Structures at the Proteolytic Processing Region of Cathepsin D*

(Received for publication, May 18, 1988)

Satoshi Yonezawa†, Takayuki Takahashi‡, Xin-juan Wang, Ricky N. S. Wong¶, Jean A. Hartsuck§, and Jordan Tang

From the Laboratory of Protein Studies, Oklahoma Medical Research Foundation, and the Department of Biochemistry and Molecular Biology, University of Oklahoma Health Science Center, Oklahoma City, Oklahoma 73104

The amino acid sequences at the “proteolytic processing regions” of cathepsin Ds have been determined for the enzymes from cows, pigs, and rats in order to deduce the sites of cleavage as well as the function of the proteolytic processing of cathepsin D. For bovine cathepsin D, the “processing region” sequence was determined from a peptide isolated from the single-chain enzyme. The COOH-terminal sequence of the light chain and the NH2-terminal sequence of the heavy chain were also determined. The processing region sequence of porcine cathepsin D was determined from its cDNA structure, and the same structure from rat cathepsin D was determined from the peptide sequence of the single-chain rat enzyme.

From sequence homology to other aspartic proteases whose X-ray crystallographic structures are known, such as pepsinogen and penicillopepsin, it is clear that the processing regions are insertions to form an extended β-hairpin loop between residues 91 and 92 (porcine pepsin numbers). However, the sizes of the processing regions of cathepsin Ds from different species are considerably different. For the enzymes from rats, cows, pigs, and human, the sizes of the processing regions are 6, 9, 9, and 11 amino acid residues, respectively. The amino acid sequences within the processing regions are considerably different. In addition, the proteolytic processing sites were found to be completely different in the bovine and porcine cathepsin Ds. While in the porcine enzyme, an Asn-Ser bond and a Gly-Val bond are cleaved to release 5 residues as a consequence of the processing; in the bovine enzyme, two Ser-Ser bonds are cleaved to release 2 serine residues. These findings would argue that the in vivo proteolytic processing of the cathepsin D single chain is probably not carried out by a specific “processing protease.”

Model building of the cathepsin D processing region conformation was conducted utilizing the homology between procathepsin D and porcine pepsinogen. The β-hairpin structure of the processing region was found to (i) interact with the activation peptide of the procathepsin D in a β-structure and (ii) place the Cys residue in the processing region within disulfide linkage distance to Cys-27 of cathepsin D light chain.

These observations support the view that the processing region of cathepsin D may function to stabilize the conformation of procathepsin D and may play a role in its activation.

Cathepsin D (EC 3.4.23.5) is an aspartic protease located in the lysosomes of all mammalian cells (see reviews in Refs. 1–3). The function of this enzyme is mainly the degradation of intracellular and endocytosed proteins (1). Cathepsin D represents one of the major endopeptidase activities in the lysosomes. Together with other exopeptidases, such as cathepsins A, B, C, H, and L, these proteases can collectively digest most proteins to small peptides and amino acids (1).

Cathepsin D has been purified and studied from many sources (1–3). The primary structure of porcine cathepsin D has been determined by protein sequencing (4), and the structure of the human proenzyme has been deduced from its cDNA structure (5). This information revealed that, structurally, cathepsin D is very similar to other aspartic proteases of the pepsin family (see review in Ref. 3), except for two major structural features. First, cathepsin D contains two oligosaccharides (6) which, as with the carbohydrates in other lysosomal enzymes, are involved in the targeting of this enzyme to lysosomes (7). Second, cathepsin D is predominantly a two-chain protein, while most other aspartic proteases are single-chain enzymes (3).

