Alternative Splicing of β-Galactosidase mRNA Generates the Classic Lysosomal Enzyme and a β-Galactosidase-related Protein*

Hans Morreau, Niels J. Galjart, Nynke Gillemans, Rob Willemsen, Gijsbertus T. J. van der Horst, and Alessandra d'Azzo

From the Department of Cell Biology and Genetics, Erasmus University, Rotterdam, The Netherlands

We have isolated two cDNAs encoding human lysosomal β-galactosidase, the enzyme deficient in GM1-gangliosidosis and Morquio B syndrome, and a β-galactosidase-related protein. In total RNA from normal fibroblasts a major mRNA of about 2.5 kilobases (kb) is recognized by cDNA probes. A minor transcript of about 2.0 kb is visible only in immunoselected polysomal RNA. A heterogeneous pattern of expression of the 2.5-kb β-galactosidase transcript is observed in fibroblasts from different GM1-gangliosidosis patients. The nucleotide sequences of the two cDNAs are extensively colinear. However, the short cDNA misses two noncontiguous protein-encoding regions (1 and 2) present in the long cDNA. The exclusion of region 1 in the short molecule introduces a frameshift in its 3' flanking sequence, which is restored by the exclusion of region 2. These findings imply the existence of two mRNA templates, which are read in a different frame only in the nucleotide stretch between regions 1 and 2. Sequence analysis of genomic exons of the β-galactosidase gene shows that the short mRNA is generated by alternative splicing. The long and short cDNAs direct the synthesis in COS-1 cells of β-galactosidase polypeptides of 85 and 68 kDa, respectively. Only the long protein is catalytically active under the assay conditions used, and it is capable of correcting β-galactosidase activity after endocytosis by GM1-gangliosidosis fibroblasts. The subcellular localization of cDNA-encoded β-galactosidase and β-galactosidase-related proteins is different.

Acid β-N-galactosidase (EC 3.2.1.23) is the lysosomal hydrolase that cleaves β-linked terminal galactosyl residues from gangliosides, glycoproteins, glycosaminoglycans, as well as a variety of artificial substrates (reviewed in Refs. 1, 2). The gene coding for the human enzyme has been localized on chromosome 3 (3). Mutations in the β-galactosidase locus cause deficient or reduced enzyme activity and pathological accumulation of undigested metabolites in lysosomes. The resulting metabolic storage diseases are GM1β-gangliosidosis and Morquio B syndrome (4–6). Among GM1-gangliosidosis patients different clinical phenotypes have been described that are classified as severe infantile, juvenile or mild infantile, and adult forms with residual β-galactosidase activity ranging from <1 to 15% of normal levels (reviewed in Refs. 1, 7).

The biosynthesis and processing of β-galactosidase have been studied in normal and mutant human fibroblasts. The enzyme is synthesized as a 65-kDa precursor translatedlympho and translatedlympho to the mature lysosomal form of 64 kDa (8). In cells of an infantile and an adult GM1-gangliosidosis patient, the precursor protein was found to be synthesized in a low amount, but no mature form could be detected (9). In a Morquio B cell strain, synthesis and processing of β-galactosidase proceed normally (9).

Lysosomal β-galactosidase has been purified to apparent homogeneity from various sources and species (reviewed in Ref. 2). In mammalian tissues (10, 11) as well as in human cultured fibroblasts (12) the majority of the active enzyme is present in a high molecular weight aggregate, and only a small fraction of the enzyme is found as monomeric 64-kDa polypeptide. It has been demonstrated that the aforementioned aggregate includes other glycoproteins; the heterodimeric 32-20-kDa "protective protein" (8, 13–15) and, under certain experimental conditions, the lysosomal neuraminidase (16). It is likely that these three glycoproteins, β-galactosidase-neuraminidase—protective protein, form a specific complex within lysosomes since they copurify by virtue of their association, and they influence each other's activity and stability (16, 17). Recently, Oshima et al. (18) have published the sequence of the lysosomal β-galactosidase, deduced from its cDNA.

We report on the cloning, sequence, and expression of two distinct cDNAs encoding the classic lysosomal form of the enzyme and a β-galactosidase-related protein with no enzymatic activity and a different subcellular localization. We provide evidence that the latter derives from alternatively spliced precursor mRNA.

EXPERIMENTAL PROCEDURES

Materials—Restriction endonucleases were purchased from the following companies: Boehringer Mannheim, Bethesda Research Laboratories (BRL), New England Biolabs, Pharmacia LKB Biotechnology Inc., and Promega Biotec. DNA polymerase, Klenow fragment, was from Promega Biotec. T4 polynucleotide kinase, avian myeloblastosis virus reverse transcriptase, M13 reverse sequencing primer, deoxy and dideoxy nucleotides, T4 DNA ligase was from BRL. The sequencing and the sequencing kit were purchased from United States Biochemical Corp. Top polymerase was from Cetus Corp. Immunoprecipitin (formalin-fixed Staphylococcus aureus cells) and prestained molecular weight markers were from BRL. Radionucleotides were obtained from Amersham Corp.: [32P]dATP and [35S]dCTP, 3000 Ci/mmol; [γ-32P]dATP, 6000 Ci/mmol,
[α-32P]dATP, >1000 Ci/mmol; [35S]methionine, >1000 Ci/mmol.
All other reagents were from standard commercial suppliers, if not specified otherwise.

