

# Lipase Maturation Factor LMF1, Membrane Topology and Interaction with Lipase Proteins in the Endoplasmic Reticulum<sup>\*[5]</sup>

Received for publication, July 27, 2009, and in revised form, September 22, 2009. Published, JBC Papers in Press, September 26, 2009, DOI 10.1074/jbc.M109.049395

Mark H. Doolittle<sup>‡§1</sup>, Saskia B. Neher<sup>¶2,3</sup>, Osnat Ben-Zeev<sup>‡§3</sup>, Jo Ling-liao<sup>§||</sup>, Ciara M. Gallagher<sup>¶4</sup>,  
Maryam Hosseini<sup>‡§||</sup>, Fen Yin<sup>‡§||</sup>, Howard Wong<sup>‡§</sup>, Peter Walter<sup>¶\*\*,\*</sup>, and Miklós Péterfy<sup>‡§||5</sup>

From the <sup>‡</sup>Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, California 90095, the <sup>§</sup>Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, California 90073, the <sup>¶</sup>Department of Biochemistry and Biophysics and the <sup>\*\*</sup>Howard Hughes Medical Institute, University of California, San Francisco, California 94158, and the <sup>||</sup>Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California 90048

Lipase maturation factor 1 (LMF1) is predicted to be a polytopic protein localized to the endoplasmic reticulum (ER) membrane. It functions in the post-translational attainment of enzyme activity for both lipoprotein lipase and hepatic lipase. By using transmembrane prediction methods in mouse and human orthologs, models of LMF1 topology were constructed and tested experimentally. Employing a tagging strategy that used insertion of ectopic glycan attachment sites and terminal fusions of green fluorescent protein, we established a five-transmembrane model, thus dividing LMF1 into six domains. Three domains were found to face the cytoplasm (the amino-terminal domain and loops B and D), and the other half was oriented to the ER lumen (loops A and C and the carboxyl-terminal domain). This representative model shows the arrangement of an evolutionarily conserved domain within LMF1 (DUF1222) that is essential to lipase maturation. DUF1222 comprises four of the six domains, with the two largest ones facing the ER lumen. We showed for the first time, using several naturally occurring variants featuring DUF1222 truncations, that *Lmf1* interacts physically with lipoprotein lipase and hepatic lipase and localizes the lipase interaction site to loop C within DUF1222. We discuss the implication of our results with regard to lipase maturation and DUF1222 domain structure.

Lipoprotein lipase (LPL)<sup>6</sup> and hepatic lipase (HL) are secreted glycoproteins that hydrolyze triglycerides sequestered in

the core of circulating lipoproteins. As such, they play important roles in lipoprotein remodeling and uptake. LPL also mediates tissue influx of fatty acids derived from triglyceride-rich lipoproteins, and both lipases have been implicated in atherosclerosis and inflammation (1–4). These functions require nascent lipase polypeptides to fold and assemble into native structures, a necessary prerequisite for enzyme activity and secretion. However, both lipases require a *trans*-acting factor to attain a functional state, as exemplified by a naturally occurring mutation in the mouse called combined lipase deficiency (*clد*). In tissues and cells homozygous for the *clد* mutation, intracellular levels of LPL mRNA and protein are normal; however, the vast majority of LPL protein (~95%) remains within the endoplasmic reticulum (ER) as misfolded (inactive) mass that is eventually degraded (5–8). HL activity is also diminished in *clد/clد* cells but to a lesser extent than LPL (9). In affected newborn mice, the resulting combined lipase deficiency causes massive chylomicronemia and neonatal death because of tissue ischemia and the absence of triglyceride-derived fatty acid influx (10). Indeed, this phenotype mirrors an LPL knock-out model (11), indicating that *clد* causes virtual abolishment of LPL function. The identity of this critical lipase maturation factor was made clear when a gene encoding a predicted transmembrane protein on mouse chromosome 17 (*Tmem112*) was found to harbor the *clد* mutation (9). The gene was renamed lipase maturation factor 1 (*Lmf1*) to reflect its role in aiding the post-translational development of enzyme function from nascent lipase polypeptides (9).

The pattern of LPL misfolding and ER retention caused by the *clد* mutation is indistinguishable from LPL that is prevented from interacting with calnexin, a well-known ER chaperone (6). However, unlike calnexin, LMF1 is not a general chaperone involved in the folding of many ER glycoproteins. Rather, it appears more client-specific, perhaps interacting with a small subset of proteins that share structural characteristics with LPL and HL. Its subcellular location seems ideally suited for such a role; *Lmf1* is a type III (multi-

fluorescence protease protection; GFP, green fluorescent protein; HEK, human embryonic kidney; HL, hepatic lipase; LMF1, lipase maturation factor 1; PL, pancreatic lipase; RT, reverse transcription; SEAP, secreted alkaline phosphatase; TAP, tandem affinity purification; TM, transmembrane; TRITC, tetramethylrhodamine isothiocyanate; YFP, yellow fluorescent protein.

\* This work was supported, in whole or in part, by National Institutes of Health Grant HL24841. This work was also supported by the Cedars-Sinai Medical Center and the United States Department of Veterans Affairs.

[5] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. S1–S4 and Table S1.

<sup>1</sup> To whom correspondence may be addressed: Veterans Affairs Greater Los Angeles Healthcare System, 11301 Wilshire Blvd., Bldg. 113, Rm. 312, Los Angeles, CA 90073. Fax: 310-268-4981; E-mail: markdool@ucla.edu.

<sup>2</sup> Supported by the Jane Coffin Childs Memorial Fund.

<sup>3</sup> Both authors contributed equally to this work.

<sup>4</sup> Supported by a long-term fellowship from the European Molecular Biology Organization.

<sup>5</sup> To whom correspondence may be addressed: Veterans Affairs Greater Los Angeles Healthcare System, 11301 Wilshire Blvd., Bldg. 113, Rm. 312, Los Angeles, CA 90073. Fax: 310-268-4981; E-mail: mpeterfy@ucla.edu.

<sup>6</sup> The abbreviations used are: LPL, lipoprotein lipase; aa, amino acid(s); CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid; DUF, domain of unknown function; eCFP, enhanced cyan fluorescent protein; endo H, endoglycosidase H; ER, endoplasmic reticulum; FPP,

## LMF1 Topology and Lipase Interaction

pass) membrane protein located to the ER (9), the site of lipase folding and assembly (12, 13). Dependent on the position and number of transmembrane (TM) domains, LMF1 is a polytopic protein divided into multiple domains that are oriented to both the cytoplasm and the ER lumen. Such a topological arrangement would place domains within the ER lumen in proximity to lipase nascent chains as they emerge from the translocon during protein synthesis.

LMF1 is also distinguished by an evolutionarily conserved domain of unknown function referred to as DUF1222 in the Pfam data base (14); this domain is found in more than 50 proteins covering a wide taxonomic range. The *clد* mutation causes a large carboxyl-terminal deletion of DUF1222 in mouse *Lmf1*, as does a nonsense mutation identified in the human homolog (Y439X). Like *clد* in mouse, Y439X causes combined lipase deficiency and hypertriglyceridemia in humans (9). Thus, both *clد* and Y439X represent loss-of-function mutations and implicate DUF1222 in lipase maturation.

In this study, we show that mouse and human LMF1 contain five TM segments that divide the protein into six separate domains; the cytoplasmic or ER luminal orientation of each domain has been determined experimentally. Notably, DUF1222 comprises four of the six domains, with the two largest oriented to the ER lumen. We provide evidence, for the first time, of the physical interaction of *Lmf1* with LPL and HL. Using *Lmf1* variants partially or completely lacking DUF1222, we narrowed the lipase interaction site to a single domain of DUF1222 oriented to the ER lumen. We discuss the implications of LMF1 topology with regard to its function as a lipase maturation factor, including the nature of the DUF1222 domain structure.

### EXPERIMENTAL PROCEDURES

**Cell Lines and Transfection**—The *clد* mutant cell line has been described previously (15); it represents a hepatocyte-derived cell line derived from 16–18-day fetal *clد/clد* mice immortalized by the SV40 T antigen. Both *clد* mutant and HEK293 cell lines were maintained at a split ratio of 1:20 in Dulbecco's modified Eagle's medium, 10% fetal bovine serum. Transfection of the *clد* mutant cell line used Effectene® transfection reagent (Qiagen) at a DNA to reagent ratio of 1:10, and HEK293 cells were transfected with FuGENE 6® transfection reagent (Roche Applied Science) according to manufacturer's instructions. Transfection was initiated 24 h after plating, and cells were harvested 24–48 h post-transfection.

