

# Molecular Cloning and Characterization of Lysosomal Sialic Acid *O*-Acetyltransferase\*

(Received for publication, February 7, 1996)

M. Jorge Guimarães<sup>‡§</sup>, J. Fernando Bazan<sup>‡</sup>, Janice Castagnola<sup>¶</sup>, Sandra Diaz<sup>¶</sup>,  
Neal G. Copeland<sup>||</sup>, Debra J. Gilbert<sup>||</sup>, Nancy A. Jenkins<sup>||</sup>, Ajit Varki<sup>¶</sup>, and Albert Zlotnik<sup>‡</sup>

From the <sup>‡</sup>DNAX Research Institute of Molecular and Cellular Biology, Palo Alto, California 94304, the <sup>¶</sup>Glycobiology Program, University of California at San Diego Cancer Center, La Jolla, California 92093-0063, and the <sup>||</sup>ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland 21702

***O*-Acetylation and de-*O*-acetylation of sialic acids have been implicated in the regulation of a variety of biological phenomena, including endogenous lectin recognition, tumor antigenicity, virus binding, and complement activation. Applying a strategy designed to identify genes preferentially expressed in active sites of embryonic hematopoiesis, we isolated a novel cDNA from the pluripotent hematopoietic cell line FDCP-mixA4 whose open reading frame contained sequences homologous to peptide fragments of a lysosomal sialic acid *O*-acetyltransferase (*Lse*) previously purified from rat liver, but with no evident similarity to endoplasmic reticulum-derived acetyltransferases. The expressed *Lse* protein exhibits sialic-acid *O*-acetyltransferase activity that is not attributable to a typical serine esterase active site. *Lse* expression is spatially and temporally restricted during embryogenesis, and its mRNA levels correlate with differences in *O*-acetyltransferase activity described in adult tissues and blood cell types. Using interspecific backcross analysis, we further mapped the *Lse* gene to the central region of mouse chromosome 9. This constitutes the first report on the molecular cloning of a sialic acid-specific *O*-acetyltransferase in vertebrates and suggests novel roles for the 9-*O*-acetyl modification of sialic acids during the development and differentiation of mammalian organisms.**

Sialic acids are a family of 9-carbon acidic sugars that constitute terminal units on diverse oligosaccharide chains of glycoconjugates in higher invertebrates and vertebrates. Modifications of the parent sialic acid *N*-acetylneuraminic acid have been described that show remarkable molecule and tissue specificity as well as developmental regulation (1–5). However, the relationship between these modifications on sialic acids and specific events in mammalian development is unclear. The availability of cDNAs encoding sialic acid-modifying enzymes is therefore expected to be relevant for the understanding of the role played by sialic acids in the biology of mammalian organ-

isms (3, 4). One modification of particular interest is the *O*-acetylation and de-*O*-acetylation of sialic acids, which determines the presence or absence of the outermost possible structure on typical *N*-linked oligosaccharides, *i.e.* *O*-acetyl esters (6). These esters have been implicated in lectin recognition, cell adhesion, tissue morphogenesis, and a variety of other biological phenomena, including tumor antigenicity, virus binding, and complement activation (1–3). Enzymes specifically capable of removing *O*-acetyl esters from the 9-position of sialic acids have been described in viruses (7–10) and in vertebrates (11–13), but no cDNA for the latter group has yet been isolated. A disulfide-linked heterodimeric protein with sialic acid-specific 9-*O*-acetyltransferase activity was purified from rat liver membrane compartments and initially called luminal sialic acid *O*-acetyltransferase (*Lse*) (11, 14) to differentiate it from a cytosolic enzyme with similar activity (12, 13). The *Lse* protein was subsequently shown to have a primarily lysosomal location (15) and has hence been renamed lysosomal sialic acid *O*-acetyltransferase (retaining the acronym *Lse*). Early attempts to clone this esterase by conventional methods were thwarted by poor immunoreactivity and extreme resistance to proteolytic digestion.<sup>1</sup> Only two amino acid sequences were obtained (14), but these were not strongly immunogenic or useful for cloning using degenerate oligonucleotides.<sup>2</sup>

We have recently described a novel strategy to identify genes differentially expressed throughout development and potentially involved in hematopoiesis (16, 17). Using this strategy, we identified a number of short cDNA fragments representing genes preferentially expressed in the yolk sac and up-regulated during the *in vitro* development of embryoid bodies (EBs)<sup>3</sup> (17). Here we describe the cloning and characterization of one of these genes, a cDNA encoding the mouse homolog of the purified rat *Lse* protein. We have expressed the *Lse* protein, studied the role of the proteolytic cleavage of the immature *Lse* protein into its disulfide-linked heterodimeric form in generating *O*-acetyltransferase activity (which we attribute to a novel enzymatic mechanism), examined the tissue distribution of the mRNA, and mapped the *Lse* gene on mouse chromosome 9. The molecular cloning of *Lse* should allow the study of the role of sialic acid esterases in development and tumorigenesis, a task that until now was impaired by the lack of any vertebrate cDNA encoding these enzymes.

## EXPERIMENTAL PROCEDURES

**Isolation of a cDNA encoding *Lse***—Total RNA derived from the hematopoietic cell line FDCPmixA4 (18), cultured as described (17), was obtained using RNazol solution (Tel-test Inc., Friendswood, TX)

\* This work was supported in part by a Veterans Administration merit review award and Grant RO1GM32373 (to A. V.) and by NCI, Department of Health and Human Services, under Contract NO1-CO-4600 with ABL. Work performed at the DNAX Research Institute of Molecular and Cellular Biology was supported by Schering-Plough Corp. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Supported by Junta Nacional de Investigação Científica e Tecnológica, Portugal (Grant Number CIÊNCIA/BD/2685/93). To whom correspondence should be addressed: DNAX Research Inst., 901 California Ave., Palo Alto, CA 94394. Tel.: 415-496-1172; Fax: 415-496-1200; E-mail: guimaraes@dnax.org.

<sup>1</sup> B. K. Hayes and A. Varki, unpublished observations.

<sup>2</sup> H. H. Higa and A. Varki, unpublished results.

<sup>3</sup> The abbreviations used are: EBs, embryoid bodies; PCR, polymerase chain reaction; kb, kilobase(s).

following the manufacturer's instructions. Poly(A)<sup>+</sup> RNA was selected from FDCPmixA4 total RNA, and a phage cDNA library was made as described (16). The PCR product Clone 165, generated by differential display by PCR (16, 17), was labeled by random priming using the Prime-it II kit (Stratagene, La Jolla, CA) and used to screen the FDCPmixA4-derived library (19). Positive clones were purified and subcloned also as described (16). The complete nucleotide sequence of Clone 165-109 was obtained as described (16) and deposited in GenBank (20) with accession number U40408.

**Constructs for Protein Expression**—Two constructs were generated for protein expression of Clone 165-109 in COS-7 cells (American Type Cell Culture) in which a tag (FLAG) sequence (21) was introduced in the Lse protein. The open reading frame of Clone 165-109 was amplified using a strategy of PCR mutagenesis (22) designed to allow the introduction of a FLAG peptide sequence either at the C terminus of the protein (LF construct) or in the middle of the two Lse subunits (LFL construct). The primers used for the Lse construct were as follows: sense primer, 5'-CACTTTGCGGCCGCGCACCATGGTTTCCCGGGG-CCTGTGTTG-3'; and antisense primer, 5'-CGTACTAGTTTACTT-GTCATCGTCGTCCTGTAGTCGATACCCCTGTGTGAAATTTG-3'. Two sets of primers for two independent reactions were used in the first round of PCR for the generation of the LFL construct: sense primer (set 1), 5'-CACTTTGCGGCCGCGCACCATGGTTTCCCGGGG-CCTGTGTTG-3'; antisense primer (set 1), 5'-GGCGTTCCAAAGGACGGAGTG-TCTTGTCCGTCCTTGTAGTCGATACCCCTGTGTGAAATTTG-3'; sense primer (set 2), 5'-CC-TTTTAGGGTTGTCCACTCTGTTGTCGACTGACTACAAGGACGACGATGACAAGGGACCGACAAGACTCCGTCCTTTGGAACGCC-3'; and antisense primer (set 2), 5'-CCCTTCATGCTCAAATTTACAC-AGGGGTATCTAAACTAGTACG-3'. The products that resulted from these two reactions were gel-extracted and amplified in a single second PCR to generate the complete LFL construct by using the sense primer from set 1 and the antisense primer from set 2. *Pfu* DNA polymerase (Stratagene) was used with 18 cycles of PCR using the following parameters: 94 °C for 30 s, 55 °C for 1 min, and 72 °C for 4 min. These constructs were cloned into the PME18X vector (DNAX) using *NofI* and *SpeI* sites incorporated into the 5'- and 3'-primers, respectively.

