PROTEINS: Structure, Function, and Genetics 32:504–514 (1998)

Structural Basis for Specificity of Papain-Like Cysteine Protease Proregions Toward Their Cognate Enzymes

Matthew R. Groves,[1] René Coulombe,[2,3] John Jenkins,[4] and Miroslaw Cygler[2,3]*
[1]*Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, Oxford, United Kingdom*
[2]*Biotechnology Research Institute, National Research Council of Canada, Montréal, Québec, Canada*
[3]*Protein Engineering Network of Centers of Excellence, Canada*
[4]*Department of Food Macromolecular Science, Institute of Food Research, Earley Gate, Whiteknights Road, Reading, United Kingdom*

ABSTRACT **Synthetic peptides corresponding to the proregions of papain-like cysteine proteases have been shown to be good and selective inhibitors of their parental enzymes. The molecular basis for their selectivity, quite remarkable in some cases, is not fully understood. The recent determination of the crystal structures of three distinct papain-like cysteine protease zymogens allows detailed structural comparisons to be made. The reasons for the specificity shown by each proregion toward its cognate enzyme are explained in terms of the three-dimensional structure of the proregion and the interface between the mature enzyme and the proregion. These comparisons reveal that insertion and substitution of amino acids within the proregion cause major rearrangement of sidechains on the enzyme/proregion interface, allowing detailed surface and charge recognition. Proteins 32:504–514, 1998.** © 1998 Wiley-Liss, Inc.

Key words: cysteine protease; zymogens; inhibition; caricain; cathepsin L

INTRODUCTION

Cysteine proteases represent a major fraction of lysosomal proteases. Their action has been implicated in a variety of intercellular and intracellular processes, including tumor invasion and metastasis,[1,2] cartilage degradation in arthritis,[3,4] Alzheimer's disease,[5] bone resorption,[6] allergic reactions,[7] intracellular protein degradation,[8] osteoporosis,[9] and prohormone processing.[10] The papain-like enzymes represent by far the most studied family of cysteine proteases, with the first crystal structure of papain being determined by Drenth et al.[11]

Due to the clinical and industrial importance of cysteine protease, much work has been carried out on the specific inhibition of these enzymes, with varying degrees of success. Most cysteine protease inhibitors, such as epoxides (E-64), chloromethyl ketones, and others, show little specificity between even distantly related members of the family (reviewed in Demuth[12] and Otto and Schirmeister[13]). Cystatins, which are a family of more specific macromolecular inhibitors of cysteine proteases,[13] are too large to be of much use for clinical or industrial applications. However, it has been shown that the prosegments of papain-like cysteine proteases display significant specificity toward their cognate enzymes.[14–16] The structural basis for this specificity of recognition is not well understood. As these proregions have a typical molecular weight of 7 to 12 kDa, they delineate a novel approach toward the rational design of specific cysteine protease inhibitors, and it is of considerable importance to understand the means by which they acquire their selectivity. The proregions consist of naturally occurring amino acids; therefore, there exists the potential for the coding sequence for the proregion alone to be inserted into a host genome. Trypsin isolated from the midgut of tobacco horn worm (*Manduca sexta*) has been shown to be inhibited in vitro by a synthetic propeptide,[17] and initial data suggest that recombinantly expressed proregions of papaya protease IV and papain inhibit the growth of the Colorado potato beetle (M. Taylor, personal communication).

Karrer et al.[18] proposed that the papain-like cysteine proteases may be subdivided into two subfamilies based on several characteristic sequence fragments and on the lengths of their respective prosegments. The canonical member of the subfamily with shorter prosegments is cathepsin B, and the length of the prosegment here is approximately 60 amino acids. Recently, crystal structures of procathepsin B of this subfamily have become available.[19–21]

The subfamily with a longer prosegment, of which papain (and mammalian cathepsin L) is the canonical member, has typically a prosegment of around 100 amino acids and also contains a conserved

*Correspondence to: Miroslaw Cygler, Biotechnology Research Institute, National Research Council of Canada, 6100 Royalmount Avenue, Montréal, Québec H4P 2R2, Canada. E-mail: mirek@bri.nrc.ca
Received 6 January 1998; Accepted 30 April 1998

Exxx**R**xxx**F**xx**N**xxx**I**xxx**N** sequence motif (hereafter known as the **ERFNIN** motif). The proregions of these two subfamilies do not show any significant sequence homology. The crystal structures of two proenzymes from this subfamily, procaricain[22] and procathepsin L,[23] have been reported elsewhere.

The structures of the proenzymes showed that despite sequence homology, the proregions of the two subfamilies are structurally similar. They share a common fold consisting of a helix-turn-strand hairpin supersecondary structure followed by an α-helix wedged into the S′ subsites of the substrate binding area. The proregion then follows along the S subsites. These structures provided insight into the mechanism of the inhibitory action of the proregions. Briefly, although the proregion extends along the substrate binding site and occupies S and S′ subsites, its orientation is opposite to that of the peptidyl substrate. This reverse orientation assures that the peptide bond near the catalytic Cys-His pair does not have a correct spatial arrangement for the catalytic events and the cleavage does not occur. At the same time, the proregion occludes the active site and shields it from access by the substrate, leading to the inhibition of enzymatic activity.

Activation of the enzyme occurs by removal of the proregion in the low pH environment (lysosome in case of cathepsins). In vitro, this is an autocatalytic process, although it is not clear if it occurs via an intermolecular or intramolecular route. Unlike in serine proteases, no conformational rearrangement occurs in the enzyme upon activation.

In many proteases the proregion plays an active role in protein folding. For example, the proregion is required for proper folding of subtilisin,[24] α-lytic protease,[25] carboxypeptidase A1,[26] and carboxypeptidase Y.[27] In the absence of their proregions these enzymes do not fold into active conformations. The available data for proteases from papain superfamily indicate that, at least for some of them, the proregion is also essential for folding of the enzyme (e.g., the proregion of cathepsin L has been shown to be crucial for the correct folding of the newly synthesized protein in vivo).[28] Experiments on folding of cysteine proteases in vitro, paralleling those performed for serine proteases, have not yet been undertaken. It is an unresolved question if the proregion on its own is folded in solution. The NMR data for the shorter proregion, that of cathepsin B, indicate that this polypeptide is indeed unfolded in solution (F. Ni, personal communication).

The proregions of proteases from the papain superfamily also fulfill other biological roles. Their presence provides stabilization of the proteases when exposed to neutral or alkaline pH environments.[28,29] The proregion of cathepsin L has been implicated in the targeting of the enzyme, as it has been shown to contain a sequence recognized by membrane proteins.[30,31]

The three-dimensional structures of the proenzymes of papain-like cysteine proteases provide the framework that is essential for the identification of structural elements defining the proregion specificity. We will compare the available structures and propose the areas of the prosegments that contain the specificity determinants. In the following discussion, we will refer to residues according to their numbers in the cathepsin L sequence unless indicated otherwise. The residues of the proregion are indicated by the suffix "p."

