We have isolated and sequenced cDNA clones corresponding to the entire coding sequences of the human lysosomal membrane glycoproteins, lamp-1 and lamp-2 (h-lamp-1 and h-lamp-2). The deduced amino acid sequences indicate that h-lamp-1 and h-lamp-2 consist of 416 and 408 amino acid residues, respectively, and suggest that 27 and 28 NH₂-terminal residues are cleavable signal peptides. The major portions of both h-lamp-1 and h-lamp-2 reside on the luminal side of the lysosome and are heavily glycosylated by N-glycans. h-lamp-1 and h-lamp-2 were found to contain 19 and 16 potential N-glycosylation sites, respectively.

The findings are consistent with the results obtained by endo-β-N-acetylglycosaminidase F treatment of h-lamp-1 and h-lamp-2 precursors, described in the preceding paper (Carlsson, S. R., Roth, J., Piller, F., and Fukuda, M. (1988) J. Biol. Chem. 263, 18911-18919). These N-glycosylation sites are clustered into two domains separated by a hinge-like structure enriched with proline and serine in h-lamp-1 or proline and threonine in h-lamp-2. The two domains of h-lamp-1 on each side of the hinge region are homologous to each other, whereas no such homology was detected between the two domains of h-lamp-2. Both proteins have one putative transmembrane domain consisting of 24 hydrophobic amino acids near the COOH terminus, and contain a short cytoplasmic segment composed of 11 amino acid residues at the COOH-terminal end. Comparison of h-lamp-1 and h-lamp-2 sequences reveal strong homology between the two molecules, particularly in the proximity to the COOH-terminal end. It is possible that this portion is important for targeting the molecules to lysosomes. These results also suggest that lamp-1 and lamp-2 are evolutionarily related. Comparison of known lamp-1 sequences among different species, on the other hand, show that human lamp-1 has more similarity to lamp-1 from other species than lamp-2. This fact, taken together with the results obtained in the preceding paper, suggests that lamp-1 and lamp-2 diverged from a putative common ancestral gene in early stages of evolution. These results also suggest that lamp-1 and lamp-2 probably have distinctly separate functions despite the fact that they share many structural features.

In the preceding paper, we described the characterization of two different human lysosomal membrane glycoproteins with Mᵋ ~ 120,000, h-lamp-1 and h-lamp-2, isolated from chronic myelogenous leukemia cells (1). The two glycoproteins appear to share some structural features and have the polypeptide portions of about ~40 kDa. They are heavily glycosylated by N-glycans, and h-lamp-1 and h-lamp-2 were shown to contain 18 and 16 N-glycans, respectively. The most striking feature is that some of the N-glycans attached to lamp-1 and lamp-2 are polylactosaminoglycans (1).

We have also succeeded in isolating cDNA clones encoding h-lamp-1 and reported its unique structure (2). However, it was also noticed that the NH₂-terminal amino acid sequence deduced from those cDNAs seemed to be truncated, when compared to that obtained on a purified protein (2).

We report here the characterization of cDNAs encoding another human lysosomal membrane glycoprotein, h-lamp-2, and the isolation of cDNA clones derived from the full-length h-lamp-1 mRNA. Comparison of the deduced amino acid sequence of h-lamp-2 with that of h-lamp-1 indicates that human lamp-1 and lamp-2 share common structural features in both gross structures and their COOH-terminal regions. Further analysis, however, reveals that lamp-1 and lamp-2 are distinct molecules which may have different functions in the lysosomal membrane.

