Structure of LEP100, a Glycoprotein That Shuttles between Lysosomes and the Plasma Membrane, Deduced from the Nucleotide Sequence of the Encoding cDNA

Douglas M. Fambrough, Kunio Takeyasu, Jennifer Lippincott-Schwarz, and Ned R. Siegel
with the technical assistance of Delores Somerville

Department of Biology, The Johns Hopkins University, Baltimore, Maryland 21218

Abstract. LEP100, a membrane glycoprotein that has the unique property of shuttling from lysosomes to endosomes to plasma membrane and back, was purified from chicken brain. Its NH$_2$-terminal amino acid sequence was determined, and an oligonucleotide encoding part of this sequence was used to clone the encoding cDNA. The deduced amino acid sequence consists of 414 residues of which the NH$_2$-terminal 18 constitute a signal peptide. The sequence includes 17 sites for N-glycosylation in the NH$_2$-terminal 75% of the polypeptide chain followed by a region lacking N-linked oligosaccharides, a single possible membrane-spanning segment, and a cytoplasmic domain of 11 residues, including three potential phosphorylation sites. Eight cysteine residues are spaced in a regular pattern through the lumenal (extracellular) domain, while a 32-residue sequence rich in proline, serine, and threonine occurs at its midpoint. Expression of the cDNA in mouse L cells resulted in targeting of LEP100 primarily to the mouse lysosomes.

S EVERAL lysosomal membrane glycoproteins have been identified by immunological methods, isolated, and their overall biochemical characteristics described (1, 6, 14, 20, 21, 27). All of these proteins bear large numbers of N-linked oligosaccharides and have apparent molecular weights (determined by SDS-PAGE) in the range 80,000 to 120,000, whereas apparent molecular weights of the core polypeptides are in the range 40,000 to 60,000. The high level of glycosylation should contribute substantially to a region of fixed negative charge on the lumenal surface of the membrane and may be related to protection of lysosomal membrane proteins from hydrolytic destruction or may have some as yet unknown biological significance. The functions of these lysosomal membrane glycoproteins are not known, but the functions may include mediation of ion permeation, export of digestion products, and facilitation of the vesicular transport that operates to deliver materials to the lysosome for digestion.

LEP100 resembles the other lysosomal membrane glycoproteins in some general physicochemical properties but appears to be unique in that it shuttles between lysosomes, endosomes, and the plasma membrane (21, 22). The distribution of LEP100 among these membrane systems is profoundly altered by agents that affect lysosomal pH (22). We have suggested that LEP100 may be involved in the shuttling process, and we have reported a remarkable congruence between patches of this protein on the plasma membrane of chloroquine-treated cells and the distribution of membrane-associated clathrin (22). Thus, we expect that some features of LEP100 reported here will prove to be general for lysosomal membrane glycoproteins, whereas other features will relate to functions unique to LEP100.

Materials and Methods

Materials

Cloning of the Encoding DNA for LEP100

LEP100 was purified from a detergent extract of chicken brain microsomes by immunoprecipitation with LEP100-IgG (21) coupled to Sepharose 4B beads. The eluted protein was subjected to SDS-PAGE, electroeluted from the gel (16), and subjected to automated Edman degradation and analysis to determine the NH$_2$-terminal amino acid sequence (15). The amino acid sequence was FDVRDSTGKVCIIANLTVAFSVEYKXXGQK, where X represents uncertain residues (see Results). A part of this sequence was used in designing an oligonucleotide that might hybridize with LEP100 mRNA and cDNA. A set of 26mer oligonucleotides containing deoxyinosine at six positions of codon third-base ambiguity (26) were designed to bind to all possible encoding sequences for the amino acid sequence TGKVCIIAN: the sequences were 5'-TTIGCIATIAT(A/G)CAIAC(C/T)TTICCIGT-3'. This set of four oligonucleotides (^{32}P-labeled with polynucleotide kinase) was used to identify several clones (2) by screening about 300,000 plaques
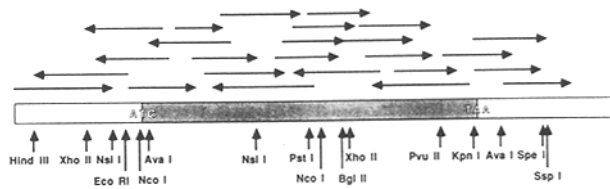
Figure 1. Nucleotide sequencing strategy. A linear map of the cDNA is depicted, with positions of initiation and termination codons and some common restriction enzyme sites. Horizontal arrows above the map indicate the direction and the length of each of the sequences read to generate the full sequence.

from unamplified chick brain cDNA libraries (30) in λgtl0 (17). Two of these clones were later determined to encode the entire LEP100 protein. The longer of these clones is described in this report.