RESULTS AND DISCUSSION

Subfamilies

The structures of procathepsin B,[19–21] procaricain,[22] and procathepsin L[23] all indicate that the mature portion of the protein is found in the active conformation, with the prosegment binding to the mainly β-sheet, C-terminal domain of the mature enzyme and to residues within the active site cleft. The proregion of the cathepsin L subfamily consists of three α-helices and a short β-strand, folded into a discrete domain.[22,23] The second helix and following β-strand run antiparallel to each other and form a hairpin structure (in the nomenclature of Cygler et al.[19]). The shorter proregions of the enzymes belonging to the cathepsin B subfamily lack the first α-helix[19,20] and are unlikely to fold as a discrete domain on their own, although the N-terminal α-helix and the β-strand are in approximately the same relative orientation as in the other subfamily and also form a hairpin (Fig. 1). The familial difference in proregion length is likely related to the presence of an additional loop in the vicinity of the active site in the procathepsin B subfamily, the so-called occluding loop, which is essential for the exopeptidase activity of this enzyme.[32–34] This loop provides additional hydrophobic and electrostatic contacts between cathepsin B and its shorter proregion but would preclude binding of the longer proregion, if it were in the same conformation as observed in the cathepsin L subfamily. In the three different crystal structures of procathepsin B that are available, the center of this loop assumes different conformations and is less well ordered than the rest of the structure. It has been suggested that the observed conformational differences reflect differences in crystal packing around this intrinsically more flexible loop.[21]

The significant local differences between the two subfamilies, that is the presence/absence of the occluding loop and associated difference in the length of the proregion, make the observed selectivity of the proregions across the subfamilies explainable in global terms, as described above. However, the observed selectivity within one subfamily requires a detailed structural interpretation and is of significant interest. We will concentrate below on the proenzymes from the larger of the two subfamilies, represented by procathepsin L and procaricain.

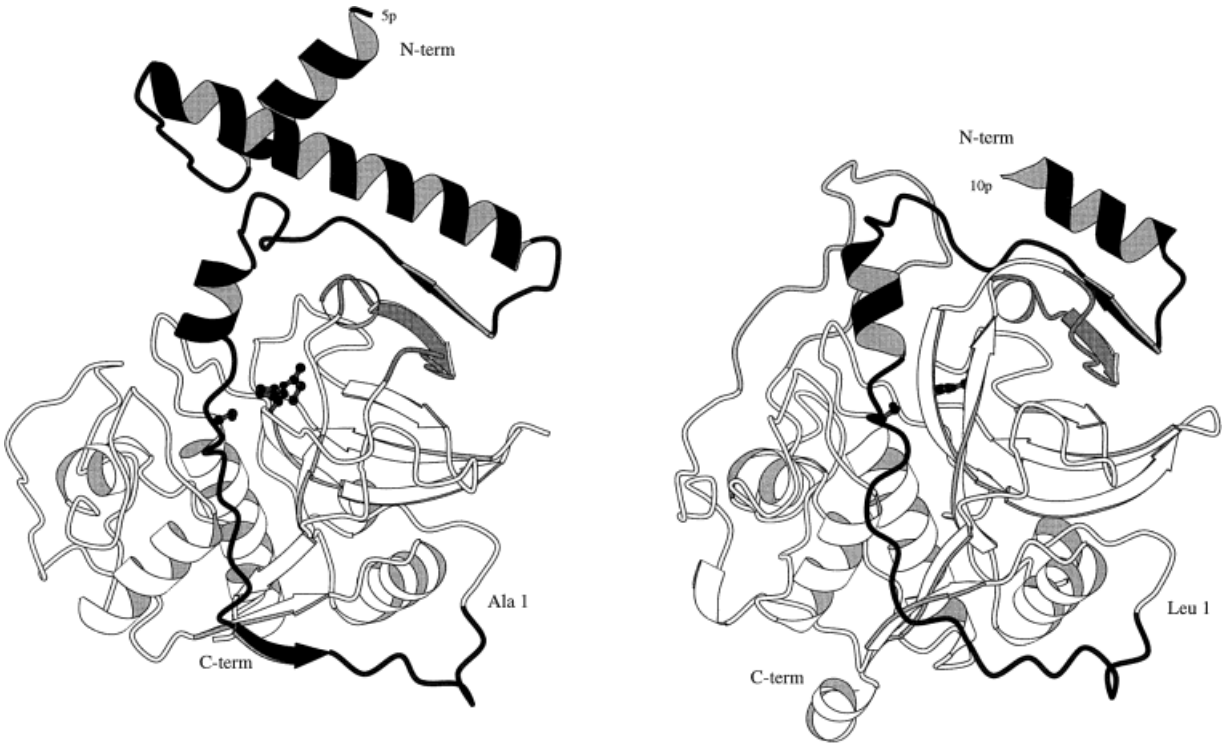

Fig. 1. Representations of procathepsin L (**left**) and procathepsin B (**right**). The proregions are shown in black, and the active site residues are shown in full in ball-and-stick representation. The N-terminal residue of the mature form is marked. The proregion binding loop is shown in dark gray, and the occluding loop in procathepsin B is shown in light gray.

Overall Comparison of Procaricain and Procathepsin L

The mature portions of the zymogen structures of caricain and cathepsin L are very similar. Using a cutoff of 3.0 Å there are 192 structurally equivalent residues (out of 216) that superimpose with an root mean square (rms) deviation on Cα's for equivalent atoms of 0.8 Å (Fig. 2). However, the electrostatic properties of the two proteins are significantly different. Calculation of the electrostatic potential along the contact surfaces between the mature enzyme and the proregion for either the mature enzyme alone or for the proregion alone, reveals that the surface of the mature cathepsin L that interacts with the prosegment has a largely negative potential, whereas the equivalent surface of caricain has a largely positive potential. The corresponding contacting surfaces of the proregions have opposite electrostatic characters to that of the mature proteins (Fig. 3). Therefore, the electrostatic features of the contacting surfaces provide one mechanism contributing to the specificity of the proregion-parental enzyme interactions.

There exist two main areas of contact between the proregion and the mature enzyme. The first of these is in the region of the loop 140–154 (138–152 papain numbering), known as the proregion binding loop (PBL).[19] The differences in electrostatic potential along the surface facing the proregion reflect, to a large extent, the differences in amino acid composition of the PBL. In particular, position 141, 142 and 147, 148 (139, 140, 145, 146 papain numbering) in this enzyme family are often occupied by charged amino acids, either basic or acidic.

The second area of contact is in the substrate binding site along the S2 to S2′ subsites. In particular, within the S′ sites the proregions of both enzymes pack onto a hydrophobic surface, which contains two conserved aromatic residues, in most cases tryptophans—Trp189, Trp193 (Trp177 and Trp181, papain numbering)—forming part of a highly conserved larger aromatic cluster, which extends in one direction to PBL (Phe143, Tyr146, Tyr/Phe151) and in the other direction to the catalytic His159. The lower than average temperature factors of residues in both structures are found in these regions, indicating that these areas represent very stable interfaces between the proregion and mature protein. Therefore, these two interfaces are likely to be the determining surfaces for specificity.

