Amino Acid Sequence of Porcine Spleen Cathepsin D Light Chain

(Received for publication, December 14, 1982)

Takayuki Takahashi 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 Sciences Center, Oklahoma City, Oklahoma 73104

The complete amino acid sequence of the light chain of cathepsin D from porcine spleen has been determined. The light chain consists of a single polypeptide chain with 97 amino acid residues. The sequence is:

\[
\begin{align*}
&10 \\
30 \\
50 \\
70 \\
\text{Y-G-S-G-L-S-G-Y-L-S-Q-D-T-V-S-P-S-N-COOH}
\end{align*}
\]

The molecular weight of the light chain was calculated from this sequence to be 10,548 (without carbohydrates). A single disulfide bond links two half-cystine residues between positions 46 and 53. A cysteine residue is located at position 27. The light chain sequence is extensively homologous to the NH2-terminal sequence of other aspartyl proteases. It shows a 50% identity with the sequence of mouse submaxillary gland renin and a 49% identity with that of porcine pepsin. A single glycosylation site is located at residue 70 of the cathepsin D light chain. This site corresponds to position 67 of pepsin by homology. The active site aspartyl residue, corresponding to Asp-32 of pepsin, is located at residue 33 in the cathepsin D light chain.

EXPERIMENTAL PROCEDURES

RESULTS

Amino Acid Sequence of Cathepsin D Light Chain—The complete amino acid sequence of the light chain is presented in Fig. 1. This sequence is constructed from the sequence of 6 tryptic peptides from carboxymethylated light chain, 12 chymotryptic peptides and 2 papain peptides from unmodified light chain. The purification scheme of these peptides is summarized in Fig. 2. The details of purification and sequence determination of these peptides are described in the Miniprint section.

The construction of the light chain sequence is as follows.

* This study was supported by Research Grant GM 20212 from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
FIG. 1. Complete amino acid sequence of cathepsin D light chain isolated from porcine spleen. ---, extent of the various fragments used to construct the sequence; —, residues identified by amino acid compositions in small peptides; —, results of carboxypeptidase A digestion; CH$_2$O, carbohydrate.

The automatic Edman degradation of the intact light chain produced the NH$_2$-terminal 38 residues (6). With this stretch of sequence and the sequence from 8 trypptic, chymotryptic, and papain peptides (T2-A, C5-A, C6-A, T1-B, C7-A, C4, C7-B2 and P2), the sequence of the NH$_2$-terminal 45 residues was fully identified and overlapped (Fig. 1). The sequence from residues 46 to 97 was established by overlapping the sequences of trypptic and chymotryptic peptides (C5-B, T2-B, C3, T4, C2, T3, C1, T1-A, C6-C, C6-B, and C7-B1). The COOH-terminal sequence was confirmed by both the sequence of peptide P1 and the results of carboxypeptidase A digestion of the light chain (Fig. 1 and Miniprint). All residues were identified by at least two independent observations in Edman degradations except for the three half-cystine residues.

The amino acid composition of the light chain as calculated from the sequence agreed well with the analyses from the native or reduced carboxymethylated light chain (Table I).

The Disulfide Bridge in Cathepsin D Light Chain—Cathepsin D light chain contains 3 half-cystine residues (Table I). Thus the light chain may contain a disulfide bond. This was confirmed when the thermolysin digest of peptide C5-B was shown to contain two disulfide-linked peptides (see Miniprint section). The sequence and the placement of peptide C5-B established that the single disulfide bridge is present between half-cystine residues at positions 46 and 53 (Fig. 1), and the third cysteine (Cys 27 in Fig. 1) exists probably as a free sulfhydryl group in native cathepsin D (see Miniprint section).