DNA Sequencing

The cDNA inserts of positive clones were subcloned into the Eco RI site in the plasmids pEMBL8 and pEMBL18 (7). DNA sequencing by the dideoxy method (28), with [³⁵S]deoxyadenosine 5′[α-thio]triphosphate, established that the cDNAs exactly encoded the NH₂-terminal amino acid sequence of LEP100. Further DNA sequencing was done on subcloned fragments and on fragments truncated by timed digestions with exonuclease III (8). We were unsuccessful in many attempts to clone the larger Eco RI fragment into the sequencing vectors pEMBL or Bluescript in an orientation suitable for exonuclease III deletions and sequencing from the 3′ to 5′ end of the encoding DNA. However, we did succeed in subcloning the full-length cDNA into pEMBL8 and 18 in both orientations after partial Eco RI digestion of the original clone in lambda. Five synthetic oligonucleotides were used as sequencing primers to obtain the nucleotide sequence in the 3′ to 5′ direction. Generally two or three sequencing runs were analyzed for each segment. Most of the cDNA, including the entire coding sequence, was determined by sequencing in both directions (Fig. 1).

Hydrophobicity and Helicity Analyses

For analysis of hydrophobicity and alpha-helicity, MacGene Plus software (Applied Genetic Technology, Fairview Park, OH) was used. These programs employ algorithms developed by Kyte and Doolittle (19) and by Garnier et al. (10), respectively.

Determination of the Number of N-linked Oligosaccharide Chains

Chicken fibroblasts were grown to confluency and then transferred to medium containing 60 μg/ml 1-deoxymannojirimycin (3, 9) and lacking methionine. After 2 h [³⁵S]methionine (150 μCi/ml) was added for 6 h. The biosynthetic intermediate of LEP100 with high-mannose N-linked oligosaccharides was isolated by immune precipitation with LEP100-IgG covalently linked to Sepharose 4B beads (21) and subjected to timed digestion with endoglycosidase H (31, 32). The digested samples were analyzed by SDS-PAGE and fluorography, and the number of N-linked oligosaccharide chains was estimated by counting the number of stepwise reductions in apparent molecular weight.

Expression of LEP100 cDNA in Mouse L Cells

An expression plasmid, pSV2CAT (12) with the restriction site 5′ to the SV40 early promotor/enhancer region changed to a Sal I site, was modified as follows. The small Eco RI to Bam HI fragment was removed with destruction of the restriction sites. Then the small Hind III to Eco RI fragment, containing the 5′ end of the CAT gene, was removed, and the polylinker from M13mp8 (25) was ligated into the Hind III–Eco RI sites. The resulting shuttle vector (30) was cut with Eco RI, and the large Eco RI fragment of the cDNA encoding LEP100 was cloned into the Eco RI site. The orientation of the insert cDNA was deduced from electrophoretic analysis of Pst I restriction enzyme digestion fragments. The initial expression clone used in the experiments reported here contained two copies of LEP100 cDNA in tandem.

Mouse Ltk⁻ cells (18) were grown in DME-MEM containing 10% FCS.

Cells were cotransfected by the calcium phosphate precipitate method (13) as described by Small and Scangos (29), with 0.1 μg of pHSVtk5′Δ0.2 DNA (24) and ~1 μg of LEP100 expression plasmid DNA per 100-mm-diameter tissue culture plate. Cultures were switched to HAT medium (23) after 24 h and maintained in this medium. Clones were picked ~2 wk later and maintained in HAT medium.

The number of LEP100 molecules expressed in the transfected cells was estimated by determining the number of binding sites for ¹²⁵I-labeled LEP100-IgG on fixed cells made permeable by treatment with saponin as previously described (21). No specific binding was found unless the cells had been cultured in medium containing 10 mM butyrate (11) to enhance expression off the SV40 early promotor. (LEP100-IgG does not bind to any antigen in nontransfected mouse L cells with or without butyrate treatment.) The distribution of LEP100 molecules in expressing cells was visualized by fluorescence microscopy after labeling fixed, permeabilized cells with FITC-LEP100-IgG. In some experiments cells were double-labeled with FITC-LEP100-IgG and mAb to LAMP1 or LAMP2, mouse lysosomal membrane proteins (6). For double immunofluorescence labeling, conditioned medium containing the anti-LAMP antibody was used first, followed by tetramethylrhodamine-labeled anti–rat IgG and FITC-LEP100-IgG.

