The amino acid sequences near the glycosylation sites and the oligosaccharide structures have been determined for the lysosomal protease cathepsin D from porcine spleen. Cathepsin D light and heavy chains were separately digested with proteases and the glycopeptides were purified. A single sequence was constructed from the amino acid sequence of the light chain glycopeptides which is: Tyr-Asn-Ser-Gly-Lys-Ser-Ser-Thr-Thr-Val-Lys-Asn(CH₂O)-Gly-Thr-Thr-Phe. A single glycopeptide sequence was also obtained for the heavy chain: Lys-Gly-Ser-Leu-Asp-Tyr-His-Asn(CH₂O)-Val-Thr-Arg-Lys-Ala-Tyr. The light chain sequence is homologous with the sequence of porcine pepsin from residues 56 to 71. The heavy chain sequence is homologous with the pepsin sequence from residues 176 to 189. Thus, the 2 oligosaccharide-linked asparagines in cathepsin D correspond to residues 67 and 183 in pepsin and other homologous aspartyl proteases. These positions are located on the surface of the crystal structures of aspartyl proteases. Five oligosaccharides linked to Asn-67 were separated and their structures determined with proton NMR. Four major oligosaccharides are structural variants from the high mannose-type each with 5 mannose residues. One of the mannose-type having 3, 5, 10 mannoses, respectively. A minor structure contained a third fucosylated GlcNAc. Three oligosaccharide structures were found linked to Asn-183. Two major oligosaccharides are of the high mannose-type each with 5 mannose residues. One of the two contains a fucosylated linked to a GlcNAc. A third, very minor oligosaccharide contains galactose.

Cathepsin D (EC 3.4.23.5) is a lysosomal aspartyl (acid) protease which has physiological functions in intracellular protein degradation (1, 2). The enzyme has been purified from various tissues and its properties are well documented (see review in Ref. 3). Structure-function studies on cathepsin D may be interesting in two different aspects. First, it is an intracellular aspartyl protease, while all the aspartyl proteases studied for structure and function relationships are extracellular enzymes, such as pepsin and penicillopepsin (see reviews in Refs. 4 and 5). Second, since cathepsin D is a lysosomal enzyme, it can serve as a model for the structure-function of other lysosomal hydrolases. For these reasons we have devised a large scale purification procedure for cathepsin D from porcine spleen and have initiated studies on the polypeptide chain arrangement (6). Porcine spleen cathepsin D (M_r = 50,000) is comprised of a light chain (M_r = 15,000) and a heavy chain (M_r = 35,000). The N-terminal amino acid sequence of the light chain is homologous to that of pepsin and other aspartyl proteases. Both chains contain carbohydrate (6).

Lysosomal enzymes, including cathepsin D, contain a recognition marker for receptor-mediated cellular endocytosis and for their packaging into the lysosome after biosynthesis (7). It is now well established that the recognition marker is mannose 6-phosphate present in the asparagine-linked high mannose-type oligosaccharide units (8-14). The enzymic system responsible for the phosphorylation of mannoses on the lysosomal enzymes has been described (15-20). N-Acetylglucosamine 1-phosphate is first transferred to an acceptor mannose. A phosphodiesterase then removes the N-acetylglucosamine. Particularly interesting was the finding of Reitman and Kornfeld (20) that the lysosomal enzymes are preferentially phosphorylated by N-acetylglucosaminylphosphotransferase, the enzyme that catalyzes the first step in this pathway. These results suggest that lysosomal enzymes may contain structural markers which mediate the phosphorylation. For these reasons, questions of oligosaccharide structures and their locations in the primary and tertiary structure of cathepsin D are interesting to pursue. The answers may provide insights into the molecular basis of physiological functions of lysosomal enzymes. In this report, we describe the amino acid sequences around the oligosaccharide units of cathepsin D, the oligosaccharide structures, and their localization in the linear and three-dimensional structure of this enzyme.