Cell Culture—Human skin fibroblasts from normal individuals, four patients with Gb4-gangliosidosis, and one obligated heterozygote were obtained from the Leiden University Medical Center (Dr. W. J. Kleyer). Fibroblasts were cultured in Dulbecco’s modified Eagle’s medium/Ham’s F10 medium (1:1 v/v) supplemented with 10% fetal bovine serum. COS-1 cells were grown in the same medium supplemented with 5% fetal bovine serum.

Probe Sequence Analysis—Human placental β-galactosidase was purified together with neuraminidase and protective protein, as described previously (17). The different components were separated by SDS-PAGE, under reducing conditions according to Hasilik and Neufeld (19). The 64-kDa β-galactosidase band was digested in situ with tosylphenylalanyl chloromethyl ketone-treated trypsin (Worthington Diagnostic Systems Inc., United Kingdom). Tryptic peptides were fractionated by HPLC (Waters 6000 System) and sequenced by automated Edman degradation on an Applied Biosystems 470A gas-phase peptide sequenator as described previously (15). For N-terminal sequence analysis, approximately 50–100 μg of the purified complex was separated as above, and the protein components were blotted onto Zeta-Probe membranes (Millipore Corp.). A filter piece containing the 64-kDa protein was excised and used as starting material for automated Edman degradation (20).

cDNA Library Screening—A human testis cDNA library in λgt11 (Clontech, Palo Alto, CA), consisting of 1 × 1010 independent clones with inserts ranging from 0.7 to 3.3 kb, was plated out at a density of 5 × 106 plaque-forming units per 90-mm plate and screened with anti-β-galactosidase antibodies as described previously (21).

Antibody-positive clones were rescanned with oligonucleotide probes labeled at the 5′ end with 32P using γ-32P and polynucleotide kinase (22). The probes were synthesized on an Applied Biosystems 381A oligonucleotide synthesizer. Hybridization and washing conditions were as described (23).

DNA Sequencing—H2Gal(S) and H2Gal(L) cDNAs and their restriction fragments were subcloned into plasmid vectors pTZ18 and pTZ19. Nucleotide sequences on both strands were obtained by the oligonucleotide synthesizer. Hybridization and washing conditions were as described previously (23).

DNA Sequencing—H2Gal(S) and H2Gal(L) cDNAs and their restriction fragments were subcloned into plasmid vectors pTZ18 and pTZ19. Nucleotide sequences on both strands were obtained by the oligonucleotide synthesizer. Hybridization and washing conditions were as described previously (23).

Isolation and Sequencing of Genomic β-Galactosidase λ Clones—A human EMBL-3 λ library (kindly provided by Dr. G. Grosved, Erasmus University, Rotterdam), derived from DNA of leukocytes of a chronic myeloid leukemia patient, was screened with the 5′ 850-bp EcoRI fragment of cDNA clone H2Gal(L). The inserts of three overlapping clones were subcloned into the plasmid vector pTZ18. Sequences of genomic exons were determined by the chain termination method on double-stranded DNA, and of Murphy and Kavanagh (25), for single-stranded DNA. M13 universals reverse primer and a synthetic oligonucleotide were used. Sequence data were analyzed using the program of Staden (26).

RNA Isolation and Northern Blot Hybridization—Total RNA was isolated from cultured fibroblasts as described (27). Poly(A) RNA, immunoprecipitated using antibodies raised against purified placental complex, was obtained following the procedure of Myerowitz and Premi (28). RNA samples were electrophoresed on a 1% agarose gel containing 0.66 M formaldehyde as described (29) and blotted onto nylon membranes (Zeta-Probe). The filter was hybridized with the cDNA probe labeled according to the procedure of Feinberg and Vogelstein (30).

Polymerase Chain Reaction—10–15 μg of total RNA and about 50 ng of polymerase RNA were reverse transcribed into single-stranded cDNA using two antisense oligonucleotide primers and avian myeloblastosis virus reverse transcriptase. Subsequently, partial cDNAs were amplified in the presence of a third sense primer and Taq polymerase as described (31), using a programmable DNA incubator (BioExcellence). Amplified material was separated on 2% agarose gels and blotted onto Zeta-Probe membranes. Filters were hybridized using either the radiolabeled oligonucleotide or the cDNA probe described previously (25). These probes were labeled as described above.

Transient Expression of β-Galactosidase cDNAs in COS-1 Cells—Subcloning of the two cDNAs into a derivative of the mammalian expression vector pc-D-X and conditions of transfections of pc-D-X constructs to COS-1 cells were described previously (15). Labeling with [35S]methionine was carried out in the presence or absence of NH4Cl (19). Radiolabeled cDNA-encoded β-galactosidase proteins were immunoprecipitated from cell extracts and medium concentrates according to the method of Proia et al. (32). Immunoprecipitated proteins were resolved on SDS-PAGE under reducing conditions. Radioactive bands were visualized by fluorography of gels impregnated with Amplify (Amersham Corp.). Apparent molecular weights were calculated with conventional marker proteins. β-Galactosidase activity in COS-1-transfected cells was measured with affinity-purified 4-methylumbelliferyl substrate using standard assay conditions (7).