**Sequence and Structural Analysis of the Lse Protein**—The original BLAST program (23) at the National Center for Biotechnology Information (accessed over the Internet at URL [http://www.ncbi.nlm.nih.gov/Recipon/bs\\_seq.html](http://www.ncbi.nlm.nih.gov/Recipon/bs_seq.html)) and an enhanced version called BEAUTY (24) developed at the Human Genome Center at the Baylor College of Medicine (<http://dot.ingen.bcm.tmc.edu:9331/seq-search/protein-search.html>) were used to comb nonredundant protein and nucleotide data bases for Lse homologs. The search for more distant relatives of Lse employed the sensitive sequence comparison strategies of Altschul *et al.* (25) and Bork *et al.* (26); banks of diagnostic sequence patterns, motifs, and profiles were also screened for faint matches to the component Lse domains (for a recent review, see Ref. 27) (<http://www.embl-heidelberg.de/~bork/pattern.html>). Predictive analysis of the Lse secondary structure principally utilized the PHD neural network program (28) accessed at the EMBL PHD Internet server (<http://www.embl-heidelberg.de/predictprotein/predictprotein.html>). Potential O-glycosylation sites were suggested by the NetOglyc program (29) (<http://www.cbs.dtu.dk/netOglyc/cbsnetOglyc.html>).

**Protein Expression of the LF and LFL Constructs**—COS-7 cells were maintained as described (17). Plasmid DNA was transfected as described (17). Cell lysates and media were collected 3 days after transfection. Lysis buffer (17) was added to the plates, which were kept on ice for 45 min. Lysates were centrifuged to eliminate cell debris (17). Supernatants of centrifuged cell lysates and sterile-filtered media from cultured cells were incubated with anti-FLAG M2 affinity gel (Kodak Scientific Imaging Systems, New Haven, CT) at 4 °C overnight and washed four times with phosphate-buffered saline. Immunoprecipitates were eluted, neutralized, and concentrated by precipitation with 24% trichloroacetic acid (Sigma) and 2% deoxycholic sodium salt (Sigma) as described (17). Pellets were eluted in 2 × sample buffer (Novex, San Diego, CA), electrophoresed on 4–20% Tris/glycine gels (Novex), and transferred to polyvinylidene difluoride membranes (Immobilon-P, Millipore Corp., Bedford, MA). Membranes were exposed to 3% nonfat milk for 1 h at 37 °C. Anti-FLAG M2 antibody (Kodak) was used in a 1:2000 dilution. Horseradish peroxidase-conjugated anti-mouse Ig (Amersham Corp.) was also used at a 1:2000 dilution, and peroxidase detection was performed with ECL detection reagents (Amersham Corp.) as recommended. M2-purified LFL construct-derived protein was run on a 8% Tris/glycine gel and processed as described (14) for amino-terminal gas-phase sequencing.

**Large-scale Purification of the COS Cell-expressed LFL Construct**

**from Cell Culture Medium**—20 dishes of COS cells were seeded at 1 × 10<sup>6</sup> cells/10-cm culture dish in minimum Eagle's medium containing 5% fetal calf serum and incubated overnight at 37 °C in 5% CO<sub>2</sub>. Cells were transfected with 10 μg of plasmid DNA/plate, with either the LF or LFL construct, using LipofectAMINE (Life Technologies, Inc.) exactly as per the manufacturer's recommendations and were left growing in 10 ml of serum-free medium. 5 h later, an equal volume of medium with 10% serum was added to the cells. ~24 h after transfection, the medium was replaced with 10 ml of fresh medium containing 5% serum and collected after 3 days. After centrifugation to remove the cells and filtration to remove any particulate material, the clarified medium was passed through a 1-ml column of Bio-Rad Bio-Gel A-agarose resin (equilibrated in Tris-buffered saline) to preclear any nonspecific binding material. A small volume of Tris-buffered saline was used to wash the column. The medium was then incubated with a 1-ml packed volume of anti-FLAG M2 affinity gel (prepared as per the manufacturer's recommendations) at 4 °C overnight with gentle mixing. The medium and resin were poured into a small column, and the medium was then re-passed through the column to maximize the binding of the protein. The resin was washed with 3 × 5 ml of Tris-buffered saline, and bound proteins were eluted by adding 1 ml of 0.1 M glycine (pH 3.3), allowing 0.5 ml of the buffer to come through, and then closing off the column and allowing it to sit for 8 min. The remainder of the eluate was collected in the same tube as the first 0.5 ml. The column was eluted with an additional 3 × 1 ml of 0.1 M glycine (pH 3.3). All collections were in tubes containing 1 M Tris (pH 8.0) to neutralize the pH immediately.

**Digestion with Endo-β-N-acetylglucosaminidase H and Peptide N-Glycosidase F**—Protein expressed from the LFL construct was prepared as described above. For peptide N-glycosidase F treatment (11, 14), 1.1-μg aliquots were boiled in 20 mM Tris (pH 8.0), 1% SDS, and 20 mM β-mercaptoethanol for peptide N-glycosidase F for 3 min to denature the protein. Reaction mixtures (45 μl) containing the denatured esterase, 1.25% Nonidet P-40, 0.25% SDS, 20 mM β-mercaptoethanol, 1 mM 1,10-phenanthroline, and 2 units of peptide N-glycosidase F in 20 mM Tris-HCl (pH 8.0) were incubated at 37 °C for 18 h. For endo-β-N-acetylglucosaminidase H treatment (11, 14), the aliquot was denatured in 0.1% SDS and 0.1 M β-mercaptoethanol as described above, and aliquots with or without endo-β-N-acetylglucosaminidase H (3 milli-units) in 45 μl of 200 mM sodium citrate (pH 5.0) were incubated at 37 °C for 18 h.

**Assay and Reprecipitation of Sialic Acid O-Acetyltransferase Activity**—Sialic acid 9-O-acetyltransferase activity was assayed and expressed in units exactly as described previously (11, 14) using [*acetyl*-9-<sup>3</sup>H]9-O-acetyl-N-acetyl-neuraminic acid as a substrate under conditions where product formation (release of [<sup>3</sup>H]acetate) was linear with time and added enzyme. Aliquots of the LFL epitope-tagged antibody-eluted preparations were incubated with either anti-FLAG M2 affinity gel or control Bio-Gel A overnight at 4 °C in 0.1 M Tris-HCl (pH 8.0) with mixing. The suspension was centrifuged to separate the supernatant from the beads. Additional buffer was added to the beads. The substrate [*acetyl*-9-<sup>3</sup>H]9-O-acetyl-N-acetyl-neuraminic acid was added to both the beads and the supernatant of each sample, and the activity was determined after 3 h at 37 °C.

**Northern Blot and PCR Analyses**—Fetal and adult mouse tissues were isolated, cell lines were cultured, and total RNA was isolated and used for Northern blot analysis as described (17). Large-scale preparations of plasmid DNA containing the differential display by PCR product Clone 165 (17) were made as described (17). Plasmid DNA was cut with *EcoRI*, gel-extracted with the QIAEX gel extraction kit (QIAGEN Inc.), and random-primed with [*α*-<sup>32</sup>P]dCTP (Amersham Corp.) as described (17). 20-μg samples of total RNA were run on formaldehyde gels (19), transferred to Nytran membranes (Schleicher & Schuell) by standard methods (19), hybridized, and washed at 65 °C as described (17).

Total RNA (5 μg) was isolated using RNazol solution from the cell lines BL/3 (30), Clone K (31), B9 (DNAX), KD83 (32), CH12 (33), A20 (34), PL1-1.4 P388D1 (35), B5A.5 PT18 (36), HC7.21 (37), HT2 (38), D1.1 (39), CDC25 (39), and D10.G4 (40). Total RNA from day 8.5 embryos, embryonic stem cells, and EBs was prepared as described (17). Reverse transcription and PCR were performed, and materials were obtained as described (17). Briefly, the PCR conditions were as follows: denaturing at 94 °C for 30 s, annealing at 55 °C for 30 s, and extension at 72 °C for 1 min for 30 cycles using a Perkin-Elmer DNA thermocycler. Primers for hypoxanthine phosphoribosyltransferase were as described (17). Primers for *Ise* were 5'-GGAGTCCCTAATACAAGGGATG-3' (sense

<sup>4</sup> S. Hudack, unpublished results.

<sup>5</sup> D. Rennick, unpublished results.

primer) and 5'-CCAAAGGGTGAGTCTCTATCAC-3' (antisense primer). The identity of the PCR products was confirmed by size on agarose gels (expected size of 494 base pairs) and hybridization with an internal primer (5'-CGGATTATAACAGGGACCTGTAC-3') after transfer to a nylon membrane (19). The probe was labeled at the 5'-end with [ $\gamma$ - $^{32}$ P]ATP (Amersham Corp.) using polynucleotide kinase (Boehringer Mannheim). Blots were hybridized in 0.5 M NaH<sub>2</sub>PO<sub>4</sub> (pH 7.2), 7% SDS, and 0.5 M EDTA (pH 8.0) at 42 °C for 24 h; washed in 6 × SSC and 0.2% SDS at 42 °C for 1 h with one change of buffer; and exposed to Kodak X-Omat film (Eastman Kodak Co.) at -80 °C.

**Southern Blot Analysis and Interspecific Mouse Backcross Mapping**—Genomic DNA from HeJ mice was purchased from the Jackson Laboratory (Bar Harbor, Maine). DNA digests were obtained using *Eco*RI, *Hind*III, and *Pst*I and processed as described (19) for Southern blot analysis using as probe entire random prime-labeled Clone 165-109 DNA.