Several studies on the biosynthesis of cathepsin D have revealed that it is synthesized as a proenzyme, procathepsin D, which is activated, most likely in the lysosomes, to a singlechain cathepsin D and subsequently to a two-chain enzyme (8–10). The proteolytic conversion of single-chain to multi-chain enzyme is apparently a phenomenon universal for all lysosomal enzymes. A large number of lysosomal enzymes have been studied in relation to biosynthesis and processing and, in all cases, proteolytic processing was observed (8, 9, 11–16). These observations collectively suggest that the proteolytic processing of lysosomal enzyme precursors may serve some important, but yet unknown, physiological function, and thus the structural basis of this processing may produce insights concerning these functions. Cathepsin D appears to be a good model for the study of the processing site structure because the amino acid sequence of the two-chain enzyme has already been determined for the pig enzyme (4). Also, the structure of human procathepsin D (5), when aligned by homology against the pig enzyme, suggests that a peptide of at least 7 residues is cut out during the conversion of a single chain to two chains. Finally, the close sequence similarity between cathepsin D and other aspartic proteases (3) suggests that the crystal structures of one of the aspartic proteases (for review, see Ref. 3) can be used to model the tertiary structure of the proteolytic processing region, which may be beneficial.
to the understanding of its functions. For these reasons, we initiated a study of the structure near the proteolytic processing site of cathepsin D in order to learn both the precursor structure and the sites of proteolytic cleavages. The "processing region" sequences of the pig and rat enzymes were determined. In addition, the structures near the processing site for both the single-chain and two-chain cathepsin D from cow were determined for comparison. These results are reported in this paper.

EXPERIMENTAL PROCEDURES

Materials

Bovine cathepsin D was purified from bovine spleen as described previously (17, 20). Cathepsin D from rat liver was purified according to the published procedure (17, 18). The porcine liver cDNA library in Agt11 bacteriophage was purchased from Clontech Laboratories. The human cathepsin D cDNA clone was a gift from Dr. John Chirgwin, University of Texas Health Science Center, San Antonio. Other reagents were the highest purity grade obtainable from commercial sources.

Methods

Protein Chemistry Methods—Bovine and rat cathepsin D preparations were separately reduced and carboxymethylated as described by Hirs (19). The bovine enzyme was then separated into the light-chain fraction and a mixture of heavy- and single-chain species by using Sephadex G-100 chromatography as previously described for the porcine enzyme (17). The conditions for proteolytic digestions, chromatographic and high performance liquid chromatography separation of peptides, amino acid analysis, and automated Edman degradation were the same as described previously (4, 6, 21).

Recombinant DNA Methods—A porcine liver cDNA library in Agt11 phage was screened for cathepsin D clones using plaque hybridization. About 2 × 10⁶ plaques were plated with Escherichia coli strain Y1088 as described by Huynh et al. (22) and screened with nick-translated human cathepsin D cDNA at 65 °C. The probe was a 1.1-kilobase EcoRI fragment containing essentially all the coding sequence of human procathepsin D. The positive clones were plaque-purified through second and third rounds of screening. The purified phage clones were grown by the plate lysate method (23) and their DNA prepared by the DEAE-cellulose slurry method of Benson and Taylor (24). The phage clones with longer inserts, in the range of 1.2-1.4 kilobases, were mapped by restriction digestions followed by polyacrylamide electrophoresis (23). Appropriate clones were selected for subcloning into bacteriophage M13 and the dideoxynucleotide termination method of sequence determination was carried out according to the method of Sanger et al. (25).

Computer Graphics Techniques—Modeling experiments were done on an Evans and Sutherland PS380 computer graphics system operating the program FRODO (26). All surface residues of porcine pepsinogen which have any atoms within 10 Å of the peptide bond between residues 90 and 91 were "mutated" to correspond to the side chains of human cathepsin D. Then the residues of the proteolytic processing region were "inserted." After a reasonable structure had been deduced by manually adjusting the model, three cycles of the
RESULTS