EXPERIMENTAL PROCEDURES

Materials

Porcine spleens were obtained from Wilson Packing Company (Oklahoma City, OK) and cathepsin D was purified from these tissues, as previously described (21). The light and heavy chains were obtained from the enzyme as previously described (21). ConA-Sepharose 4B (Lot 5761) and Sephadex gels were obtained from Pharmacia. Methyl-a-D-glucopyranoside (Lot 500076) and thermolysin (Lot 100540) were obtained from Calbiochem. Methyl-a-D-glucopyranoside (Lot 500076) and thermolysin (Lot 100540) were obtained from Calbiochem. 1,1'-Tosylamido-2-phenyl-ethyl chloromethyl ketone-trypsin (Lot 343970) and a-chymotrypsin were obtained from Worthington. D-Mannose, myo-inositol, D-xylene, a-D-fucose, D-galactose, a-D-glucose, N-acetylglucosamine, N-acetylgalactosamine, N-glucosamine HCl and NaBH₄, were obtained from Sigma. Trail and pyridine were obtained from Pierce. Acetic anhydride was obtained from Supelco, a-Mannosidase (jack bean) was a gift of Dr. Yu-Teh La. The HPLC solvents were obtained from Waters Associates. Sequencer reagents were purchased from Beckman and amino acid analyzer reagents were purchased from Dionex. All other reagents used were the highest purity obtained commercially and used without further purification.

The abbreviations used are: ConA, concanavalin A; HPLC, high performance liquid chromatography; PTH, phenylthiohydantoin.

* This study was supported by Research Grants GM-20210 and GM-25703 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.

† Recipient of Research Career Development Award AM00525 from the National Institutes of Health.

(Received for publication, June 18, 1982)
Oligosaccharides of Lysozyme Cathepsin D

Methods

Proteolytic Digestions—Heavy chain, light chain, or their glycopeptides were digested separately with trypsin, a-chymotrypsin, and thermolysin in a substrate-enzyme ratio of 50:1 (w/w). The protein concentration was 1 mg/ml in one of the two buffers: 0.1 M NaH2PO4, pH 7.9; or 0.05 M Tris-HCl, pH 7.9. The incubations were carried out at 37 °C for the time durations specified in the text.

ConA-Sepharose Column Chromatography—ConA-Sepharose 4B was equilibrated and washed with a buffer (Ris-Rad) of 0.1 M NaH2PO4, pH 7.0, or 0.05 M Tris-HCl, pH 7.5, containing 1% NaCl and 1 mM each of NaCl, MgCl2, and MnCl2. In all experiments, a ratio of 10 mg of sample to 1 ml of ConA-Sepharose was used. After loading the peptide sample, the column was washed through with the same buffer at a flow rate of 3.5 ml/h until the absorbance at 280 nm fell to baseline. The glycopeptides were eluted with 0.2 M methyl-a-0-glucopyranoside in the same buffer with the same flow rate. A sharp peak eluted, which was monitored by absorbance at 215 nm; it was collected (3 ml/tube), desalted in a column of Sephadex G-15, eluted with 0.1 M NaH2PO4, pH 7.5, and lyophilized.

HPLC Separation of Glycopeptides—Glycopeptides were dissolved in 0.01 M potassium phosphate, pH 7.5. HPLC separation was performed on a reverse-phase column of Supelcosil with LC-18 packing (0.46 x 25 cm, from Supelco) using a Waters Associates HPLC. The elution was done under constant flow of 1.2 ml/min with a linear gradient from 0.01 M potassium phosphate, pH 7.5, with increasing content of acetonitrile in the same buffer. The peptide content in the effluent was monitored at 215 nm. In preparative HPLC, up to 750 μl of sample was injected and the material under the peptide peaks collected and desalted for further studies. All collections were done manually. A given peptide peak, the elution was usually started from the beginning of the fast rising recording line of a peak and was stopped at the end of the fast declining line of the peak. For peaks which were closely spaced (such as in Fig. 3), the collections were essentially between the lowest points of the valleys. Minor contaminating material, which was observed as shoulders in an analytical run (e.g. LC4 in Fig. 3 and HTh2 in Fig. 4), was excluded from the collections.