Uptake Studies in Human Cells—The preparation of conditioned media used in uptake studies and the experimental conditions were as reported (16). Human recipient cells were from an infantile Gb4- gangliosidosis patient (Fig. 5, patient 11). They were seeded on 6-well plates 4 days before addition of conditioned media. The uptake was carried for a further 3 days. Cells were harvested by trypsinization and homogenized by vortexing in double-distilled water. Enzyme activities were measured in cell homogenates using 4-methylumbelliferyl substrates (7).

Indirect Immunofluorescence—For light microscopy, COS-1 cells were transfected with pCDH-Ga constructs as above, but omitting the labeling step. Twelve hours before harvesting, transfected cells were reseeded at a low density on coverslips. Fixation and immunolabeling were performed according to Ref. 33 using anti-β-galactosidase antibodies and goat anti-rabbit IgG2 conjugated with fluorescein in the second incubation step.

RESULTS

Partial Amino Acid Sequence and Isolation of Antibodies—The β-galactosidase, neuraminidase, protective protein complex was purified from human placenta, and its components were separated by SDS-PAGE under reducing conditions. The 64-kDa β-galactosidase, electroeluted from the gel, was used to raise monospecific polyclonal antibodies in rabbit. This antibody preparation, tested in biosynthetic labeling experiments and Western blots, precipitates both mature and precursor forms of β-galactosidase (data not shown). In addition, a gel slice containing the 64-kDa protein was digested in situ with trypsin, and the resulting peptides were fractionated by reverse-phase HPLC. Five of the oligopeptides were subjected to automated Edman degradation, but only three of them gave an unambiguous amino acid sequence (Fig. 1A). We also sequenced the N terminus of intact mature 64-kDa β-galactosidase. A stretch of 18 amino acid residues was obtained in this case (Fig. 1A, N-ter).

Isolation and Characterization of cDNA Clones—One tryptic peptide sequence (T3) and the N-terminal sequence were used to synthesize two oligonucleotide probes complementary to the mRNA (Fig. 1B). Probe 1, a unique 45-mer, was constructed on the basis of codon usage frequencies in mammalian proteins, whereas probe 2, a 17-mer, was degenerated. A human testis λgt11 cDNA expression library was first screened with anti-β-galactosidase antibodies. Several recombinant clones were isolated and rescanned with both oligonucleotide probes. One clone, λH2Gal39, with a total insert size of 1.7 kb, carried an internal EcoRI site which released, upon digestion with EcoRI, two fragments of 500 and 1200 bp, hybridizing with probes 1 and 2, respectively. These results supported the identity of the cDNA and defined its orientation. Partial nucleotide sequencing of this cDNA revealed the presence of a putative ATG translation start codon, but the absence of a polyadenylation signal. In vitro translation of total RNA from cultured fibroblasts established a molecular mass of about 73 kDa for the non-glycosylated β-galactosidase preproprotein (5). Therefore, λH2Gal39 could not contain the entire coding region of β-galactosidase preprotein. Rescreening of the library with this cDNA probe yielded a clone, λH2Gal(L), that consisted of a 5′ EcoRI fragment of 850 bp and a 3′ fragment of 1550 bp. Both cDNAs were subcloned into pTZ18 and pTZ19, subjected to restriction endonuclease analysis, and

2 G. T. J. van der Horst, unpublished data.
sequestered using the dyeoxy chain termination method (24). In Fig. 2 a compendium of the partial restriction maps of the two cDNAs is depicted together with the nucleotide sequencing.

sequences of the two cDNAs from their internal EcoRI site toward the 3' end are identical, except that H3Ga39 misses the last 412 nucleotides including 94 bp of coding sequence and the 3' -untranslated region with the polyadenylation signal.

Although there is no direct proof that the 3' ends of the mRNAs specifying the two cDNAs are the same, S1 nuclease protection analysis of this region did not reveal the presence of differentially spliced transcripts (data not shown). Therefore, it is likely that H3Ga39 is a partial cDNA truncated at the 3' end. In contrast, a comparison of the 5' ends of the two clones revealed significant differences. The EcoRI fragment encompassing the 5' end of H3Ga39 is 338 nucleotides shorter than the corresponding fragment of H3Ga(L). The missing sequences comprise two stretches (boxed in Fig. 3), one of 212 nucleotides, between positions 295 and 508, and one of 181 nucleotides, between positions 602 and 784 (referred to as regions 1 and 2, respectively). Sequences immediately flanking these regions are completely identical in the two clones. If translation starts at the common ATG initiation codon, the exclusion of region 1 causes a -1 frameshift in the open reading frame of H3Ga39 which is reverted by a +1 frameshift due to the exclusion of region 2. In order to obtain a full length cDNA bearing the short 5' end, we have substituted the 3' end EcoRI fragment of H3Ga39 for the H3Ga(L) fragment. The resulting cDNA construct, H3Ga(S), has an open reading frame of 1608 nucleotides, which starts at the same ATG (position 51) and is interrupted by the same stop codon as the long cDNA. These surprising findings imply the existence of two β-galactosidase mRNA templates encoding proteins that are translated in different frames in the 95-nucleotide stretch between the two regions.