DISCUSSION

Evidence presented above established the amino acid sequence of cathepsin D light chain from porcine spleen. We feel that the evidence supporting this sequence is sufficient for the following reasons. (a) With the exception of two
TABLE I

<table>
<thead>
<tr>
<th>Amino acid composition of light chain of porcine spleen cathepsin D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>Aspartic acid</td>
</tr>
<tr>
<td>Threonine</td>
</tr>
<tr>
<td>Serine</td>
</tr>
<tr>
<td>Glutamic acid</td>
</tr>
<tr>
<td>Proline</td>
</tr>
<tr>
<td>Glycine</td>
</tr>
<tr>
<td>Alanine</td>
</tr>
<tr>
<td>Half-cystine</td>
</tr>
<tr>
<td>Carboxymethyl-cysteine</td>
</tr>
<tr>
<td>Valine</td>
</tr>
<tr>
<td>Methionine</td>
</tr>
<tr>
<td>Isoleucine</td>
</tr>
<tr>
<td>Leucine</td>
</tr>
<tr>
<td>Tyrosine</td>
</tr>
<tr>
<td>Phenylalanine</td>
</tr>
<tr>
<td>Lysine</td>
</tr>
<tr>
<td>Histidine</td>
</tr>
<tr>
<td>Arginine</td>
</tr>
<tr>
<td>Tryptophan</td>
</tr>
<tr>
<td>Total residues</td>
</tr>
</tbody>
</table>

*a Acid hydrolysates were performed on unmodified light chain for 24, 48, and 72 h. The values of threonine and serine were obtained by extrapolations to zero time of hydrolysis. Values of valine and isoleucine were from 72-h hydrolysates. Value of arginine in the light chain was less than 0.05 residue per molecule and was taken as 0.

*b Obtained by the method of Hirs (18).

'Obtained by iodoacetate treatment of light chain without reduction as described under "Experimental Procedures."

4 aspartic acid and 5 glutamine residues.

2 glutamic acid and 3 glutamic residues.

* Determined as cysteic acid after performic acid oxidation.

4.26 3.92 4.80 5.82

Lysine | 4.25 | 4.65 | 4.60 | 5 |

Histidine | 4.26 | 3.92 | 3.99 | 4 |

Arginine | 0 | 0 | 0 | 0 |

Tryptophan | 1.88h | NDh | ND | 2 |

Total residues | 97 | | | |

A comparison of the amino acid sequence of the light chain with those of mouse submaxillary gland renin (10, 11) and porcine pepsin (12) reveals a striking sequence homology (Fig. 3). Cathepsin D light chain contains 59% identical residues with the NH2-terminal sequence of mouse submaxillary gland renin and has 49% identity with the NH2-terminal sequence of porcine pepsin. Homology with other aspartyl proteases not shown in Fig. 3 is also extensive. For example, the light chain sequence has 51% identity with the NH2-terminal sequence of bovine chymotrypsinogen (13) and 30% identity with the NH2-terminal sequence of penicillopepsin (14). The alignment between cathepsin D light chain and renin does not require the creation of any gap (Fig. 3). The alignment between light chain and pepsin requires the introduction of only a two-residue gap at positions 47A and 47B (Fig. 3). A high degree of homology likely exists between the cathepsin D heavy chain and the corresponding regions in other aspartyl proteases. The sequence near the oligosaccharide site in cathepsin D heavy chain is homologous to part of the peptide positions (residues 27 and 46), all residues were identified by at least two independent observations from different peptides. The complete sequence is fully overlapped. The only one-residue overlap occurs at tyrosine residue 59. We believe this to be sufficient evidence for the overlap at this position because the total tyrosine contents were in complete agreement with those calculated from the sequence (Table I). Also, all the tryptic and chymotryptic peptides of the light chain were accounted for in our peptide isolation. (c) The amino acid compositions agreed reasonably well between that derived from the sequence and from the amino acid analyses (Table I). The near one-residue discrepancies of the serine and threonine contents were probably produced from the errors in extrapolation of these amino acid analyses to zero time hydrolysis.

Download from http://www.jbc.org/ by guest on October 24, 2017
sequence (9). Other sequences from the heavy chain produced in our laboratory² so far include regions corresponding to pepsin residue numbers 195–235, 248–278, and 311–327. These sequences are homologous to other aspartyl proteases to a similar degree as shown in Fig. 3 for the light chain alignment. These facts suggest that the three-dimensional structure of cathepsin D light chain must be closely homologous with those of other aspartyl proteases.