**EXPERIMENTAL PROCEDURES**

Isolation of the Glycoprotein Fraction Containing Human Lamp-1 and Lamp-2 and Their Derived Peptides—The glycoprotein fraction, which is a mixture of h-lamp-1 and h-lamp-2, was purified as described previously (1). Peptides were generated by CNBr fragmentation of the glycoprotein fraction after reduction and S-carboxymethylation. Reduced and carboxymethylated glycoprotein (~250 μg of protein) was cleaved with CNBr (50 mg/ml in 70% formic acid) for 20 h. After lyophilization, the sample was dissolved in 0.1% trifluoroacetic acid and filtered through a 0.45-μm Millipore filter. The filtered sample was applied on a Vydac TP214 C18 wide pore reverse phase column (4.6 mm × 20 cm), previously equilibrated with 0.1% trifluoroacetic acid. After washing with 10 ml of 0.1% trifluoroacetic acid, the column was eluted with 40 ml consisting of a gradient (0-60%) of acetonitrile in 0.08% trifluoroacetic acid. Flow rate was 1 ml/min, and the absorbance of the eluate was monitored at 215 nm (Fig. 1). Three major peaks, obtained after HPLC, were further purified by gel filtration on Sephacryl S-200. Each peak was lyophilized, dissolved in 0.5 ml of 0.1 M NH₄HCO₃, and applied on a column (1 x...
120 cm) of Sephacryl S-200, equilibrated and eluted with 0.1 M NH₄HCO₃. Flow rate was 6 ml/h, fractions (2 ml) were collected (Fig. 2). Among these peptides, four fractions designated J₁, K₁, K₂, and M₂, were sequenced by automated Edman degradation using an Applied Biosystems Model 470A gas phase amino acid sequencer. Phenylthiobiotindantoine amino acid derivatives were identified by HPLC as described (3). Similar peptides were isolated from trypsin digests by Sephacryl S-200 gel filtration followed by reverse phase HPLC.

Oligonucleotide Probes—Synthetic oligonucleotides were synthesized on an Applied Biosystems 350A oligonucleotide synthesizer. Sequencing primers were used directly without further purification. For cDNA library screening, oligonucleotide probes, which are based on amino acid sequences using most probable codon usage (4, 5), were synthesized. The oligonucleotides were purified on reverse phase HPLC and end-labeled with [γ-³²P]ATP (>7000 Ci/mmol, Du Pont-New England Nuclear) using T₄ polynucleotide kinase (Bethesda Research Laboratories). The radioactive oligonucleotides were used for screening of human fetal fibroblast cell line (IMR-90) (6) and human placental (7) λgt11 cDNA libraries. S₁ nuclease was employed in the synthesis of the IMR-90 library, whereas the placental library was made without the use of S₁ nuclease. The λgt11 cDNA libraries were generously provided by Drs. Krusius, Ruoslahti, and Millan at the La Jolla Cancer Research Foundation.

In situ colony hybridization was performed essentially according to the method of Woods et al. (8). In brief, the filter paper was prehybridized at 60 °C (for cDNA probe) or at 42 °C (for oligonucleotide probe) in 6 × SSC, 1 × Denhardt's solution containing 0.25% SDS and 100 µg/ml sheared and denatured salmon sperm DNA (for denatured salmon sperm DNA, see Ref. 9). The filters were then hybridized in 50 mM potassium phosphate, pH 6.5, 6 × SSC, 1 × Denhardt's, 50 µg/ml of Escherichia coli RNA for 20–22 h. The hybridization temperature was 45 °C for probes on the basis of the most probable codon usage, 50 °C for exact oligonucleotide probes, and 60 °C for cDNA inserts. The filters were washed at room temperature for 40 min to 1 h in 2 × SSC, 0.1% SDS. If necessary, the filters were then washed further at 50 °C in 6 × SSC, 0.1% SDS for exact oligonucleotide probes or at 60 °C in 0.1 × SSC, 0.1% SDS for cDNA inserts. The filters were then washed at 50 °C for 15–40 min. The background radioactivity was monitored by a hand-held Geiger counter (series 900 mininmonitor, Mini-Instruments, Ltd., purchased from Research Products International Corp., Mount Prospect, IL).

In order to obtain cDNAs encoding h-lamp-1 from the placental library, previously described (1, h-lamp-A-4), (2) was used as probe for initial screening. cDNA probes were made by random-oligonucleotide-primed extension (10) as described previously (11). For the later stages of screening, differential hybridization to oligonucleotide probes specific to the h-lamp-1 or h-lamp-2 cDNA sequence and end-labeled as described above, was performed. The oligonucleotide probes were synthesized according to the method of Woods et al. (8). The λgt11 cDNA library was isolated, based on their hybridization to oligonucleotide A. The predicted amino acid sequence of h-lamp-1, based on this cDNA clone (X-h-lamp-A-4), was described previously (2), and oligonucleotide A was denoted as oligonucleotide 1 in the previous report (2). This glycoprotein is a mouse counterpart m-lamp-1 than h-lamp-2, which was isolated by oligonucleotide probe B (as described below).

