Molecular Cloning and Characterization of Lysosomal Sialic Acid O-Acetylylesterase*

(Received for publication, February 7, 1996)

M. Jorge Guimarães§, J. Fernando Bazant, Janice Castagnola¶, Sandra Diaz¶, Neal G. Copeland, Debra J. Gilbert, Nancy A. Jenkins, Ajit Varki†, and Albert Zlotnik‡

From the DNAX Research Institute of Molecular and Cellular Biology, Palo Alto, California 94304, the Glycobiology Program, University of California at San Diego Cancer Center, La Jolla, California 92093-0063, and the 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 amplification. 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-acetylylesterase (Lse) previously purified from rat liver, but with no evident similarity to endoplasmic reticulum-derived acetylylesterases. The expressed Lse protein exhibits sialic-acid O-acetylylesterase 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-acetylylesterase 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-acetylylesterase 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 organisms (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-acetylylesterase activity was purified from rat liver membrane compartments and initially called luminal sialic acid O-acetylylesterase (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-acetylylesterase (retaining the acronym Lse). Early attempts to clone this esterase by conventional methods were thwarted by poor immunoreactivity and extreme resistance to proteolytic digestion. Only two amino acid sequences were obtained (14), but these were not strongly immunogenic or useful for cloning using degenerate oligonucleotides.2

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 in vitro development of embryoid bodies (EBs)3 (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-acetylylesterase 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 FDCP-mixA4 (18), cultured as described (17), was obtained using RNAzol solution (Tel-test Inc., Friendswood, TX)
following the manufacturer's instructions. Poly(A) + 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 FD-

CPmixA4-derived library (19). Potential clones were purified and sub-

cloned (16). The DNA 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 cDNA (data not shown). The PCR amplification used the sense primer 5'-CACTTGGGCGCGCAATGGTTTTCGGGGGCGGTCGGTGTGGG-3' and antisense primer 5'-CGTACATGTTTTACTTGTACCTGGCATACCCGGG-3'. Two sets of primers with independent reactions were used in the first round of PCR for the generation of the LFL construct: sense primer (set 1), 5'-CACTTTGGGCGCGCAATGGTTTTCGGGGGCGGTCGGTGTGGG-3'; antisense primer (set 1), 5'-GGGCTTCAAGAAGGAGGAGGATCCTTCTGCTGCTTGGCTGCTTGGCCTGACAGAGAAGAGAGAGACCTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
Lysosomal Sialic Acid O-Acetyleresterase

primer and 5'-CCAAAGGTGGTGCCTGTTACTAG-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'-CGGATTACCACTGTTACG-3') after transfer to a nylon membrane (19). The probe was labeled at the 5'-end with [γ-32P]ATP (Amersham Corp.) using polynucleotide kinase (Boehringer Mannheim). Blots were hybridized in 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 EcoRI, HindIII, and PstI and processed as described (19) for Southern blot analysis using as probe entire random primer-encoded Clone 165-109 DNA.

Interspecific progeny were generated by mapping (C57BL/6J × Mus musculus) females and C57BL/6J males as described (41). A total of 205 N2 mice were used to map the 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 were prepared with Hybond-N+ (Amersham Corp.). The probe, an ~2.7-kb fragment of mouse cDNA, was labeled with [α-32P]dCTP using a random primer labeling kit (Stratagene); washed in 0.5 × NaHPO4 (pH 7.2), 7% SDS, and 0.1% BSA at 65 °C. Major fragments of 7.4 and 5.7 kb were detected in TaqI-digested M. musculus DNA. The presence or absence of the 5.7-kb TaqI M. musculus 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 leukaemia oncogene 1 (ets1), thymus cell antigen 1 (thyl), and dopamine receptor 2 (drd2), has been reported previously (42). Recombination distances were calculated as described (43) using the computer program MAPMAKER. 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 under conditions of expression was a novel 1.5-kb cDNA containing a different polyadenylation site at the 3′ end of the message (44). Further analysis of this clone, designated Clone 165-8, revealed that 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 databases, 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 carboxy-terminal domains of Lse (equivalent to the small and large subunits of Fig. 1B) were then used to separately comb sequence and structure pattern databases; in addition, diverse prediction algorithms (28, 45) were applied to the Lse domain sequences in order to classify their respective prospective type II and to delineate α-helical and β-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 carboxy-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 joshi (46) (Fig. 1D). Interestingly, the sequence of the bacterial open reading frame diverges 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 pair 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; [3H]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-acetyleresterases (48). A diverse number of esterase enzymes that contain the catalytic serine motif compose a structural superfamily called the α/β-hydrolases (49). These enzymes conserve a catalytic triad of Ser, Asp, and His residues positioned at the carboxy-terminal end of a parallel β-sheet that forms the core of the hydrolase scaffold; similar active-site clefts are found in other doubly wound hydrolases. Sequence relationships between different α/β-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 α- and β-secondary structure (data not shown).