Amino Acid Analysis—The amino acid analyses were performed according to the method of Spackman et al. (29) with a Durrum Model D300 amino acid analyser using a standard 90-min program provided by the manufacturer. Protein and peptide samples were hydrolyzed in 5.7 M HCl for 20 h at 110 °C in evacuated, sealed tubes. For the analysis of glucosamine, the elution time for the first buffer (pH 3.23) was shortened by about 15 min (to 34 min). The glucosamine eluted before lysine and was well separated from galactosamine. Authentic samples of glucosamine were used in the standards for quantitation. Appropriate blanks and standards were also run for quantitation.

NMR Sample Preparation—Desalted glycopeptides from HPLC fractionation were lyophilized and dissolved in 500 μl of 90% atom % 2H2O (Aldrich). Samples with excessive residual water signals were glycolyzed. Some samples were treated prior to lyophilization with CNBr to remove possible contamination with other glycopeptides. However, line widths were not found to be affected by this treatment. Concentrations of glycopeptides for NMR ranged from 0.054 to 5.2 mM. A trace of acetone was added to provide an internal reference, and its position taken to be 2.225 ppm from 2,2-dimethyl-2-silapentane-5-sulfonate.

NMR Spectra—Pulsed Fourier transform 1H NMR spectra were recorded at 300 MHz with a Cryomagnet Systems Inc. 70/50 magnet and 5-mm probe. A Nicolet 1180 computer was used with custom-built quadrature phase detection rf electronics. D2O served as the lock signal. Several hundred transients were typically time-averaged and processed with exponential weighting corresponding to 0.5- to 1-Hz line broadening. The residual water resonance was suppressed by presaturation during the delay between pulses. Spectra were acquired at temperatures of 23 and 38 °C. The chemical shift of the residual water peak is temperature-dependent, revealing obscured nearby peaks upon a temperature change.

RESULTS

Amino Acid Sequence of Glycopeptides from the Light Chain—The light chain of cathepsin D contains an average of 2 glucosamine and 5 mannose residues (Table I). Since most lysosomal enzymes contain oligosaccharides of the asparagine-linked high mannose-type (35), these data suggest a single glycosylation site in the light chain. Light chain was digested with a-chymotrypsin and trypsin. The glycopeptides and some non-glycopeptides were purified as summarized in Fig. 1. The details of peptide purification and sequence determination are described in the miniprint. From the peptide structures, the amino acid sequence near the glycosylation site was constructed as shown in Fig. 2.

Amino Acid Sequence of Glycopeptides from the Heavy Chain—The glycopeptides from the heavy chain were purified from trypsin, chymotryptic, and thermolytic digestion (Fig. 1) and their sequences determined (see Miniprint). These peptide structures give rise to a single amino acid sequence around the glycosylation site in the heavy chain (Fig. 2). Because the glucosamine analysis of the heavy chain produced about 2.6

2 Portions of the "Results" (Figs. S1 to S10 and Tables S1 and SII) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 950 Rockville Pike, Bethesda, MD 20814. Request Document No. 82 M-1634, cite the authors, and include a check or money order for $4.80 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.
### TABLE I

| Sugar contents of light chain, heavy chain, and the glycopeptides from porcine spleen cathepsin D |
|---------------------------------------------------------------|---------------------------------|
| Sugar contents                                              | Light chain glycopeptides       |
|                                                             | Heavy chain glycopeptides       |
|                  | LC1   | LC2   | LC3   | LC4   | LC5   | HTh1  | HTh2  | HTh3  |
| Glucosamine      | 2.14a | 2.57a | 1.69b | 1.59b | 2.63b | 1.75b | 1.79b | 1.79b |
| Mannose          | 4.81  | 6.66  | 6.20  | 5.04  | 3.41  | 3.14  | 4.84  | 4.80  | 5.23  |
| Fucose           | 0.00  | 0.00  | 0.00  | 0.00  | 0.00  | 0.00  | 0.00  | 0.40  | 0.00  |
| Galactose        | 0.00  | 0.00  | 0.00  | 0.00  | 0.00  | 0.00  | 0.00  | 1.25  |       |
| Per cent of total| 20.3  | 29.2  | 18.9  | 5.3   | 26.6  | 55.8  | 34.9  | 3.3   |       |