H-lamp-1 was previously called lamp-A (2). Although the amino acid predicted from the h-lamp-A-4 cDNA was a nearly perfect match for the amino acid sequence of the isolated synthetic oligonucleotide A, the predicted amino acid sequence of h-lamp-1, based on this cDNA clone (X-h-lamp-A-4), was described previously (2), and oligonucleotide A was denoted as oligonucleotide 1 in the previous report (2).

RESULTS AND DISCUSSION

Construction of Oligonucleotide Probes—Initially, the glycoprotein fraction containing both human lamp-1 and lamp-2 was used to generate peptides, since it was not clear that this fraction contained both glycoproteins. The reverse phase HPLC of the CNBr fragments provided three major peptide peaks (J₁, K₁, and M₂), as shown in Fig. 1. Each peak was further purified by gel filtration on Sephacryl S-200. Fraction J₁ provided one peptide, fraction K₁ provided two peptides, and fraction M₂ provided one peptide (Fig. 2). The amino acid sequence of each peptide was obtained as summarized in Table I. We chose J₁ and K₁ peptides to construct oligonucleotide probes A and B. It was later found that peptide J₁ was derived from lamp-1 protein and K₁ was derived from lamp-2 protein (see below).

A mixture of two probes was used to screen a λgt11 library of IMR-90 cDNA. In the initial screening, a total of nine positive clones were obtained. When these clones were tested for reactivity of oligonucleotide probe A and probe B separately, it was discovered that each clone reacted with only one of the probes (Fig. 3). We therefore tested if the cDNA clones could be grouped into two sets. Labeled cDNA inserts from clone 5 and clone 9, respectively, were used for hybridization to cDNAs from clones 1 to 9. As shown in the right side of Fig. 3, the results clearly indicate that those cDNA clones represent two different sets of molecules and each clone reacts with only those within the same set of cDNAs. During these experiments, clone 3 was found to be unrelated to either set and was not tested further.

Isolation of cDNAs Encoding of the Full-length of h-lamp-1—For the first attempt, cDNAs isolated from the IMR-90 library were isolated, based on their hybridization to oligonucleotide A. The predicted amino acid sequence of h-lamp-1, based on this cDNA clone (λ-h-lamp-A-4), was described previously (2), and oligonucleotide A was denoted as oligonucleotide 1 in the previous report (2). This glycoprotein is called h-lamp-1 because it was found to have more similarity to the mouse counterpart m-lamp-1 than h-lamp-2, which was isolated by oligonucleotide probe B (as described below).

H-lamp-1 was previously called lamp-A (2). Although the amino acid predicted from the λ-h-lamp-A-4 cDNA was a nearly perfect match for the amino acid sequence of the isolated synthetic oligonucleotide A, the predicted amino acid sequence of h-lamp-1, based on this cDNA clone (X-h-lamp-A-4), was described previously (2), and oligonucleotide A was denoted as oligonucleotide 1 in the previous report (2).

RESULTS AND DISCUSSION

Construction of Oligonucleotide Probes—Initially, the glycoprotein fraction containing both human lamp-1 and lamp-2 was used to generate peptides, since it was not clear that this fraction contained both glycoproteins. The reverse phase HPLC of the CNBr fragments provided three major peptide peaks (J₁, K₁, and M₂), as shown in Fig. 1. Each peak was further purified by gel filtration on Sephacryl S-200. Fraction J₁ provided one peptide, fraction K₁ provided two peptides, and fraction M₂ provided one peptide (Fig. 2). The amino acid sequence of each peptide was obtained as summarized in Table I. We chose J₁ and K₁ peptides to construct oligonucleotide probes A and B. It was later found that peptide J₁ was derived from lamp-1 protein and K₁ was derived from lamp-2 protein (see below).

A mixture of two probes was used to screen a λgt11 library of IMR-90 cDNA. In the initial screening, a total of nine positive clones were obtained. When these clones were tested for reactivity of oligonucleotide probe A and probe B separately, it was discovered that each clone reacted with only one of the probes (Fig. 3). We therefore tested if the cDNA clones could be grouped into two sets. Labeled cDNA inserts from clone 5 and clone 9, respectively, were used for hybridization to cDNAs from clones 1 to 9. As shown in the right side of Fig. 3, the results clearly indicate that those cDNA clones represent two different sets of molecules and each clone reacts with only those within the same set of cDNAs. During these experiments, clone 3 was found to be unrelated to either set and was not tested further.