Results and Discussion

Identification of the cDNA Encoding LEP100

LEP100 was purified from a detergent extract of chick brain microsomes by immune precipitation followed by SDS-PAGE (see Materials and Methods). The NH₂-terminal amino acid sequence of LEP100, determined by sequential automated Edman degradation and product analysis, was Phe-AspValArgAspSerThrGlyLysValCysIleIleAlaXLeuThrValAla-PheSerValGluTyrLysXXGlyGlnLys. Position 15 was deduced to be Asn on the basis of failure to identify any residue with or without reduction and alkylation of the protein before Edman degradation, and positions 26 and 27 were suspected to be Cys or Ser. (These predictions were confirmed by the nucleotide sequence.) With Asn assigned to position 15, the sequence includes a potential glycosylation signal (AsnLeuThr), and the lack of a signal for Asn₁₅ would suggest that it might be completely in the N-glycosylated form. If this were the case, the NH₂-terminal end of the molecule would be on the lumenal side of the lysosomal membrane.

A set of 26-mer oligonucleotides based on a part of the NH₂-terminal amino acid sequence and containing deoxyinosine at six positions of codon ambiguity was used to screen an unamplified chick brain cDNA library in λgtl0. Two positive clones containing cDNA inserts of ~2.0 and 2.2 kb and each containing an internal Eco RI site were found; the longer clone appeared to include all of the sequence of the shorter clone. The longer cDNA was subcloned into the plasmids pEMBL8 and pEMBL18 (7) as full-length insert in both orientations and as Eco RI fragments. Initial nucleotide sequencing by the dideoxy method revealed that the larger Eco RI fragment contained the nucleotide sequence that encoded precisely the amino acid sequence determined for the NH₂-terminal end of LEP100.