* Determined with the Elson-Morgan method (30).
* Glucosamine was determined with an amino acid analyzer. The glucosamine content of HTh3 was the result of a single analysis.
* Analyzed by alditol acetate method (33) in gas-liquid chromatography.
* Determined with cysteine-H2SO4 method (30).

**FIG. 1.** Flow chart of glycopeptide purification from the light and heavy chains of cathepsin D.

**LIGHT CHAIN:**

- Chymotryptic
- Tryptic

**HEAVY CHAIN:**

- Chymotryptic
- Tryptic
- Thermolysin

**FIG. 2.** Amino acid sequence of glycopeptides isolated from cathepsin D. The double-headed arrows under the sequences represent the isolated peptides with their peptide numbers. The right-handed arrows represent PTH identifications from the Edman degradation.

residues/mol, the question is raised as to whether the heavy chain contains more than one glycosylation site. However, the following reasons support the single site observations. First, reduced and carboxymethylated heavy chains produced 2 peptides derived from the same sequence as shown in Fig. 2 (see Miniprint). Second, a thermolysin digest of heavy chain was fractionated in HPLC with a full gradient designed to elute all the peptides. Only a single glucosamine-containing peak was found in the elution pattern (see Miniprint). In addition, we found all three proteolytic digests of heavy chain...
produced glycopeptides derived from the same sequence. We concluded that only a single glycosylation site is present in heavy chain.

**Oligosaccharide Structures**—The α-chymotrypsin digest of the light chain produced 5 glycopeptides, LC1 to LC5, which separate in HPLC (Figs. 1 and 3). As discussed in the Mini-print, these peptides have the same amino acid composition and the sequence Val-Lys-Asn(CH$_2$O)-Gly-Thr-Thr-Phe. They each apparently contain 2 glucosamines (Table I). The mannose numbers are sufficiently different so that these peptides are assumed to differ in either sugar number or structure, which was the apparent basis for HPLC separation. In a similar manner, 3 glycopeptides, HTH1 to HTH3, from a thermolysin digest of heavy chain were separated by HPLC (Figs. 1 and 4). Again, these 3 peptides have the same amino acid sequence Leu-Asp-Tyr-His-Asn(CH$_2$O) (see Mini-print), but contain different sugars. As shown in Table I, the glucosamine and mannose contents are similar, but peptide HTH2 contains fucose and peptide HTH3 contains galactose.

The major oligosaccharide structures were studied by proton NMR. Light chain fractions LC1, LC2, LC3, and LC5 gave spectra characteristic of high mannose glycopeptides (36–39). Fig. 5 shows spectra of these species in the region of the anomeric (H-1) and mannose H-2 resonances. Because the samples were specifically not digested with pronase, these glycopeptide spectra also contain resonance from the 7 amino acid residues. Fortunately, the extra peaks do not obscure any key sugar resonances. Beginning with the smallest chain (LC5), progressive assignments of sugar peaks can be made based on published data (36–39). A characteristic H-2 doublet ($\gamma_1=10$ Hz) is found at 5.03 ppm for the asparagine-linked GlcNAc while a smaller coupling constant and broader peak identify the H-1 resonance of the next GlcNAc at 4.62 ppm. Mannose H-1 peaks are easily picked out by their very small (unresolved in these spectra) coupling constants. The core mannose anomeric proton is at 4.775 ppm, on the upfield shoulder of the water peak in these spectra, but is clearly resolved in spectra taken at 38 °C (not shown). The mannose residue linked α(1-3) to the core sugar has H-1 at 5.098 ppm while the anomeric proton of the α(1-6) linked group is at 4.915 ppm. These values are within 0.01 ppm of the positions expected based on spectra of known glycopeptides (Table II). Between 3.9 and 4.3 ppm are found resonances of mannose H-2. They also serve as guides for oligosaccharide sequencing in...
Oligosaccharides of Lysosomal Cathepsin D

Values given are in parts/million from 2,2-dimethyl-2-silapentane-5-sulfonate measured relative to acetone at 2.225 ppm. Numbers in parentheses are chemical shifts for authentic compounds previously reported, or were estimated based on reported values for similar glycopeptides (38).