Bovine Cathepsin D Single-chain Processing Region Sequence and Proteolytic Cleavage Sites—Bovine spleen cathepsin D contains about equal amounts of single-chain and two-chain species (20). To obtain information on the bovine cathepsin D single-chain processing region sequence and the cleavage sites which produce the two-chain enzyme, three sets of sequences were obtained: the processing region sequence from the single chain, the COOH-terminal sequence of the light chain, and the NH2-terminal sequence of the heavy chain. To determine the first two, bovine cathepsin D, which contained both the single- and two-chain species, was subjected to proteolysis, peptide purification, and sequence determination. The results summarized in Fig. 1 show that the sequence of a 30-residue peptide, P3, is highly homologous with the COOH-terminal sequence of porcine cathepsin D light chain and thus must have been derived from the same region of the bovine light chain. Two peptides were derived from the connecting sequence of the single-chain enzyme. A chymotryptic peptide P1C9 contained 25 residues (Fig. 1), and a thermolysin peptide P1Th contained part of peptide P1C9 and was 16 residues long (Fig. 1). Since these two peptides contain the COOH-terminal sequence of the light chain (from P3) and extended for 12 more residues, they were judged to be derived from the processing region sequence of the single-chain enzyme. The NH2-terminal sequence of the heavy chain was determined from automatic Edman degradation of the sample containing the heavy and the single chains. It gives rise to two sets of sequences. However, since the NH2-terminal sequence of the single chain and the light chain are identical and are known (20), the second sequence can be identified by subtraction. As shown in Fig. 1, the NH2-terminal sequence of the heavy chain was found in the processing region sequence in peptide P3. This further substantiated that the origin of this peptide is the proteolytic processing region of the single-chain enzyme.

Porcine Cathepsin D Single-chain Connecting Sequence—Since the two-chain porcine cathepsin D sequence has been determined (4), the COOH-terminal sequence of the light chain as well as the NH2-terminal sequence of the heavy chain are known (4). It was decided that, in this case, the processing region sequence could be most readily determined from the nucleotide sequence of porcine cathepsin D cDNA. A porcine liver cDNA Agtll library (6.4 × 10^6 independent clones) was screened with a human cathepsin D cDNA probe and 23 positive clones from the library were isolated, mapped, and partially sequenced. While the details of these results are included in the Miniprint, the pertinent results are summarized in Fig. 1, which shows the sequence obtained including that of the processing region. The identity of this region is established by the presence of the known COOH-terminal sequence of the light chain and the NH2-terminal sequence of the heavy chain (4). There is a discrepancy of a single residue in the sequences determined with the two different methods. The COOH-terminal sequence of porcine cathepsin D light chain determined using protein chemistry was... Pro-Cys-Asn-Ser-Ala-Leu-Ser-Oly-Vai.


3 Portions of this paper (including part of "Results," Figs. S1–S10, and Tables S1 and SII) are presented in miniprint at the end of this paper. The abbreviations used are: HPLC, high performance liquid chromatography; Con A, concanavalin A; kb, kilobase; SDS, sodium dodecyl sulfate; PTH, phenylthiodantoin. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.

4 S. Yonezawa, unpublished results.
high degree of sequence similarity to the corresponding region of the porcine and bovine enzymes. The sequence of peptide PC1 is shown in Fig. 1. Peptide PC2 is 1 residue shorter at the COOH terminus. By sequence similarity of peptide PC1 to the sequences of the processing region obtained from other species, it is obvious that this peptide must have originated from the proteolytic processing region of single-chain rat cathepsin D.

DISCUSSION

Current results from bovine cathepsin D structural studies have revealed the sequence of the proteolytic processing region, generated from the peptide sequence of the single-chain enzyme, and the proteolytic processing positions, generated from the terminal sequences of the peptides derived from the light and heavy chains, respectively (Fig. 1). For porcine cathepsin D, the proteolytic processing region was deduced from the cathepsin D cDNA sequence (Fig. 1). Since the amino acid sequences of the two-chain cathepsin D was determined previously (4), the proteolytic cleavage positions in porcine cathepsin D are now identified. Two additional sequences of the corresponding proteolytic processing regions of rat and human cathepsin Ds are also known. The former was obtained from peptide sequences of rat cathepsin D (Fig. 1), and the latter was reported in the cDNA sequence of human cathepsin D (5). There is good evidence that both enzymes are proteolytically processed (8, 13, 27), and the sites of proteolysis are unknown.