Proregions of Procaricain and Procathepsin L

The prosegments of procaricain and procathepsin L share the same fold, namely short N-terminal helix α1, long helix α2, strand β1, turn, short helix α3, and an extended C-terminal fragment with a short strand

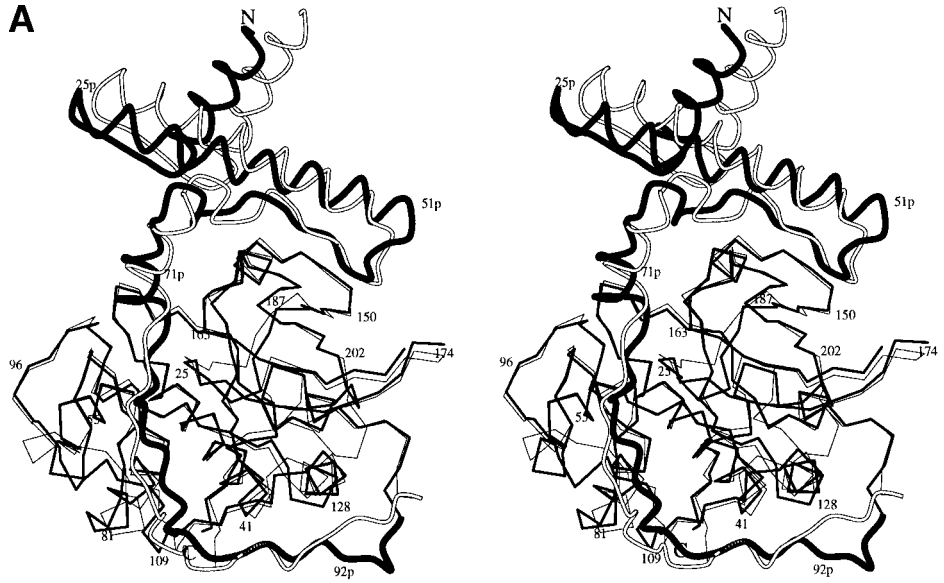

Fig. 2. **A:** Superposition of Cα traces of procathepsin L and procaricain. The transformation matrix was based on the mature portions of the two enzymes. Procathepsin L is shown in thick black lines. The prosegments are shown in shaded lines. **B:** Structure-based sequence alignment of procathepsin L and procaricain. Sequence of propapain was added as a reference. Residues that are conserved in these three sequences are marked with stars. The beginning of the proregion (past the signal peptide) and the beginning of the mature enzyme are marked with an arrow. Residues forming the active site are marked with plus signs.

β2 near the end. The first ~65 residues of the prosegments show good spatial correspondence, although there is some difference in the position of the C-terminal fragment starting from helix α3. The superposition of the prosegments, based on the backbone atoms of helices α1, α2, and the following β1-strand only, is shown in Figure 4. The rms deviation for the backbone atoms of 48 structurally equivalent residues is 1.0 Å. The superposition shows clearly that there are local differences in the loops connecting helices α1 and α2 (loop1) and those between helix α2 and strand β1 (loop2), originating from insertions/deletions. Specifically, loop1 in procathepsin L is one residue shorter than that in procaricain, with a deletion at position 24p. On the other hand, loop2 is four residues longer in procathepsin L, with the insertion centered at position 52p. The angle between helices α1 and α2 differs slightly between the two proregions; in procathepsin L it is 108°, and in procaricain this angle is 100°. The packing of helices α1 and α2 is determined by the interactions of three aromatic residues at positions 12p, 15p, and 35p. The character of residues in these three positions is conserved without exception in all known proregion sequences from the cathepsin L subfamily. The stability of this core is further supported by salt bridges and hydrogen bonds between highly conserved residues: Arg21p . . . Asp65p . . . Tyr23p and Glu27p . . . Arg31p. Some of these residues come from two previously identified, highly conserved motifs: **ERFNIN**[18] and **GNFD**.[35]

The relative orientation of helix α2 and strand β1 is similar in both prosegments despite a large difference in the length of the loop connecting them. The two residues that to a large extent stabilize this arrangement are Ile42p, through hydrophobic contacts, and Asn46p (procathepsin L numbering), which forms two hydrogen bonds to the backbone of strand β1. Both of them are part of the **ERFNIN** motif. A very similar stabilization of the hairpin is observed in the proregion of cathepsin B, namely Asn18p forms identical hydrogen bonds to the β1-strand.

The largest differences occur around helix α3, which is translated by half a turn and tilted by approximately 20° when the two proregions are compared. The C-terminal tail follows roughly the same path but makes different excursions. A short β2 strand is formed in both cases with the corresponding region on the mature enzyme. This shift allows the amino acids of helix α3 to make different interactions with the surface of the mature enzymes. These interactions will be discussed along with the interactions made by the proregion in the vicinity of the S2′ pocket.

The ERFNIN Motif

The structural significance of the **ERFNIN** motif has been mentioned elsewhere.[22,23] Its primary role is to restrain the three-dimensional arrangement of the three α-helices and a β-strand of the proregion into a discrete globular domain. The contribution of most of these residues has been described in the references above. The highly conserved Asn38p is involved in the interactions with the distal loop