<table>
<thead>
<tr>
<th>Glycopeptide</th>
<th>GlcNAc(a)</th>
<th>GlcNAc(b)</th>
<th>Man(a)</th>
<th>Man(b)</th>
<th>Man(c)</th>
<th>Man(d)</th>
<th>Man(e)</th>
<th>Fucose</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC5</td>
<td>5.028</td>
<td>4.62</td>
<td>4.775</td>
<td>5.098</td>
<td>4.915</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.61)</td>
<td>(4.78)</td>
<td>(5.09)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LC3</td>
<td>5.03</td>
<td>4.60</td>
<td>4.78</td>
<td>5.095</td>
<td>4.875</td>
<td>5.095</td>
<td>4.91</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.61)</td>
<td>(4.78)</td>
<td>(5.095)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LC2</td>
<td>5.022</td>
<td>4.60</td>
<td>4.775</td>
<td>5.09</td>
<td>4.868</td>
<td>5.09</td>
<td>5.149</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.61)</td>
<td>(4.78)</td>
<td>(5.09)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LC1</td>
<td>5.02</td>
<td>4.60</td>
<td>4.77</td>
<td>5.35</td>
<td>5.055</td>
<td>4.87</td>
<td>5.09</td>
<td>5.15</td>
</tr>
<tr>
<td></td>
<td>(4.61)</td>
<td>(4.78)</td>
<td>(5.352)</td>
<td>(5.049)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HTh1</td>
<td>5.037</td>
<td>(\sim)4.6</td>
<td>4.78</td>
<td>5.19</td>
<td>4.876</td>
<td>5.10</td>
<td>4.914</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.61)</td>
<td>(4.78)</td>
<td>(5.09)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HTh2</td>
<td>5.065</td>
<td>(\sim)4.6</td>
<td>4.778</td>
<td>5.105</td>
<td>4.877</td>
<td>5.105</td>
<td>4.914</td>
<td>4.815</td>
</tr>
<tr>
<td></td>
<td>(4.61)</td>
<td>(4.78)</td>
<td>(5.09)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(a\) Numbering scheme shown in Fig. 6.

\(b\) Reported values for this group are based on pronase-digested peptides and generally range from 5.05 to 5.09 ppm (36). Some variations are expected for cathepsin D glycopeptides since Asn is flanked by other residues.

1\(^H\) NMR (36, 38). Chemical shifts for these peaks in LC5 and the other glycopeptides are all within 0.02 ppm of expected values (38). Glycopeptides LC3, LC2, and LC1 are progressively higher mannose chains of 5, 6, and 7 residues based on integration of H-1 peaks in their NMR spectra and chemical analysis (Table I). Each fraction has exactly 2 N-acetyl methyl groups with peaks at 2.01, 2.06, and a shoulder at 2.075 ppm. No other sugar methyls appear in the spectrum nor is there any trace of peaks for N-acetylneuraminic acid or fucose in the 0-3 ppm region. Apparently half the LC4 sample is a mannose. A doublet in the upfield region at 4.54 ppm may arise from H-1 of GlcNAc \(\beta(1-4)\) linked mannose H-1 and H-2 peaks match very closely expected chemical shifts for the structures given in Fig. 6 (Table II).