A fluorescence protease protection (FPP) assay was performed on HeLa cells grown in 6-well plates in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and 1% L-glutamine. Cells were transfected with Lipofectamine 2000 (Invitrogen). 30 h post-transfection, cells were trypsinized and plated at 20% confluency in 35-mm poly-D-lysine-coated glass-bottomed plates (MatTek) for FPP analysis.

**RT-PCR and Site-directed Mutagenesis**—RT-PCR was performed as described (9, 16) using RNA obtained from the liver, muscle, adipose tissue, kidney, brain, and testis of non-*clد* (wild type) animals. To eliminate amplification from genomic DNA templates, RT-PCR was performed on cDNA using an exon-spanning primer in each primer set (Fig. 5A; primer A). The

sequences of primers A, B, and C used for RT-PCR amplification of the 1-2-3-3A, 1-2-3\*-3A and 1-2-3\* splice variants of *Lmf1* are available on request, as are the mutagenesis primers for *Lmf1*<sup>N396A</sup>, *Lmf1*<sup>N430A</sup>, *Lmf1*<sup>InsNA243</sup> and the insertion of all additional ectopic glycan attachment sites (see below). Mutagenesis was performed using the QuikChange® site-directed mutagenesis kit (Stratagene) according to manufacturer's instructions.

**Expression Constructs and *Lmf1* Variant Sequences**—Human LPL, HL, and pancreatic lipase (PL) cDNAs were subcloned into the pcDNA6 expression vector (Invitrogen) containing a carboxyl-terminal V5 epitope tag as described (12, 17). For experiments using lipase affinity purification, a tandem affinity purification (TAP) tag was synthesized for in-frame integration into an AgeI site occurring just after the V5 epitope tag of pcDNA6 (18). After transfection, the resulting expressed LPL, HL, and PL proteins contained a carboxyl-terminal V5-TAP tag consisting of the V5 epitope followed by a single calmodulin-binding peptide domain and a tobacco etch virus (TEV) protease site and ending in two adjacent IgG-binding domains derived from protein A (18).

All mouse and human *LMF1* cDNA sequences, except for green fluorescent protein (GFP) fusions, were subcloned into the pcDNA3.1 expression vector (Invitrogen) containing an amino-terminal c-Myc epitope tag (9). The following mouse *Lmf1* sequences were used: *Lmf1*<sup>wt</sup>, encoding the full-length protein (574 amino acids); *Lmf1*<sup>clد</sup>, truncating *Lmf1* at position 361 with the added peptide sequence GKPATQQLPTTPAP at its carboxyl terminus (9); *Lmf1* lacking DUF1222, truncating *Lmf1* at position 171 (*Lmf1*<sup>ΔDUF</sup>) with the added peptide sequence GAQGAITSYPQSRTERE at its carboxyl terminus; *Lmf1*<sup>N396A</sup> and *Lmf1*<sup>N430A</sup>, removing N-linked glycan attachment sites at positions 396 and 430, respectively; *Lmf1*<sup>InsNA243</sup>, with an insertion of NA at position 243 of *Lmf1*<sup>N430A</sup> to complete an ectopic <sup>243</sup>NAT<sup>245</sup> glycan attachment site. Additional ectopic glycan attachment sites were inserted at the following positions of mouse *Lmf1*<sup>N430A</sup>: 28, 94, 101, 175, 188, and 350. The following human *LMF1* sequences were used: *LMF1*<sup>wt</sup>, encoding the full-length protein (567 amino acids); and *LMF1*<sup>N428A</sup>, removing the N-linked glycan attachment site at position 428.

For GFP fusion constructs, human *LMF1* was PCR-amplified from human cDNA clone 100062174 (Open Biosystems) and inserted into pcDNA3.1/CT-GFP-TOPO (Invitrogen) and pcDNA3.1/NT-GFP-TOPO (Invitrogen) according to manufacturer's instructions. Additionally, human *LMF1* cDNA was truncated at positions in the protein corresponding to amino acid 95, 120, 200, 250, and 350 and then inserted into pcDNA3.1/CT-GFP-TOPO (Invitrogen). Control plasmids expressing fluorescent proteins located on the luminal or cytoplasmic side of the ER membrane (eCFP-CD3δ and CD3δ-YFP) were a kind gift from J. Lippincott-Schwartz.

To normalize for transfection efficiency in the LMF1 functional assay, a secreted human placental alkaline-phosphatase (SEAP) reporter gene in the pM1 expression vector was used (X-extremeGENE®, Roche Applied Science). All expression vector plasmids were prepared using the EndoFree maxi kit (Qiagen) according to the manufacturer's instructions. Diluted

plasmid solutions were quantitated using a NanoDrop 2000 spectrophotometer (Thermo Scientific).

**LMF1 Functional Assay**—Details of the Lmf1 functional assay are reported elsewhere (19). The *cld* mutant cell line was co-transfected with an LMF1 test sequence along with a mixture of LPL and SEAP. At 24 h post-transfection, a sample of medium was taken for SEAP activity measurements, and heparin was added to a final concentration of 10 units/ml. Two h later, cells were washed with phosphate-buffered saline and lysed in detergent-containing buffer (0.2% sodium deoxycholate, 10 units/ml heparin, 10 mM Tris-HCl, pH 7.5). After sonication and centrifugation, supernatants were used for the measurement of LPL activity and detection of LMF1 protein levels by Western blot analysis (see below).

SEAP activity was measured using the SEAP reporter assay kit (InvivoGen) according to manufacturer's instructions. LPL activity was measured using a lecithin-stabilized radiolabeled triolein (glycerol tri[9,10(*n*)-3H]oleate) substrate as described (20). Levels of LPL activity and LMF1 protein were normalized to SEAP activity to account for assay-to-assay variations in transfection efficiency.

**Lipase Affinity Purification**—The details of tandem affinity purification are described elsewhere (18, 21). In experiments using lipase-TAP constructs (Fig. 7), the calmodulin affinity step was used alone. Briefly, HEK293 cells transfected with lipase-TAP constructs were lysed in 2% CHAPS, 0.15 M NaCl, 10 mM Tris-HCl, pH 8.0; they were then applied to calmodulin affinity resin (Stratagene) and incubated overnight at 4 °C with constant mixing. Bound lipase-TAP proteins were eluted by Ca<sup>2+</sup> chelation using 0.5% CHAPS, 0.15 M NaCl, 10 mM 2-mercaptoethanol, 1 mM Mg acetate, 1 mM imidazole, 20 mM EGTA, 10 mM Tris-HCl, pH 8.0. SDS and glycerol were added to the eluates at final concentrations of 2 and 10%, respectively, and the eluates were then subjected to Western blot analysis for detection of lipase and LMF1 proteins using the V5 and c-Myc epitope tags (see below).

**Western Blot Analysis, Immunocytochemistry, and FPP Assay**—SDS-PAGE of cell lysates was performed using 7 or 8% Tris acetate or Tris-glycine preformed gels (Invitrogen) as described (22). Proteins were transferred onto polyvinylidene difluoride paper and blocked using SuperBlock (Thermo Scientific) or 5% nonfat milk. The lipase-V5 epitope tag was detected using horseradish peroxidase-conjugated mouse anti-V5 monoclonal antibody (Invitrogen) at a dilution of 1:2,500 in SuperBlock or nonfat milk. The Lmf1-Myc tag was similarly detected using a biotinylated rabbit anti-Myc polyclonal antibody (GenTex) at 1:2,500 followed by horseradish peroxidase-conjugated streptavidin (Invitrogen) at 1:50,000. After washing, bound horseradish peroxidase was visualized using a chemiluminescent substrate (ECL+<sup>®</sup>, GE Healthcare), and images were captured on Hyperfilm ECL<sup>®</sup> (GE Healthcare). Quantitation of protein bands was carried out by densitometric scanning of Western blot films using NIH Image 1.63 software.

For endoglycosidase (endo H) experiments, 20 μg of cell lysates was denatured in 0.5% SDS, 50 mM sodium phosphate buffer, pH 5.9; then 2 μl of *N*-endoglycosidase H (Roche Applied Science) or water (–endo H) was added to the denatured lysates. After overnight incubation at 37 °C, SDS sample

buffer was added to final concentrations of 2% SDS, 2% 2-mercaptoethanol, 10% glycerol, 0.004% bromphenol blue, 10 mM Tris-HCl, pH 8.0, and the mixture was placed in a boiling water bath for 2 min. The samples were subjected to SDS-PAGE and Western blotting as described above.