Interspecific progeny were generated by mapping (C57BL/6J × *Mus spretus*)F<sub>1</sub> females and C57BL/6J males as described (41). A total of 205 N<sub>2</sub> mice were used to map the *Lse* locus (see "Results" for details). DNA isolation, restriction enzyme digestion, agarose gel electrophoresis, Southern blot transfer, and hybridization were performed essentially as described (19). All blots were prepared with Hybond-N<sup>+</sup> (Amersham Corp.). The probe, an ~2.7-kb fragment of mouse cDNA, was labeled with [ $\alpha$ - $^{32}$ P]dCTP using a random prime labeling kit (Stratagene); washing was done to a final stringency of 1.0 × SSCP and 0.1% SDS at 65 °C. Major fragments of 7.4 and 5.7 kb were detected in *Taq*I-digested *M. spretus* DNA. The presence or absence of the 5.7-kb *Taq*I *M. spretus*-specific fragment was followed in backcross mice.

A description of the probes and restriction fragment length polymorphisms for the loci linked to *Lse*, including E26 avian leukemia oncogene 1 (*ets1*), thymus cell antigen 1 (*thy1*), and dopamine receptor 2 (*drd2*), has been reported previously (42). Recombination distances were calculated as described (43) using the computer program SPRETUS MADNESS. Gene order was determined by minimizing the number of recombination events required to explain the allele distribution patterns.

## RESULTS

**Molecular Cloning of a Novel cDNA Encoding Mouse Lse**—In a previous report, we described a strategy directed toward the understanding of early hematopoietic development by isolating cDNAs preferentially expressed in the yolk sac, up-regulated during the *in vitro* development of EBs, and expressed in the pluripotent hematopoietic cell line FDCPmixA4 (17). Among the PCR products identified with the above-described pattern of expression was a novel 165-base pair DNA sequence designated Clone 165 (17). To isolate its corresponding complete cDNA, we screened a FDCPmixA4 cell line-derived cDNA library using Clone 165-derived DNA as a probe. The representation of Clone 165 cDNA in this phage library was ~1/17,000 (after screening ~5 × 10<sup>5</sup> independent clones). Positive clones were purified and subcloned into the pZL1 vector, and their size and restriction digest pattern were analyzed. A 2.7-kb insert, designated Clone 165-109, represented the longest cDNA among the clones isolated. Its complete sequence was determined as well as that of another clone with a slightly different restriction endonuclease digestion pattern. The latter clone, designated Clone 165-8, was found to represent a partial 1.5-kb cDNA containing a different polyadenylation site at the 3'-untranslated region (Fig. 1A). The predicted peptide sequence of Clone 165-109 indicated a 531-amino acid protein that appeared to be the mouse homolog of the heterodimeric disulfide-bonded Lse protein purified from rat liver (11, 14). Pulse-chase studies of the Lse protein in rat hepatoma cells had shown that the mature dimeric form was derived from the gradual processing and internal cleavage of a single-chain protein precursor (14) (Fig. 1B) into an ~58-kDa heterodimeric protein with two subunits of ~38 and ~28 kDa (11, 14). As indicated in Fig. 1C, the protein sequences of the mouse and rat Lse proteins are highly similar.

Lse is a lysosomal glycoprotein that traverses the endoplasmic reticulum-Golgi pathway during biosynthesis (14, 15). The existence of a secretory signal peptide in mouse Lse is sug-

gested by the presence of an N-terminal hydrophobic sequence (Fig. 1A); amino-terminal peptide sequencing reveals that the amino-terminal processing site of the mature Lse protein is located between Gly ↓ Ile at positions 23 and 24 from the first Met (data not shown). Thus, the amino terminus of the mouse Lse protein corresponds precisely to that of the rat Lse small subunit peptide fragment (Fig. 1C). Eight potential N-glycosylation sites (motif of Asn-X-Ser/Thr) are found scattered in the mature Lse sequence (Fig. 1A); three potential O-glycosylation sites (threonines at positions 37, 45, and 51) are also suggested by the NetOglyc neural network program (29).

**Relationship of Lse to Existing Esterase Families**—Aside from the close relationship of embedded sequences in mouse Lse to the rat Lse peptide fragments (Fig. 1C), a determined search of nonredundant sequence data bases, including the burgeoning bank of human-expressed sequence tags (44), failed to uncover clear homologs of the Lse enzyme. Taking a cue from the proteolytic processing and functional domain division of the Lse chain (11, 14), the amino- and carboxyl-terminal domains of Lse (equivalent to the small and large subunits of Fig. 1B) were then used to separately comb sequence and structure pattern data banks; in addition, diverse prediction algorithms (28, 45) were applied to the Lse domain sequences in order to classify their prospective fold types and to delineate  $\alpha$ -helical and  $\beta$ -strand segments, en route to a desired structural description of the Lse component folds and catalytic apparatus. However, only a faint but significant match to the 266-residue C-terminal domain of Lse was detected with the carboxyl-terminal portion of an open reading frame of unknown function at the 3'-end of a gene cluster involved in porphyrin biosynthesis from the anaerobic bacterium *Clostridium josui* (46) (Fig. 1D). Interestingly, the sequence of the bacterial open reading frame divulges an additional, likely incomplete, 159-amino acid N-terminal domain that appears to be unrelated to the mouse Lse amino-terminal (small) subunit. This different evolutionary pairing of domains suggests, first, that the heteromeric Lse enzyme is composed of two distinct globular protein modules and, second, that the function of the large Lse subunit can be married to other enzymatic schemes.

The catalytic activity of Lse has been shown to be abrogated by the serine active-site inhibitors diisopropyl fluorophosphate and diethyl-*p*-nitrophenyl phosphate; [ $^3$ H]diisopropyl fluorophosphate specifically labels the small enzyme subunit (11). However, inspection of the N-terminal domain sequence of Lse does not reveal the typical active-site sequence (Gly-X-Ser-X-Gly) of serine active-center esterases (47) or the derivative Gly-Asp-Ser-Arg-Thr/Ser signature of influenza and coronavirus sialic acid 9-O-acetyltransferases (48). A diverse number of esterase enzymes that contain the former catalytic serine motif compose a structural superfamily called the  $\alpha/\beta$ -hydrolases (49). These enzymes conserve a catalytic triad of Ser, Asp, and His residues positioned at the carboxyl-terminal end of a parallel  $\beta$ -sheet that forms the core of the hydrolase scaffold; similar active-site clefts are found in other doubly wound  $\alpha/\beta$ -folds (50). Sequence relationships between different  $\alpha/\beta$ -hydrolase enzymes are typically detected only after structural superposition of equivalent topological features (49). Comparison of the presumed Lse catalytic (small) subunit with this fold family must then rely on structural considerations. Predictive algorithms (28, 45) accordingly suggest that both Lse subunits are composed of alternating  $\alpha$ - and  $\beta$ -secondary structure (data not shown).

**Lse Is a Glycoprotein Detectable in Both the Intra- and Extracellular Compartments**—Lysosomal extracts from mouse liver have an activity corresponding to the rat Lse protein, *i.e.* concanavalin A-binding 9-O-acetyltransferase activity (data not

A

1 AATCAGGATCTTCACAACAATGGTTTCCCCGGGSCCTGTGTTGGGATAGTCTGCTCATTAATCGCGGAGTCAGCAGAAAGTCAGGTTATTGGTTTTGCTTTGCTTCATACATCGATAA 120  
 -23 M V S P G P V F G I V L L I I A R V S R S A G I G F R F A S Y I D N 11

121 TTACATGGTGTGCAGAAAGACCTTCAGGGGCTGTGATCTGGGGCTTTGGTACACCTGGAGCCAGTGCAGTGCACCTTGTGCCAAGTCCAGAAACCATCATGAAAGATGACCAG 240  
 12 Y M V L Q K E P S G A V L Q W F G P T V T V T L C Q G Q E T I M K K V T S 15

241 TGTGAAAGAACCCCTCAACACCTGGATGGTGTACTGGACCCATGAAGCCTGGGGGACCTTTTGAAGTGTGGCCACACAGACTTTGGGGACAATGAATTTACCCTGAGATCCATGA 360  
 52 V K E P S N T N M V V L D P M K P G G P F E V M A Q Q T L G T M N F T L R V H D 91

361 TGTCTTATTTGGAGATGTCTGGCTTTGCAGTGGGCAGAGTAACATGCAGATGACTTTTCCACAGATCTTTAAGCCCTCAAAGGAGTGTCCGACACTGCTGCCTACAGCTGTGCGCAT 480  
 92 V L F G D V W L C S G Q S N M Q M T V S Q I F N A S K E L S D T A A Y Q S V R I 131

481 CTTCTGTCTCCCTCACCACATCAGAGGAGGTTGGATGACCTTACTGAGTTCACCTGTCATGTCCTCAAGCCACCCGAGGAACTTTGGGCCATGGAAATTTCACTTACATGTCTGC 600  
 132 F S V S L T Q S E E E L D D L E V D L S W S K P T A G N L G H G N F T Y M S A 171

601 GGTGTGCTGGCTCTTTGGCGTTACCTGTATGACACTCTGCAATATCCCATGGGTTGGTCTCTCCAGTTCGGGTTGGGAGCTACATTAAGTCTGGTATCTAGAAGGACACTGAAAGC 720  
 172 V C W L F G R Y L Y D T L Q Y P I G L V S S S W G G T Y I E V W S S R R T L K A 211