Isolation of cDNAs Encoding of the Full-length of h-lamp-1—For the first attempt, cDNAs isolated from the IMR-90 library were isolated, based on their hybridization to oligonucleotide A. The predicted amino acid sequence of h-lamp-1, based on this cDNA clone (λ-h-lamp-A-4), was described previously (2), and oligonucleotide A was denoted as oligonucleotide 1 in the previous report (2). This glycoprotein is called h-lamp-1 because it was found to have more similarity to the mouse counterpart m-lamp-1 than h-lamp-2, which was isolated by oligonucleotide probe B (as described below).

H-lamp-1 was previously called lamp-A (2). Although the amino acid predicted from the λ-h-lamp-A-4 cDNA was a nearly perfect match for the amino acid sequence of the isolated
Molecular Cloning of Human Lysosomal Membrane Glycoproteins

**Fig. 2.** Sephacryl S-200 gel filtration of peptides which were separated by HPLC. J (A), K (B), and M (C) peptides, isolated after HPLC in Fig. 1, were separately subjected to Sephacryl S-200 gel filtration as described under "Experimental Procedures." Fractions were pooled as indicated by horizontal bars. Other fractions were found to contain negligible amounts of peptides by amino acid analysis. The vertical arrows indicate the elution positions of the void volume.

**TABLE I**

<table>
<thead>
<tr>
<th>Peptides</th>
<th>Amino acid sequences</th>
<th>Amino acid residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>J1</td>
<td>N-VTVTLHDAIQRALS-SFSRGETRC<em>EQDRPP</em>·P*</td>
<td>h-lamp-1, 136-170</td>
</tr>
<tr>
<td>K1</td>
<td>ALQL·ITQDKVASVINPN</td>
<td>h-lamp-2, 210-225</td>
</tr>
<tr>
<td>K2</td>
<td>SFVY·LSDTHLP·AS</td>
<td>h-lamp-1, 89-104</td>
</tr>
<tr>
<td>M1</td>
<td>FTVRYETT·KTYKTV</td>
<td>h-lamp-2, 21-36</td>
</tr>
<tr>
<td>T,c</td>
<td>EKPEAGTY</td>
<td>h-lamp-2, 186-195</td>
</tr>
<tr>
<td>SD</td>
<td>DPAFK</td>
<td>h-lamp-1, 287-301</td>
</tr>
</tbody>
</table>

h-lamp-1

**Fig. 3.** Blot hybridization analysis of cDNA clones for human lamp-1 and lamp-2 isolated from the IMR-90 cDNA library. λgt11 clones (1–9) and control λgt11 (λ) were cut by EcoRI and resolved by electrophoresis in 1% agarose gels. Following transfer to nitrocellulose, the blot was sequentially hybridized to oligonucleotide probes and cDNA inserts. The hybridization to oligonucleotide probes was performed at 45°C for A and 50°C for B, and low-stringency wash was performed as described under “Experimental Procedures.” Inserts from λ5 and λ9 clones were excised by EcoRI and purified by agarose gel electrophoresis. Those cDNA inserts (850 base pairs from λ5 and 700 base pairs from λ9) were sequentially hybridized to the same blot.

**Fig. 4.** The amino acid sequence deduced from the phL1-15B clone is consistent with that obtained from purified h-lamp-1 glycoprotein purified from chronic myelogenous leukemia cells (2).

To determine if another cDNA could be isolated that more closely matches the protein sequence, a placental cDNA library was screened using λ-h-lamp-A-4 cDNA as probe. Positive clones were further tested by the oligonucleotides synthesized according to the sequences of h-lamp-1 (nucleotides 334 to 358 in Fig. 4) and h-lamp-2 (nucleotides 146 to 193 in Fig. 6). These oligonucleotides were chosen on the basis that the nucleotide sequence of the NH2-terminal half of the glycoproteins is more characteristic to each gene than that in the COOH-terminal half. The cDNA clones, which preferentially reacted with the h-lamp-1 oligonucleotide, were isolated and sequenced. Fig. 4 shows the nucleotide sequence of one of the clones, P-hL1-15B.