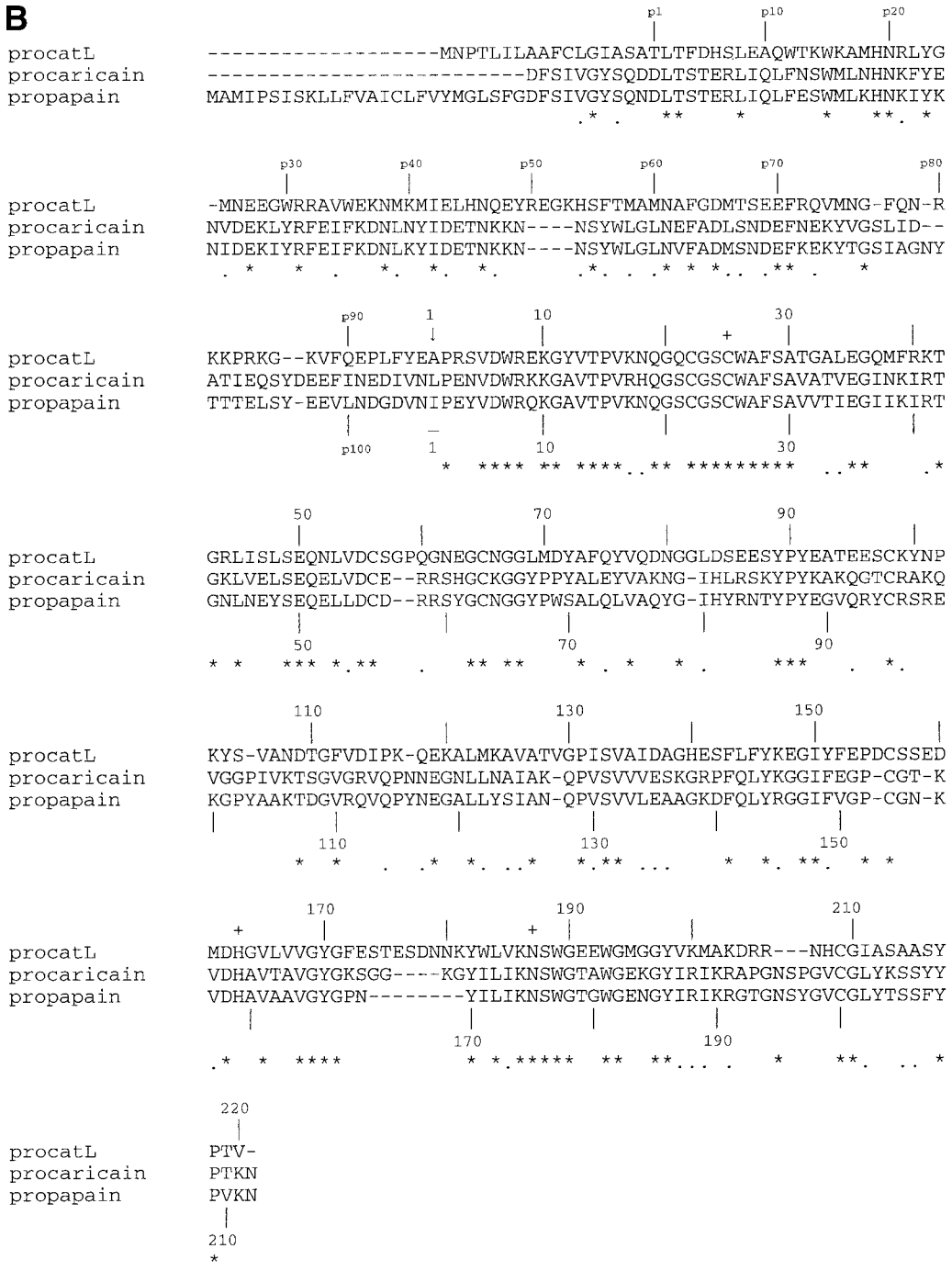

Figure 2. (Continued.)

between strand β1 and helix α3 (loop3) of the proregion: it forms a hydrogen bond to the backbone of residue 64p.

The preceding residue, Lys37p (procathepsin L) or its equivalent Asp48p in procaricain, makes contacts with residues of the mature protease. There appears to be a correlation between the charge on this residue and the hydrophilicity of the contacting surface on the mature enzyme. This choice of Lys37p (procathepsin L) or Asp48p (procaricain) could be a determinant in the recognition of a highly charged residue in the region of the PBL and a mechanism for specific subdivision of papain-like cysteine proteases into proteins with positive, negative, and uncharged residues at the 139 position (papain numbering). This position is on the N-terminal side of the helix

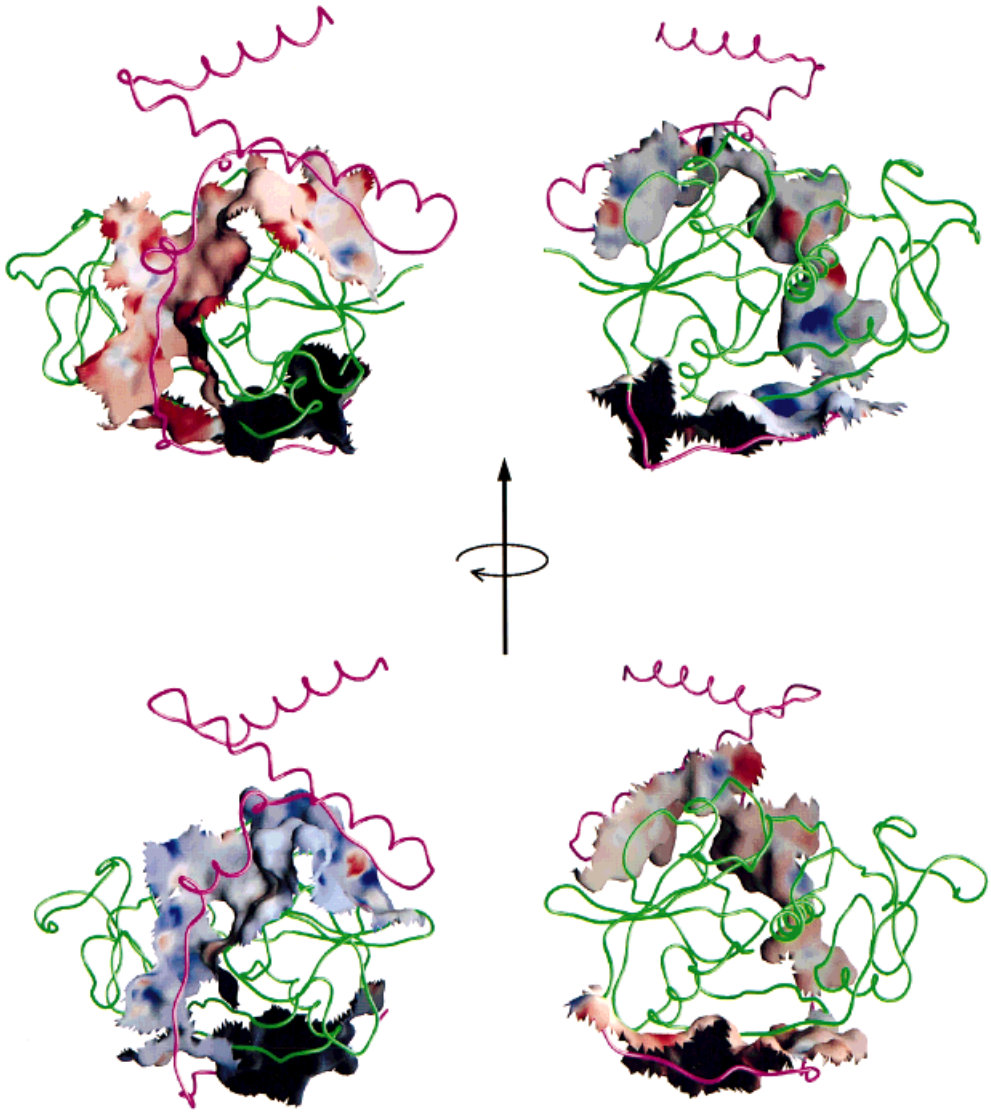

Fig. 3. The potential distribution on the contact surfaces between the proregion and the mature enzyme. Procathepsin L is shown on the top, and procaricain is shown on the bottom. Left view shows the contact surface of the mature protease; right view shows the contact surface of the proregion. The external sides of the surfaces are colored according to the value of the potential; the inner sides are in dark colors. Negative potential is in red, and positive potential is in blue. The proteins are shown as coils, green for the mature enzyme and purple for the proregion.

which caps the C-terminal domain of the papain-like enzymes.