The information described above permits a comparison of the proteolytic processing sequences of cathepsins from four species and the cleavage locations from two of these, as summarized in Fig. 2. This comparison, in turn, permits a closer examination of the possible structure and function relationships of the processing regions in cathepsin D molecules. The sequences in Fig. 2 are presented in a "hairpin" β-structure because in the homology alignment with the sequences of other aspartic proteases, the proteolytic processing region of the cathepsin Ds occurs as a sequence addition (3), and this addition is located at the end of the β hairpin loops in the crystal structures of aspartic proteases (28–31). As can be seen in Fig. 2, the corresponding sequences in porcine pepsin and penicillopepsin are highly conserved in cathepsins while the "proteolytic processing regions" in cathepsin Ds are absent in pepsin and penicillopepsin. The comparisons in Fig. 2 clearly show that the processing regions in the cathepsin Ds are variable both in size and in sequence. For rat cathepsin D, this region has 6 residues. In the pig and cow enzymes, this region has 9 residues and, in human cathepsin D, 11 residues are inserted. The only conserved sequence among the four processing regions is Pro-Cys at the amino-terminal end of the inserted sequence. There appears to be a high content of serine residues in this region of the various cathepsin Ds. However, upon close examination, the sequences involving these serines are quite different. The fact that these serine residues appear in a different part of the insertion region further reveals their structural differences. One of the possible functional hypotheses for the region is that it may be involved in recognition, for example for intracellular sorting or targeting (7, 32), and this possibility is piqued by both the location of the processing region on the molecular surface (see below for more discussion of the conformational aspects) and by its unique presence in cathepsin D, but not other aspartic proteases. However, the lack of uniformity in size and sequence of the processing regions from different cathepsin Ds suggests that it is unlikely to be involved in targeting recognition, which could be expected to be better conserved in the evolution of the enzymes.

Current results also permit an examination of the specificity in the proteolytic processing of cathepsin D. The proteolytic processing of the porcine cathepsin D single chain results in the removal of 6 residues to form the two-chain enzyme (Fig. 2). The minimal cleavage sites are those between Asn-
Ser and Val-Gly bonds in the sequence -Pro-Cys-Asn-Ser-Ala-Leu-Ser-Gly-Val-Gly-. The proteolytic processing of the bovine cathepsin D single chain results in the removal of 2 central serine residues in a region with the sequence of Ser-Ser-Ser-Ser-Ser-Ser (Fig. 1). The sites of the proteolytic cleavages must be Ser-Ser bonds (Fig. 2). Bovine cathepsin D differs from the porcine enzyme in that one-half of it is present in the spleen as the unprocessed single-chain form (20), while the porcine enzyme is almost completely processed. This could be interpreted as showing that the Ser-Ser bond is less favored in the specificity of the processing enzyme. It is interesting to note that rat liver lysosomal cathepsins B and H are both processed into two chains and in both cases, as in the case of porcine cathepsin D, this involves the cleavage of an asparaginyl-X bond. Rat single-chain cathepsin B is cleaved in a processing sequence of -His-Thr-Asn-Arg-Val-Asn-Val- (33) to form the two-chain enzyme without the Gly-Arg dipeptide (34). This processing involves at minimum the cleavage of the Asn-Gly and Arg-Val bonds. Rat cathepsin H is also a two-chain enzyme and the and of the chains in the single-chain cathepsin H can be placed by homology to papain (33). Thus, it is possible to deduce that the proteolytic processing must also involve an Asn-X bond (34). In addition, the processing of the β-chain of human β-hexosaminidase involves an Asn-Lys bond (35). Since there appears to be a high incidence of lysosomal enzyme processing involving cleavage of an Asn-X bond, it is tempting to suggest that asparaginyl bonds may be one of the initial points in the processing specificity for the lysosomal enzymes. In the case of porcine cathepsin D and rat cathepsin B processing, the initial asparaginyl bond cleavages could be followed by the aminopeptidase activity of cathepsins C (36) and H (37) to produce the mature two-chain enzymes. The specificity for asparaginyl bond cleavage does not appear to be universal for all cathepsin Ds, however. In the "proteolytic cleavage regions" of rat and human cathepsin D, no asparagines are present. In addition, some of the processed bonds in cathepsin B (38) and β-hexosaminidase (35, 39) do not involve an asparagine residue. In the case of the bovine enzyme, there is an Asn-Pro bond which is resistant to cleavage and the bovine single-chain cathepsin D may be present in large amounts due to the absence of a cleavage Asn-X bond representing the primary specificity of the processing enzyme. This discussion also places uncertainty on the physiological function of lysosomal cathepsin D processing, since in some species the single-chain enzymes are cleaved rather poorly. For this reason, the possible roles of the proteolytic processing regions in the tertiary structures of cathepsin D and procathepsin D were examined as described below.