There are three half-cystine residues in the light chain. The disulfide bridge between residues 46 and 53 of the light chain (Fig. 1) corresponds to the known disulfide bridge between residues 45 and 50 in pepsin (12) (Fig. 3). An insertion of residues 47A and 47B in the aligned light chain and renin sequences (Fig. 3) occurs between these two disulfide-linked half-cystines. This part of the structure is known to be on the surface of the molecules in all the known crystal structures of aspartyl proteases (14–16). Therefore, this two-residue insertion in this region will probably not seriously affect the homology between the tertiary structures of these enzymes.

Evidence suggests that residue 27 is likely to be a cysteine in the native enzyme (see Miniprint section). This free sulfhydryl group may be nonessential for enzyme activity since it is fully masked in the fully active purified enzyme. In the available crystal structures of aspartyl proteases (14–16), the corresponding residue to cysteine 27 is located on the surface of the molecule away from the substrate binding cleft. These comparisons seem to support the above suggestion.

We have previously reported the sequence near the oligosaccharide attachment sites in cathepsin D light and heavy chains (9). The reported glycopeptide sequence in the light chain was confirmed in the current work. In addition, the alignment of the sequence of the entire light chain against the pepsin sequence (Fig. 3) confirmed that the glycosylation site in cathepsin D light chain corresponds with residue 67 in pepsin (Fig. 3). Our previous report established that there are 5 different high mannose-type oligosaccharides linked to this position in cathepsin D (9).

The question of whether the COOH terminus of cathepsin D light chain is the site of cleavage in the processing of the single chain enzyme is not certain. Based on the alignment in Fig. 3, the COOH terminus of the light chain corresponds to residue 94 in pepsin. This part of the structure is located on the surface of all known crystal structures of aspartyl proteases (14–16) and would be easily accessible to a processing protease. However, the possibility cannot be excluded that an original cleavage site occurs further toward the COOH terminus of the single chain which may be followed by the hydrolysis by a carboxypeptidase to the current COOH terminus of the light chain.

We have suggested that the differences in isoelectric points of cathepsin D isoforms may be due to their microheterogeneity in amino acid sequences. It is interesting to note that no microheterogeneity in amino acid sequence was observed in the light chain. This is in contrast with the apparent microheterogeneity observed near the NH²-terminal region of the heavy chain (6).

Acknowledgments—We wish to thank Dr. Robert Delaney for helpful suggestions in the preparation of this manuscript and Abdul Dehedarani and Azar Fesmire for their able technical assistance.

REFERENCES


Additional references are found on p. 6439.

Porcine Spleen Cathepsin D Light Chain Sequence

FRACTION NUMBER

MINUTES

ABSORBANCE at 280 nm

MINUTES

ABSORBANCE at 215 nm

% ACETONITRILE
Porcine Spleen Cathepsin D Light Chain Sequence

Fig. 63. HPLC of T1 fraction. HPLC was performed as described in the legend of Fig. 6. Dotted lines indicate the collected fractions.

Fig. 64. Chromatography of chromatographic phase of cathepsin D light chain. 0.1 mg of cathepsin D was treated as described in Fig. 6. Samples were collected from the column at 10 min intervals. The fractions were analyzed by HPLC using a system of gradient elution. The collected fractions are indicated by horizontal lines.

Fig. 65. HPLC of T1 fraction. Dotted lines indicate the collected fractions.

Fig. 66. HPLC of T1 fraction. Dotted lines indicate the collected fractions.

Fig. 67. HPLC of T1 fraction. Dotted lines indicate the collected fractions.

Fig. 68. HPLC of T1 fraction. Dotted lines indicate the collected fractions.

Fig. 69. HPLC of T1 fraction. Dotted lines indicate the collected fractions.
Amino acid sequence of porcine spleen cathepsin D light chain.
T Takahashi and J Tang