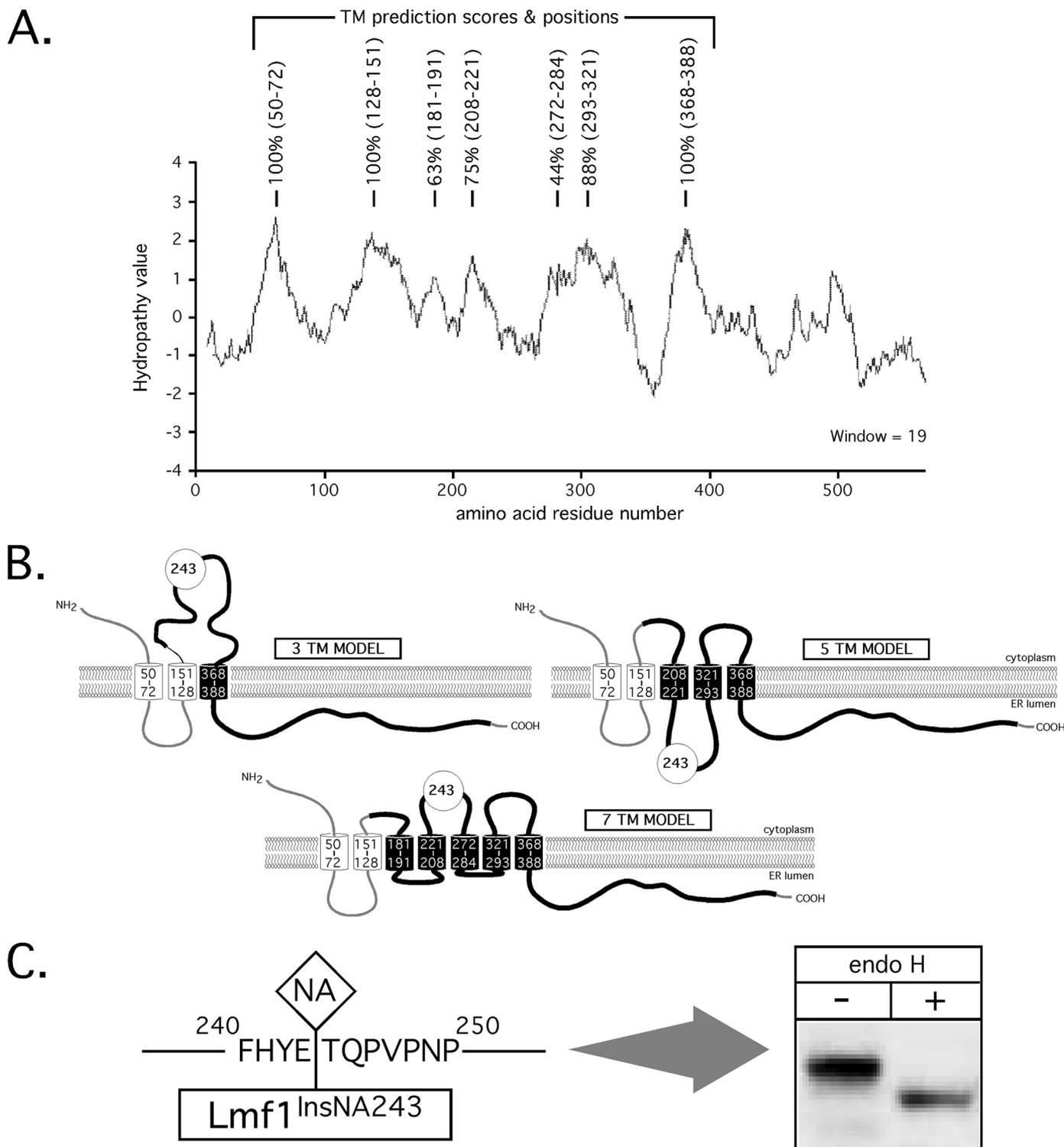
ER localization experiments of Lmf1 protein variants used confocal laser scanning fluorescence microscopy performed on HEK293 cells transfected using the FuGENE 6 transfection reagent (Roche Applied Science) as described (9, 16). Lmf1-Myc was detected with anti-Myc-TRITC antibody (1:100, Santa Cruz Biotechnology), and endogenous calnexin was visualized using rabbit anti-calnexin primary (1:50, Stressgen) and goat anti-rabbit IgG-fluorescein isothiocyanate secondary (1:100, Santa Cruz Biotechnology) antibodies. Images were processed using Adobe Photoshop software (Adobe Systems, Inc.). For the FPP assay, transfected HeLa cells on 35-mm poly-D-lysine-coated glass-bottomed plates (MatTek) were used for visualization of all GFP-LMF1 fusion constructs and of eCFP-CD3δ and CD3δ-YFP in live cells. The FPP assay was performed as described (23) with the following exceptions. Cells were treated with 30 mM digitonin (Calbiochem) for 80 s before treatment with 4 μM trypsin. Live cells were imaged on a Zeiss LSM510 Meta confocal microscope. ImageJ ([rsb.info.nih.gov/ij/](http://rsb.info.nih.gov/ij/)) was used to analyze the integrated density of individual cells across the time course to construct graphs of GFP signal intensity per time after trypsin addition.

**LMF1 Topology Prediction Methods**—LOCATE, a mammalian protein localization data base, used the Membrane Organization Prediction Data (MemO) automated pipeline to assign a total of seven α-helical TM domains for mouse and human LMF1 derived from data supplied by five individual predictors: HMMTOP, TMHMM v2.0, SVM-TM v3.0, MEMSAT, and DAS (24). Predicted α-helical TM sequences for mouse and human LMF1 were also obtained with three additional prediction methods: TMpred, SOUSUI, and PSORTII.

## RESULTS

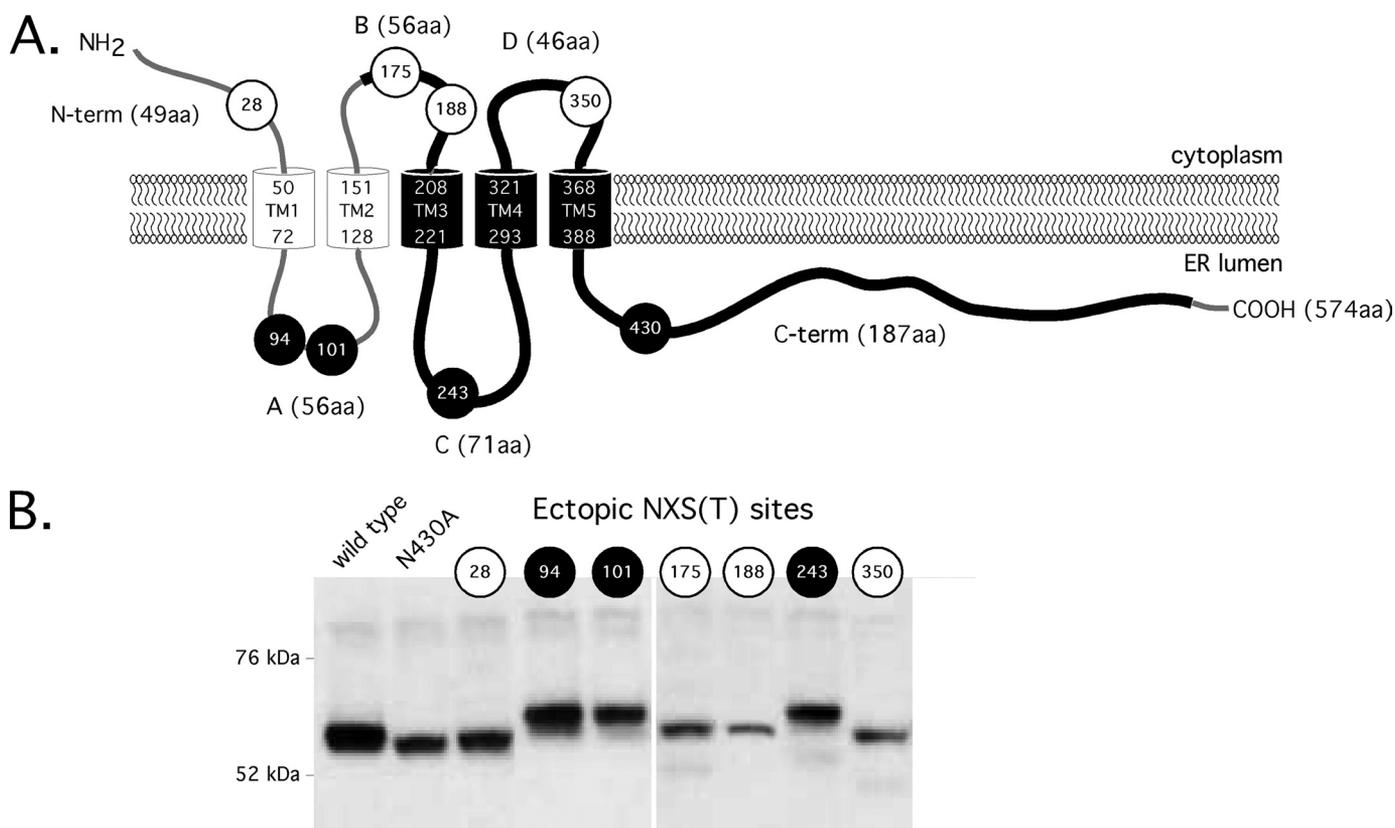
**Membrane Topology of LMF1**—Mouse and human LMF1 are predicted to be polytopic membrane proteins for which the topology in the ER is determined by the position and number of TM domains. The LOCATE data base (24) uses MemO, a high throughput, automated pipeline, to orient a protein with respect to the membrane. α-Helical TM prediction was performed by a MemO consensus method utilizing five publicly available predictors (listed under “Experimental Procedures”). The MemO consensus assigned a total of seven TM domains to both mouse and human LMF1 in the same locations for both orthologs (supplemental Table S1). We added three publicly available prediction methods in addition to the five used by MemO (listed under “Experimental Procedures”) and scored a particular TM domain with regard to how often it was predicted in all eight methods for each LMF1 ortholog. Three of the seven TM domains assigned by MemO scored 100%, indicating that for both mouse and human LMF1 proteins, all eight methods were in complete agreement (supplemental Table S1). Of the remaining four TM domains assigned by MemO, three were in agreement at least 60% of the time (88, 75, and 63%), and the last domain was predicted with a score of 44%.