To verify whether these two mRNAs arise by alternative splicing, we have isolated genomic λ clones spanning the area of interest. The entire sequence of the exons encoding nucleotides 296–784 in H3Ga(L) cDNA (Fig. 3) was determined. In Fig. 4 the exons involved are schematically shown together with their exon/intron boundaries. Region 1 in the long cDNA is encoded by two exons of 151 and 61 bp, respectively, and region 2 by one exon of 181 bp. A separate exon specifies the 95-bp sequence between these two regions. The exact mapping of the different exons within the gene has not been determined. These results confirm that the two β-galactosidase transcripts derive from alternative splicing of the precursor mRNA.

Predicted Primary Sequences of β-Galactosidase and β-Galactosidase-related Proteins—As shown in Fig. 3, the two cDNA clones encode polypeptides of 677 and 546 amino acids, respectively, which have the first 82 N-terminal residues in common. These are followed, in the predicted sequence of the long cDNA-encoded β-galactosidase, by two noncontiguous sequences (boxed in Fig. 3) of 71 amino acids (residues 83–153) and 61 amino acids (residues 154–215), which do not occur in the short protein, referred to as β-galactosidase-related, because of splicing out of regions 1 and 2. Consequently, a unique stretch of 32 amino acids is found in the β-galactosidase-related protein (residues 89–114), which is different from the sequence between regions 1 and 2 (residues 154–185) in the long molecule.

All tryptic peptides as well as the N-terminal sequence of 64-kDa placental β-galactosidase are found in the amino acid sequence deduced from the cDNAs (Fig. 3, thick line). The only disagreement is at residue 1 of T1 where the experimentally determined residue is aspartic acid (Fig. 1A), whereas the amino acid predicted from the nucleotide sequence of the two cDNAs is threonine. Both cDNA-encoded proteins start with a putative signal peptide which is characterized by an
FIG. 3. Combined nucleotide and predicted amino acid sequences of β-galactosidase cDNAs. The numbers above the nucleotide sequence refer to HPGα(L) cDNA. Numbers on the right and on the left specify the amino acids of the long and short β-galactosidase-predicted sequences, respectively. Amino acid sequences corresponding to tryptic peptides and N terminus of the mature placental enzyme are indicated with a thick underline. Potential N-linked glycosylation sites are indicated with a thin underline. The nucleotide and deduced amino acid sequences of regions 1 and 2 are boxed. The stretch of 32 amino acid residues in the short protein, translated from another frame, because of exclusion of regions 1 and 2, is shown below the corresponding sequence of the long β-galactosidase. A putative polyadenylation signal is indicated with a double underline.
the signal peptide, and it is followed by 18 amino acids in the predicted primary sequence (Fig. 3, thin line). The most probable site for signal peptidase cleavage is Gly-23 (34). Seven potential N-linked glycosylation sites are present. The glycosylation site at position 26 is located immediately after the N-terminal region including a positively charged residue (Arg-7), a highly hydrophobic core, and a polar C-terminal domain. The most probable site for signal peptidase cleavage is Gly-23 (34). Seven potential N-linked glycosylation sites are present. The glycosylation site at position 26 is located immediately after the signal peptide, and it is followed by 18 amino acids (residues 29–46) that are colinear with the chemically determined N terminal of the purified placental enzyme. The predicted M, of unglycosylated β-galactosidase and β-galactosidase-related protein, including the signal peptide, are 76,091 and 60,552, respectively. Their amino acid sequences were compared with other sequences present in the NBRF (release 19.0, December 31, 1988) and EMBL (release 18, February 1989) data base. No significant homology was found.

**RNA Hybridization Studies**—The HβGa(L) cDNA insert was labeled by random priming and used to probe total and polysomal RNA isolated from cultured fibroblasts of normal individuals, four GMI-gangliosidosis patients, and one heterozygote. As shown in Fig. 5, an mRNA of about 2.5 kb is the major transcript detected in normal fibroblasts. The same hybridization pattern was obtained with total human testis RNA (data not shown). When immunoselected polysomal RNA is applied a faint minor band of about 2.0 kb becomes visible. It is clear that this 2.0-kb species is present in a much lower amount than the long mRNA. This difference in amount is also reflected by the amount of respective cDNA clones found in the library (1 versus 12).

The 2.5-kb mRNA is also detected in total RNA from fibroblasts of the adult GMI-gangliosidosis patient (Fig. 5). However, the three infantile forms of the disease exhibit a very different expression pattern. In the first patient (I), a faint broad band is visible. In some gels this band can be resolved into two, one of which is slightly larger than 2.5 kb (data not shown). The mother of this patient displays a hybridizing band of normal size but somewhat less intense. There is no detectable β-galactosidase transcript in the second infantile patient (II), whereas in the third patient (III) the 2.5-kb mRNA is present in a much lower quantity than in controls. The Northern blot was rehybridized with a probe recognizing the glyceraldehyde-3-phosphate dehydrogenase mRNA (35). Signals of equivalent intensity corresponding to this 1.2-kb message were detected in all samples (data not shown). Taken together these results demonstrate that different mutations must be involved in apparently similar GMI-gangliosidosis clinical phenotypes.