The amino acid sequence deduced from the P-hL1-15B clone is consistent with that obtained from purified h-lamp-1 protein. Starting from the initiating methionine, a hydrophobic segment of 27 amino acids is present. Following this, the sequence beginning at the alanine residue exactly matches the NH2-terminal sequence of the isolated human lamp-1. The amino acid sequence surrounding residue 1 (15) and the presence of arginine near the initiation methionine are consistent with the peptide segment of 27 amino acids serving as a leader peptide (16). The cDNA sequence also includes a large segment of 3′-end untranslated region. These two portions were missing in the previous clone, perhaps due to S1 protein, the NH2-terminal sequence deduced from cDNA analysis was different from that obtained on a human lamp-1 glycoprotein purified from chronic myelogenous leukemia cells (2).
**Fig. 4.** The nucleotide sequence and amino acid sequence of the human lamp-1 cDNA clone, P-hL1-15B, isolated from placental cDNA library. The positions that were shown are indicated by two dotted lines. Two regions of the sequence corresponding to the serine/proline-rich hinge region are indicated by boxes. The NH2-terminal amino acid is indicated by a solid line, and potential glycosylation sites are indicated by dots. The consensus sequence for polyadenylation signal is not underlined. The repeated sequences in the 3'-untranslated region are underlined. The NH2-terminal amino acid is indicated by a thick line. The sequence used making a probe was underlined with a dashed line. The positions that were shown as blank during amino acid sequencing are circled. Cysteine residues are boxed. Molecular Cloning of Human Lysosomal Membrane Glycoproteins
The tandemly repeated nucleotide sequence (B) and the alignment of the tandemly repeated nucleotide sequence (A) and the alignment of the tandemly repeated nucleotide sequence (B). A, the nucleotide sequence is presented both on the horizontal and the vertical axes. Dots, which represent matched sequences, are obtained by the Microgenie (Beckman) program. B, the tandemly repeated sequences (nucleotides 1688-1805) are aligned. The nucleotides are numbered as described in Fig. 4.

nuclease digestion of nascent cDNA during the IMR-90 library construction. All the other structural features are the same as those shown in the cDNA clone isolated from the IMR-90 library (2). Thus, we believe that a full-length cDNA clone for human lamp-1 has been isolated. This conclusion is supported by the fact that the size of human lamp-1 mRNA was found to be about 2.2 kilobases (2). It is noteworthy that the amino acid sequence contains 19 potential N-glycosylation sites, whereas that of h-lamp-2 is enriched by proline and serine like the IgA-cy1 chain hinge region. Furthermore, the hinge-like structure in h-lamp-1 is enriched by proline and threonine. It is of interest to note that the hinge region of IgD heavy chain is enriched by proline and threonine.

When the presence of repeating sequences was tested in the whole cDNA sequence by the matrix method, extensive repeating sequences were detected in the 3'-untranslated region of h-lamp-1 by computer search (Fig. 5A). Each of the repeats is essentially composed of GAGGGGTGGGGGTGCCGCT-CTCTCT, and the sequence is tandemly repeated nine times (Fig. 5B and the dotted line in Fig. 4). It is not known if the repeating sequences play any role in transcription or translation of h-lamp-1. It has been shown that the Ha-ras gene has tandem repeats of 28 base pairs in a position 3' to the coding region (17), and those repeated sequences were shown to have an enhancer activity (18, 19). Furthermore, it has been shown that the length of such tandem repeats are highly variable according to individuals, and those variations serve as a basis for restriction fragment length polymorphisms (20).

It is noteworthy, however, that the tandem repeats in the Ha-ras genes are present in introns, whereas the tandem repeats in the h-lamp-1 gene are located in transcribed, 3'-untranslated sequences. It will be interesting to determine if any function is associated with these repeats in the h-lamp-1 mRNA.