Nucleotide Sequence of the cDNA and Deduced Amino Acid Sequence of LEP100

The nucleotide sequence of the cDNA and the deduced amino acid sequence of LEP100 are shown in Fig. 2. The amino acid sequence begins with an 18-residue signal peptide and is followed by 396 residues, the first 30 of which exactly match the NH₂-terminal amino acid sequence obtained for the mature protein, confirming the identification

```
-466
-405 TGTATTACTAAGCTTTAAACACGATGCCTGGCATATGCAGGGTGTGAACAACGCAATACTCAACCCTACAGGACTCTACGATTGGACAGCTCCCCTGTTTGATATTTTGATGCTCTGGCAGCTCGTTAGGAGCTG -271
-270 GTACTTTTTTAAAATGGGTGTTATTTTTATTTACATTTTTTTGTAAAGTGATTTCAGTCTCTTCTGTTGGAGGTCCAGAGAGATCCTGTTACTGCACATGTGTACAATATAGGTCTCAATCTGCTGATCTCTGCT -136
-135 TAACTGGCTGGGGGTTGTACACTTAAGGATCAGTTCTTATGTATGCATTGCCTGTAACAATGCTAATAAAGACACTTTTTTTTCTGTAAAAAAAAGGAATTCCCTTCTGCCGCCGCCTCGCACCTCCTGCCCCGCC -1
```

```
  1 ATG GGC GGC GCT GCT CGG GCC GTC CTG CTC GGC TTT TTA CAG GCT TCC TCT TCA TTT GAC GTG AGA GAT TCG ACA GGC AAA GTC TGC ATA ATT GCT AAC CTG 102
    Met Gly Gly Ala Ala Arg Ala Val Leu Leu Gly Phe Leu Gln Ala Ser Ser Ser|Phe Asp Val Arg Asp Ser Thr Gly Lys Val Cys Ile Ile Ala Asn Leu
                                                                                                                          △

103 ACA GTA GCC TTC TCA GTG GAA TAC AAA AGC AGT GGG CAA AAG CAG TTT GCA CAC TTC TTT CTT CCC CAA AAT GCC ACA TCA CAA TCA CAC AGC TCG TGT GGT 204
    Thr Val Ala Phe Ser Val Glu Tyr Lys Ser Ser Gly Gln Lys Gln Phe Ala His Phe Phe Leu Pro Gln Asn Ala Thr Ser Gln Ser His Ser Ser Cys Gly
                                                                                              △

205 GAA GGT AAC ACA TCT CAT CCA ATT CTA GCA TTG AGT TTT GGA GCA GGA CAC TTG ATA AGT CTG AAT TTC TCA AAA ACT CTG GAT AAG TAC CAA GTT GAA GAG 306
    Glu Gly Asn Thr Ser His Pro Ile Leu Ala Leu Ser Phe Gly Ala Gly His Leu Ile Ser Leu Asn Phe Ser Lys Thr Leu Asp Lys Tyr Gln Val Glu Glu
             △

307 CTA ACT TTC CAC TAC AAT CTG TCT GAT GAA ACT TTA TTC CCT AAT GCA ACC GAA GGA AAA GTG ATG GTA GCG ACA CAA AAA TCT GTT ATT CAG GCA CGT ATT 408
    Leu Thr Phe His Tyr Asn Leu Ser Asp Glu Thr Leu Phe Pro Asn Ala Thr Glu Gly Lys Val Met Val Ala Thr Gln Lys Ser Val Ile Gln Ala Arg Ile
                      △                                      △

409 GGC ACA GAG TAC AGA TGC ATC AAC TCC AAG TAC GTC CGT ATG AAG CAT GTA AAC ATT ACT TTC AGC AAC GTG ACT TTG GAA GCC TAT CCC ACA AAT GAC ACT 510
    Gly Thr Glu Tyr Arg Cys Ile Asn Ser Lys Tyr Val Arg Met Lys His Val Asn Ile Thr Phe Ser Asn Val Thr Leu Glu Ala Tyr Pro Thr Asn Asp Thr
                                                              △                △                                            △

511 TTC AGT GCA AAC AAG ACA GAA TGT AGG GAA GAT ATG GTC TCT ACT ACC ACT GTA GCA CCA ACA ACT CCA AAA CAC GCT ACT AGC CAG GTT CCA ACA ACT AGC 612
    Phe Ser Ala Asn Lys Thr Glu Cys Arg Glu Asp Met Val Ser Thr Thr Thr Val Ala Pro Thr Thr Pro Lys His Ala Thr Ser Gln Val Pro Thr Thr Ser
                 △                                                         △         △

613 CCA GCA CCA ACT GCA GCA CCC TCA AGC CCC GCA GTT GGT AAA TAC AAC GTG ACT GGT GCA AAT GGA ACC TGT GTC CTT GCC TCC ATG GGG TTA CAG CTT AAC 714
    Pro Ala Pro Thr Ala Ala Pro Ser Ser Pro Ala Val Gly Lys Tyr Asn Val Thr Gly Ala Asn Gly Thr Cys Val Leu Ala Ser Met Gly Leu Gln Leu Asn
                                                       △                     △

715 ATC ACC TAT GTA AAA AAA GAT GAA AAG ATG GGG TTA GAT CTG CTG AAT TTC ATA CCA CAT AAT ACA AGT GCT TCT GGA ATG TGT GAG AGT ACA TCT GCC TTC 816
    Ile Thr Tyr Val Lys Lys Asp Glu Lys Met Gly Leu Asp Leu Leu Asn Phe Ile Pro His Asn Thr Ser Ala Ser Gly Met Cys Glu Ser Thr Ser Ala Phe
                                                                           △

817 TTG AAT CTG GCT TTT GAG AAA ACA AAG ATC ACC TTC CAC TTT GTA TTG AAT GCA AGC AGT GAG AAA TTC TTC CTG CAA GGT GTG AAT GTC AGC ACA ACT CTG 918
    Leu Asn Leu Ala Phe Glu Lys Thr Lys Ile Thr Phe His Phe Val Leu Asn Ala Ser Ser Glu Lys Phe Phe Leu Gln Gly Val Asn Val Ser Thr Thr Leu
                                                                        △                              △

919 CCT TCT GAA GCA AAA GCT CCA ACG TTT GAA GCA TCG AAT GAC AGC ATG AGT GAA TCG AGA GCT ACG GTG GGG AAC TCC TAC AAG TGC AGC GCA GAG GAG AAT 1020
    Pro Ser Glu Ala Lys Ala Pro Thr Phe Glu Ala Ser Asn Asp Ser Met Ser Glu Ser Arg Ala Thr Val Gly Asn Ser Tyr Lys Cys Ser Ala Glu Glu Asn
                                                       △

1021 TTC CAG GTC ACA GAC AAA GCC CTT GTC AAT GTA TTT AAT GTC CAG GTC CAG GCT TTC AAG GTT GAT GGA GAC AAG TTT GGG GCA ATG GAA GAA TGT CAG CTG 1122
     Phe Gln Val Thr Asp Lys Ala Leu Val Asn Val Phe Asn Val Gln Val Gln Ala Phe Lys Val Asp Gly Asp Lys Phe Gly Ala Met Glu Glu Cys Gln Leu

1123 GAT GAA AAC AAC ATG CTG ATC CCT ATA ATA GTT GGT GCA GCC CTT GCC GGT CTG GTT CTG ATT GTC CTG ATC GCT TAC TTG ATT GGC AGA AAG AGG AGC CAC 1224
     Asp Glu Asn Asn Met Leu Ile Pro Ile Ile Val Gly Ala Ala Leu Ala Gly Leu Val Leu Ile Val Leu Ile Ala Tyr Leu Ile Gly Arg Lys Arg Ser His
                            M 1

1225 GCT GGG TAC CAA ACC ATT TAA TGACTTGCACTAAATTCCAGTGTAGATGAAAAGCAATTGCAAGAAGCTGGAAAAACAAACAAACAAACCACCCCTATCTCTCCCCGAGCTAAGATTGATATCGTAAG 1352
     Ala Gly Tyr Gln Thr Ile ***

1353 GGAACACCTTTTCTTGCAAATTGCTTCTAAAGGTCTGCTTTATTAAATGTGAAATCATCTTGCAGCATTTAACTATGCACAAAAATAACTTCTGAAAATCCTGGTGTTTAATTTTGCTAACTAGTTGAAATATTT 1487
1488 ACCCACTTAGAAACAAGCTAAAAGTGTTTGGTTTTTTTTTTTTTTTAAGGATGATTTTTCTTTTTTGAATTGGCATATGGTAAAGTCAGCATAAAGATGCTAATTT.........3' 1592
```
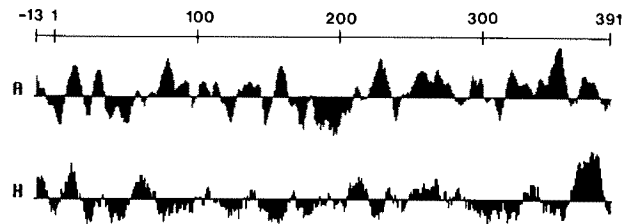
Figure 2. Nucleotide sequence and deduced amino acid sequence of a chick brain cDNA encoding LEP100. Nucleotide residues are numbered in the 5'-to-3' direction: the first residue of the translation initiation triplet ATG is given the number 1, and nucleotides 5' to this residue are indicated by negative numbers. The sequence of only a portion of the 3' untranslated region of the cDNA is reported; the unreported segment 3' to that shown in the figure is less than 100 bases long and does not contain a poly-A tail. (Genomic cloning reveals, in fact, that the Eco RI site at the 5' end of the cDNA is a natural site in the genomic sequence rather than due to Eco RI linker addition in construction of cDNA libraries [unpublished observations].) The 17 glycosylation sites are marked by triangles, cysteine residues are indicated by ovals, and potential phosphorylation sites by rectangles. The signal peptide is boxed, and the putative membrane-spanning segment is indicated by *M1*.