**Detection of Two mRNA Transcripts by PCR Amplification**—Since it is difficult to visualize the small mRNA molecule on Northern blots, we decided to use the polymerase chain reaction (PCR) to increase the detection level and to screen specific regions of β-galactosidase mRNA(s) for the presence or absence of regions 1 and 2. The strategy applied in these experiments is depicted in Fig. 6B. Three oligonucleotide primers were designed according to distinct complementary DNA sequences present in the two β-galactosidase clones (sequences are given in the legend to Fig. 6). Their positions, flanking or within regions 1 and 2, were chosen to direct the synthesis and amplification of cDNA fragments representative for the two different mRNA species. Total RNA from cultured fibroblasts and from human testis as well as polysomal mRNA from fibroblasts were reverse transcribed into

![Exon/Intron Boundaries](image)

**Fig. 4.** Exons involved in alternative splicing, generating human β-galactosidase long and short mRNAs. Boxes represent separate exons; numbers specify their length in base pairs. Interrupted lines depict intronic sequences of unknown size. The sequences of all exons have been determined except for the 5' end of the left-most exon and the 3' end of the right-most exon, as indicated by the vertical zig-zag lines. Sequences are identical to the corresponding HβGa(L) cDNA sequence, but only intron/exon boundaries (A to D) are shown. Small arrows specify splicing events generating the long β-galactosidase mRNA; large arrows indicate the mode of splicing giving rise to the small β-galactosidase transcript.

![Northern blot analysis of fibroblast RNA](image)

**Fig. 5.** Northern blot analysis of fibroblast RNA. Total and polysomal fibroblast RNA from three normal individuals and total fibroblast RNA from an adult and three infantile GMI-gangliosidosis patients as well as one heterozygote were fractionated on a formaldehyde-agarose gel and probed with the HβGa(L) cDNA. The sizes of the two β-galactosidase transcripts are indicated. Exposure time was 2 days.
Alternative Splicing of β-Galactosidase mRNA

Fig. 6. Detection of two mRNAs for β-galactosidase by PCR amplification. A, total (T. Fibr.) and polysomal fibroblast (P. Fibr.) RNA and total testis (T. Testis) RNA were used to synthesize single-stranded cDNAs that were subsequently subjected to 24 rounds of amplification. Amplified products were separated on 2% agarose gels, blotted, and hybridized with type-specific probes. E. coli tRNA was included as control. Sizes of the amplified fragments are indicated. Exposure times were 1 h for lanes 1-6 and 30 min for lanes 7 and 8. B, the EcoRI fragment at the 5' end of both cDNAs is shown. The triangles represent regions 1 and 2; numbers correspond to the nucleotide positions at their 5' and 3' ends. Solid bars 1, 2, 3 are the primers used for cDNA syntheses and PCRs. Arrows indicate their sense and antisense orientations. The sequence of the antisense primer 1 is 5'-AACATCTGCTGTTGCTGCTG-3'; of the antisense primer 3 is 5'-ACATTTCAGGAATGTTTTATGTGCT-3'. Sense primer 2 is 5'-TGTAACAGAGCTCCCTGCTGAA-3'. Cross-hatched bars designate a 90-bp PstI probe and two 20-mer oligonucleotide probes. The sequence of the 5'-oligonucleotide probe is 5'-CCATCCAGACATTACCTGGC-3'; of the 3' probe is 5'-AAACAGTGTCGAGCAGCAAT-3'.

Single-stranded cDNA using either antisense primer 1 or 3. The polymerase chain reactions were subsequently performed by adding the sense primer 2. Escherichia coli tRNA was used in separate reactions as a negative control. Amplified material was separated on agarose gels and Southern-blotted. In order to unequivocally distinguish between amplified fragments originating from the short or the long mRNA, type-specific probes were used (Fig. 6B, cross-hatched bars). Two 20-mers were synthesized on the basis of sequences of the H2Ga(S) cDNA, which are colinear with the 10 nucleotides flanking each end of regions 1 and 2 of H2Ga(L) (sequences are given in the legend to Fig. 6). These 20-mers hybridize, under stringent conditions, only to the cDNA fragment derived from the short mRNA. On the other hand the cDNA fragment specifying the long mRNA is detected by a 90-bp PstI probe present in region 2. As shown in Fig. 6A, fragments of 169 and 498 bp, representing the short and the long mRNA, respectively, are amplified in all samples and are identical in the two tissues tested (lanes 1-3 and 6-8). The identity of much fainter smaller bands present in lanes 6 and 8 is unknown. No hybridizing bands are visible in the tRNA lanes (lanes 4 and 5). It is noteworthy that the aforementioned cDNA fragments can also be amplified from polysomal RNA. This implies that the short transcript undergoes translation.