721 CTGTGGAGTCCCTAATACAGGGATGAAAGAGTGGTGCAGCCGAGATAAAGCCCATGAGGAATGAATGTAACAGTGAAGGAGTCCATGCTCCTTTAGGGTTGTCCTTCTGTCTGTGT 840  
 212 C G V P N T R D E R V G Q P E I K P M R N E C N S E E S S C P F R V V P S V R V 251

841 GACTGGACCGACAAGACACTCCGCTCTTTGGAACGCCATGATCCATCCACTGAAAACATGACTTTGAAAGGGTGGTGGTACCAGGAGAGTCCAATCGGATATAACAGGGACCT 960  
 252 T G P T R H C S V L W N A M I H P L Q N M T L K K G V V W Y Q G E S N A Y K N R D L 291

961 GTACACCTGCATGTTTCTGAACTTGAAGACTGGCCGACCTTCCACTACGGCTCCAGGAGTGCAGAGATCGTTCCTTCCATTTGGGTTTCCAGCTGTCTCATACATGTT 1080  
 292 Y T C M F P E L I E D W R Q T F H Y G S Q G Q T D R F P F P G F V Q L S S Y M L 331

1081 AAAGAATCATCGGATATGGATTTCCAGAGATCCGTTGGCATCAAATCGACACTTTGGCCATGTCPCCAACCCCAAGATGCCTAATACTTTTATGGCTCGCTATTGACCTCTGTGA 1200  
 332 K N S S D Y G F P E I R W H Q T A D F G H V P N P K M P N T F M A V A I D L C D 371

1201 TAGAGACTCACCTTTGGAAAGTATCCACCTCGAGACAAACAGACTGTGGCCTATCGTCTGCACTTGGGTGCTGGAGCTGTGGCATATGGTGAAGAAAATTTGACCTTTCAAGGACCACT 1320  
 372 R D S P P G S I H P R D K Q T V A Y R L H L G A R A V A Y G E K N L T F Q G P L 411

1321 ACCTAAGAAGATAGAATCTTGGCTTCCAAATGGCCTGCTCAACCTCACATATGACAGGAAATCCAGGTGCAAAATGCAAGATAACAAAACATTTGAGATCTCCTGTTGTAGTAGCCGTC 1440  
 412 P K K I E L L A S N G L L N L T Y D Q E I Q V Q M Q D N K T F E I S C C S D R H 451

1441 TTGCAATGGCTCCAGCTCCTGTTAACACTTTTCCACCCAGACTGTGATTTGGATCTCAATGCTTGTCTGGCACAGTGGTGTCTGCTATGCTTGGACACCATGGCCCTTGTGA 1560  
 452 C K W L P A P V N T F S T Q T L I L D L N A L G T V V A V R Y A W T T W P C E 491

1561 ATATAAGCAATGTGCTGTTTACCACACCAGCAGTATGTTGCCAGCTCCTCCCTCAATGCTCAAATTTACACAGGGGTATCTAAGGACATTGGAGTAGTGCCTCTTCTATACAAG 1680  
 492 Y K Q C A V Y H T S S M L P A P P F I A Q I S H R G I 518

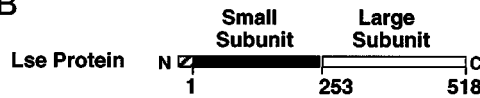
1681 GGCAATCTTTAGTTTGGGTATGACGAGACTTTAAAAGATGACTGCTGTTTTTAGGCCAATGGGATAAACCCCTMTTGAATAGCTCCAGCTAGCATATGGCCCTCCAGTATTCTGAC 1800

1801 ATAAGCACAGGCTAGCAGATCTTGAATGCATCATCTCTGCTCTAGTTAACCTAAGCCCTGTGCTGGTCTAATTTCTCACTATATCACTGGCCCTTACTAGCATTCACTTCTTGG 1920

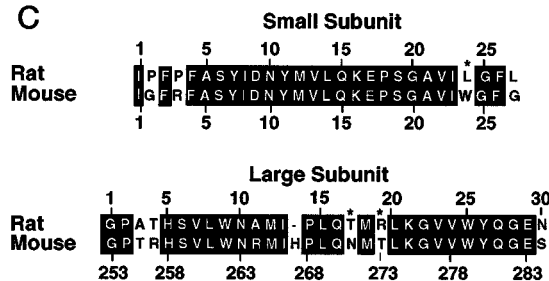
1921 TATTAGTGGGAGTAATATGCCCTATTCGTGCCCTTCTTCACTACTCTGGACTGTGGACTCTGGACCACGAAGGACTTTAAAGGCTGTGGAGGCTGGTCTTTGTGTGTTCTTTACAA 2040

2041 CCCCTGAAAACGGCTATGCAAGTTTCCATGCTGTTCTCTCTGTAACAACATGCTCTTGTCAAACCTTACTACCCCATGACAGGCTTTTCAGGTTGCCTGTGACAGATCATGTGTTTTA 2160  
 2161 TCACTAGGCTCTCCAAAGCTGTTGTGAGCCTAATCACTCTAAAGGGAATCTGTGCATATTGCCATGTGCTAAGGAGTGGCCTTAAGGAAGAAAGAGGGCCACTGCAGGTTCTCAGG 2280  
 2281 CAGAGGTTGGGAAATCACAGGCACAGGGACACAGGAAGTAGATGTGCTTGGGGCAATAGCTCTTTAAATGTAAGTCTTAGGTCCTGAGGTAAGTTCACTAAGCCTAAGCCTAA 2400  
 2401 TTCCCAGGAAGGAGGAGTTAGGCTGAAAGTTACAGACATGGCTTATAGACCTTACAGGTTGATTTTGTAGTTTACCATTGGTTAAAGACAGCTGCTAGCTGTGAAAGTGCACCTCT 2520  
 2521 AGAACTGGAAATAAGTCCAGCAGTACATACACTAATCCAGCACTCAGGAGGCGAGGTTGAGCTGTTCTACAGAAAGGCTCTAGGATAGCAAGTACCTAGAGAAACACAGCTCT 2640  
 2641 TGATCTCTCTGCCCCACAAAGAGAAAAAATAAAAAA 2680

B



C

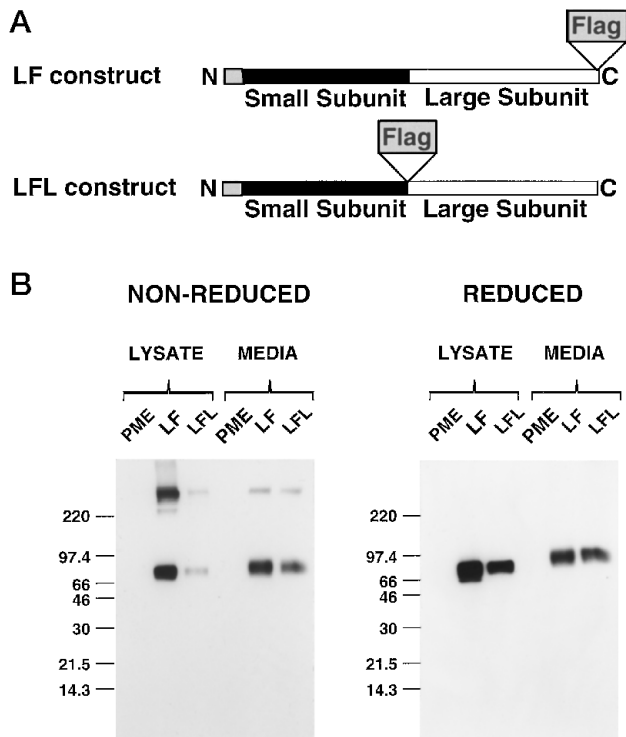


D

Mo LSE 253 G P T R H S V L W N A M I H P L Q N M T L K G V V W Y O G E S N A D Y N R D L Y T C M F P E L I E D W R Q T F H Y G S [6] F P F G F V O L S S Y M  
 Cl ORF 160 V Q W R P L G L Y N G M I A P V T S Y A I K G F I W Y Q G E A N T K . N P V G Y E N L L K A L I S D W R O K W G M G N . L P F L Y V Q L P N F M

Mo LSE 331 L K [23] P K M P N T F M A V A I D L C [4] P F G S I H P R D K Q T V A Y R L H L G A R A V A Y G E K N L T F C G P L P [65] V V A V R Y A W T T W P C [28]  
 Cl ORF 230 E A [19] L S V P N T G M I V T I D L G E W N D I H P S N K K D V G F R L A L A A M K T V Y G D K T I T A Y G P M Y [52] P A V V R Y A W A D N P Q [21]

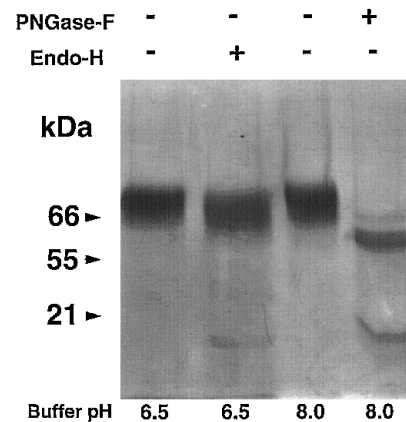
FIG. 1. Identification of a novel cDNA that encodes mouse Lse. A, nucleotide sequence of Clone 165-109. The rectangular box includes the signal peptide sequence. The arrow shows the beginning of the peptide sequence that is homologous to the larger subunit of the heterodimeric rat Lse protein. The arrowhead identifies an alternative polyadenylation site found in a different clone (Clone 165-8). Underlined residues represent potential N-glycosylation sites. B, diagrammatic representation of the mouse Lse protein. The hatched box represents the signal peptide, and numbers represent amino acid residues positions. C, alignment between the peptide sequence of rat Lse (14) and the deduced peptide sequence of mouse Lse. \*, residues in the rat Lse sequence for which inconclusive results had been reported in earlier sequencing (14). D, comparison of the mouse Lse small subunit sequence (Mo LSE) and the C-terminal domain from a Clostridium open reading frame (Cl ORF) (46). Identical residues are in reverse lettering; chemically conserved amino acids are in black boxes. The length of gaps in the compressed alignment are indicated by numbers (of residues) in brackets.