of Asn at position 15 and Ser at positions 26 and 27. Note that whereas nucleotides are numbered from the translation start codon, the aminoacyl residues are numbered from the phenylalanine that is the NH_2-terminal residue of the mature protein. A search for homologous DNA and protein sequences in the Los Alamos National Laboratory GenBank failed to uncover any significant homologies.

The signal peptide displays all of the features normally associated with leader peptides of eukaryotes (5). There is a single Arg in the NH_2-terminal region, a seven–amino acid "hydrophobic region," and the last five residues constitute a more hydrophilic region with Ser at positions −3, −2, and −1, fitting the "−3,−1 rule."

Pattern of N-Glycosylations

As shown in Fig. 2, there are 17 potential N-glycosylation sites in the LEP100 molecule. An estimate of the actual number of N-linked oligosaccharide chains on the LEP100 molecule was made from analysis of the products of partial deglycosylation by endoglycosidase H. For this determination, a high-mannose intermediate form of LEP100 labeled with [35S]methionine was isolated from chick fibroblasts and subjected to timed digestion by endoglycosidase H (see Materials and Methods). The digested samples were ana-

lyzed by SDS-PAGE and fluorography, and the number of high-mannose N-linked oligosaccharide chains was estimated by counting the number of stepwise reductions in apparent molecular weight of the LEP100 molecule as individual oligosaccharide chains were removed (Fig. 3). The entire experiment was done three times. Although at the high molecular weight end of this digestion series the digestion products were never well resolved, presumably due to heterogeneity in the size of the high-mannose N-linked oligosaccharides, nevertheless, it appears that there are approximately 17 steps in the series. In the absence of endoglycosidase H digestion, the high-mannose intermediate form of LEP100 always appeared predominantly as a single band on fluorographs of SDS-PAGE, suggesting little heterogeneity in total glycosylations per molecule. These results are consistent with the deduced amino acid sequence and with the hypothesis that each of the 17 possible sites for N-glycosylation is indeed usually glycosylated.

Other Features of the Primary Structure of LEP100

In membrane and secretory proteins with multiple cysteine residues there are usually intrachain disulfide bridges. Eight cysteine residues occur in the LEP100 sequence, distributed in a remarkably regular pattern in which pairs of cysteines

Figure 3. Estimate of number of N-linked oligosaccharides. A high-mannose form of LEP100 labeled with [^{35}S]methionine was isolated from cultured chick fibroblasts and subjected to timed digestion with endoglycosidase H. Digestion conditions were 34 mU/ml endoglycosidase H in 0.1% SDS, 2% 2-mercaptoethanol, 60 mM sodium acetate, pH 5.8, at 37°C. At intervals up to 8 h, aliquots were removed and incubated for 2 min in SDS sample buffer at 100°C to stop the enzymatic reaction. Enzyme-treated samples were analyzed by SDS-PAGE and fluorography. In the graph, the position of each discernable band in the digestion series relative to the position of the undigested form is plotted against band number. The resulting curve is extrapolated to the position of the undigested band, yielding an estimate of N-linked oligosaccharides. The inset photograph shows a fluorograph of part of one digestion series, with positions of countable bands and extrapolated positions marked. In this example, there is some heterogeneity in apparent mobility of the undigested LEP100.