**FIGURE 2. Lmf1 has five TM domains.** *A*, MemO-predicted TM domains are positioned with respect to a hydropathy plot of mouse Lmf1. The scores refer to the percentage of times that a given TM domain is predicted in both mouse and human LMF1 orthologs across all eight prediction methods (see supplemental Table S1). Hydropathy values  $\geq 1.6$  are consistent with TM domains. *B*, models of Lmf1 membrane topology are illustrated using three, five, or seven TM domains. The three-TM model is built with domains having a top score of 100%; the five-TM model includes two additional domains having the next highest scores (88 and 75%); and the seven-TM model includes all MemO-predicted domains (supplemental Table S1). The position of amino acid 243 is given as a reference in all three models. The cylinders represent the TM domains having an  $\alpha$ -helical structure; DUF1222 is indicated in black. *C*, the dipeptide NA sequence was inserted after position 243 of mouse Lmf1 to complete an ectopic NAT glycan attachment site (Lmf1<sup>InsNA243</sup>); the single, naturally occurring glycosylation site at position 430 was removed (N430A; see supplemental Fig. S2). After transfection of HEK293 cells, the cell lysates were treated with (+) or without (-) endo H, which removes N-linked glycans of the high mannose type. Lmf1<sup>InsNA243</sup> was visualized after Western blot analysis of the treated lysates using the c-Myc epitope tag. The mobility of Lmf1<sup>InsNA243</sup> treated with endo H was equivalent to unglycosylated Lmf1 (Lmf1<sup>N430A</sup>; see supplemental Fig. S2).

## LMF1 Topology and Lipase Interaction



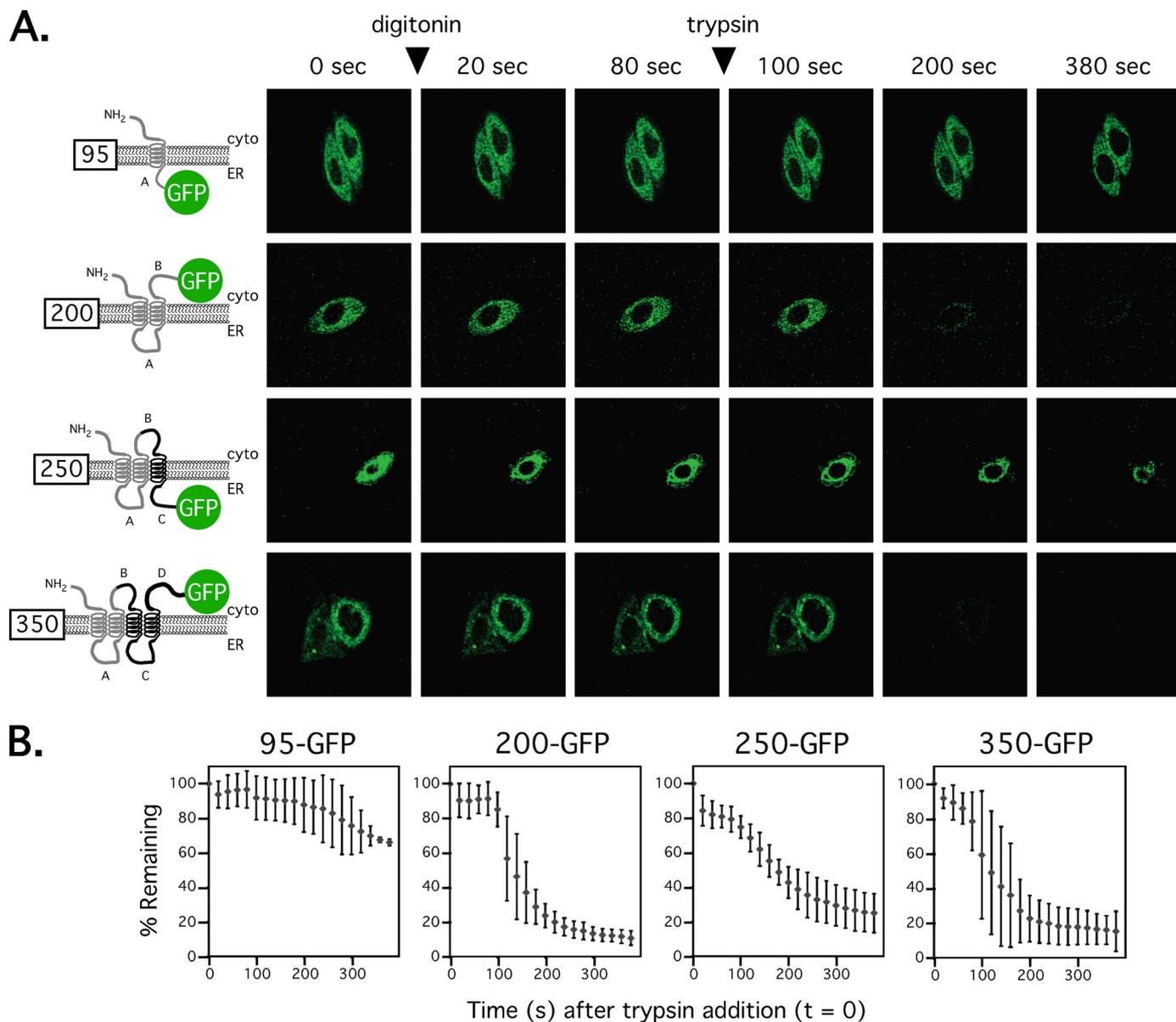
**FIGURE 3. Confirmation of mouse *Lmf1* membrane topology.** *A*, on the basis of the five-TM model, ectopic *N*-linked glycan sites were positioned in the predicted amino-terminal domain at site 28, in loop A at sites 94 and 101, in loop B at sites 175 and 188, in loop C at site 243, and in loop D at site 350. The carboxyl-terminal domain has a single, naturally occurring *N*-linked glycan attached at site 430 (see Fig. S2). All constructs with ectopic sites had the site at position 430 removed (N430A). The approximate lengths of the various domains are indicated. *B*, glycan utilization of the ectopic sites was visualized by decreased mobility of *Lmf1* as compared with the N430A control, which is not glycosylated. *Solid circles* represent sites that were utilized for glycosylation after transfection into HEK293 cells; *open circles* represent sites that were not glycosylated. In separate transfection experiments, glycosylation of only those ectopic sites predicted to be within the ER lumen (sites 94, 101, and 243) was confirmed by endo H digestion (Fig. S3).

To limit possible models of LMF1 topology based on MemO prediction results, the cytoplasmic and/or ER luminal orientation of the amino and carboxyl termini of LMF1 was determined. Besides identifying the location of LMF1 termini, this analysis constrains models of LMF1 topology to having either an even or an odd number of TM domains. To determine orientation, a FPP assay was employed (25), which measures in real time the trypsin proteolysis of fluorescent tags fused to the terminus of a membrane protein of interest. Digitonin, which is added to cells transfected with such a fusion construct, selectively permeabilizes the plasma membrane, and thus fluorescent tags oriented to the cytoplasm by fusion to the end of a membrane protein are accessible to trypsin digestion. In contrast, tags oriented to the ER lumen are protected from trypsin, as the cholesterol-poor ER membrane is relatively resistant to digitonin (26).

