,H.L. and Chapman,H.A. (1996) Essential role for cathepsin S in MHC class II-associated invariant chain processing and peptide loading. *Immunity*, **4**, 357–366.

Riese,R.J., Mitchell,R.N., Villadangos,J.A., Shi,G.P., Palmer,J.T., Karp,E.R., De Sanctis,G.T., Ploegh,H.L. and Chapman,H. (1998) Cathepsin S activity regulates antigen presentation and immunity. *J. Clin. Invest.*, **101**, 2351–2363.

Roche,P.A. and Cresswell,P. (1991) Proteolysis of the class II-associated invariant chain generates a peptide binding site in intracellular HLA-DR molecules. *Proc. Natl Acad. Sci. USA*, **88**, 3150–3154.

Rodriguez,G.M. and Diment,S. (1995) Destructive proteolysis by cysteine proteases in antigen presentation of ovalbumin. *Eur. J Immunol.*, **25**, 1823–1827.

Šali,A. and Blundell,T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.

Strnad,J. *et al*. (1989) Molecular cloning and characterization of a human adenocarcinoma/epithelial cell surface antigen complementary DNA. *Cancer Res.*, **49**, 314–317.

Strubin,M., Berte,C. and Mach,B. (1986) Alternative splicing and alternative initiation of translation explain the four forms of the Ia antigen-associated invariant chain. *EMBO J.*, **5**, 3483–3488.

Stubbs,M.T., Laber,B., Bode,W., Huber,R., Jerala,R., Lenarčič,B. and Turk,V. (1990) The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *EMBO J.*, **9**, 1939–1947.

Turk,B., Turk,V. and Turk,D. (1997) Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. *Biol. Chem.*, **378**, 141–150.

Turk,D. (1992) Weiterentwicklung eines programmes für molekülgraphik und electronendichte- manipulation und seine anwendungen auf verschiedene protein-strukturaufklärungen. PhD Thesis, Technische Universität München, Germany.

Turk,D., Podobnik,M., Kuhelj,R., Dolinar,M. and Turk,V. (1996) Crystal structures of human procathepsin B at 3.2 and 3.3 Angstroms resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide. *FEBS Lett.*, **384**, 211–204.

Turk,D., Gunčar,G., Podobnik,M. and Turk,B. (1998) Revised definition of substrate binding sites of papain-like cysteine proteases. *Biol. Chem.*, **379**, 137–147.

Yamashita,M. and Konagaya,S. (1996) A novel cysteine protease inhibitor of the egg of chum salmon, containing a cysteine-rich thyroglobulin-like motif. *J. Biol. Chem.*, **271**, 1282–1284.

Received October 14, 1998;
revised and accepted December 15, 1998