Figure 4. Alpha-helicity and hydrophobicity plots of LEP100. The deduced amino acid sequence of LEP100 was analyzed with Mac-Gene Plus software that employs the algorithms of Garnier et al. (10) and Kyte and Doolittle (19). Calculations were made without consideration of the numerous N-linked oligosaccharides. (*A*) Alpha-helicity plot. (*H*) Hydrophobicity plot. In these plots the probability of alpha-helical structure and the hydrophobicity indices are plotted for the center position of a window of 11 amino acids that is shifted in single amino acid steps through the sequence. Thus the first position that can be plotted is the sixth aminoacyl residue of the signal peptide (−13 in the numbering system that assigns 1 to the NH$_2$-terminal phenylalanine of the mature protein), and there is no calculation for the last five residues at the COOH-terminus.

have intrapair spacings of 38, 36, 38, and 37 residues, respectively, whereas interpair spacings are 75, 50, and 69. We suggest that pairs of adjacent cysteine residues may form disulfide bridges that participate in stabilization of a set of four hydrophilic domains. In this hypothetical model of higher-order structure the first three such domains are glycosylated whereas the fourth, closest to the membrane-spanning segment, is unique in lacking N-linked oligosaccharides.

In the center of the amino acid sequence there is a 32-residue segment rich in Ser and Thr and containing seven proline residues (positions 172, 175, 183, 187, 189, 193, and 196). In the helicity plot (10) shown in Fig. 4, this central region shows a markedly low probability of containing alpha-helical structure. This central region of the LEP100 polypeptide separates pairs of the hypothetical domains mentioned above and might be a region of flexibility in the protein.

A hydrophobicity plot (19), based upon calculations that ignore the N-linked oligosaccharides, is shown in Fig. 4. It seems probable that the single very hydrophobic region near the COOH-terminus is a membrane-spanning region. Previously we reported evidence that LEP100 is an integral membrane protein (21). The evidence included (*a*) demonstration that LEP100 is a glycoprotein that could only be isolated from membrane fractions by detergent extraction, and (*b*)

demonstration that LEP100 partitions into the detergent phase of a Triton X-114 extract. Because all of the N-glycosyl-ated sites surely occur on the lumenal side of the lysosomal membrane, any transmembrane segment within the glycosyl-ated NH$_2$-terminal 75% of the molecule would have to pass twice through the membrane. No such segment exists. The only candidate for a membrane-spanning segment is the 25-amino acid stretch from positions 361 to 385. Just past this putative membrane-spanning segment there is a set of three positively charged amino acids that presumably constitute a "stop-transfer" sequence (4) that may serve during biosynthesis to prevent extrusion of the nascent polypeptide into the lumen of the endoplasmic reticulum and thenceforth to stabilize LEP100 in the lipid bilayer. Given that LEP100 is an integral membrane protein and that hydrophobic segment 361–385 is the only candidate for a membrane-spanning segment, we conclude that this segment is a bona fide membrane-spanning one.

There is a very short amino acid sequence from the membrane-spanning segment to the COOH-terminus, constituting the cytoplasmic domain. We had expected a larger cytoplasmic domain for the LEP100 molecule because we postulate that such a domain might be involved in membrane–cytosol interactions and/or membrane–membrane recognition events. The codistribution of membrane-associated clathrin and LEP100 in chloroquine-treated cells (22) suggested the possibility that LEP100 might interact directly with clathrin or with some cytosolic component that links LEP100 and clathrin. This possibility is not excluded by our findings, although we feel that the lack of a bulkier protein mass on the cytosolic face diminishes the likelihood of this possibility being correct. The cytosolic domain of LEP100 contains three potential phosphorylation sites: one serine, one threonine, and one tyrosine. We reported previously (21) that we could not detect any phosphorylation of LEP100 isolated from tissue cultured fibroblasts. In those cells ~98% of LEP100 molecules occur in lysosomal and endosomal membranes. It may be that the state of phosphorylation is related to the distribution of LEP100 in the cells. If this were

NH₂

COOH

Figure 5. Cartoon depicting some aspects of LEP100 structure. Each circle represents an amino acid in the mature LEP100 molecule. NH₂- and COOH-terminal ends are indicated, and the membrane-spanning segment is represented as a projected helix. The proline-rich central segment is kinked to indicate its probable nonhelical, non-beta-sheet structure. Symbols representing N-linked oligosaccharide chains are attached to the asparagine residues of the glycosylation signals.

the case, then LEP100 might be detectably phosphorylated in chloroquine-treated cells where ~30% of LEP100 molecules are in the plasma membrane subtended by patches of clathrin (22).