While it is clearly a mixture, the N-acetyl methyl region contains a proportion of HThl as a cross-contaminant), but it is an interesting mixture with a broad range of chemical shifts and assignments for anomeric protons of cathepsin D glycopeptides

The minor peak (LC4) from HPLC of light chain glycopeptides HThl and HTh2 produced the NMR spectra seen in Fig. 5. HThl closely matches the spectrum of LC3, except for different peptide signals. On this basis, it is assigned the Man\(\alpha\)-GlcNAc-Asn structure shown in Fig. 6. HTh2 is similar to HThl in many respects (and in fact, contains a proportion of HThl as a cross-contaminant), but key differences are seen. In HTh2 the H-1 doublet for the first GlcNAc is shifted downfield from 5.037 to 5.065 ppm and a new peak \(\langle J = 5 \text{ Hz} \rangle\) appears at 4.815 ppm. In the upfield region, a methyl doublet is found at 1.13 ppm with an area of 1.8 protons (0.6 methyl groups). The extra peaks indicate the presence of fucose, and in fact HTh2 has 0.4 mol of fucose/mol of glycopeptide by chemical analysis (Table I), close to the ratio found by NMR (0.6).

DISCUSSION

Porcine spleen cathepsin D contains two glycosylation sites, one on each chain. Since they are the first sequences determined around the glycosylation sites of a lysosomal enzyme, we have examined these sequences for possible functionally related residues. The threonyl residue two away from the glycosylated Asn, residue (Fig. 2) is the normal glycosylation signal for N-linked oligosaccharides (41). The lysyl and seryl residues 7 and 5 residues to the left of the glycosylated Asn (Fig. 2) are in common in the two sequences. Whether these similarities have any functional significance is not clear at the present.

Since glucosamine analysis of heavy chain gave a slightly higher value (Table I), we considered the possibility of the
presence of a second glycosylation site in the heavy chain. However, in three proteolytic digests from the heavy chain, we obtained glycopeptides from the same original sequence.

The reduced and carboxymethylated heavy chain was digested with trypsin. Again, the same glycopeptides were found (see Miniprint section). Finally, the heavy chain was digested with thermolysin and the resulting peptides were separated in reverse-phase HPLC with a gradient designed to elute all peptides. Only a single glycopeptide was found (see Miniprint). From these results we concluded that the heavy chain contains only a single glycosylation site.

Since we have previously shown that cathepsin D is structurally homologous with pepsin and other aspartyl proteases (6), the sequence near the oligosaccharide sites of cathepsin D were compared to the porcine pepsin sequence (42). As shown in Fig. 7, the light chain glycopeptide sequence from Fig. 2 aligns well by homology with the segment of residues 56-72 of pepsin. The heavy chain sequence from Fig. 2 aligns with residues 171-189 of pepsin (Fig. 7). These are the only significant homologous sequences found between the cathepsin D glycopeptides and pepsin; and the relatedness seems reasonably strong. Thus the two glycosylation positions correspond to residues 67 and residue 183 in pepsin. (These two sites in cathepsin D will be referred to as Asn-67 and Asn-183, respectively. This is consistent with the convention in which the active-site residues of aspartyl proteases are referred to as Asp-32 and Asp-215, according to the pepsin residue numbers). These results provide additional structural evidence that the primary structure of cathepsin D is closely homologous with pepsin and other aspartyl proteases. It is interesting to note that porcine pepsin is phosphorylated at a serine at residue 68. The significance for the closeness in glycosylation.

...
and phosphorylation positions in these two enzymes is not clear at present.