Predicted Primary Sequence and Structure of h-lamp-2—
cDNA clones reacting with the oligonucleotide B were isolated, and three independent clones were found as shown in Fig. 3. Among them, clones 2 and 5 were found to contain cDNA corresponding to the full-length coding region of h-lamp-2. The sequence of clone 2, I-hL2-2, is shown in Fig. 6B. The identical sequence was obtained from clone 5. This cDNA contains nucleotide sequences encoding a 408-amino acid polypeptide, which contains a putative signal sequence and transmembrane domain (Fig. 6B). Analogous to h-lamp-1, a considerably hydrophobic domain of 28 amino acids is present following the initiation methionine. The sequence following this domain exactly matches the NH2-terminal amino acid sequence determined from purified mature protein, starting from the leucine residue. The amino acid sequence surrounding residue 1 conforms with the leader peptide cleavage site (–3, –1) rule (15), and the presence of 1 arginine residue close to the initiation methionine also supports the idea that the peptide segment of 28 amino acids is probably a leader peptide, which is cleaved from the mature protein (16). The deduced amino acid sequence of h-lamp-2 has a hydrophobic segment (amino acid 346 to residue 369) of 24 amino acids near the COOH-terminus. This sequence is flanked on the COOH-terminal side by a putative cytoplasmic tail of approximately 10 amino acid residues. The amino acid sequence indicates that there are 16 potential N-glycosylation sites in h-lamp-2. Five of the 16 putative glycosylation sites are within the sequenced CNBr fragments. Each of these at positions 4, 10, 21, 30, and 214 gave a blank signal in amino acid sequencing, indicating that these asparagine residues are modified by N-linked saccharides. The partial and complete digestions of h-lamp-2 by endo-β-N-acetylgalactosaminidase F provided evidence that h-lamp-2 contains 16 N-linked saccharides, as shown in the preceding paper (1), and the results indicate that all of 16 N-glycosylation sites identified in the deduced amino acid sequence are utilized in the mature protein.

Beginning at the amino acid residue 171, a characteristic of a 20-amino acid-long domain with a high content of proline (6/20) and threonine plus serine (8/20) is found in the h-lamp-2 sequence. This segment corresponds to the previously identified hinge region of h-lamp-1, which in turn was found to have a homology to the IgA-α chain hinge structure (2, 21). Furthermore, the hinge-like structure in h-lamp-1 is enriched by proline and serine like the IgA-α chain hinge region, whereas that of h-lamp-2 is enriched by proline and threonine. Both immunoglobulin hinge regions are glycosylated by O-linked oligosaccharides (23, 24). It is, therefore, most likely
that the O-glycans found in human lamp-1 and lamp-2, shown in the preceding paper (1), are attached around these hinge regions. This hinge-like structure in h-lamp-2 appears to separate the molecule into two domains. Each domain contains 8 N-linked carbohydrates and 4 cysteine residues. The latter residues probably form disulfide bonds, since the molecule could not be digested well by trypsin unless it was first reduced with 2-mercaptoethanol followed by S-carboxymethylation.

Another initiation methionine signal (nucleotides 9–11 in Fig. 6B) is present at 90 nucleotides upstream from the putative initiation methionine, but this sequence is possible due to an artifact of the cDNA library construction, considering that it is next to the linker. This upstream methionine would produce a leader peptide with a comparable size to that of h-lamp-2, strongly suggesting that the ATG at nucleotides 100–102 codes for the initiation methionine. Further studies on genomic structure will be essential to clarify these points.

Homology between Human lamp-1 and lamp-2—Based on the deduced amino acid sequences of human lamp-1 and lamp-2, the homology between these proteins was examined. The results are shown in Fig. 7A and can be summarized as follows: 1) all of the cysteine residues are aligned well between human lamp-1 and lamp-2 sequences; 2) six of the N-glycosylation sites are at identical positions and four additional N-glycosylation sites are in close proximity; and 3) the homology is more prominent in the second half of the COOH-terminal domain (40.5% matches) than the first half of the NH2-terminal domain (31.5% matches). In particular, the homology is most prominent in the transmembrane and cytoplasmic portion. In contrast, the leader peptides, which are hydrophobic, have minimal similarity. In total, the two proteins share 36.7% matched sequences.

When homologous sequences between human lamp-1 and lamp-2 were examined, the results shown in Fig. 7B were obtained. The first homology lies in the hinge-like structure. In fact, this homology extends to the residue 177 (lamp-1) and 186 (lamp-2), when the nucleotide sequences are compared. The second and third homologous sequences are present around the first cysteine residue after the hinge structure. The fourth homologous sequence encompasses the transmembrane and cytoplasmic portion of the hinge segment. This homology structure plays an important role shared by both proteins.