**FIG. 2. Expression of the LF and LFL constructs in COS-7 cells.** *A*, diagrammatic representation of the LF and LFL constructs showing the locations where the FLAG sequence was introduced into the Lse protein coding region. *B*, Western blot analysis of immunoprecipitates from lysates and media of COS-7 cells transiently transfected with the expression vector without insert (PME18X (PME)) and with the LF and LFL constructs. The immunoprecipitates, run under both reducing and nonreducing conditions, were obtained and probed with the anti-FLAG M2 antibody. The precursor peptide obtained with both constructs was not processed into its heterodimeric disulfide form even 80 h after transfection.

shown). However, the previously described polyclonal and monoclonal antibodies against the rat liver Lse protein (14) do not react with these extracts (data not shown). To monitor the production of the Lse protein, we therefore epitope-tagged Clone 165-109 with a FLAG peptide sequence (21) at the C terminus of the predicted open reading frame of the cDNA (LF construct). We also generated a construct for protein expression of mouse Lse in which the FLAG epitope was introduced immediately before the N terminus of the larger Lse subunit (LFL construct) (see Figs. 1 (A-C) and 2A). This allowed us to use the M2 monoclonal antibody, which recognizes the FLAG sequence, both for purification and detection of the protein by Western blot analysis. The LFL construct should also allow proteolytic cleavage using enterokinase, an enzyme for which a recognition site is available at the C terminus of the FLAG sequence, yielding a heterodimeric disulfide-bonded protein similar to the mature rat protein (Fig. 2A).

Following transient transfection of COS cells with these constructs, the cell lysates and media were immunoprecipitated with agarose-conjugated anti-FLAG antibody and subjected to Western blotting with anti-FLAG antibody after separation by nonreducing SDS-polyacrylamide gel electrophoresis. As shown in Fig. 2B (left panel), both constructs yielded an ~75-kDa protein. However, no major change in molecular mass was seen after reduction, indicating that the bulk of the expressed protein was not cleaved into the two subunits in COS cells (Fig. 2B, right panel). We observed an ~250-kDa protein band under nonreducing conditions that disappeared under reducing conditions and likely represents aggregates of LF and LFL proteins. We also observed that the LFL protein was less immu-

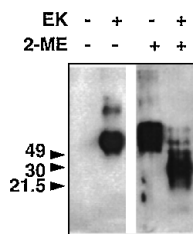


**FIG. 3. Expressed protein carries N-linked oligosaccharides.** The secreted protein encoded by the LFL construct was purified from the spent medium of transfected cells as described and concentrated, and aliquots were subjected to digestion with endo- $\beta$ -*N*-acetylglucosaminidase H (*Endo-H*) and peptide *N*-glycosidase F (*PNGase-F*) as described under "Experimental Procedures." Treated samples and untreated controls were run on a reducing SDS-polyacrylamide gel and stained with silver. The molecular mass markers indicate the positions of known standards.

noreactive under nonreducing conditions (Figs. 2B and 4), possibly due to the internal location of the FLAG peptide sequence in this protein. The difference in molecular mass from that predicted by the cDNA open reading frame (calculated polypeptide molecular mass is 58.441 kDa using the PEPTIDE-SORT software program from the Genetics Computer Group package) could be due to post-translational modifications (e.g. glycosylation) of the polypeptide by the COS cells. Indeed, as shown in Fig. 3, treatment of the purified secreted LFL protein with endo- $\beta$ -*N*-acetylglucosaminidase H (which cleaves high mannose and hybrid-type *N*-linked oligosaccharides) resulted in a small but reproducible reduction in apparent molecular mass, and treatment with peptide *N*-glycosidase F (which cleaves most known *N*-linked oligosaccharides) gave a further shift to an apparent molecular mass of ~62 kDa. Other post-translational modifications might account for the additional differences in molecular mass. The remaining discrepancy in molecular mass could be explained by other post-translational modifications or by failure of the peptide *N*-glycosidase F digestion to go to completion because of the extreme difficulty in denaturing the Lse protein.<sup>1</sup> Both possibilities are supported by the fact that the protein band remains somewhat diffuse, even after peptide *N*-glycosidase F treatment. The appearance of a small amount of low molecular mass material after peptide *N*-glycosidase F treatment likely represents exposure of the polypeptide backbone to a trace protease that was previously hindered by the oligosaccharide chains.

The protein purified after expression of the LFL construct was subjected to N-terminal gas-phase sequencing. The sequence obtained (Ile-Gly-Phe) corresponded to amino acids 24-26 of the open reading frame, confirming the identity of the expressed product and revealing the most likely exact location of the signal peptide cleavage site.

*Lse Has Sialic Acid O-Acetyltransferase Activity*—Rat Lse is gradually processed many hours after synthesis (presumably in the lysosome) into an ~58-kDa heterodimeric protein with two subunits of ~38 and ~28 kDa (11, 14). However, Western analysis under reducing conditions did not show cleavage of mouse Lse into two subunits in the COS cells following transfection with either the LF or LFL construct, even when cell harvesting was performed more than 80 h after transfection (Fig. 2B, right panel). This could be due to an inability of COS cell proteases to process the polypeptide precursor in the lyso-



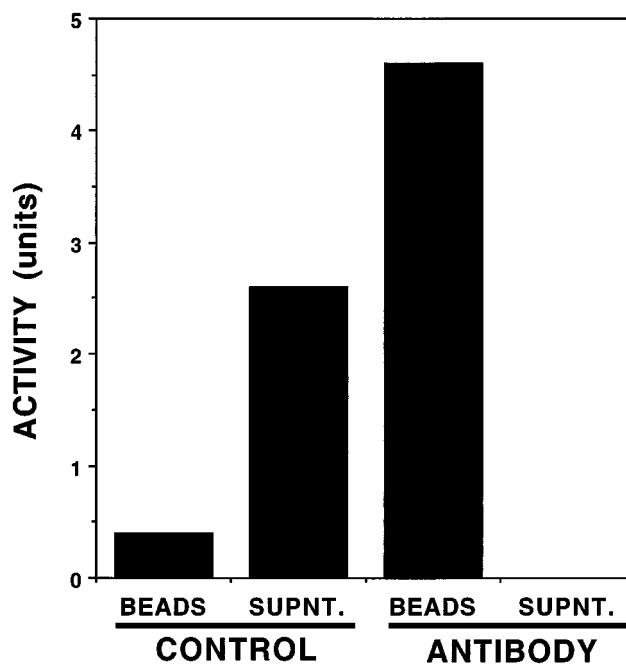
**FIG. 4. Enterokinase treatment of epitope-tagged LFL proteins expressed by transfected COS cells.** The secreted protein encoded by the LFL construct was purified from the spent medium of transfected cells as described and concentrated, and aliquots of 0.375  $\mu$ g were subjected to digestion with 2 units of enterokinase following the manufacturer (Biozyme, San Diego, CA) instructions. Reactions were boiled in sample buffer with or without reducing agents, separated by SDS-polyacrylamide gel electrophoresis, and blotted on Bio-Rad Trans-Blot filters, which were probed with anti-FLAG antibody. Note that the unreduced and uncleaved LFL protein is less immunoreactive, but can be detected upon longer exposure (data not shown). *EK*, samples treated with enterokinase; *2-ME*, samples treated with  $\beta$ -mercaptoethanol.

some or to an interference with the accessibility of the protease cleavage sites due to the introduction of the FLAG sequences. A similar finding was made when the secreted protein was analyzed (Fig. 2*B*, right panel), but this is in keeping with the fact that the secreted protein from rat hepatoma cells is also not cleaved (14).

To decrease the possibility of endogenous sialic acid esterase contamination and due to the lack of endopeptide cleavage of the intracellular protein expressed in COS-7 cells, we studied the anti-FLAG antibody-purified preparations derived from the medium of cells expressing the LFL construct to detect activity against [*O*-acetyl- $^3$ H]sialic acids. The growth medium from untransfected COS-7 cells yielded no activity (data not shown). The secreted LFL protein was obtained in microgram quantities and in apparently pure form (see Fig. 3 for an example). This molecule did not undergo spontaneous cleavage en route to being secreted (Fig. 4), but instead underwent the expected cleavage with enterokinase treatment *in vitro*. After enterokinase treatment, a diffuse subunit band of 35 kDa was detectable. This species likely corresponds to the small subunit because the FLAG peptide that is immunoreactive with the anti-FLAG antibody is expected to remain associated with the carboxyl-terminal end of the amino-terminal subunit.