Existing structural evidence indicates that cathepsin D and pepsin are homologous in primary structure. We have previously shown that the two-chain cathepsin D is derived from a single chain enzyme (6). The light chain corresponds to the \( \text{NH}_2 \)-terminal region while the heavy chain corresponds to the remaining \( \text{COOH} \)-terminal portion of the single chain enzyme (6). The \( \text{NH}_2 \)-terminal sequence, 39 residues, of the cathepsin D light chain is highly homologous with the \( \text{NH}_2 \)-terminal region of pepsin (6). We have now a tentative sequence of the light chain which has a homology with pepsin extended through to the \( \text{COOH} \) terminus of an approximately 95-residue structure.\(^3\) This places the \( \text{NH}_2 \) terminus of the heavy chain around residue 92 of pepsin. In spite of the microheterogeneity in the \( \text{NH}_2 \)-terminal region of the heavy chain (6), moderate homology can be seen with the region of pepsin from residue 92 to 109 (42).\(^3\) Additionally, Keilova (43) has obtained a peptide sequence of the diazo compound-reactive site of cathepsin D which is highly homologous with the sequence around active-site aspartyl-215 in pepsin. We have shown that this site is located in the heavy chain of cathepsin D (6).

Since the amino acid sequence of cathepsin D is homologous to pepsin and other aspartyl proteases, the tertiary structure of cathepsin D and these enzymes should be closely homologous. (The tertiary structural homologies are usually closer than the primary structural homology in the available com-

\(^3\) T. Takahashi and J. Tang, unpublished results.
Oligosaccharides of Lysosomal Cathepsin D

Comparisons. Between the two lobes of penicillopepsin, for example, only the short sequences around the active-site aspartyls are homologous; but the tertiary structures of the two lobes contain 35% of topologically equivalent residues (44). Since the tertiary structure of cathepsin D has not been determined, it is justifiable to examine the glycosylation positions using the available homologous crystal structures. In all four available crystal structures of aspartyl proteases, pepsin (45), penicillopepsin (46), Endothia parasitica acid protease (47), and Rhizopus chinensis acid protease (47), positions 67 and 183 (pepsin residue numbers) are located on the surface of the molecules, as should be expected for glycosylation sites. It is interesting that both glycosylation sites are located at the crossover points in the upper strain of a twisted β-structure (corresponding to the cross-point of a letter “a”).

Aspartyl proteases are known to contain two halves highly homologous in tertiary structure (44). The two glycosylation positions are located one in each half of the molecule. However, position 67 in the NH2-terminal lobe does not correspond to position 183 in the COOH-terminal lobe.

Asn-67 in the light chain is apparently attached to 5 different sugar structures. Four of these (in peptides LC1, LC2, LC3, and LC5) are in similar proportions ranging from about 19 to 29% of the total (Table I). These proportions were closely reproducible in three different preparations. Although the mannose contents of peptides LC1 and LC2 appear somewhat too low (Table I), the analytical data for mannose in these five peptides were generally in agreement with the proton NMR results, which clearly demonstrated that these peptides differ in the numbers and structures of mannose (Figs. 5 and 6). A fifth oligosaccharide linked to the light chain, peptide LC4, represents about only 5% of the total. Because of the low yield, the NMR spectrum from this peptide was not as clear-cut as those for the four major oligosaccharide structures. However, the existing evidence from sugar analysis (Table I) and from NMR studies suggest the presence of an extra GlcNAc linked to a mannose. Several of such high mannose-type structures are known (40). In the heavy chain, there are three types of oligosaccharide chains linked to Asn-183. All three contain 5 mannoses (Table I) with the same arrangement for HTH1 and HTH2 (Fig. 6) and presumably HTH3 (which had a yield too low for NMR studies), but two of these contain additional fucose or galactose. The presence of a fucose residue linked to a high mannose oligosaccharide, as in peptide HTH2, has not been reported for lysosomal enzymes. A similar structure with 4 mannoses and a fucose has been reported for a lima bean lectin by Misaki and Goldstein (48). The higher mannose content of peptide HTH3 was too low for multiple-sample analyses of glucosamines at different hydrolysis times. The value of 2 glucosamines in this peptide may be too low since most galactose residues in asparagine-linked oligosaccharides are attached to a third N-acetylglucosamine.