Binding to the PBL Region

The proregion binding loops of caricain and cathepsin L have very similar conformations. The 15 residue long loops contain only 5 identical amino acids, yet they align to an rms deviation of 1.0 Å on main chain atoms. The main contacts occur between the aromatic residue, Phe56p (Tyr63p in procaricain), and the highly conserved aromatic cluster in the depression in the center of the PBL. Also, the residue two positions downstream (Met58p) is always an aliphatic residue and is pointing toward the center of the PBL. Although the van der Waals contacts made by the aromatic residue (56p) are similar in both proteins, there are also local differences. In particular, the hydroxyl oxygen of Tyr63p in procaricain forms a hydrogen bond to the carbonyl oxygen of Gly151 (2.3 Å). This hydrogen bond cannot be formed in procathepsin L. The conformation of the Gly151-Pro152 peptide bond within PBL is quite different in the two proenzymes: it is *cis* in procaricain (and also in the mature caricain[36]), whereas the corresponding Glu153-Pro154 bond in procathepsin L assumes the *trans* conformation.

Alignment of the two proenzyme structures based on the local superposition of their proregion binding

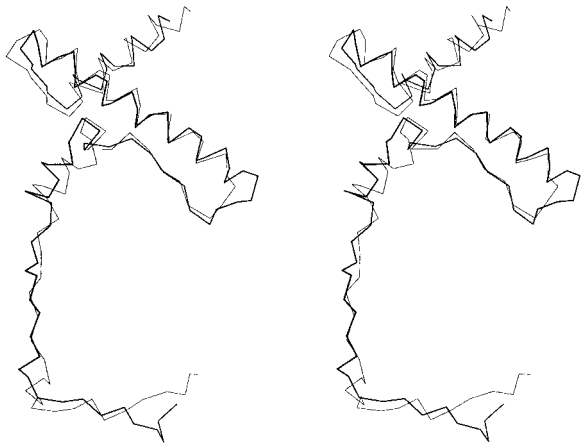

Fig. 4. Local superposition of the backbones of the prosegments of procathepsin L (thick lines) and procaricain (thin lines) based on the backbone atoms of helices α1, α2, and strand β1.

loops is quite instructive from the selectivity viewpoint. The potential interactions between the proregion of caricain and the PBL of cathepsin L indicate that with the exception of Trp64p, the procaricain peptide 62p-66p (55p-59p in procathepsin L) makes a good fit with the surface of the PBL of cathepsin L (Fig. 5). The arrangement of the prosegment on the PBL brings position 57p in procathepsin L (Trp64p in procaricain) into a close contact with residue 148 in cathepsin L (146 in caricain/papain). Analysis of the types of sidechains occurring in these two positions in the sequences of other proteins from the cathepsin L subfamily leads to several conclusions. First, the sidechain on the prosegment (57p) is usually large, whereas that on the enzyme (148) is small. In particular, tryptophan on the prosegment is accompanied by a glycine on the enzyme. Second, a charged sidechain at enzyme's position 148 usually coexists with an opposite charge at position 57p on the prosegment. One can conclude that binding of an aromatic residue from the prosegment to the center of the PBL, while providing significant binding energy, does not provide specificity due to its highly conserved character. On the other hand, the sidechain in position 57p (64p in caricain) clearly provides some specificity either by its size and/or by its charge.

Binding to the S Sites

The prosegment residue closest to the catalytic cysteine in procathepsin L is Gly77p. However, in procaricain there is an insertion of Ser85p between the corresponding Gly84p and Leu86p (Fig. 6). This insertion in procaricain within the region of the prosegment that binds to the S1 site impacts mostly the binding to the S′ sites (see below), while the binding to the S sites is less affected. The largest difference is at the site of insertion, the S1 site, where the peptide of procaricain penetrates less deeply into the substrate binding cleft of caricain than does the equivalent peptide in procathepsin L (Fig. 6b). However, both proenzymes bind to the S2 and S3 sites in a very similar fashion. The insertion causes a small, ~1 Å, shift of the caricain proregion toward the C-terminus but does not greatly affect the penetration of the hydrophobic sidechain into the S2 pocket, the main specificity determinant for the substrate. These contacts are augmented by two hydrogen bonds to Gly68 (in procaricain one of them, to the corresponding Gly 66, is somewhat weaker), and the next sidechain occupies the same volume in the S3 site. These interactions provide an anchor for the prosegment near the active site. In the immediate surroundings of the catalytic cysteine there are more pronounced differences, particularly that the carbonyl oxygen of Asn76p of procathepsin L is placed in the oxyanion hole, but neither the carbonyl oxygen of the corresponding Val83p of procaricain nor oxygen of the equivalent Cys42p in the two structures of procathepsin B are pointing to the oxyanion hole. There is no hydrogen bonding to the active site residues in procathepsin L, whereas in the structure of procaricain (His159Gly, active site mutant) the proregion hydrogen bonds to the active site cysteine and is within hydrogen bonding distance of the modeled active site histidine. In both structures, proregion residues within the substrate binding cleft display higher temperature factors than those of residues comprising the active site cleft. This is indicative of a higher degree of flexibility of the S_1'-S_2 peptide in both structures.

Residues C-terminal to those occupying the S3 site display no immediate similarity, except for residue Phe89p in procaricain and corresponding Phe99p in procaricain, which bind to the equivalent hydrophobic pockets on the enzyme's surfaces. The character of the residues lining the pocket (Ile115, Ala121, Ala125, Val129, and Tyr217 in procathepsin L) is reasonably well conserved in other enzymes from this family.

Helix α3 and the S′ Sites

The one residue insertion (Ser85p) in the part of the procaricain proregion that occupies the S1 site, relative to that of procathepsin L, is accompanied by a shift in the position of helix α3 in the former protein away from the active site by approximately 4 Å. This positional difference involves almost no rotational component around the helix axis (Fig. 2a). The side of helix α3 directed toward the enzyme contains large hydrophobic sidechains at two consecutive turns: Met75p (Tyr82p in procaricain), which enters the S2′ site, and Phe71p (Phe78p in procaricain). In the aligned sequences of the cysteine protease proregions, the character of these two sidechains is well conserved. They point into a hydrophobic pocket on the enzyme's surface, at the bottom and side of which are highly conserved Trp189 and