Constructs were prepared with the GFP fused to either the amino or carboxyl terminus of human LMF1 (GFP-LMF1 and LMF1-GFP, respectively). The fusion constructs exhibited LMF1-specific activities comparable with a wild type control (supplemental Fig. S1), indicating a native-like topology. When the FPP assay was applied to both constructs (Fig. 1, *A* and *C*), GFP at the amino terminus of LMF1 was clearly accessible to trypsin proteolysis, whereas it remained protected when fused to the LMF1 carboxyl terminus. Thus, the amino and carboxyl

termini of LMF1 were oriented oppositely to the cytoplasm and the ER lumen, respectively, indicating that LMF1 must have an odd number of TM domains. The ability of the FPP assay to identify the orientation accurately was validated using a known type 1 membrane protein, CD3 $\delta$  (Fig. 1, *B* and *C*).

The FPP assay and the MemO prediction analyses were used together to model LMF1 membrane topology based on an odd number (3, 5, or 7) of TM domains. Fig. 2*A* shows the seven MemO-predicted TM domains superimposed over a hydrophathy plot of mouse *Lmf1* (27). As expected, the predicted TM domains were localized over peaks with hydrophathy scores of >1.6 (window = 19), indicative of membrane-spanning regions (27). Fig. 2*B* shows a three-TM model of *Lmf1* constructed using the three predicted TM domains scoring 100% and a five- and seven-TM model including predicted TM domains having successively lower scores (supplemental Table S1);  $\alpha$ -helical TM domains are represented as *cylinders*. For reference, amino acid position 243 is *highlighted* in Fig. 2*B*, showing that this residue is within a loop directed to the cytoplasm in the three- and seven-TM models of *Lmf1*; in contrast, it is positioned on a loop facing the ER lumen in the five-TM model. The domain of unknown function (DUF) 1222 is indicated as a *black line*, whereas TM domains within DUF1222 are *solid black cylinders* (Fig. 2*B*).



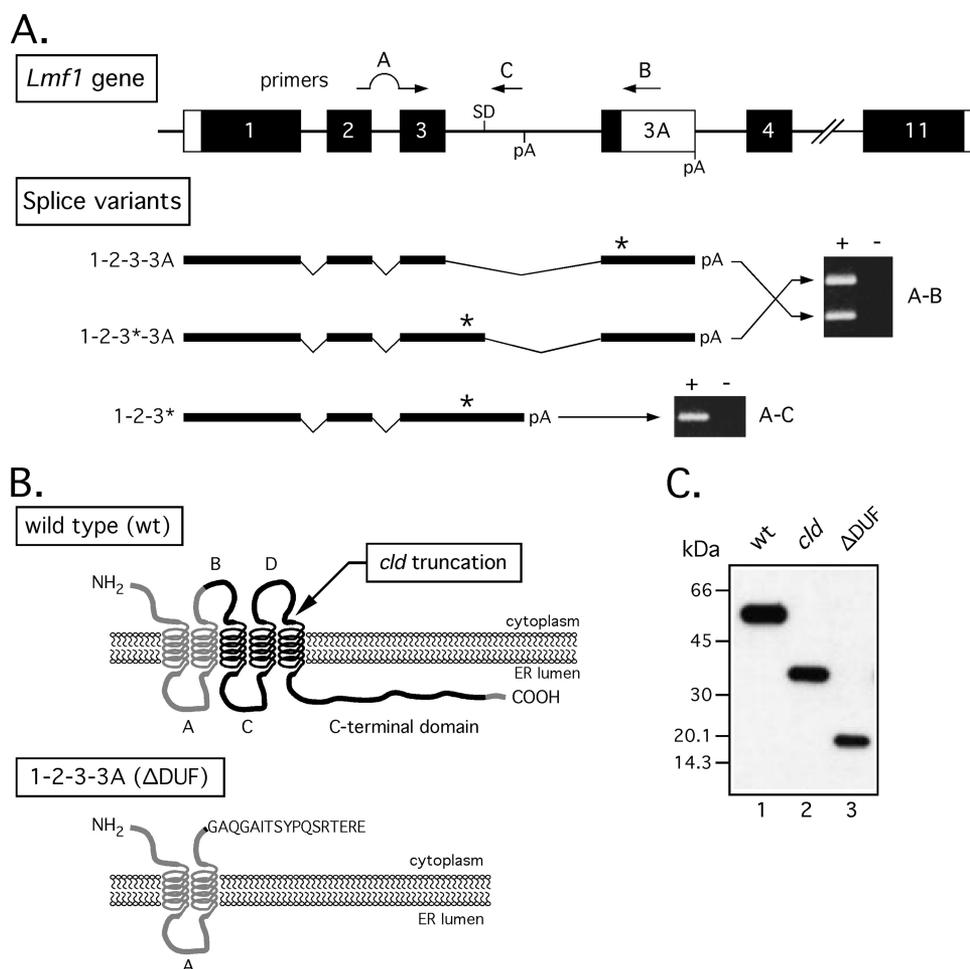
**FIGURE 4. Confirmation of human LMF1 membrane topology.** *A*, the FPP assay was applied to carboxyl-terminal GFP fusion constructs of human LMF1 truncated at positions 95 (loop A), 200 (loop B), 250 (loop C), and 350 (loop D). On the basis of the five-TM model, a diagram of each construct is provided with the predicted orientation of GFP with regard to its location in the cytoplasm (*cyto*) or ER lumen (*ER*). *B*, graphs of fluorescent signal intensity versus assay time are shown for all GFP constructs. See the legend for Fig. 1 for details.

To differentiate between the three possible models, an ectopic glycan attachment site was introduced at position 243 of mouse Lmf1 (Fig. 2C, Lmf1<sup>InsNA243</sup>), which has its single, naturally occurring glycosylation site at position 430 removed (supplemental Fig. S2, Lmf1<sup>N430A</sup>). Because glycosylation can occur only within the ER lumen, Lmf1<sup>InsNA243</sup> was transfected into HEK293 cells and tested for the presence of attached glycan chains of the high mannose type, indicative of ER-based glycosylation (28). Western blot analysis of the expressed Lmf1<sup>InsNA243</sup> protein showed a band that had increased mobility when treated with endo H (Fig. 2C), which selectively removes high mannose chains from glycoproteins. Thus, site 243 is glycosylated, indicating its exposure to the oligosaccharyltransferase complex in the ER lumen (28). The model consistent with this result is the one based on five TM domains

(Fig. 2B); in this model only, position 243 resides on a loop oriented to the ER lumen.

A similar strategy, using additional inserted ectopic glycan attachment sites, was used to verify mouse Lmf1 topology based on the five-TM model (Fig. 3A). Placement of ectopic sites on the model are shown in Fig. 3A, including the single, naturally occurring site at position 430 (supplemental Fig. S2). Notably, *N*-linked glycosylation at site 430 places the C-terminal domain of mouse Lmf1 within the ER lumen, consistent with the ER localization of the carboxyl terminus of human LMF1 (Fig. 1A). Like position 430, ectopic sites placed on loops A and C, also predicted to be oriented lumenally (Fig. 3A), were similarly glycosylated (Fig. 3B). In contrast, site 28, predicted to be within the N-terminal domain of mouse Lmf1, was not utilized for glycosylation, consistent with the localization of the amino

## LMF1 Topology and Lipase Interaction



**FIGURE 5. Coding variants of *Lmf1* lacking DUF1222.** *A*, three naturally occurring splice variants of the *Lmf1* gene are illustrated. The *Lmf1* gene has 11 exons plus alternative exon 3A; filled boxes represent coding sequences. The location of primers used to detect the splice variants and alternative splice donor (*SD*) and polyadenylation (*pA*) sites are indicated. The exon composition of splice variants 1-2-3-3A, 1-2-3\*-3A, and 1-2-3\* is provided. An asterisk denotes in-frame translation termination codons. To the right of each splice variant is the amplified RT-PCR product using the indicated primer sets (A-B and A-C) on a cDNA mix obtained from liver, muscle, adipose tissue, kidney, brain, and testis of wild type mice; + and - signs refer to the addition of reverse transcriptase. *B*, illustration of the *Lmf1* protein variants and their topology based on the five-TM model. The full-length wild type protein and the location of the *cld*-based truncation are shown. The DUF1222 domain is highlighted in black. The lower panel shows the predicted protein product resulting from the 1-2-3-3A splice variant ( $\Delta$ DUF); the 17 amino acids listed at its C terminus represent those coded by alternative exon 3A. *C*, Western blot analysis of HEK293 cells after transfection with wild type (*wt*), *cld*, and  $\Delta$ DUF *Lmf1* expression constructs.

terminus of human LMF1 to the cytoplasm (Fig. 1A). Like site 28, all additional ectopic sites predicated to be within loops facing the cytoplasm were not glycosylated (Fig. 3, A and B). Notably, functional analysis of all ectopic glycan mutants showed activity as high as the wild type control (data not shown), indicating that native topology was retained. Separate transfection experiments confirmed the glycosylation status of all ectopic sites (supplemental Fig. S3).