Homology among lamp-1 Molecules—To determine if human lamp-1 is more similar to lamp-1 from other species or to human lamp-2, the deduced amino acid sequence of human lamp-1 was compared to available mouse lamp-1 (25) and chicken LEP100 (29) sequences. The results of this comparison are shown in Fig. 8, and it can be seen that the alignment of these sequences is extensive. It was found that the human lamp-1 and mouse lamp-1 sequences have 66.1% identity, while 51.5% identical residues were found in the human lamp-2.2 and 186 (lamp-2), when the nucleotide sequences are compared. The second and third homologous sequences are present around the first cysteine residue after the hinge structure. The homology structure plays an important role shared by both proteins.

M. A. Williams, and M. Fukuda, unpublished results.
FIG. 7. Alignment of the amino acid sequences of human lamp-1 and lamp-2 (A) and homology of the human lamp-1 and lamp-2 amino acid sequences (B). The alignment of two amino acid sequences and homology between them were tested by using the Microgenie (Beckman) program. In A, identical amino acids are boxed, N-glycan attachment sites are indicated by asterisks, and cysteine residues are circled. In B, the amino acid residues that differ are marked by asterisks.

FIG. 8. Alignment of the lamp-1-related molecule from human, mouse, and chicken origins. The amino acid sequences of human lamp-1 (present study), mouse lamp-1, m-lamp-1 (25), and chicken LEP100 (26) are compared for alignment. The identical amino acids are boxed.
Fig. 9. Comparison of the NH2-terminal domain (N) and the COOH-terminal domain (C) of human lamp-1 sequence. The human lamp-1 amino acid sequences of residues 1 to 169 (N domain) and 191 to 354 (C domain) were compared by the Microgenie alignment program. Identical residues are boxed, cysteine residues are circled, and N-glycosylation sites are indicated by asterisks. Note that all cysteine residues are aligned, and they have 22.4% matches in total.

A

B

Fig. 10. Depicted structure of human lamp-1 (A) and lamp-2 (B). N-Linked carbohydrate moieties and hinge-like structures are indicated by \( \gamma \) and \( \gamma' \), respectively. It is assumed that the N-glycans attached closely to transmembrane are polyactosaminoglycan, and those at the NH2-terminal end are typical complex N-linked saccharides. N-Glycans in the third loop may be of intermediate size. Each presumptive loop is made by a disulfide bond. Cysteine residues connected to each other are tentatively assigned. The majority of the molecules reside in the luminal side of lysosomes. The membrane is indicated by a shaded area. Although this figure does not include O-glycan attachment sites, it is likely that they are clustered around hinge-like structures as shown in IgA \( \alpha \)-chain (23) and IgD \( \kappa \)-chain (24).

Close inspection of the deduced human lamp-1 amino acid sequence indicated that internal repeats exist between the NH2-terminal half and the COOH-terminal half of the luminal domain, which are separated by a hinge-like structure. As shown in Fig. 9, it is apparent that two domains of human lamp-1 constitute internal repeats with a good alignment of all cysteine residues. Interestingly, attempts to locate internal repeats in the deduced amino acid sequence on human lamp-2 failed to reveal such a structure, and no alignment of cysteine residues was achieved. These results further support the hypothesis that the lamp-1 and lamp-2 genes have undergone extensive changes since diverging from the putative ancestor gene. These results also suggest that the ancestor gene for lamp-1 and lamp-2 may have evolved from duplication of the precursor gene which would have encoded one domain consisting of about 100 amino acids, including 4 cysteine residues. It is conceivable that a gene segment coding for the hinge-like structure was inserted between duplicated genes and the transmembrane and cytoplasmic portions were added independently at the COOH-terminal of the gene. Studies on the genomic structures of lamp-1 and lamp-2 will be essential to make further hypothesis on the evolution of these related glycoproteins.