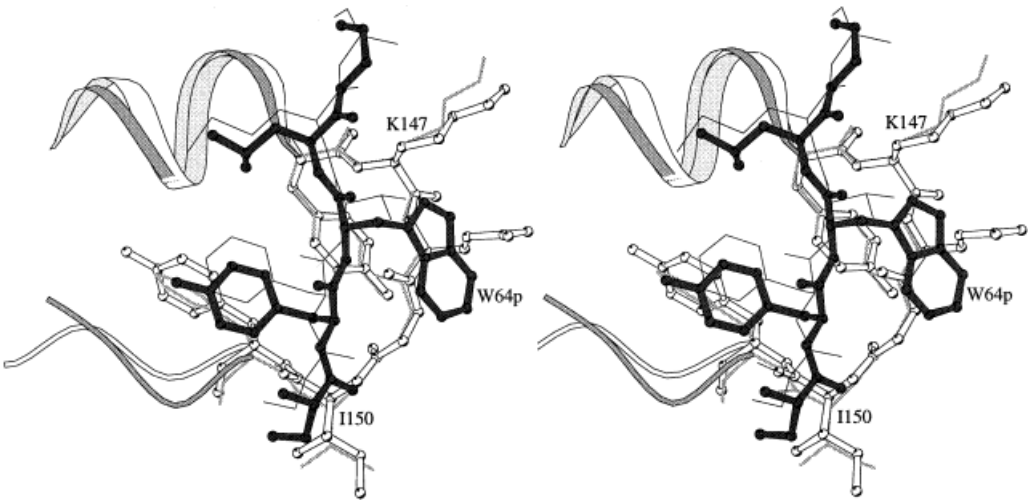

Fig. 5. Superposition of the PBL of cathepsin L (open lines) with the prosegment fragment of procaricain (filled lines). The superposition is based on the backbone atoms of the two proregion binding loops. The PBL of caricain (gray lines) and proregion of cathepsin L (thin black lines) are shown for comparison. Numbering for PBL refers to cathepsin L, and for proregion refers to procaricain.

Trp193 (Fig. 7). The shift in the helix results in a difference in the direction from which these sidechains approach the pocket, particularly evident for the Phe71p in procathepsin L and its counterpart Phe78p in procaricain (Fig. 7). In addition, an aromatic residue from loop3, Phe63p (Phe70p in procaricain, also highly conserved), participates in this hydrophobic cluster. The conformation of the loop around the latter residue is not identical in the two proregions but seems to be driven by the need to sequester this aromatic residue into a hydrophobic environment.

We generated hybrid computer models to test the possible contribution of this part of the proregion to the selectivity. To this end, we superimposed the proenzymes based on the best superposition of their mature portions and then combined the mature enzyme of each enzyme with the proregion from its counterpart. In the case of cathepsin L with a docked caricain proregion, the sidechain of Phe78p is within 1.5 Å of Leu144 (cathepsin L), indicating potential steric clash. Leu144 is flanked on the other side by Phe145 and left with very little room to move. Analogous residues in caricain are Gln143 and Leu143, with glutamine protruding less on the surface and making good van der Waals contacts with Phe78p. The replacement of Met75p in procathepsin L, which penetrates into the S2′ site, with a tyrosine in procaricain appears not to interfere with binding (Fig. 7a). In the alternative model, that of the caricain with the docked proregion of cathepsin L, there are no particularly bad contacts within this region. However, Phe63p is partially solvent-exposed, which could have a destabilizing effect on the interface. There is only one close contact, that between Phe71p and Ser136 (caricain, 2.7 Å).

CONCLUSIONS

The recent determination of the three-dimensional structures of three proenzymes from the papain superfamily provides the first opportunity to investigate the molecular basis for the proregion selectivity for the cognate enzyme. Although the selectivity across the subfamilies is not too surprising due to differences in the enzyme (occluding loop) and in proregion lengths, the selectivity within the cathepsin L subfamily has more subtle underpinnings.

The comparison of the structures of procaricain and procathepsin L indicates that the binding of the prosegment is governed by the interactions in two key areas: aromatic ring stacking to the center of the PBL accompanied by the formation of a short β-sheet, and the packing of the sidechains into the S and S′ sites. The proregion inhibits the enzymatic activity by occupying the substrate binding site. Its cleavage does not occur because it lines the binding site in the direction opposite to that of a peptidyl substrate. The interactions within the S sites with concomitant hydrogen bonds to the highly conserved Gly68 are similar in the investigated structures. Although these interactions provide significant binding energy,[37] they do not differ greatly between the two enzymes. The anchoring role of proregion residues that occupy the S2-S3 sites is confirmed also by the structure of the wild type procathepsin L in which the Cys25 is oxidized. The position of the prosegment near this cysteine is somewhat different from that in