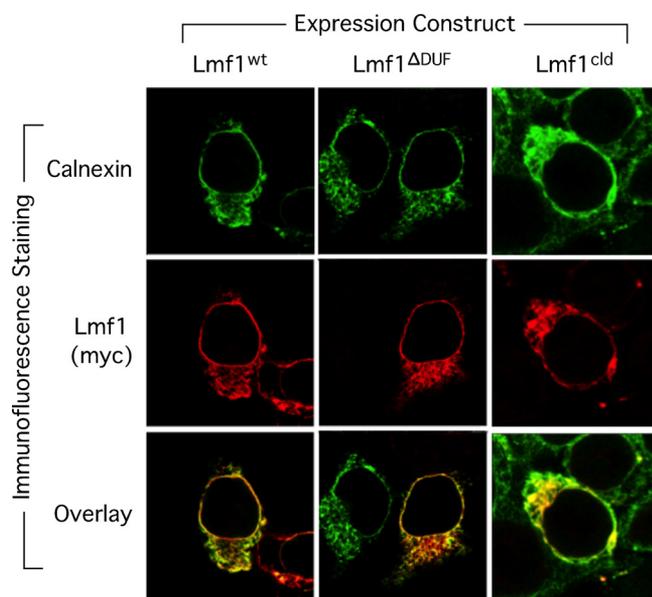
Finally, the five-TM model was confirmed by using human LMF1-GFP fusion constructs in the FPP assay (Fig. 4). Carboxyl-terminal GFP fusions were made with human LMF1 truncated at sites within loops predicated by the five-TM model; these included ER-facing loop A (95-GFP), cytoplasm-facing loop B (200-GFP), ER-facing loop C (250-GFP), and cytoplasm-facing loop D (350-GFP). As shown in Fig. 4, the predicated location of each loop was confirmed by the FPP

assay, with cytoplasm- and ER-facing loops showing accessibility or resistance to trypsin proteolysis, respectively. Thus, all results using ectopic glycan tags and FPP analysis on both mouse and human LMF1 constructs are consistent with a five-TM model of topology, subdividing LMF1 into three domains facing the cytoplasm (the amino-terminal domain and loops B and D) and three oriented to the ER lumen (loops A and C and the carboxyl-terminal domain). The smallest-to-largest domains of LMF1 are the amino-terminal domain (~28 aa), loop D (~46 aa), loops A and B (~56 aa), loop C (~71 aa), and the large carboxyl-terminal domain (~186 aa). The evolutionarily conserved DUF1222 begins in loop B and extends through most of the carboxyl-terminal domain. Indeed, as discussed below, those regions comprising DUF1222 are the most likely to have functional roles in lipase maturation.

**Coding Variants of *Lmf1* Affecting DUF1222**—The mouse *Lmf1* gene has 11 exons encoding a protein that is 574 amino acids in length (9). However, we determined the existence of several naturally occurring *Lmf1* splice variants that eliminate exons 4–11 encoding DUF1222 (Fig. 5A). Two of these variants (1-2-3-3A and 1-2-3\*) are present in the dbEST data base, and all three have been detected by RT-PCR (Fig. 5A) and confirmed by sequencing. All splice variants comprise exons 1–3, and two include an alternative exon

3A ending in a polyadenylation signal. One of these latter variants (1-2-3\*-3A) uses a cryptic splice donor (Fig. 5A, *SD*) site that results in the inclusion of some intronic sequence downstream of exon 3. The intronic sequence is also present in the variant without exon 3A (1-2-3\*), as it uses an alternative downstream polyadenylation site (Fig. 5A, *pA*). All encode *Lmf1* proteins lacking the DUF1222 domain ( $\Delta$ DUF), and splice variant 1-2-3-3A has the largest coding potential (188 amino acids).

Fig. 5B illustrates the predicted topology of wild type (full-length) *Lmf1* and the 1-2-3-3A splice variant, referred to hereafter as  $\Delta$ DUF, based on the five-TM model described above. DUF1222 begins in loop B and extends throughout loops C and D and most of the carboxyl-terminal domain (Fig. 5B, black line). In contrast, the  $\Delta$ DUF variant includes the amino-terminal domain, loop A, and the part of loop B preceding DUF1222 (Fig. 5B, gray line). The truncation of *Lmf1* caused by the *cld*



**FIGURE 6. Lmf1 variants are localized to the ER.** Expression constructs for full-length Lmf1<sup>wt</sup>, *cl<sup>d</sup>*-truncated Lmf1<sup>cl<sup>d</sup></sup>, and the DUF1222-truncated splice variant Lmf1<sup>ΔDUF</sup> were transfected into HEK293 cells. Using fluorescently labeled antibodies, endogenous calnexin was visualized in all cells; the expressed Lmf1 c-Myc epitope tag was detected only in transfected cells. Calnexin is a type 1 ER membrane protein with a classical ER retention signal.

mutation eliminates the last transmembrane domain and all of the downstream carboxyl-terminal sequence (Fig. 5B, arrow) and encodes a protein that is 375 amino acids in length. Fig. 5C shows that, when transfected into HEK293 cells, *cl<sup>d</sup>*-truncated Lmf1 (lane 2) and the ΔDUF splice variant (lane 3) expressed protein products of the expected molecular mass, indicating that these variants express stable proteins.

To verify the importance of DUF1222 in lipase maturation, Lmf1<sup>wt</sup>, Lmf1<sup>ΔDUF</sup>, and Lmf1<sup>cl<sup>d</sup></sup> protein variants were examined for their ability to restore lipase activity in mutant *cl<sup>d</sup>* cells. Although LPL transfected into mutant *cl<sup>d</sup>* cells is inactive and is retained within the ER, co-transfection of LPL with fully functional Lmf1 can restore LPL activity to wild type levels (9, 19). Such an assay was used to determine the ability of Lmf1<sup>ΔDUF</sup> and Lmf1<sup>cl<sup>d</sup></sup> variants to restore LPL activity in mutant *cl<sup>d</sup>* cells. Lmf1 without DUF1222 (Lmf1<sup>ΔDUF</sup>) failed to restore LPL activity (supplemental Fig. S4), confirming the importance of this evolutionarily conserved domain in lipase maturation. In fact, removal of the carboxyl-terminal domain of DUF1222 by the *cl<sup>d</sup>* truncation (Lmf1<sup>cl<sup>d</sup></sup>) was enough to cause loss of function (supplemental Fig. S4). Importantly, the loss of function caused by the DUF1222 deletion or *cl<sup>d</sup>* truncation was not due to dislocation of Lmf1 from the ER, the site of lipase maturation. Fig. 6 shows that Lmf1<sup>ΔDUF</sup> and Lmf1<sup>cl<sup>d</sup></sup>, along with the wild type control (Lmf1<sup>wt</sup>), co-localized with calnexin, a type 1 membrane protein containing a classical ER retention signal. Thus, although ER localization of Lmf1 remains unimpaired when DUF1222 is absent, this evolutionarily conserved domain is essential in carrying out the function of lipase maturation.