Fig. 7. Transient expression of pCDHβGa constructs in COS-1 cells. The pCDHβGa(L) and pCDHβGa(S) cDNA constructs in the sense (se, lanes 1-3 and 7-9) and antisense (a, lane 4) orientations were transfected to COS-1 cells. A mock transfection was carried out without the addition of DNA (lane 5). n.t. indicates not transfected (lane 6). After 48 h cells were incubated with [35S]methionine for an additional 16 h and without NH4Cl (+ and -). Labeled proteins from cells and media were immunoprecipitated with anti-β-galactosidase antibodies, analyzed on a 12% SDS-polyacrylamide gel, and visualized by fluorography. Molecular sizes were calculated by comparison with protein markers. Exposure time for lanes 1-6 was 1 day and for lanes 7-9 was 1 week.
fected cells also do not show any cross-reactive bands. We also tested whether the cDNA-encoded proteins were able to correct β-galactosidase activity in \( G_{\text{M1}} \)-gangliosidosis cells. For this purpose, medium from COS-1 cells transfected with sense or antisense pCDHβGa constructs as well as medium from mock-transfected cells were collected and concentrated. As shown in Table II, COS-1 cell-derived 86-kDa precursor taken up by \( G_{\text{M1}} \)-gangliosidosis cells corrects β-galactosidase activity. In a similar uptake experiment carried out using radiolabeled secretions from COS-1-transfected cells, we could demonstrate that the 85-kDa precursor and the 68-kDa β-galactosidase-related protein were taken up by the mutant cells, but only the 85-kDa precursor was further processed to the mature 64-kDa form (data not shown).

In order to determine the intracellular distribution of the two proteins, indirect immunofluorescent staining was performed on transfected cells using anti-β-galactosidase antibodies and fluorescein-labeled second antibodies (Fig. 8). A perinuclear labeling of β-galactosidase is observed in COS-1 cells transfected with the pCDHβGa(L)-sense construct (Table I). Using the same assay conditions, the β-galactosidase-related molecule is apparently not active.

**Table I**

| Activity of β-galactosidase in COS-1 cells after transfection with pCDHβGa plasmid DNAs |
|-----------------------------------------------|
| Plasmid                | mU/mg protein |
| pCDHβGa(L)-sense       | 8.9           |
| pCDHβGa(S)-sense       | 1.4           |
| pCDHβGa(S)-anti        | 1.8           |
| Mock-transfected       | 1.7           |
| Not transfected        | 1.5           |

*One milliunit of enzyme activity is defined as the activity that releases 1 nmol of 4-methylumbelliferone per min.*

**Table II**

| Correction of β-galactosidase activity in \( G_{\text{M1}} \)-gangliosidosis fibroblasts after uptake of COS-1 cell-derived β-galactosidase precursor |
|----------------------------------------------------------------------------------------------------------------------------------|
| Transfection in COS-1 cells | Activity in \( G_{\text{M1}} \)-gangliosidosis fibroblasts |
|--------------------------------|-----------------------------------------------|
| pCDHβGa (L) sense             | 234                                           |
| pCDHβGa (S) sense             | 6.6                                           |
| pCDHβGa (S) sense             | 7.6                                           |
| Mock                           | 12.3                                          |

**Tables**

We have isolated and characterized two distinct cDNA clones encoding human lysosomal β-galactosidase and a β-galactosidase-related protein. In total RNA from normal human fibroblasts, a major mRNA of 2.5 kb is recognized by cDNA probes. A minor transcript of about 2.0 kb is detectable only in immunoselected polysomal RNA. The 2.5-kb β-galactosidase mRNA is also present in fibroblasts from the adult GM1-gangliosidosis patient, but it is either absent or reduced in amount in cells from three patients with the infantile form of the disease. The pattern of expression of this β-galactosidase mRNA in patients I and II is consistent with data from immunoprecipitation studies that established the absence of cross-reactive material for β-galactosidase in fibroblasts from these patients.2 Apparently, other infantile GM1-gangliosidosis patients, not yet analyzed at the molecular level, do synthesize β-galactosidase-related protein (36). The adult and the third infantile patient studied here were previously reported to synthesize a β-galactosidase precursor that did not get phosphorylated (37). This might still hold true for these two patients, but the assumption made by Hoogven et al. (37) that all GM1-gangliosidosis variants are phosphorylation mutants is not substantiated by the results presented here. Patients I and II, for instance, may represent splicing and/or promoter mutants. Obviously, different or even the same clinical phenotypes are caused by distinct genetic lesions, and further studies are needed to define the clinical and biochemical heterogeneity observed in \( G_{\text{M1}} \)-gangliosidosis patients.

The nucleotide sequences of the two cDNAs comprise open reading frames that begin at a common ATG translation initiation codon and terminate at the same stop codon. However, HβGa(L) is 393 bp longer than HβGa(S). Its nucleotide sequence is colinear with the human placental β-galactosidase cDNA recently isolated by Oshima et al. (18). The only sequence differences we find are at nucleotide positions 79 (T instead of C), 650 (G instead of C), and 651-653 (CGC instead of GCG), resulting in the following amino acid changes: Leu-10 instead of Pro-10 and Arg-201 instead of Ala-201. These discrepancies may represent true allelic variations and/or mistakes introduced by cDNA cloning procedures. The sequence of the short cDNA is virtually identical to the former, but it misses two noncontiguous protein-encoding sequences, regions 1 and 2, present in the long clone. Furthermore, the exclusion of region 1 in this cDNA introduces a frameshift in its 3′-flanking sequence which is subsequently restored by the exclusion of region 2. These unusual findings imply the existence of two distinct mRNA templates which, most remarka-
Alternative Splicing of β-Galactosidase mRNA