Since the amino acid sequences of cathepsin D and procathepsin D are highly similar to other aspartic proteases and their zymogens (3), the possible conformation of the proteolytic processing regions was modeled in a computer graphic system, and the interactions of this region with the rest of the enzyme or zymogen molecule were examined using as a basis the high resolution X-ray crystal structures of penicillopepsin (30), for cathepsin D, and porcine pepsinogen, for procathepsin D. In the latter, the proteolytic processing region was found to be located near the activation peptide (pro-part) and considerable interaction between the two regions appeared likely. The likely conformation of the processing region involved in this interaction was explored by model building. Our preferred conformation for the processing region of the procathepsin D is a hairpin loop on the surface of the molecule parallel with a region of the activation peptide, as shown in Fig. 3 for human procathepsin D. There are several factors favoring the stability of this conformation. (a) Stabilizing interactions between the activation peptide and the processing region include two β-structure hydrogen bonds and other side-chain interactions (Fig. 3). (b) The half-cystine residue located at the beginning of the processing region is near another half-cystine residue unique to cathepsin D (residue 27 in porcine cathepsin D, Ref. 4) which also is present in the human and bovine enzymes (5, 20). In this conformation shown in Fig. 3, these 2 half-cystines are within disulfide linkage distance of each other. We suggest that they are disulfide-bonded. (c) There is no polar side chain of cathepsin D buried under the added structure of the processing region of Fig. 3. The only charged residue involved (Glu-) can be positioned in such a way that its carboxyl group is accessible to the solvent. These observations suggest that the processing region probably brings about a structural stabilization of the procathepsin D tertiary structure. This possibility prompts two additional questions to which there is no definitive answer at the present. First, why do procathepsins D require the additional processing region for the stabilization of theirzymogen structures which is absent in other aspartic protease zymogens. Second, what is the role, if any, for the processing region in the activation of procathepsin D in the acid environment of the lysosomes.

Acknowledgments—We wish to thank Dr. John Chirgwin for a clone of human cathepsin D cDNA, Mr. A. D. Dehedarani and Azar Dashi for technical assistance, and Paul Chambers for assistance in computer graphics.

REFERENCES

* J. A. Hartseuck and J. Remington, manuscript in preparation.
Proteolytic Processing Region of Cathepsin D


SUGGESTED BOOKS


2. Suggested books on "Proteolytic Regions of Bovine Cathepsin Dungee" in Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY


7. Suggested books on "Proteolytic Regions of Bovine Cathepsin Dungee" in Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY


10. Suggested books on "Proteolytic Regions of Bovine Cathepsin Dungee" in J. Biol. Chem. 261, 4288–4292
Proteolytic Processing Region of Cathepsin D

Peptide P7 (from C-terminal terminus of bovine cathepsin D light chain):

Peptide P10 (from "processing region" of bovine cathepsin D single chain):

Table 1:

<table>
<thead>
<tr>
<th>Amino Acid Composition of Peptides P1, P3 and P10</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>Asp</td>
</tr>
<tr>
<td>Thr</td>
</tr>
<tr>
<td>Ser</td>
</tr>
<tr>
<td>Glu</td>
</tr>
<tr>
<td>Gly</td>
</tr>
<tr>
<td>Ala</td>
</tr>
<tr>
<td>Asn</td>
</tr>
<tr>
<td>Arg</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

* Relative to His.