**Lipase Interaction with Lmf1**—Although the mechanisms underlying Lmf1-induced lipase maturation are unknown, a first step in identifying the process is to determine whether Lmf1 physically interacts with lipase proteins in the ER. To

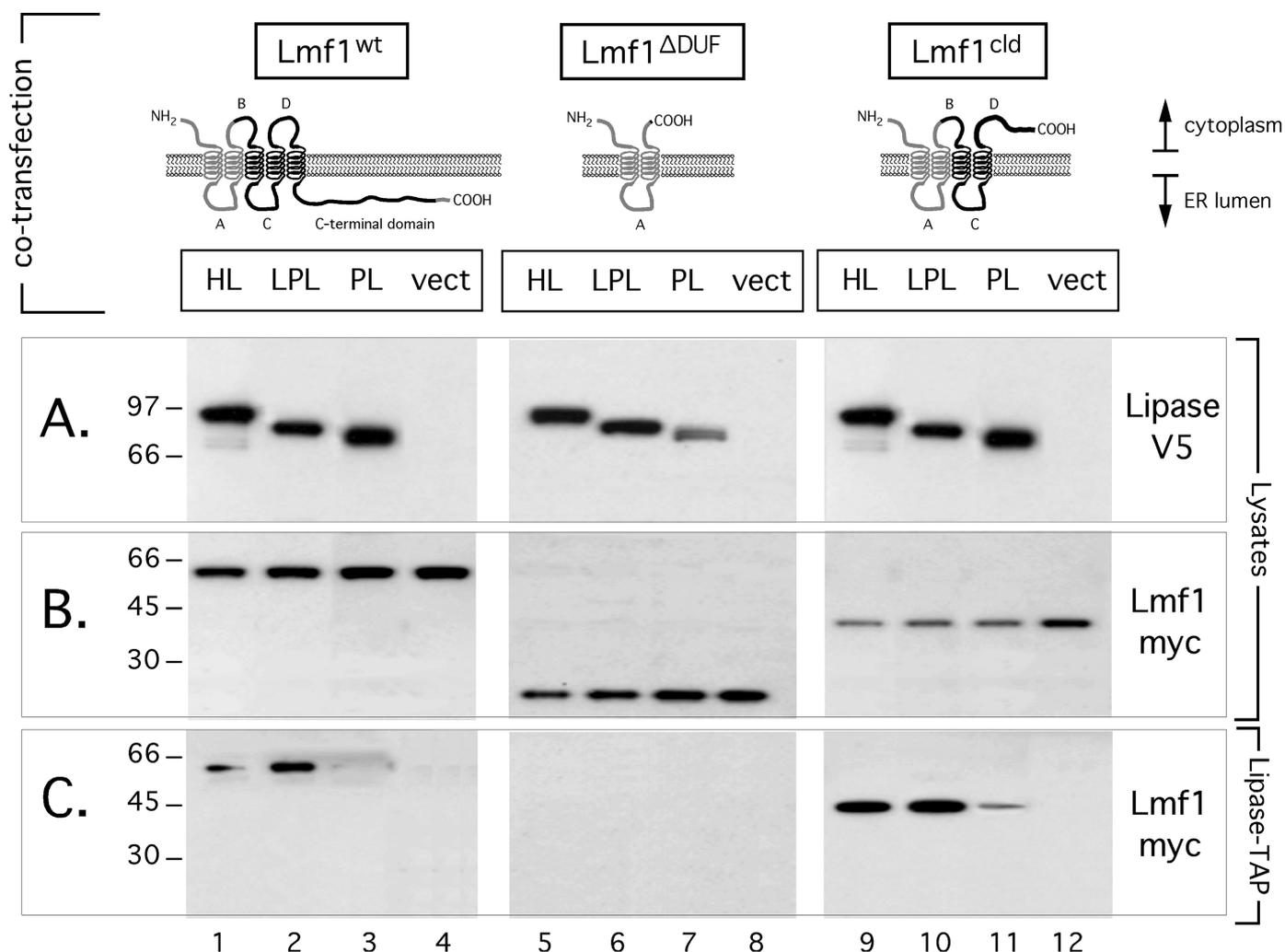
detect Lmf1-lipase interaction, lipase constructs fused to a TAP tag were used to achieve lipase affinity purification under conditions favoring the retention of protein-protein interactions (18). Three lipase-TAP constructs were used to evaluate Lmf1-lipase interaction: LPL, HL, and PL. Although all three are members of the lipase gene family (29), only maturation of LPL and HL is dependent on Lmf1 (5, 6, 9).

Lmf1<sup>wt</sup>, Lmf1<sup>cl<sup>d</sup></sup>, and Lmf1<sup>ΔDUF</sup> expression constructs were co-transfected in HEK293 cells along with each lipase-TAP construct and a vector only control. Fig. 7 shows that the various lipase-TAP proteins (panel A) and Lmf1 protein variants (panel B) were co-expressed in transfected HEK293 cell lysates; these lysates were then used for affinity purification of lipase proteins. The resulting lipase-TAP isolates were subjected to Western blot analysis and visualized using the Lmf1-specific c-Myc epitope tag (Fig. 7C). Notably, the Lmf1<sup>wt</sup> protein was shown to co-isolate with HL and LPL but only very poorly with PL (Fig. 7, lanes 1–3), indicating that Lmf1 interacts best with the two lipases dependent on Lmf1 for their maturation. In contrast, the DUF-deleted Lmf1 protein (Lmf1<sup>ΔDUF</sup>) failed to co-isolate with any lipase-TAP protein, verifying the placement of the lipase interaction site within DUF1222. When loop C was reinstated as the only ER-facing domain of DUF1222 (Lmf1<sup>cl<sup>d</sup></sup>), Lmf1-lipase interaction was restored (Fig. 7, lanes 9–11). Thus, loop C contains the lipase interaction site but cannot function independently of the large carboxyl-terminal domain in carrying out the function of lipase maturation (Fig. 6A).

## DISCUSSION

The accuracy of TM prediction is limited by the many diverse factors affecting TM usage and insertion during polytopic protein folding. Indeed, some polytopic proteins, such as the cystic fibrosis transmembrane conductance regulator, can assume multiple topological forms (30–33). In this study, we used a number of predictors to locate putative TM domains within LMF1; these predictions were based on Kyte-Doolittle hydrophathy plots (27), the dense alignment surface (DAS) method (34), and machine-learning algorithms (hidden Markov models such as TMHMM and HMMTOP), which are trained by analyzing the residues that tend to occupy defined regions in integral membrane proteins (31). However, the training set used by such programs is limited because of sparse information regarding eukaryotic membrane proteins; indeed, hydrophobic cores within soluble regions can be misidentified as TM domains, whereas short TM domains may be overlooked entirely (31). Because of these limitations, we used a number of TM prediction methods to interrogate two mammalian LMF1 orthologs (mouse and human). We assumed that for a given TM domain, a high degree of agreement among different methods in two homologs would provide a more accurate outcome than relying on a single method alone. We started with the predictors employed by the LOCATE data base (24) and used the resulting MemO-derived consensus of seven TM domains as our basis for model building. We compared these MemO predictors along with three others to calculate an alignment score for the seven predicted TM domains (supplemental Table S1) and then used the scores in a hierarchical manner to build additional models based on three and five TM domains. In all cases,

## LMF1 Topology and Lipase Interaction



**FIGURE 7. Loop C of Lmf1 is the site for lipase interaction.** Lmf1 expression constructs encoding full-length ( $Lmf1^{wt}$ ), *cl*-truncated ( $Lmf1^{cl}$ ), and DUF1222-truncated ( $Lmf1^{\Delta DUF}$ ) protein variants were co-transfected individually in HEK293 cells with LPL, HL, PL, or empty vector (*vect*). The expressed lipase protein was fused to a carboxyl-terminal TAP used for lipase purification. Lipase affinity purification was done under nonreducing conditions to preserve protein-protein interactions. Both total cell lysates (A and B) and lipase-TAP isolates (C) were subjected to Western blot analysis, and expressed lipase and Lmf1 protein were specifically detected using a V5 and c-Myc epitope tag, respectively. The positions of molecular mass markers (kDa) are given to the left of A–C. The topology of the various Lmf1 protein variants is included for reference. Taking into account the recovery of Lmf1 from the TAP affinity step, the calculated percentages of wild type Lmf1 bound to HL, LPL, and PL are 2.6, 3.4, and 0.5%, respectively. These percentages increased to 13.9, 13.5, and 1.0%, respectively, when the *cl*-truncated Lmf1 construct was used in the experiment.

we found the high degree of homology between mouse and human LMF1 (88% amino acid similarity) reflected in the conservation of the predicted TM domains, which were essentially identical between the two orthologs (supplemental Table S1).

Because of this conservation, we tested our models experimentally using mouse and human LMF1 in an interchangeable manner. Our strategy of employing tagging methods at LMF1 termini and within predicated loops is a proven means of testing topological models of polytopic proteins located in the plasma membrane (35, 36), mitochondria (37), and ER (38–40). The agreement in results between LMF1 orthologs using two different techniques (FPP analysis and ectopic glycan tagging) provided confidence in the five-TM model as a reliable representation of mammalian LMF1 topology. This representative model provides a structural basis for determining the functions associated with the six domains identified in our study, half of which faced the cytoplasm (the amino-terminal domain and loops A and D) or the ER lumen

(loops A and C and the carboxyl-terminal domain). This model also provides a first look at the arrangement of the DUF1222, an evolutionarily conserved domain located toward the carboxyl-terminal end of many hypothetical integral membrane proteins in bacteria and eukaryotes (14). Indeed, LMF1 is the first example of this family with a known function and, because of this study, an experimentally tested topology. It should be noted, however, that the boundary limits given for each of the TM domains in LMF1 are estimates. This is because the accuracy of TM boundaries assigned by prediction methods has been difficult to assess, as determining TM boundaries by biochemical and structural approaches has proven difficult (41). As shown in supplemental Table S1, for a given TM domain, different prediction methods give boundary limits that can vary by as much as 10–13 amino acids. We used the boundaries provided by the mouse MemO consensus to define the limits of the TM domains in our model of LMF1 (Fig. 3A).