By sequencing genomic β-galactosidase clones, we could demonstrate that nucleotides 296-784 of the 2.5-kb mRNA, spanning regions 1 and 2 as well as their intermediate sequence, are encoded by four separate exons. As shown by the sequence of the exon/intron borders, all four exons obey the GT/AG rule (38). These results strongly indicate that the short mRNA is generated by a differential splicing process that involves three exons. An increasing number of genes are known to create protein diversity through the use of differential splicing (reviewed in Ref. 39). Among lysosomal proteins in transiently transfected COS-1 cells, a form that was not detected previously. Whether or not this protein has a defined biological function is not known. Although both polypeptides are recognized by the antibodies, the β-galactosidase-related protein is not catalytically active under the assay conditions used. The same holds true for the short β-glucuronidase protein (40). Furthermore, even though both cDNA-encoded proteins, 85 and 68 kDa, are as efficiently endocytosed by GM1-gangliosidosis fibroblasts, only the 85-kDa precursor is further processed intracellularly and corrects β-galactosidase activity.

The subcellular localization of COS-1-derived β-galactosidase and β-galactosidase-related proteins is different. The long β-galactosidase has a clear lysosomal distribution, whereas the short molecule is found only in the perinuclear region. The latter is likely to reach the Golgi apparatus, since it is secreted into the extracellular space even without the addition of NH4Cl. The differential subcellular distribution of the two proteins might explain their distinct catalytic behavior. Further studies are needed to define the function and substrate specificity of the β-galactosidase-related protein. It will be of interest to analyze the domains that are either missing or different in the two polypeptides.

This work together with our studies on the other components of the complex, the protective protein and neuraminidase, will enable us to gain more insight in the fine mechanisms of mutual cooperation between these lysosomal glycoproteins.

Acknowledgments—We are very grateful to Professor H. Galjaard for continuous support and Dr. Gerard Grosfeld for crucial advice and suggestions. We would like to thank Alan Harris of the Laboratory of Protein Structure, National Institute for Medical Research, London, and Dr. Gary Hathaway of the Biotechnology Instrumentation Facility, University of California, Riverside, for their expert help with the protein sequencing; Nike Soekarman for assistance on the PCR experiment; Sjoerf van Baal for help with the computer data analysis; Erik Bonten for technical assistance. We also thank Mirko Kuit and Pim Visser for the photographic work and Jeannette Lokker and Nellie van Sluijsdam for typing and editing the manuscript.

REFERENCES
1. O'Brien, J. S. (1989) in The Metabolic Basis of Inherited Disease (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds) pp. 1797-1806, McGraw-Hill, New York
2. Conzelmann, E., and Sandhoff, K. (1987) Adv. Enzymol. 60, 89-217
3. Shows, T. B., Scaffar-Wolf, L. R., Brown, J. A., and Meisler, M. (1979) Somatic Cell Genet. 5, 147-158
4. Okada, S., and O'Brien, J. S. (1988) Science 260, 1002-1004
5. O'Brien, J. S., Gugler, E., Giedion, A., Wiessman, U., Herschkowitz, N., Meijer, C., and Leroy, J. (1976) Clin. Genet. 9, 495-504
6. Grobe, H., Krins, M., Schmidberger, H., von Figura, K., Harzer, K., Kresse, H., Paschke, E., Sewell, A., and Ullrich, K. (1980)

sidase cDNAs code for 677 and 546 amino acids, respectively, with the first 23 residues in common representing a typical signal peptide (34). Both proteins carry seven potential N-linked glycosylation sites at identical positions. One of them is located immediately after the signal sequence and precedes the N-terminal sequence of mature 64-kDa placental β-galactosidase. From its location we can infer that the substantial proteolytic processing of the 85-kDa β-galactosidase precursor observed in human fibroblasts (8) as well as in mouse kidney cells and macrophages (41, 42) must occur nearly exclusively at the C terminus.

The two cDNAs direct the synthesis in COS-1 cells of immunoprecipitable polypeptides, which are also recovered extracellularly. The molecular mass of the long protein, 85 kDa, is in agreement with the apparent size of β-galactosidase glycosylated precursor immunoprecipitated from human fibroblasts (8). The 68-kDa protein derived from the short cDNA is a form that was not detected previously. Whether or not this protein has a defined biological function is not known. Although both polypeptides are recognized by the antibodies, the β-galactosidase-related protein is not catalytically active under the assay conditions used. The same holds true for the short β-glucuronidase protein (40). Furthermore, even though both cDNA-encoded proteins, 85 and 68 kDa, are as efficiently endocytosed by GM1-gangliosidosis fibroblasts, only the 85-kDa precursor is further processed intracellularly and corrects β-galactosidase activity.

The subcellular localization of COS-1-derived β-galactosidase and β-galactosidase-related proteins is different. The long β-galactosidase has a clear lysosomal distribution, whereas the short molecule is found only in the perinuclear region. The latter is likely to reach the Golgi apparatus, since it is secreted into the extracellular space even without the addition of NH4Cl. The differential subcellular distribution of the two proteins might explain their distinct catalytic behavior. Further studies are needed to define the function and substrate specificity of the β-galactosidase-related protein. It will be of interest to analyze the domains that are either missing or different in the two polypeptides.