Table 2:

<table>
<thead>
<tr>
<th>Amino Acid Composition of Peptide G1 and G2-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1-2</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Asp</td>
</tr>
<tr>
<td>Thr</td>
</tr>
<tr>
<td>Ser</td>
</tr>
<tr>
<td>Glu</td>
</tr>
<tr>
<td>Gli</td>
</tr>
<tr>
<td>Arg</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

* Relative to Ile. Nearest integer in parentheses.

Figure 2:

Chromatographic separation of (a) light chain carboxy-terminal peptide P7, and (b) peptides P1 and P3 from bovine cathepsin D single chain on a column of Sephacryl S-300. For (a) bovine cathepsin D was prepared, carboxymethylated, the light chain isolated from a column of Sephacryl S-300. The light chain was digested with trypsin and the glycoprotein precipitated with a 2:1 titer of anti cathepsin D. Fractions of 0.5 ml were collected. The glycoprotein was precipitated with 0.5 ml 2.5 M (NH₄)₂SO₄, and the glycopeptide was digested using a proteolytic enzyme and the resulting glycopeptide was recovered with the co-A-Sepharose column before chromatography on the Sephacryl S-300 column. The elution conditions were identical to those in (a). Peptides P1 and P3 (solid circles) were eluted near fractions 44 and 45, respectively.

Figure 3:

Chromatographic separation of (a) light chain carboxy-terminal peptide P7, and (b) peptides P1 and P3 from bovine cathepsin D single chain on a column of Sephacryl S-300. For (a) bovine cathepsin D was prepared, carboxymethylated, the light chain isolated from a column of Sephacryl S-300. The light chain was digested with trypsin and the glycoprotein precipitated with a 2:1 titer of anti cathepsin D. Fractions of 0.5 ml were collected. The glycoprotein was precipitated with 0.5 ml 2.5 M (NH₄)₂SO₄, and the glycopeptide was digested using a proteolytic enzyme and the resulting glycopeptide was recovered with the co-A-Sepharose column before chromatography on the Sephacryl S-300 column. The elution conditions were identical to those in (a). Peptides P1 and P3 (solid circles) were eluted near fractions 44 and 45, respectively.

Figure 4:

HPLC chromatographic separation of (a) carboxy-terminal peptide P1, and (b) carboxy-terminal peptides P1 and P3 from bovine cathepsin D single chain. Peptide P1 from Sephacryl S-300 colon elution chromatography (Fig. 2) was digested with diisopropylfluorophosphatase (pH 8.0, 0.1 M NaCl, 0.2 M CaCl₂, 0.05 M DTT) and subjected to HPLC separation on a reverse-phase column (Fig. 4a) and subjected to HPLC separation on a reverse-phase column (Fig. 4b). The elution conditions and the column system were identical. Peptide P1 (see Fig. 4a) was eluted with diisopropylfluorophosphatase by essentially the same conditions described above. The peak of P1 was indicated by an arrow.
Proteolytic Processing Region of Cathepsin D

FORCINE CATHEPSIN D SEQUENCES:

Single chain (cDNA) sequence:


Fig. 4a. Chromatography of synthetic peptide fragments of rat cathepsin D on a column of Sephadex G-50. Each peptide was reduced, carboxypeptidase Y treated, and reacted with isopropylamine before isolation. The sample was dialyzed against 0.05 M potassium phosphate buffer, pH 7.4, and the peptide mixture subjected to reverse phase HPLC on a 3.4 mm x 14 cm Sephadex G-50 column eluted with 0.1 M acetic acid. Fractions under the horizontal bar were sequenced.

Fig. 4b. Chromatography of synthetic peptide fragments on a preparative column of Sephadex G-50. The recovered pool from Fig. 4a was dialyzed with isopropylamine and chromatographed again on a 3.4 mm x 14 cm Sephadex G-50 column as described in the legend of Fig. 4a. Fractions under the horizontal bar were pooled, lyophilized, and subjected to ESI mass spectrometry.

Fig. 4c. Yield of PTH-amino acids in the automated Edman degradation of peptide P1 (a) and P2 (b).