The importance of DUF1222 in lipase maturation has been implied by loss-of-function mutations that truncate this domain at its carboxyl-terminal end. The *clد* mutation represents the most severe truncation because of introgression of a polyadenylation site occurring in the long term repeat of an inserted murine retrovirus in intron 7 of the mouse *Lmf1* gene (9). It removes all of the downstream sequence starting from position 361, eliminating the fifth TM domain (aa 368–388) and the entire carboxyl-terminal domain (aa 389–574). However, despite this large deletion, the Lmf1<sup>clد</sup> protein can be expressed (Fig. 5C) and remains localized to the ER membrane (Fig. 6B). Moreover, the carboxyl-terminal GFP fusion of human LMF1 truncated at position 350 (near position 361) showed a normal topology based on the expected cytoplasmic orientation of the truncation site within loop D (Fig. 4). This result suggests that *clد* does not affect Lmf1 topology upstream of the truncation site. Notably, Lmf1<sup>clد</sup> still retains loop C and its associated lipase interaction site, although it is incapable of functioning as a lipase maturation factor. In fact, Lmf1<sup>clد</sup> appears to bind more LPL and HL than does the wild type protein (see legend for Fig. 7). This may reflect lack of an adjacent carboxyl-terminal domain in Lmf1<sup>clد</sup> and suggests that although lipases may initially bind to loop C, subsequent interaction with the large carboxyl-terminal domain absent in Lmf1<sup>clد</sup> may in fact trigger their eventual release. Indeed, it is intriguing to speculate that the juxtaposition of loop C and the carboxyl-terminal domain may promote cooperation between these domains to facilitate the related functions of lipase binding and maturation.

In conclusion, our model of LMF1 topology provides a rational basis for pursuing structure/function studies to identify the specific roles of individual domains in lipase maturation. It may also be useful in identifying potential sites for cytoplasm-based post-translational modifications (e.g. phosphorylation) that may in turn regulate LMF1 activity.

*Acknowledgment*—We greatly appreciate the gift of eCFP-CD3δ and CDδ-YFP constructs from Dr. J. Lippincott-Schwartz.

## REFERENCES

- Hasham, S. N., and Pillarisetti, S. (2006) *Clin. Chim. Acta* **372**, 179–183
- Mead, J. R., Irvine, S. A., and Ramji, D. P. (2002) *J. Mol. Med.* **80**, 753–769
- Perret, B., Mabile, L., Martinez, L., Tercé, F., Barbaras, R., and Collet, X. (2002) *J. Lipid Res.* **43**, 1163–1169
- Santamarina-Fojo, S., González-Navarro, H., Freeman, L., Wagner, E., and Nong, Z. (2004) *Arterioscler. Thromb. Vasc. Biol.* **24**, 1750–1754
- Scow, R. O., Schultz, C. J., Park, J. W., and Blanchette-Mackie, E. J. (1998) *Chem. Phys. Lipids* **93**, 149–155
- Briquet-Laugier, V., Ben-Zeev, O., White, A., and Doolittle, M. H. (1999) *J. Lipid Res.* **40**, 2044–2058
- Olivecrona, T., Chernick, S. S., Bengtsson-Olivecrona, G., Paterniti, J. R., Jr., Brown, W. V., and Scow, R. O. (1985) *J. Biol. Chem.* **260**, 2552–2557
- Davis, R. C., Ben-Zeev, O., Martin, D., and Doolittle, M. H. (1990) *J. Biol. Chem.* **265**, 17960–17966
- Péterfy, M., Ben-Zeev, O., Mao, H. Z., Weissglas-Volkov, D., Aouizerat, B. E., Pullinger, C. R., Frost, P. H., Kane, J. P., Malloy, M. J., Reue, K., Pajukanta, P., and Doolittle, M. H. (2007) *Nat. Genet.* **39**, 1483–1487
- Reue, K., and Doolittle, M. H. (1996) *J. Lipid Res.* **37**, 1387–1405
- Coleman, T., Seip, R. L., Gimble, J. M., Lee, D., Maeda, N., and Semenkovich, C. F. (1995) *J. Biol. Chem.* **270**, 12518–12525
- Ben-Zeev, O., and Doolittle, M. H. (2004) *J. Biol. Chem.* **279**, 6171–6181
- Ben-Zeev, O., Mao, H. Z., and Doolittle, M. H. (2002) *J. Biol. Chem.* **277**, 10727–10738
- Sonnhammer, E. L., Eddy, S. R., and Durbin, R. (1997) *Proteins* **28**, 405–420
- Boedeker, J. C., Doolittle, M. H., and White, A. L. (2001) *J. Lipid Res.* **42**, 1858–1864
- Péterfy, M., Phan, J., and Reue, K. (2005) *J. Biol. Chem.* **280**, 32883–32889
- Ben-Zeev, O., Doolittle, M. H., Davis, R. C., Elovson, J., and Schotz, M. C. (1992) *J. Biol. Chem.* **267**, 6219–6227
- Doolittle, M. H., Ben-Zeev, O., Bassilian, S., Whitelegge, J. P., Péterfy, M., and Wong, H. (2009) *J. Lipid Res.* **50**, 1173–1184
- Yin, F., Doolittle, M. H., and Péterfy, M. (2009) *J. Lipid Res.*, in press
- Briquet-Laugier, V., Ben-Zeev, O., and Doolittle, M. H. (1999) *Methods Mol. Biol.* **109**, 81–94
- Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Séraphin, B. (2001) *Methods* **24**, 218–229
- Doolittle, M. H., Ben-Zeev, O., and Briquet-Laugier, V. (1998) *J. Lipid Res.* **39**, 934–942
- Lorenz, H., Hailey, D. W., Wunder, C., and Lippincott-Schwartz, J. (2006) *Nat. Protoc.* **1**, 276–279
- Fink, J. L., Aturaliya, R. N., Davis, M. J., Zhang, F., Hanson, K., Teasdale, M. S., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., and Teasdale, R. D. (2006) *Nucleic Acids Res.* **34**, D213–D217
- Lorenz, H., Hailey, D. W., and Lippincott-Schwartz, J. (2006) *Nat. Methods* **3**, 205–210
- Liscum, L., and Munn, N. J. (1999) *Biochim. Biophys. Acta* **1438**, 19–37
- Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
- Helenius, A., and Aebi, M. (2004) *Annu. Rev. Biochem.* **73**, 1019–1049
- Wong, H., and Schotz, M. C. (2002) *J. Lipid Res.* **43**, 993–999
- Dowhan, W., and Bogdanov, M. (2009) *Annu. Rev. Biochem.* **78**, 515–540
- Ott, C. M., and Lingappa, V. R. (2002) *J. Cell Sci.* **115**, 2003–2009
- Skach, W. R. (2007) *J. Cell Biol.* **179**, 1333–1335
- Turner, R. J. (2003) *J. Membr. Biol.* **192**, 149–157
- Cserző, M., Wallin, E., Simon, I., von Heijne, G., and Elofsson, A. (1997) *Protein Eng.* **10**, 673–676
- Kast, C., Canfield, V., Levenson, R., and Gros, P. (1995) *Biochemistry* **34**, 4402–4411
- Kast, C., Canfield, V., Levenson, R., and Gros, P. (1996) *J. Biol. Chem.* **271**, 9240–9248
- Gonzalez-Baro, M. R., Granger, D. A., and Coleman, R. A. (2001) *J. Biol. Chem.* **276**, 43182–43188
- Joyce, C. W., Shelness, G. S., Davis, M. A., Lee, R. G., Skinner, K., Anderson, R. A., and Rudel, L. L. (2000) *Mol. Biol. Cell* **11**, 3675–3687
- Lin, S., Cheng, D., Liu, M. S., Chen, J., and Chang, T. Y. (1999) *J. Biol. Chem.* **274**, 23276–23285
- Man, W. C., Miyazaki, M., Chu, K., and Ntambi, J. M. (2006) *J. Biol. Chem.* **281**, 1251–1260
- Deber, C. M., Wang, C., Liu, L. P., Prior, A. S., Agrawal, S., Muskat, B. L., and Cuticchia, A. J. (2001) *Protein Sci.* **10**, 212–219