It is clear from the structural studies that several oligosaccharide structures are found at each glycosylation site. A question is raised as to the origin of the structural microheterogeneity within a glycosylation site. One possible source is the difference in oligosaccharide chain processing during the biosynthesis of cathepsin D. It is well known that in the biosynthesis of the N-linked oligosaccharides, glucoses and mannoses are removed from the original precursor structure (40). Different degrees of mannose removal and selective glycosylation with N-acetylglucosamine, galactose, and fucose during the subsequent processing could result in the structural microheterogeneity. The second explanation is that most of the oligosaccharide structures are degradation products from high mannose-type oligosaccharides within the lysosomes after synthesis, glycosylation, mannose phosphorylation, and lysosome packaging. The fact that no phosphorylated mannose was observed in all the oligosaccharides also appears to have been due to hydrolysis taking place within the lysosomes. Hasilik and Neufeld (49) have demonstrated that cathepsin D and several other lysosomal hydrolases are phosphorylated on the oligosaccharides during biosynthesis. Since the purification of cathepsin D in this work was carried out in the presence of a phosphatase inhibitor (21), the absence of detectable phosphate was likely due to in vivo hydrolysis.

It should be noted that mannose structural microheterogeneity is present only at the Asn-67 site. The oligosaccharides linked to Asn-183 have the same 5-mannose structure. Even though the tertiary structure of cathepsin D is a factor and may limit the degree of mannose degradation at the Asn-183 site, a specific processing mechanism for mannose at this site cannot be ruled out at the present.

Acknowledgments—We wish to thank Dr. Stuart Kornfeld and Dr. Marc Reitman for helpful discussions during the course of this work and constructive criticism of this manuscript, Dr. Yu-Teh Li for a sample of α-mannosidase, Dr. Kenneth W. Jackson for his kind
assistance in the methods of HPLC and sequencing, and Abdul Deh-jaroni and Azar Fejsi for their able technical assistance.

REFERENCES

32. Reinhold, V. N. (1972) Methods Enzymol. 25, 244-249
35. Marinetti, G. V. (1962) J. Lipid Res. 3, 1-20
Oligosaccharides of Lysosomal Cathepsin D

The results presented in this paper describe the detailed characterization of oligosaccharides from the lysosomal cathepsin D. A high-performance liquid chromatography (HPLC) analysis was performed on the purified enzyme, and the oligosaccharides were isolated and characterized. The structure of the oligosaccharides was determined by mass spectrometry and NMR analysis.

Figures:

Figure 1: Chromatography of oligosaccharides from Lysosomal Cathepsin D.

Table 1: Oligosaccharide Composition

<table>
<thead>
<tr>
<th>Type</th>
<th>DP</th>
<th>Molar Mass (kDa)</th>
<th>Charge</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table above shows the oligosaccharide compositions of Type A and Type B, with their respective molar masses, charges, and compositions.

Discussion:
The analysis of oligosaccharides was determined to be an essential step in understanding the function of lysosomal enzymes. The oligosaccharides were isolated and characterized, and their compositions were determined. This information is crucial for understanding the enzymatic activity and substrate specificity of lysosomal enzymes.

Conclusion:
The study presented in this paper provides valuable insights into the structural and functional properties of lysosomal cathepsin D. The characterization of the oligosaccharides will be essential in further studies to elucidate the role of cathepsin D in various biological processes.
Oligosaccharides of Lysosomal Cathepsin D

Figure 1: HPLC separation of oligosaccharides from the lysosomal fraction of cathepsin D heavy chain. The results are described in the Materials section.

Figure 2: HPLC separation of oligosaccharides from the lysosomal fraction of cathepsin D heavy chain. The results are described in the Materials section.

Figure 3: HPLC separation of oligosaccharides from the lysosomal fraction of cathepsin D heavy chain. The results are described in the Materials section.

Figure 4: HPLC separation of oligosaccharides from the lysosomal fraction of cathepsin D heavy chain. The results are described in the Materials section.
Oligosaccharide units of lysosomal cathepsin D from porcine spleen. Amino acid sequence and carbohydrate structure of the glycopeptides.

T Takahashi, P G Schmidt and J Tang


Access the most updated version of this article at http://www.jbc.org/content/258/5/2819

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/258/5/2819.full.html#ref-list-1