Lse is a Glycoprotein Detectable in Both the Intracellular and Extracellular Compartments—Lysosomal extracts from mouse liver have an activity corresponding to the rat Lse protein, i.e. concanavalin A-binding 9-O-acetyleresterase activity (data not shown).
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 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 (CI 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.
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 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.

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 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 LF 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 reduced conditions and likely represents aggregates of LF and LFL proteins. We also observed that the LFL protein was less immunoreactive 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-β-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. 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-Acetylacesterase 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 lysosome.
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 μg 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 β-mercaptoethanol.

FIG. 5. Reprecipitation of the sialic acid O-acetylemesterase 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 2 μl 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-acetylemesterase activity as described under “Experimental Procedures.”

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. 2B, 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-9-3H]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-acetylemesterase 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 region 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 N2a, 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. 6B). The pluripotent myeloid precursor cell line NF560 (51), the virus-transformed erythroleukemic cell line MEL (52), the macrophage cell lines B5A and P388D1, and the mast cell line PT18 (Fig. 6B) 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 D11.1, 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 cordisone-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. 6B (lanes 22-25), we could...
detect in RNA 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 ESs (day 6 ESs). To exclude the possibility that the signals could be due to genomic DNA contamination of 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; M. Muscie, skeletal muscle; A. Fat, perivisceral abdominal fat; FDCP, hematopoietic cell line FDCP-mixA4; STO, fibroblastic cell line STO; N,a, neuronal cell line N,a. The 28 S RNA control used for loading is indicated. B, reverse transcription-PCR analysis of Ise expression in blood cells, day 8.5 embryos, embryonic stem cells, and EBs. Lane 1, no cDNA added; lane 2, reverse transcription control (FDCPmixA4A 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, biotinylated 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 KDB3, 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 PT16; lane 16, cortisone-treated thymocytes; lanes 17-21, HC7 cell clone derived from CD4-CD8-negative thymocytes, HT2 T-helper clone, T-helper-1 clone D1L1, and T-helper-2 clones CDC25 and D10G4, respectively; lane 22, embryonic stem cells; lanes 23-25, day 3, day 6, and day 9 ESs. HPRT, hypoxanthine phosphoribosyltransferase.

Chromosomal Mapping of Mouse Ise—Southern blot analysis of mouse Ise with three different enzymes allowed us to identify a pattern consistent with the presence of a single Ise gene that extends over an ~15-kb genomic DNA sequence (data not shown). The mouse chromosomal location of Ise was determined by interspecific backcross analysis using progeny derived from matings of [(C57BL/6J X M. spretus)F1, 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 Ise 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 database of human linkage information maintained by the William H. Welch Medical Library of The Johns Hopkins University (Baltimore).
centimorgans ± S.E.) are as follows: bta1, 3.8 ± 1.5; lse, 6.8 ± 2.0; thy1, 1.2 ± 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). One 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 extra-cellular 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 disopropyl 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 [3H]diisopropyl fluorophosphate (11). The notable absence of a hydrophobic-like catalytic Ser motif (Gly-X-Ser-X-Gly) in Lse is tempered by the recent finding of a bona fide α/β-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 hydrophobic Ser motif (Gly-Gln-Ser, residues 102–104) is evident in Lse, forming a loop between a strongly predicted α-strand and an α-helix; a potential active-site His residue is present at position 163 in a Gly-His-Gly sequence similar to other α/β-hydrolase His motifs (49). Alternatively, the Lse catalytic subunit could fold with an αβ-topology distinct from α/β-hydrolases. For example, the bacterial response regulator protein CheB, a Ser active-center methyltransferase, has a doubly wound α/β-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 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-β-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 N2a. 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-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 CD22β 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. Kastelian 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.