The protein preparation derived from the LFL construct showed intrinsic sialic acid *O*-acetyltransferase activity (950 units/mg), and only a small increase was seen after enterokinase treatment (1300 units/mg). To confirm that the activity seen was not due to a cellular contaminant in the purified preparation, the protein was reprecipitated with anti-FLAG antibody immobilized on agarose. As shown in Fig. 5, the antibody-agarose beads specifically reprecipitated the activity. These results confirm that the murine cDNA that we isolated encodes a protein that is structurally and functionally similar to the previously described rat Lse protein.

*Expression of Lse during Mouse Development Is Spatially and Temporally Restricted: Lse Is Not Expressed in All Blood Cell Types*—We had previously observed two differently sized mRNAs (2.7 and 4.3 kb) reactive with Clone 165 in the yolk sac at day 8.5 of fetal development (17). Here we analyzed the expression of *Lse* mRNA at days 11.5 and 15 of fetal development to evaluate the expression of *Lse* in intraembryonic sites of hematopoiesis. We confirmed a higher abundance of the 4.3-kb message and the same proportional distribution of the two different messages mentioned above in all tissues analyzed (data not shown). At day 11.5, the yolk sac still is a site where the *Lse* mRNA is still abundant, but this changes by day 15 (Fig. 6). Interestingly, the day 11.5 aorta-gonada-mesonephros re-

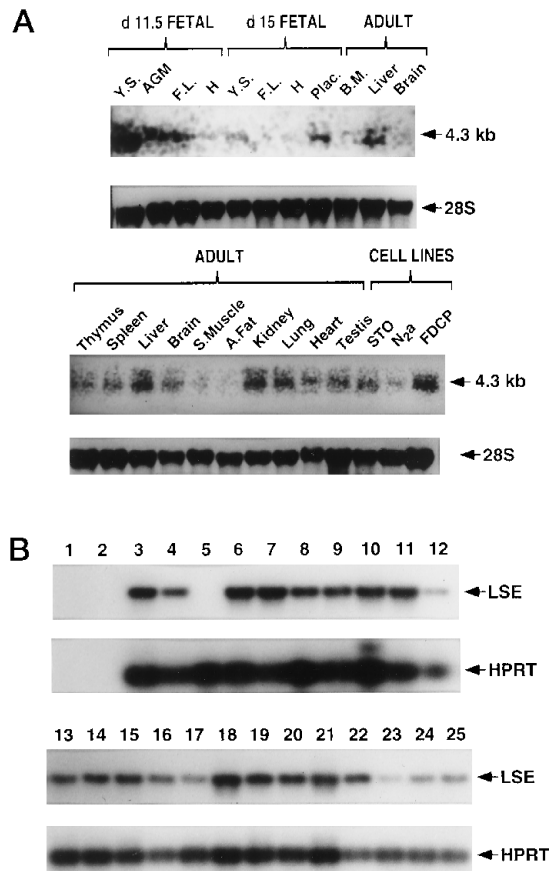


**FIG. 5. Reprecipitation of the sialic acid *O*-acetyltransferase activity purified from COS cells with the LFL construct.** Portions of the secreted protein encoded by the LFL construct (purified from the spent medium of transfected cells) were incubated with 20  $\mu$ l of anti-FLAG-Sepharose beads or control Bio-Gel A-agarose beads overnight with mixing. After spinning out the beads, the supernatant (*SUPNT.*) and the beads were assayed separately for sialic acid *O*-acetyltransferase activity as described under "Experimental Procedures."

gion and fetal liver, both active sites of embryonic hematopoiesis (17), express *Lse* more abundantly than the head primordium of the embryo. We also found a decrease in *Lse* mRNA in fetal liver from days 11.5 to 15, whereas at the adult stage, this was found to be the organ where the message is most abundant. The tissue distribution of the *Lse* mRNA in the adult generally parallels that reported for the rat *Lse* activity (11), with liver, testis, and kidney being locations of high expression, whereas skeletal muscle, adipose tissue, and heart have lower levels of *Lse* transcripts.

Given the high expression levels of the *Lse* mRNA in the hematopoietic cell line FDCPmixA4 relative to the fibroblastic cell line STO and the neuronal cell line N<sub>2</sub>a, we sought to analyze the expression of *Lse* mRNA by reverse transcription-PCR in a variety of blood cell-derived lines representing both lymphoid and myeloid lineages (Fig. 6*B*). The pluripotent myeloid precursor cell line NFS60 (51), the virus-transformed erythroleukemic cell line MEL (52), the macrophage cell lines B5A and P388D1, and the mast cell line PT18 (Fig. 6*B*) were all found to express the *Lse* message. Expression was also found in all the T-lymphocyte populations that were analyzed, including the T-helper-1 clone D1.1, the T-helper-2 clones D10.G4 and CDC25, and cortisone-treated thymocytes, as well as in the pre-B-cell line Clone K and a variety of B-lymphoma-derived cell lines. However, the bipotential B-cell and macrophage precursor cell line BL/3 did not express the *Lse* mRNA even after repeated 30-cycle PCR experiments. The placental stromal cell line PL1.1 and the thymic stroma of cortisone-treated mice express *Lse* mRNA.

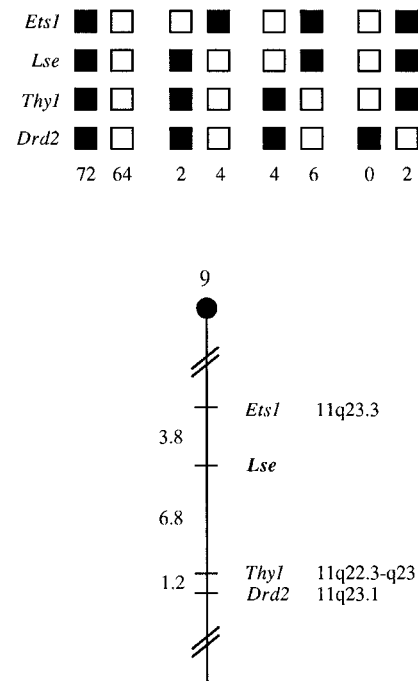
In a previous report, we detected, using Northern blot analysis, an up-regulation of the *Lse* mRNA levels by day 6 of the development of EBs (17). We decided to analyze the expression of *Lse* by reverse transcription-PCR to try to detect expression at earlier stages of the *in vitro* development of embryonic stem cells into EBs. As shown in Fig. 6*B* (lanes 22–25), we could



**FIG. 6. Northern and reverse transcription-PCR analyses of the expression of mouse *Lse*.** *A*, Northern analysis in fetal and adult tissues. The *Lse* mRNA is abundant in embryonic sites of active hematopoiesis at day 11.5 (yolk sac (Y.S.), aorta-gonada-mesonephros region (AGM), and fetal liver (F.L.)) as compared with the head primordium (H) of the embryo at the same stage of development. By day 15, there is a down-regulation of the expression of the message in the tissues analyzed. High levels of expression were also found in the placenta (Plac.) and in a variety of adult tissues, with the highest expression in liver, kidney, and testis. B.M., bone marrow; S.Muscle, skeletal muscle; A.Fat, perivisceral abdominal fat; FDCP, hematopoietic cell line FDCP-mixA4; STO, fibroblastic cell line STO; N<sub>2</sub>a, neuronal cell line N<sub>2</sub>a. The 28 S rRNA control used for loading is indicated. *B*, reverse transcription-PCR analysis of *Lse* expression in blood cells, day 8.5 embryos, embryonic stem cells, and EBs. Lane 1, no cDNA added; lane 2, reverse transcriptase control (FDCPmixA4 RNA to which no reverse transcriptase was added during the reaction); lane 3, late day 8.5 embryo; lane 4, day 8.5 yolk sac; lane 5, bipotential B-cell and myeloid cell line BL/3; lane 6, interleukin-7-responsive pre-B-cell line Clone K; lane 7, B-cell line B9; lanes 8–10, B-cell lymphoma-derived cell lines KD83, CH12, and A20, respectively; lane 11, placenta-derived cell line PL1-1; lane 12, thymic stroma from cortisone-treated mice; lanes 13 and 14, macrophage cell lines P388D1 and B5A, respectively; lane 15, mast cell line PT18; lane 16, cortisone-treated thymocytes; lanes 17–21, HC7 cell clone derived from CD4<sup>+</sup>CD8<sup>-</sup>negative thymocytes, HT2 T-helper clone, T-helper-1 clone D1.1, and T-helper-2 clones CDC25 and D10.G4, respectively; lane 22, embryonic stem cells; lanes 23–25, day 3, day 6, and day 9 EBs. HPRT, hypoxanthine phosphoribosyltransferase.

detect *Lse* expression in embryonic stem cells and an apparent decrease in mRNA levels upon the onset of differentiation of these cells followed by an up-regulation at later stages of the development of EBs (day 6 EBs). To exclude the possibility that the signals could be due to genomic DNA contamination of the RNA, we determined that genomic DNA from HeJ mice produces a distinctive ~600-base pair PCR product, but not the 494-base pair product observed with these various cDNAs (data not shown).