Fig. 5 is a cartoon that incorporates many features of LEP100 and also some of the speculations about its higher-order structure mentioned previously.

Expression of LEP100 in Mouse L Cells

The large Eco RI fragment of LEP100 cDNA was cloned into an expression vector derived from pSV2CAT (12, 30) and cotransfected into mouse Ltk⁻ cells together with a plasmid

containing the herpes virus tk gene (24). No expression was detected transiently, but many tk⁺ clones were isolated in HAT medium. Expression of LEP100 could be activated in some of these clones by maintaining them in medium containing 10 mM butyrate (11). Under these conditions expression approached maximal levels after ~48 h, at which time expression was comparable with that in chick fibroblasts, as judged by the binding of ¹²⁵I-labeled LEP100-IgG to fixed, permeabilized cells (data not shown).

When expressed in the mouse L cells, LEP100 is correctly targeted to lysosomes (Fig. 6). This was confirmed by double-label immunocytochemistry. Fixed, permeabilized

Figure 6. Stable expression of LEP100 in mouse L cells. Mouse Ltk⁻ cells were cotransfected with an expression plasmid carrying the cDNA encoding LEP100 together with a plasmid containing the encoding DNA for herpes virus thymidine kinase. Cells were selected in HAT medium and clones expressing LEP100 were identified after induction of LEP100 expression by 10 mM sodium butyrate. Cells growing on cover slips were fixed for 10 min with 1% formaldehyde fixative and then double-labeled with anti-LAMP1 mAb and LEP100-IgG for fluorescence microscopy as described in Materials and Methods. Phase (*A*) and fluorescence (*B* and *C*) micrographs of the same cell illustrate the distribution of mouse lysosomes defined by LAMP1 (*B*) and the distribution of LEP100 (*C*). Bar in C, 10 μm.

cells of the transfected mouse L cell lines were labeled with mAb against the lysosomal membrane proteins LAMP1 and LAMP2 (6) followed by tetramethylrhodamine-labeled goat anti–rat IgG and FITC-LEP100-IgG. Codistribution of intracellular fluorescence from the anti–LAMP antibodies and LEP100-IgG was found. Furthermore, in the presence of 100 μM chloroquine, LEP100 redistributed until about half of the molecules were in the mouse L cell plasma membrane (unpublished observations of Paul Mathews in our laboratory). Thus, the expression experiments verify the identification of the LEP100 cDNA (based initially upon nucleotide sequence match-up with the NH₂-terminal amino acid sequence of LEP100) and provide a basis for study of relationships between LEP100 structure and its targeting to lysosomes and shuttling between membrane systems.

References

1. Barriocanal, J. D., J. S. Bonifacino, L. Yuan, and I. V. Sandoval. 1986. Biosynthesis, glycosylation, movement through the Golgi system, and transport to lysosomes by an N-linked carbohydrate-independent mechanism of three lysosmal integral membrane proteins. *J. Biol. Chem.* 261: 16755–16763.
2. Benton, W. D., and R. W. Davis. 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. *Science (Wash. DC).* 196:180–182.
3. Bischoff, J., L. Liscum, and R. Kornfeld. 1986. The use of 1-deoxymannojirimycin to evaluate the role of various α-mannosidases in oligosaccharide processing in intact cells. *J. Biol. Chem.* 261:4766–4774.
4. Blobel, G. 1980. Intracellular protein topogenesis. *Proc. Natl. Acad. Sci. USA.* 77:1496–1500.
5. Briggs, M. S., and L. M. Gierasch. 1986. Molecular mechanisms of protein secretion: the role of the signal sequence. *Adv. Protein Chem.* 38:109–180.
6. Chen, J. W., T. Murphy, M. Willingham, I. Pastan, and J. T. August. 1985. Identification of two lysosomal membrane glycoproteins. *J. Cell Biol.* 101:85–95.
7. Dente, L., G. Cesareni, and R. Cortese. 1983. pEMBL: a new family of single stranded plasmids. *Nucleic Acids Res.* 11:1645–1655.
8. Donelson, J. E., and R. Wu. 1972. Nucleotide sequence analysis of deoxyribonucleic acid. VII. Characterization of *Escherichia coli* exonuclease III activity for possible use in terminal nucleotide sequence analysis of duplex deoxyribonucleic acid. *J. Biol. Chem.* 247:4661–4668.
9. Fuhrmann, U., E. Bause, G. Legler, and H. Ploegh. 1984. Novel mannosidase inhibitor blocking conversion of high mannose to complex oligosaccharides. *Nature (Lond.).* 307:755–758.
10. Garnier, J., D. J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* 120:97–120.
11. Gorman, C. M., and B. H. Howard. 1983. Expression of recombinant plasmids in mammalian cells is enhanced by sodium butyrate. *Nucleic Acids Res.* 11:7631–7648.
12. Gorman, C. M., L. Moffat, and B. H. Howard. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* 2:1044–1051.
13. Graham, F. L., and A. J. van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology.* 52:456–467.
14. Ho, M.-K., and T. A. Springer. 1983. Tissue distribution, structural characterization, and biosynthesis of MAC-3, a macrophage surface glycoprotein exhibiting molecular weight heterogeneity. *J. Biol. Chem.* 258: 636–642.
15. Hunkapiller, M. W., R. M. Hewick, W. J. Dreyer, and L. E. Hood. 1983. High-sensitivity sequencing with a gas-phase sequenator. *Methods Enzymol.* 91:399–413.