This work together with our studies on the other components of the complex, the protective protein and neuraminidase, will enable us to gain more insight in the fine mechanisms of mutual cooperation between these lysosomal glycoproteins.
Alternative Splicing of β-Galactosidase mRNA

Am. J. Hum. Genet. 32, 258–272

7. Galjaard, H. (1980) Genetic Metabolic Disease, Early Diagnosis and Prenatal Analysis, Elsevier Science Publishers B. V., Amsterdam

8. d’Azzo, A., Hoogeveen, A. T., Reuser, A. J. J., Robinson, D., and Galjaard, H. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 4535–4539

9. Hoogeveen, A. T., Graham-Kawashima, H., d’Azzo, A., and Galjaard, H. (1984) J. Biol. Chem. 259, 1974–1977

10. Norden, A. G. W., Tennant, L., and O’Brien, J. S. (1974) J. Biol. Chem. 249, 7969–7976

11. Yamamoto, Y., Fujis, M., and Nishimura, K. (1982) J. Biochem. (Tokyo) 92, 15–21

12. Hoeksema, H. L., van Diggelen, O. P., and Galjaard, H. (1979) Biochim. Biophys. Acta 566, 72–79

13. Hoogeveen, A. T., Verheijen, F. W., and Galjaard, H. (1983) J. Biol. Chem. 258, 12143–12146

14. Willemsen, R., Hoogeveen, A. T., Sips, H. J., van Dongen, J. M., and Galjaard, H. (1986) Eur. J. Cell Biol. 40, 9–15

15. Galjart, N. J., Gillemans, N., Harris, A., van der Horst, G. T. J., Verheijen, F. W., Galjaard, H., and d’Azzo, A. (1988) Cell 54, 755–764

16. Verheijen, F. W., Palmeri, S., Hoogeveen, A. T., and Galjaard, H. (1986) Eur. J. Biochem. 149, 315–321

17. van der Horst, G. T. J., Galjart, N. J., d’Azzo, A., Galjaard, H., and Verheijen, F. W. (1989) J. Biol. Chem. 264, 1317–1322

18. Oshima, A., Tsuji, A., Nagao, Y., Sakuraba, H., and Suzuki, Y. (1988) Biochem. Biophys. Res. Commun. 157, 238–244

19. Hasilik, A., and Neufeld, E. F. (1986) J. Biol. Chem. 255, 4937–4945

20. Aebersold, R. H., Teplov, D. B., Hood, L. E., and Kent, S. B. H. (1986) J. Biol. Chem. 261, 4229–4238

21. Huynh, T. V., Young, R. A., and Davis, R. W. (1985) in DNA Cloning: A Practical Approach (Glover, D. M., ed) Vol. I, pp. 49–78, IRL Press Ltd., Oxford

22. Maniatis, T., Fritsch, E. F., and Sambrook, J. (eds) (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

23. Wood, W. I., Gitschier, J., Lasky, L. A., and Lawn, R. M. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 1585–1588

24. Sanger, F. G., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467

25. Murphy, G., and Kavanagh, T. (1988) Nucleic Acids Res. 16, 5198

26. Staden, R. (1986) Nucleic Acids Res. 14, 217–231

27. Auffray, C., and Rougeon, F. (1980) Eur. J. Biochem. 107, 303–314

28. Myerowitz, R., and Proia, R. L. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 5394–5398

29. Fourney, R. M., Miyakoshi, J., Day, III, R. S., and Paterson, M. C. (1988) Focus 10, 5–7

30. Feinberg, A. P., and Vogelstein, B. (1983) Anal. Biochem. 132, 6–13

31. Hermans, A., Gow, J., Selleri, L., von Lindern, M., Hagemeyer, A., Wiedemann, L. M., and Grosveld, G. (1988) Leukemia 2, 629–633

32. Proia, R. L., d’Azzo, A., and Neufeld, E. F. (1984) J. Biol. Chem. 259, 3350–3354

33. Galjaard, H. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 4535–4539

34. Uytterhoeven, R., van Diggelen, O. P., and Galjaard, H. (1984) J. Biol. Chem. 259, 259, 3350–3354

35. Feinberg, A. P., and Vogelstein, B. (1983) Anal. Biochem. 132, 6–13

36. Maniatis, T., Fritsch, E. F., and Sambrook, J. (eds) (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

23. Wood, W. I., Gitschier, J., Lasky, L. A., and Lawn, R. M. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 1585–1588

24. Sanger, F. G., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467

25. Murphy, G., and Kavanagh, T. (1988) Nucleic Acids Res. 16, 5198

26. Staden, R. (1986) Nucleic Acids Res. 14, 217–231

27. Auffray, C., and Rougeon, F. (1980) Eur. J. Biochem. 107, 303–314

28. Myerowitz, R., and Proia, R. L. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 5394–5398

29. Fourney, R. M., Miyakoshi, J., Day, III, R. S., and Paterson, M. C. (1988) Focus 10, 5–7

30. Feinberg, A. P., and Vogelstein, B. (1983) Anal. Biochem. 132, 6–13

31. Hermans, A., Gow, J., Selleri, L., von Lindern, M., Hagemeyer, A., Wiedemann, L. M., and Grosveld, G. (1988) Leukemia 2, 629–633

32. Proia, R. L., d’Azzo, A., and Neufeld, E. F. (1984) J. Biol. Chem. 259, 3350–3354