**Chromosomal Mapping of Mouse *Lse***—Southern blot analysis of mouse *Lse* with three different enzymes allowed us to identify



**FIG. 7. *Lse* was placed on mouse chromosome 9 by interspecific backcross analysis.** The segregation patterns of *Lse* and flanking genes in 154 backcross animals that were typed for all loci are shown at the top of the figure. For individual pairs of loci, more than 154 animals were typed (see "Results"). Each column represents the chromosome identified in the backcross progeny that was inherited from the (C57BL/6J × *M. spretus*)F<sub>1</sub> parent. The black boxes represent the presence of a C57BL/6J allele, and the white boxes represent the presence of a *M. spretus* allele. The number of offspring inheriting each type of chromosome is listed at the bottom of each column. A partial chromosome 9 linkage map showing the location of *Lse* in relation to linked genes is shown at the bottom of the figure. Recombination distances between loci in centimorgans are shown to the left of the chromosome, and the positions of the corresponding loci in human chromosomes, where known, are shown to the right. References for the human map positions of loci cited in this study can be obtained from Genome Data Base, a computerized data base of human linkage information maintained by the William H. Welch Medical Library of The Johns Hopkins University (Baltimore).

a pattern consistent with the existence of a single *Lse* gene that extends over an ~15-kb genomic DNA sequence (data not shown). The mouse chromosomal location of *Lse* was determined by interspecific backcross analysis using progeny derived from matings of [(C57BL/6J × *M. spretus*)F<sub>1</sub> × C57BL/6J] mice. This interspecific backcross mapping panel has been typed for over 1900 loci that are well distributed among all the autosomes as well as the X chromosome (41). C57BL/6J and *M. spretus* DNAs were digested with several enzymes and analyzed by Southern blot hybridization for informative restriction fragment length polymorphisms using a mouse cDNA probe. The 5.7-kb *TaqI* *M. spretus* restriction fragment length polymorphism (see "Experimental Procedures") was used to follow the segregation of the *Lse* locus in backcross mice. The mapping results indicate that *Lse* is located in the central region of mouse chromosome 9, linked to *ets1*, *thy1*, and *drd2*. Although 154 mice were analyzed for every marker and are shown in the segregation analysis (Fig. 7), up to 162 were typed for some pairs of markers. Each locus was analyzed in pairwise combinations for recombination frequencies using the additional data. The ratios of the total number of mice exhibiting recombinant chromosomes to the total number of mice analyzed for each pair of loci and the most likely gene order are as follows: centromere, *ets1*, 6:158, *Lse*, 11:162, *thy1*, 2:162, *Drd2*. The recombination frequencies (expressed as genetic distances in

centimorgans  $\pm$  S.E.) are as follows: *ets1*,  $3.8 \pm 1.5$ , *lse*,  $6.8 \pm 2.0$ , *thy1*,  $1.2 \pm 0.9$ , *drd2*. The central region of mouse chromosome 9 is syntenic with human chromosome 11q (summarized in Fig. 7), suggesting that human *lse* will reside on chromosome 11q as well.

We have compared our interspecific map of chromosome 9 with a composite mouse linkage map that reports the location of many uncloned mutations (provided by Mouse Genome Data Base, a computerized data base maintained at the Jackson Laboratory). There are four known mouse mutations that map in a similar region of chromosome 9: *luxoid* (*lu*), *variable spotting* (*vs*), *rough fur* (*rf*), and *rough coat* (*rc*). Of these, DNA was available for the first three, from the Jackson Laboratory. Southern blots of restriction digests similar to those of mice with identical background but no abnormal phenotype showed no obvious genomic differences in their DNAs (data not shown).

#### DISCUSSION

We have isolated a cDNA that encodes the sialic acid-specific modifying enzyme Lse and showed that mouse Lse has the sialic acid esterase activity originally described for the purified rat protein. Interestingly, and as previously shown for the rat Lse protein (14), mouse Lse is also partially secreted. Future studies will explore if the Lse protein secreted into the extracellular compartment can be taken up by cells and rendered active by proteolytic digestion and if the uncleaved native secreted Lse protein is enzymatically competent. In this regard, it would be interesting to know if this lysosomal protein carries the mannose 6-phosphate recognition marker found on most other soluble lysosomal enzymes with a similar itinerary (53).

All sialic acid esterase activities studied to date can be irreversibly inactivated by diisopropyl fluorophosphate (48), indicating the presence of a Ser active-site mechanism analogous to the extended family of serine esterases and proteases (47). Prior studies with rat Lse have intimated that the small subunit (here shown to be the amino-terminal domain) was covalently modified by [ $^3$ H]diisopropyl fluorophosphate (11). The notable absence of a hydrolase-like catalytic Ser motif (Gly-X-Ser-X-Gly) in Lse is tempered by the recent finding of a *bona fide*  $\alpha/\beta$ -hydrolase fold in a myristoyl-acyl carrier protein-specific thioesterase from *Vibrio harveyi* where the active-site Ser residue is embedded in an Ala-Ala-Ser-Leu-Ser motif (54). In this vein, an imperfect hydrolase Ser motif (Gly-Gln-Ser, residues 102–104) is evident in Lse, forming a loop between a strongly predicted  $\beta$ -strand and an  $\alpha$ -helix; a potential active-site His residue is present at position 163 in a Gly-His-Gly sequence similar to other  $\alpha/\beta$ -hydrolase His motifs (49). Alternatively, the Lse catalytic subunit could fold with an  $\alpha/\beta$ -topology distinct from  $\alpha/\beta$ -hydrolases. For example, the bacterial response regulator protein CheB, a Ser active-center methylesterase, has a doubly wound  $\alpha/\beta$ -fold distinct from the hydrolase superfamily with a convergent catalytic triad of Ser, His, and Asp (55). As other mammalian homologs of Lse are characterized, it may be possible both to sharpen the predicted secondary structure of Lse and to winnow the number of catalytically relevant residues by conservation analysis (28, 56), allowing more accurate investigation of specific candidate motifs for enzymatic activity.

Our lack of knowledge of the precise biological functions of this enzyme makes it difficult to predict the phenotype that might be expected for an alteration in its activity in the intact animal. *lse* mapped in a region of the composite map that includes the mouse mutations *luxoid* (*lu*), *variable spotting* (*vs*), and *rough fur* (*rf*). A preliminary screen of these mutants (*lu*, *vs*, and *rf*) failed to show gross alterations of the *lse* genomic region. However, this result does not rule out minor mutations that must be investigated at the nucleotide sequence level. In

this regard, it is important to note that earlier studies (57) reported that the extent of 9-O-acetylation of murine erythrocyte sialic acids was affected by an unknown locus on chromosome 9; one possibility is that this locus corresponds to *lse* and that allelic variations in the enzyme are involved in this polymorphism.

As a lysosomal enzyme, the Lse protein might be involved only in the terminal degradation of O-acetylated sialic acids. On the other hand, the differential expression of Lse mRNA throughout both *in vivo* and *in vitro* development suggests a role for the de-O-acetylation of sialic acids in developmental processes. If so, this would explain why the cellular enzyme binds well to concanavalin A, while the secreted form shows only a small shift with endo- $\beta$ -N-acetylglucosaminidase H digestion. In this regard, transgenic mice constitutively overexpressing a viral protein with de-O-acetylation activity at the cell surface have been found to arrest development at the two-cell stage embryo or to show developmental abnormalities at later stages (58). More important, we have observed low levels of *lse* mRNA upon induction of embryonic stem cell differentiation, followed by an increase in *lse* mRNA levels at later stages of the *in vitro* development of EBs (Fig. 6B) (16). An interesting aspect to note is the preferential expression of the message in active sites of hematopoiesis during fetal development: the yolk sac, aorta-gonada-mesonephros region (17), and fetal liver. Moreover, the pluripotent hematopoietic cell line FDCPmixA4 shows higher levels of the message than the fibroblastic cell line STO or the neuronal cell line N<sub>2</sub>a. It will be challenging to find out what is the role played by this sialic acid modification in the yolk sac or during the early stages of the *in vitro* development of EBs. It is possible that by influencing the hydrophobicity, conformation, or structure of glycoconjugates, de-O-acetylation may influence the cellular interactions that occur during embryogenesis and possibly have a role in the development of blood and/or primordial germinative cells as these cells migrate from the yolk sac to home to their intraembryonic compartments.

Analysis of *lse* expression in blood cell lines showed that the message is widely distributed among various blood cell populations. However, the finding of a bipotential B-cell and myeloid progenitor that does not express the message points to the possibility that *lse* may be differentially expressed throughout successive steps of the ontogeny of different blood cell types, namely B- and T-lymphocytes. Indeed, selective expression of O-acetylated sialic acids has been reported in subpopulations of thymocytes and peripheral leucocytes (59–63). Further studies involving *in situ* staining techniques to analyze the expression of *lse* in lymphoid organs and purified populations of lymphocytes will be required to characterize these differences. Moreover, immunohistological studies of lymphoid tissues show that potential ligands capable of mediating CD22 $\beta$ -dependent B-cell adhesion events are masked by 9-O-acetylation of sialic acids on specific cell types and are regionally distributed (62). Specifically, it was found that masking of CD22 $\beta$  ligands by 9-O-acetylation is differentially regulated on the same cell type between two different lymphoid organs. The availability of *lse* cDNA presented here, in conjunction with studies on the differential activity of the enzymes involved in the synthesis of 9-O-acetylated sialic acids, will help to explore the involvement of 9-O-acetyl esters in the molecular mechanisms by which lymphocytes within lymphoid tissues are segregated into distinct microdomains (62). Furthermore, T-cells of patients with various malignancies have been reported to acquire O-acetylation, and the addition of a single O-acetyl group to a cell surface-associated ganglioside can create an unusual cell type-specific antigen (63). The limited exploration of these matters



is in great part due to the lack of mammalian cDNAs encoding the enzymes that regulate O-acetylation. This report will therefore constitute an important advance in the study of the role of de-O-acetylation of sialic acids in various aspects of mammalian biology, including development, hematopoiesis, and tumorigenesis.