Depicted Structure of Human lamp-1 and lamp-2—We propose that human lamp-1 and lamp-2 have the structures depicted in Fig. 10, based on their deduced amino acid sequences. The data discussed above suggest that each glycoprotein contains four loops, with each presumptive loop formed by a disulfide bond and each loop consisting of 36 and 39 amino acid residues. The distance between the first loop (NH2-terminal) and the second loop is the 74 residues in both h-lamp-1 and h-lamp-2. The distance between the third and fourth loop is 68 residues for h-lamp-1 and 63 for h-lamp-2. Alternatively, it is possible that loops are formed between the second and third cysteine and between the sixth and seventh cysteine. In this case, the first loop consists of 74 residues (Cys92 to Cys167 in lamp-1) or Cys91 to Cys166 in lamp-2), and the second loop consists of 68 residues (Cys167 to Cys235 in lamp-1) or 63 residues (Cys235 to Cys298 in lamp-2). The size of loop is in a range that is observed in immunoglobulin superfamilies sequences (27). Furthermore, the COOH-terminal end of these putative loops contains a consensus sequence, Tyr-X-Cys, found in immunoglobulin-related domains (27). However, we detected no further homology between lamps and immunoglobulin-related domains in the sequences preceding and following the cysteine residues. If the cysteine residues in lamps are linked to form loops similar to those observed in immunoglobulin superfamilies, it would suggest a potential receptor function for lamp molecules (27). It will, therefore, be important to determine how disulfide bonds are formed in lamps in order to understand the functional significance of these molecules.

The glycoproteins are divided roughly in half by the presence of proline/serine- or proline/threonine-rich hinge-like regions. The NH2-terminal half consists of 169 amino acid residues, and the COOH-terminal half contains 155 amino acid residues both in h-lamp-1 and h-lamp-2. Both contain one transmembrane domain consisting of 24 hydrophobic amino acid residues and an approximately 11-residue cytoplasmic tail.

In these aspects, human lamp-1 and lamp-2 have strikingly similar structures. In addition, it is evident that human lamp-1 and lamp-2 share very homologous structures in the third disulfide loop in Fig. 10, which begins at residue 200 or 201 (Fig. 7B). The functional significance of the third loop is not currently known, but its appearance in two glycoproteins, which seemingly diversified early in evolution, suggests that this structure plays an important role in the synthesis or function of lamp-1 and lamp-2. Similarly, very homologous sequences are present in the transmembrane and cytoplasmic segments. Specific signals have been identified for proteins remaining in the endoplasmic reticulum (28, 29), those staying in the Golgi (30), and those transported to the nucleus (31). These signals have been found to reside in the COOH-terminal portion (28, 29, 31) or transmembrane and cytoplasmic regions of specific proteins (30). The data presented here suggest that the homologous transmembrane and cytoplasmic segments of lamp-1 and lamp-2 may serve as a molecular
signal to target these glycoproteins to lysosomes.

The deduced amino acid sequences of human lamp-1 and lamp-2 show that these glycoproteins have some differences in their sites for N-linked saccharides. H-lamp-1 has only one N-glycosylation site close to the transmembrane portion, whereas h-lamp-2 has four N-glycosylation sites close to the transmembrane portion. This difference could be important for the content of polyolactosaminoglycans in the lamp-1 and lamp-2. We recently discovered that the distance of glycosylation sites from the membrane may determine the mode of glycosylation (32). When human chorionic gonadotropin α-chain was fused with the transmembrane plus cytoplasmic segment of vesicular stomatitis virus-G protein, the fused protein acquired polyolactosaminoglycan, whereas the parent α-chain and vesicular stomatitis virus-G protein express only typical N-linked saccharides (33, 34). We proposed that, by bringing the glycosylation sites close to the transmembrane portion, a reduction in the glycosylation sites to the membrane is involved in the acquisition of polyolactosaminoglycan-containing glycoprotein. It will be interesting to further test this hypothesis by engineering and expressing altered forms of human lamp-1 and lamp-2.

Human lamp-1 and lamp-2 differ in amino acid sequences as described above. It is likely that the amino acid sequence of lamp-2 has been conserved during evolution, although this cannot be tested at this time since cDNA for lamp-2 has not been isolated from other species. It is highly possible that lamp-1 and lamp-2 have distinct functions which emerged as non-homologous structures contribute to any function of these lysosomal membrane glycoproteins.

Acknowledgments—We would like to thank Drs. Chen and August for communicating unpublished results on mouse lamp-1, Dr. Der for useful discussion, Drs. Krusius, Ruoslahti, and Millan for providing cDNA libraries, Mat Williamson for amino acid sequencing, John Knight for oligonucleotide synthesis, Dr. Mark Williams for critical reading of the manuscript, and Tami Clevenger for secretarial assistance.

REFERENCES