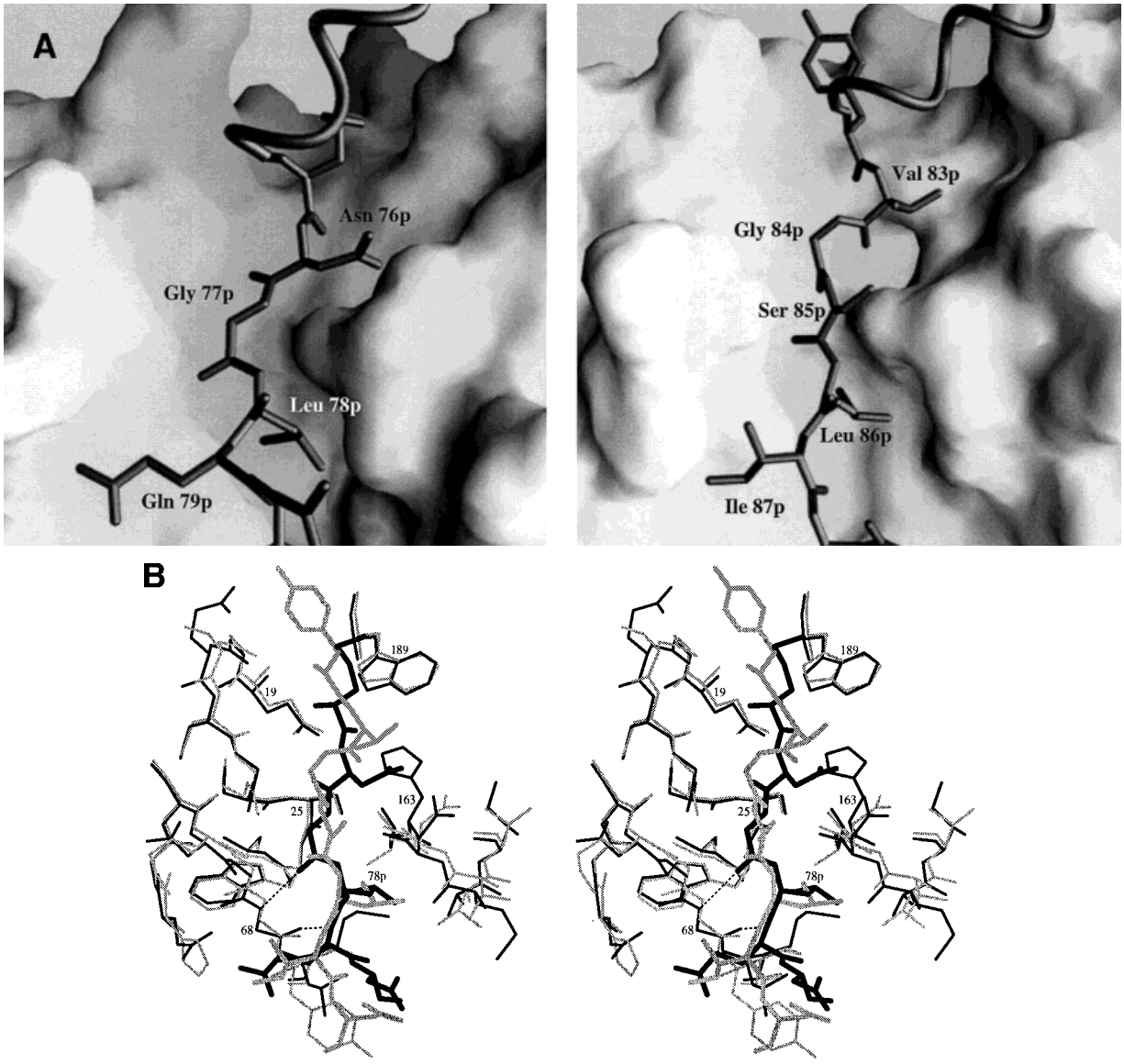

Fig. 6. Comparison of the proregions of cathepsin L and caricain in the vicinity of the active site. **A:** Surface representations of cathepsin L and caricain with their proregions shown as sticks, **B:** Stereo view of the superimposed S-S′ sites and the proregion fragments lining the active site. The superposition is based on the backbones of mature enzymes. Procathepsin L is shown in black lines, and procaricain is shown in gray lines.

the previously determined mutant, but only the disposition of helix α3 is affected; the two anchor sites (S sites and binding to the center of PBL) remain unchanged (R. Coulombe, M.Cygler, unpublished data). The parts of the proregions spanning these two sites form numerous contacts with the enzyme (especially in the S′ sites), show differences in relative orientation toward the enzymes, and vary in sequence. Thus, it is the most likely region of the proregion to contain the principal specificity determinants toward the cognate protease.

This comparison suggests that there are several factors involved in specificity of recognition between the prosegment and its parent protease. An important consideration has to be given to charge distribution differences between various mature enzymes. The prosegments display compensatory charge distributions to those of the mature enzymes, and in particular positions 57p and 148 seem to be either bearing compensatory charges or have hydrophilic characters. In this respect, the portion of the proregion C-terminal to the active site, although it does not contribute substantially to the binding/inhibition constant toward the cognate enzyme, provides compensating charges to those on the contacting surface of the enzyme and may, therefore, contribute to the proenzyme's specificity. The specific interactions between the prosegment and the enzyme are

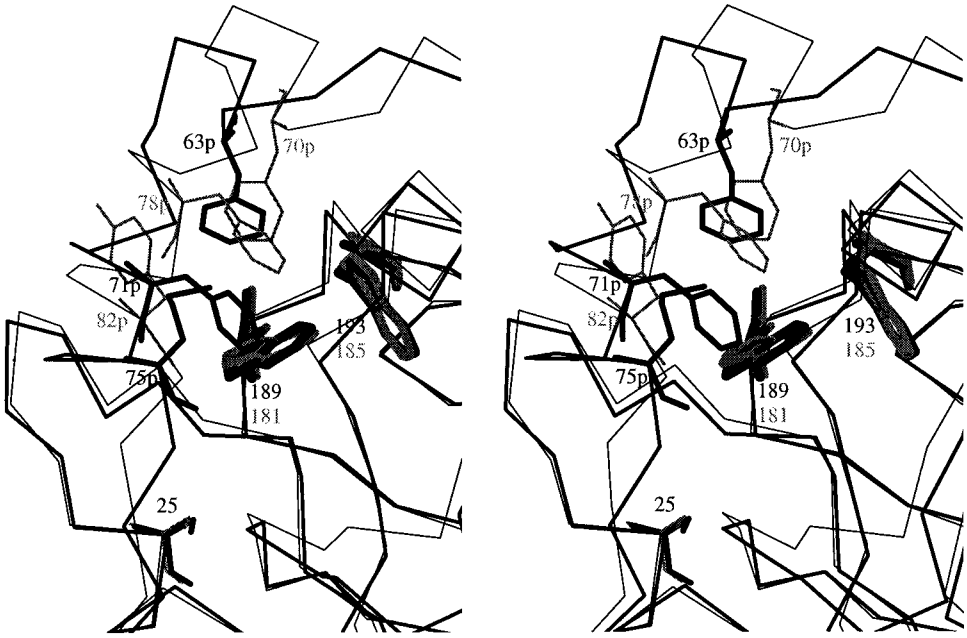

Fig. 7. The hydrophobic pocket on the S′ side of the active site. Highly conserved tryptophans delineating this site are shown in full. Superposition is based on the mature enzymes. Procathepsin L is shown in thick black lines, and procaricain is shown in thin and gray lines.

located mostly within the PBL and the S′ sites. The hydrophobic packing interactions around the S2′ pocket are of particular importance. Our analysis shows that the proenzymes also have an unusual way to modify their binding to the enzyme, through an insertion of a residue into the part that lines the center of the substrate binding cleft. By this means, the spatial disposition of helix α3 relative to the active site of the enzyme can be modified to fit to the enzyme's surface in the S′ sites. Although the first view of the basis for proregion specificity has been gained, several other proenzyme structures will have to be determined to obtain an understanding of the full complexity of this recognition process.