**Acknowledgments**—We thank Professors J. M. Pina Cabral and A. Coutinho for guidance. We also thank Dr. T. McClanahan for kindly providing the blood cell-derived cDNAs; Drs. F. Lee and R. A. Kastelein for advice; B. Devaux for critical reading of the manuscript; D. Campbell, D. Gilbert, and D. A. Barnhart for excellent assistance; and Dr. G. Hardiman for encouragement.

## REFERENCES

- Schauer, R. (1982) *Cell Biology Monograph*, Vol. 10, Springer-Verlag, New York
- Schauer, R. (1991) *Glycobiology* **1**, 449–452
- Varki, A. (1992) *Glycobiology* **2**, 25–40
- Varki, A. (1993) *Glycobiology* **3**, 97–130
- Roth, J. (1993) *Histochem. J.* **25**, 687–710
- Diaz, S., Higa, H. H., Hayes, B. K., and Varki, A. (1989) *J. Biol. Chem.* **264**, 19416–19426
- Herrler, G., Rott, R., Klenk, H.-D., Muller, H.-P., Shukla, A. K., and Schauer, R. (1985) *EMBO J.* **4**, 1503–1506
- Vlasak, R., Luytjes, W., Spaan, W., and Palese, P. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 4526–4529
- Vlasak, R., Muster, T., Lauro, A. M., Powers, J. C., and Palese, P. (1989) *J. Virol.* **63**, 2056–2062
- Herrler, G., and Klenk, H.-D. (1991) *Adv. Virus Res.* **40**, 213–234
- Higa, H. H., Manzi, A., and Varki, A. (1989) *J. Biol. Chem.* **264**, 19435–19442
- Schauer, R., Reuter, G., and Stoll, S. (1988) *Biochimie (Paris)* **70**, 1511–1519
- Schauer, R., Reuter, G., Stoll, S., and Shukla, A. K. (1989) *J. Biochem. (Tokyo)* **106**, 143–150
- Butor, C., Higa, H. H., and Varki, A. (1993) *J. Biol. Chem.* **268**, 10207–10213
- Butor, C., Griffiths, G., Aronson, N. N., Jr., and Varki, A. (1995) *J. Cell Sci.* **108**, 2213–2219
- Guimarães, M. J., Lee, F., Zlotnik, A., and McClanahan, T. (1995) *Nucleic Acids Res.* **23**, 1832–1833
- Guimarães, M. J., Bazan, J. F., Zlotnik, A., Wiles, M. V., Grimaldi, J. C., Lee, F., and McClanahan, T. (1995) *Development (Camb.)* **121**, 3335–3346
- Ford, A. M., Healy, L. H., Bennett, C. A., Navarro, E., Spooncer, E., and Greaves, M. F. (1992) *Blood* **79**, 1962–1971
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Benson, D. A., Boguski, M., Lipman, D. J., and Ostell, J. (1994) *Nucleic Acids Res.* **22**, 3441–3444
- Hopp, T. P., Prickett, K. S., Price, U., Libby, R. T., March, C. J., Cerretti, P., Urdal, D. L., and Conlon, P. J. (1988) *Bio/Technology* **6**, 1205–1210
- Vallette, F., Mege, E., Reiss, A., and Adanski, M. (1989) *Nucleic Acids Res.* **17**, 723–733
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410
- Worley, K. C., Wiese, B., and Smith, R. F. (1995) *Genome Res.* **5**, 173–184
- Altschul, S. F., Boguski, M. S., Gish, W., and Wootton, J. C. (1994) *Nat. Genet.* **6**, 119–129
- Bork, P., Ouzounis, C., and Sander, C. (1994) *Curr. Opin. Struct. Biol.* **4**, 393–403
- Bork, P., and Gibson, T. (1996) *Methods Enzymol.* **266**, 162–184
- Rost, B., and Sander, C. (1996) *Ann. Rev. Biophys. Biomol. Struct.* **25**, in press
- Hansen, J. E., Lund, O., Engelbrecht, J., Bohr, H., Nielsen, J. O., Hansen, J. E. S., and Brunak, S. (1995) *Biochem. J.* **308**, 801–813
- Rennick, D., Jackson, J., Moulds, C., Lee, F., and Yang, G. (1989) *J. Immunol.* **142**, 161–166
- Hunt, P., Robertson, D., Weiss, D., Rennick, D., Lee, F., and Witte, O. N. (1987) *Cell* **48**, 997–1007
- Everson, M. P., Brown, C. B., and Lilly, M. B. (1989) *Blood* **74**, 1472–1476
- Arnold, L. W., LoCasio, N. J., Lutz, P. M., Pennell, C. A., Klapper, D., and Haughton, G. (1983) *J. Immunol.* **131**, 2064–2073
- Kim, K. J., and Dejoy, S. Q. (1986) *Immunology* **59**, 15–21
- Bauer, S. R., Holmes, K. L., Morse, H. C., III, and Potter, M. (1986) *J. Immunol.* **136**, 4695–4699
- Udem, B. J., Torphy, T. J., Goldman, D., and Chilton, F. H. (1990) *J. Biol. Chem.* **265**, 6750–6758
- MacNeil, I., Kennedy, J., Godfrey, D. I., Jenkins, N. A., Masciantonio, M., Mineo, C., Gilbert, D. J., Copeland, N. G., Boyd, R. L., and Zlotnik, A. (1993) *J. Immunol.* **151**, 6913–6923
- Watson, J. (1979) *J. Exp. Med.* **150**, 1510–1519
- Lederer, J. A., Liou, J. S., Todd, M. D., Glimcher, L. H., and Lichtman, A. H. (1993) *J. Immunol.* **152**, 77–86
- Deibel, M. R., Jr., and Killar, L. M. (1988) *Lymphokine Res.* **7**, 469–477
- Copeland, N. G., and Jenkins, N. A. (1991) *Trends Genet.* **7**, 113–118
- Wilkie, T. M., Chen, D. J., Moore, K. J., Yu, L., Simon, M. I., Copeland, N. G., and Jenkins, N. A. (1993) *Genomics* **18**, 175–184
- Green, E. L. (1981) *Genetics and Probability in Animal Breeding Experiments*, pp. 77–113, Oxford University Press, New York
- Boguski, M. S. (1995) *Trends Biochem. Sci.* **20**, 295–296
- Bazan, J. F. (1996) *Proteins* **24**, 1–17
- Fujino, E., Fujino, T., Karita, S., Sakka, K., and Ohmiya, K. (1995) *J. Bacteriol.* **177**, 5169–5175
- Brenner, S. (1988) *Nature* **334**, 528–530
- Hayes, B. K., and Varki, A. (1989) *J. Biol. Chem.* **264**, 19443–19448
- Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J. L., Verschuere, K. H. G., and Goldman, A. (1992) *Protein Eng.* **5**, 197–211
- Branden, C. I. (1980) *Q. Rev. Biophys.* **13**, 317–338
- Bourette, R. P., Mouchiroud, G., Ouazana, R., Morlé, F., Godet, J., and Blanchet, J.-P. (1993) *Blood* **81**, 2511–2520
- Marks, P. A., and Rifkind, R. A. (1979) *Annu. Rev. Biochem.* **47**, 419–448
- Kornfeld, S. (1992) *Annu. Rev. Biochem.* **61**, 307–330
- Lawson, D. M., Derewenda, U., Serre, L., Ferri, S., Sztittner, R., Wei, Y., Meighen, E. A., and Derewenda, Z. S. (1994) *Biochemistry* **33**, 9382–9388
- West, A. H., Martinez-Hackert, E., and Stock, A. M. (1995) *J. Mol. Biol.* **250**, 276–290
- Zvelebil, M. J., and Sternberg, M. J. (1988) *Protein Eng.* **2**, 127–138
- Varki, A., and Kornfeld, S. (1980) *J. Exp. Med.* **152**, 532–544
- Varki, A., Hooshmand, F., Diaz, S., Varki, N. M., and Hedrick, S. M. (1991) *Cell* **65**, 65–74
- Kamerling, J. P., Makovitzky, J., Schauer, R., Vliegenthart, J. F. G., and Wember, M. (1982) *Biochim. Biophys. Acta* **714**, 351–355
- Knip, B., Claus, C., Peter-Katalinic, J., Monner, D. A., Dippold, W., and Nimtz, M. (1995) *J. Biol. Chem.* **270**, 30173–30180
- Zimmer, G., Suguri, T., Reuter, G., Yu, R. K., Schauer, R., and Herrler, G. (1994) *Glycobiology* **4**, 343–350
- Sjoberg, E. R., Powell, L. D., Klein, A., and Varki, A. (1994) *J. Cell Biol.* **126**, 549–562
- Cheresh, D. A., Reisfeld, R. A., and Varki, A. P. (1984) *Science* **225**, 844–846