16. Hunkapiller, M. W., E. Lujan, F. Ostrander, and L. E. Hood. 1983. Isolation of microgram quantities of proteins from polyacrylamide gels for amino acid sequence analysis. *Methods Enzymol.* 91:227-236.

17. Huyhn, T. V., R. A. Young, and R. W. Davis. 1985. Constructing and screening cDNA libraries in λgt10 and λgt11. *In* DNA Cloning. A Practical Approach. Vol. 1. D. M. Grover, editor. IRL Press, Oxford. 49-78.

18. Kit, S., D. R. Dubbs, L. J. Piekarski, and T. C. Hsu. 1963. Deletion of thymidine kinase activity from L cells resistant to bromodeoxyuridine. *Exp. Cell Res.* 31:297-312.

19. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105-132.

20. Lewis, V., S. A. Green, M. Marsh, P. Vihko, A. Helenius, and I. Mellman. 1985. Glycoproteins of the lysosomal membrane. *J. Cell Biol.* 100:1839-1847.

21. Lippincott-Schwartz, J., and D. M. Fambrough. 1986. Lysosomal membrane dynamics: structure and interorganellar movement of a major lysosomal membrane glycoprotein. *J. Cell Biol.* 102:1593-1605.

22. Lippincott-Schwartz, J., and D. M. Fambrough. 1987. Cycling of the integral membrane glycoprotein, LEP100, between plasma membrane and lysosomes: kinetic and morphological analysis. *Cell.* 49:669-677.

23. Littlefield, J. W. 1964. Selection of hybrids from matings of fibroblasts in vitro and their presumed recombinants. *Science (Wash. DC).* 145:709-710.

24. McKnight, S. L. 1980. The nucleotide sequence and transcription map of the herpes simplex virus thymidine kinase gene. *Nucleic Acids Res.* 8:5949-5964.

25. Messing, J. 1983. New M13 vectors for cloning. *Methods Enzymol.* 101: 20-78.

26. Ohtsuka, E., S. Matsuki, M. Ikehara, Y. Takahashi, and K. Matsubara. 1985. An alternative approach to deoxyoligonucleotides as hybridization probes by insertion of deoxyinosine at ambiguous codon positions. *J. Biol. Chem.* 260:2605-2608.

27. Reggio, H., D. Bainton, E. Harms, E. Coudrier, and D. Louvard. 1984. Antibodies against lysosomal membranes reveal a 100,000 molecular weight protein that cross-reacts with purified H^+,K^+ ATPase from gastric mucosa. *J. Cell Biol.* 99:1511-1526.

28. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463-5467.

29. Small, J., and G. Scangos. 1984. Expression of cotransferred genes in mouse L cells. *Gene Anal. Tech.* 1:13-17.

30. Takeyasu, K., M. M. Tamkun, N. R. Siegel, and D. M. Fambrough. 1987. Expression of hybrid $(Na^+ + K^+)$-ATPase molecules after transfection of mouse Ltk⁻ cells with DNA encoding the β-subunit of an avian brain sodium pump. *J. Biol. Chem.* 262:10733-10740.

31. Tamkun, M. M., and D. M. Fambrough. 1986. The $(Na^+ + K^+)$-ATPase of chick sensory neurons: studies on biosynthesis and intracellular transport. *J. Biol. Chem.* 261:1009-1019.

32. Tarentino, A. L., and F. Maley. 1974. Purification and properties of an endo-β-*N*-acetylglucosaminidase from *Streptomyces griseus. J. Biol. Chem.* 249:811-816.