REFERENCES

1. Yagel, S., Warner, A.H., Nellans, H.N., Lala, P.K., Waghorne, C., Denhardt, D.T. Suppression by cathepsin L inhibitors of the invasion of amnion membranes by murine cancer cells. Cancer Res. 49:3553–3557, 1989.
2. Sloane, B.F., Honn, K.V. Cysteine proteinases and metastasis. Cancer Metastasis Rev. 3:249–263, 1984.
3. Esser, R.E., Angelo, R.A., Murphey, M.D., et al. Cysteine proteinase inhibitors decrease articular cartilage and bone destruction in chronic inflammatory arthritis. Arthritis Rheum. 37:236–247, 1994.
4. Esser, R.E., Watts, L.M., Angelo, R.A., Thornburg, L.P., Prior, J.J., Palmer, J.T. The effects of fluoromethyl ketone inhibitors of cathepsin B on adjuvant induced arthritis. J. Rheumatol. 20:1176–1183, 1993.
5. Cataldo, A.M., Paskevich, P.A., Kominami, E., Nixon, R.A. Lysosomal hydrolases of different classes are abnormally distributed in brains of patients with Alzheimer disease. Proc. Natl. Acad. Sci. USA 88:10998–11002, 1991 [erratum in 89:2509, 1992].
6. Vaes, G., Delaisse, J.M., Eeckhout, Y. Relative roles of collagenase and lysosomal cysteine-proteinases in bone resorption. Matrix Suppl. 1:383–388, 1992.
7. Kalsheker, N.A., Deam, S., Chambers, L., Sreedharan, S., Brocklehurst, K., Lomas, D.A. The house dust mite allergen Der p1 catalytically inactivates alpha 1- antitrypsin by specific reactive centre loop cleavage: A mechanism that promotes airway inflammation and asthma. Biochem. Biophys. Res. Commun. 221:59–61, 1996.
8. Bond, J.S., Butler, P.E. Intracellular proteases. Annu. Rev. Biochem. 56:333–364, 1987.
9. Votta, B.J., Levy, M.A., Badger, A., Bradbeer, G., Dodds, R.A., James, I.E., Thompson, S., Bossard, M.Y., Carr, T., Connor, J.R., Tomaszek, T.A., Szewczuk, L., Drake, F.H., Weber, D.F., Gowen, M. Peptide aldehyde inhibitors of cathepsin K inhibit bone resorption both in vitro and in vivo. J Bone Miner. Res. 12:1396–1406, 1997.
10. Azaryan, A.V., Hook, V.Y. Distinct properties of prohormone thiol protease (PTP) compared to cathepsins B, L, and H: Evidence for PTP as a novel cysteine protease. Arch. Biochem. Biophys. 314:171–177, 1994.
11. Drenth, J., Jansonius, J.N., Koekoek, R., Swen, H.M., Wolthers, B.G. Structure of papain. Nature 218:929–932, 1968.
12. Demuth, H.U. Recent developments in inhibiting cysteine and serine proteases. J. Enzym. Inhib. 3:249–278, 1990.
13. Otto, H.H., Schirmeister, T. Cysteine proteases and their inhibitors. Chem. Rev. 97:133–171, 1997.
14. Fox, T., de Miguel, E., Mort, J.S., Storer, A.C. Potent slow-binding inhibition of cathepsin B by its propeptide. Biochemistry 31:12571–12576, 1992.
15. Taylor, M.A.J., Baker, K.C., Briggs, G.S., et al. Recombinant pro-regions from papain and papaya proteinase IV-are selective high affinity inhibitors of the mature papaya enzymes. Protein Eng. 8:59–62, 1995.
16. Carmona, E., Dufour, E., Plouffe, C., et al. Potency and selectivity of the cathepsin L propeptide as an inhibitor of cysteine proteases. Biochemistry 35:8149–8157, 1996.
17. Taylor, M.A., Lee, M.J. Trypsin isolated from the midgut of the tobacco hornworm, Manduca sexta, is inhibited by

synthetic pro-peptides in vitro. Biochem. Biophys. Res. Commun. 235:606–609, 1997.
18. Karrer, K.M., Peiffer, S.L., DiTomas, M.E. Two distinct gene subfamilies within the family of cysteine protease genes. Proc. Natl. Acad. Sci. USA 90:3063–3067, 1993.
19. Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A.C., Mort, J.S. Structure of rat procathepsin B: model for inhibition of cysteine protease activity by the proregion. Structure 4:405–416, 1996.
20. Turk, D., Podobnik, M., Kuhelj, R., Dolinar, M., Turk, V. 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. 384:211–214, 1996.
21. Podobnik, M., Kuhelj, R., Turk, V., Turk, D. Crystal structure of the wild-type human procathepsin B at 2.5 A resolution reveals the native active site of a papain-like cysteine protease zymogen. J. Mol. Biol. 271:774–788, 1997.
22. Groves, M.R., Taylor, M.A.J., Scott, M., Cummings, N.J., Pickersgill, R.W., Jenkins, J.A. The prosequence of procaricain forms an alpha-helical domain that prevents access to the substrate-binding cleft. Structure 4:1193–1203, 1996.
23. Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J.S., Cygler, M. Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. EMBO J. 15:5492–5503, 1996.
24. Bryan, P., Wang, L., Hoskins, J., et al. Catalysis of a protein folding reaction: mechanistic implications of the 2.0 A structure of the subtilisin-prodomain complex. Biochemistry 34:10310–10318, 1995.
25. Baker, D., Sohl, J.L., Agard, D.A. A protein-folding reaction under kinetic control. Nature 356:263–265, 1992.
26. Phillips, M.A., Rutter, W.J. Role of the prodomain in folding and secretion of rat pancreatic carboxypeptidase A1. Biochemistry 35:6771–6776, 1996.
27. Ramos, C., Winther, J.R., Kielland-Brandt, M.C. Requirement of the propeptide for in vivo formation of active yeast carboxypeptidase Y. J. Biol. Chem. 269:7006–7012, 1994.
28. Tao, K., Stearns, N.A., Dong, J., Wu, Q.L., Sahagian, G.G. The proregion of cathepsin L is required for proper folding, stability, and ER exit. Arch. Biochem. Biophys. 311:19–27, 1994.
29. Mach, L., Mort, J.S., Glossl, J. Noncovalent complexes between the lysosomal proteinase cathepsin B and its propeptide account for stable, extracellular, high molecular mass forms of the enzyme. J. Biol. Chem. 269:13036–13040, 1994.
30. McIntyre, G.F., Erickson, A.H. The lysosomal proenzyme receptor that binds procathepsin L to microsomal membranes at pH 5 is a 43-kDa integral membrane protein. Proc. Natl. Acad. Sci. USA 90:10588–10592, 1993.
31. McIntyre, G.F., Godbold, G.D., Erickson, A.H. The pH-dependent membrane association of procathepsin L is mediated by a 9-residue sequence within the propeptide. J. Biol. Chem. 269:567–572, 1994.
32. Musil, D., Zucic, D., Turk, D., et al. The refined 2.15 A X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. EMBO J. 10:2321–2330, 1991.
33. Turk, D., Podobnik, M., Popovic, T., et al. Crystal structure of cathepsin B inhibited with CA030 at 2.0-A resolution: A basis for the design of specific epoxysuccinyl inhibitors. Biochemistry 34:4791–4797, 1995.
34. Illy, C., Quraishi, O., Wang, J., Purisima, E., Vernet, T., Mort, J.S. Role of the occluding loop in cathepsin B activity. J. Biol. Chem. 272:1197–1202, 1997.
35. Vernet, T., Tessier, D.C., Chatellier, J., et al. Structural and functional roles of asparagine 175 in the cysteine protease papain. J. Biol. Chem. 270:16645–16652, 1995.
36. Pickersgill, R.W., Rizkallah, P., Harris, G.W., Goodenough, P.W. Determination of the structure of papaya protease omega. Acta Crystallogr. B. 47:766–771, 1991.
37. Chen, Y., Plouffe, C., Menard, R., Storer, A.C. Delineating functionally important regions and residues in the cathepsin B propeptide for inhibitory activity. FEBS Lett. 393:24–26, 1996.