STRUCTURAL AND FUNCTIONAL ANALYSIS OF LYSOSOMAL ß-GALACTOSIDASE AND ITS RELATION TO THE PROTECTIVE PROTEIN

Aan mijn ouders

STRUCTURAL AND FUNCTIONAL ANALYSIS OF LYSOSOMAL ß-GALACTOSIDASE AND ITS RELATION TO THE PROTECTIVE PROTEIN

Structurele en functionele analyse van lysosomaal ß-galactosidase en de relatie tot het 'protective protein'

Proefschrift

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Professor Dr. C.J.Rijnvos
en volgens besluit van het College van Dekanen.

De openbare verdediging zal plaatsvinden op
woensdag 17 juni 1992 om 15^{45} uur

door

Johannes Morreau

geboren te Hilversum

Rodopi

Amsterdam 1992

Promotiecommissie

Promotor: Prof. Dr. H. Galjaard

Co-promotor: Dr. A. d'Azzo

Overige leden: Prof. Dr. D. Bootsma
Prof. Dr. A. Westerveld
Prof. Dr. B. Wieringa

Dit proefschrift werd bewerkt binnen de vakgroep Celbiologie en Genetica van de faculteit der Geneeskunde en Gezondheidswetenschappen van de Erasmus Universiteit Rotterdam.

CONTENTS

PUBLICATIONS IN SECTION 2.3

(1) Alternative splicing of ß-galactosidase mRNA generates the classic lysosomal enzyme and a ß-galactosidase-related protein (1989).
Morreau, H., Galjart, N.J., Gillemans N., Willemsen,R., van der Horst, G.T.J., and d 'Azzo, A.
J.Biol.Chem. **264**, 20655-20663

(2) Organization of the gene encoding human lysosomal ß-galactosidase (1991).
Morreau, H., Bonten, E.J., Zhou, X.Y., and d'Azzo, A.
DNA Cell Biol. **10** , 495-504

(3) Aberrantly spliced ß-galactosidase mRNA with a twenty nucleotide intronic insertion causes an infantile form of G_{M1}-gangliosidosis (1992).
Morrone A., Morreau,H., Zhou, X.Y., Zammarchi E., Kleyer W.J., Galjaard,H. and d'Azzo, A.
(Submitted for publication)

(4) Human lysosomal protective protein has cathepsin A-like activity distinct from its protective function (1991).
Galjart, N.J., Morreau, H., Willemsen, R., Gillemans, N., Bonten, E.J., and d'Azzo, A.
J.Biol.Chem. **266**, 14754-14762

(5) Human lysosomal protective protein: Glycosylation, intracellular transport and association with ß-galactosidase in the endoplasmic reticulum (1992).
Morreau, H., Galjart, N.J., Willemsen, R., Gillemans, N., Zhou, X.Y., and d'Azzo A.
J.Biol.Chem., in press

SCOPE OF THE THESIS

Lysosomal ß-galactosidase is the glycosidase, that cleaves ß-linked galactosyl moieties from a variety of natural and synthetic substrates. In normal tissues of various species this enzyme appears to associate with two other hydrolases, N-acetyl-α-neuraminidase and the protective protein. Mutations at the ß-galactosidase locus on chromosome 3 are the basis of two lysosomal storage disorders, G_{M1}-gangliosidosis and Morquio B syndrome. The scope of our experimental work is to gain more insights into the structural and functional properties of these proteins, their mutual relation in lysosomes and their involvement in genetic disorders. Knowledge on the primary structure of lysosomal protective protein has recently brought to the identification of its multifunctional activities.

This thesis deals with the isolation and characterization of the cDNAs encoding the classic lysosomal ß-galactosidase and a ß-galactosidase-related protein. The latter is derived from alternatively spliced pre-mRNA, as deduced from analysis of the genomic organization of the ß-galactosidase gene. These studies include the identification of the ß-galactosidase promoter region.

In order to introduce this part of the experimental work, the thesis starts with an overview of what is known to date about promoter elements of genes encoding lysosomal enzymes and their putative mode of regulation. The occurrence of alternative splicing among pre-mRNAs coding for other lysosomal enzymes is briefly reviewed. The thesis continues with the description of the genetic lesions found in two patients with the severe infantile type of G_{M1}-gangliosidosis. Curiously this mutation, affecting both alleles in the patients, impairs correct splicing of premRNA molecules.

The final part of the thesis describes on one hand the two separable roles of the protective protein: a protective function towards ß-galactosidase and neuraminidase and a catalytic activity resembling that of lysosomal cathepsin A. On the other hand it gives new meaning to the association of ß-galactosidase and protective protein, that seems to be important for the correct transport of ß-galactosidase out of the endoplasmic reticulum and to the lysosomes.

CHAPTER I

1.1 TRANSCRIPTION OF GENES ENCODING LYSOSOMAL ENZYMES

1.1.1 General features of transcription elements in eukaryotic genes

Transcription is a very conserved and regulated process by which from a variety of genes different RNAs are generated. The formation of an mRNA template is started by the action of one of the RNA polymerases at a transcription initiation site, due to the activation of promoter and enhancer elements by a network of regulatory factors that bind in a sequence specific manner.

Promoters are composed of cis-acting control elements, that need the interaction with transacting factors to initiate transcription. They are usually located in relatively fixed positions with respect to the transcription initiation site(s) and may contain a variety of specific promoter elements One of these modules is the TATA-box that is mostly present at position -25, upstream of the transcription initiation site and tends to be located within GC-rich sequences (for reviews see Breathnach and Chambon,1981 ; Dynan and Tjian, 1985, Mc Knight and Tjian, 1986). In combination with the TATA element a CCAAT box can be encountered at position -80. GC boxes are often found as multiple copies within a promoter. CCAAT and GC boxes may function in either orientation, whereas a TATA box directs transcription in one orientation from a specific initiation site. In addition to these conventional "initiators", many other sequence specific motifs have been identified as potential promotor elements (Mitchel and Tjian, 1989; Sen and Baltimore, 1986; Garcia *et al.*, 1987; Hurst *et al.*,1987, Smale and Baltimore, 1989).

In absence of a TATA box in the promoter, multiple start sites of transcription may be encountered like it is observed in the majority of ubiquitously expressed housekeeping genes. Their features will be described in the next paragraph.

Enhancers are regulatory elements, whose presence increase the transcription level of nearby genes in an orientation independant manner. Contrary to promoter elements, their distance to transcription initation sites may vary tremendously (Banerji *et al.*, 1981; Moreau *et al.*, 1981). However, a distinction between promoter and enhancer sequences is not always clear. Transcription regulation is needed in cell type specific gene expression, during embryonal development, tissue differentation and under specific (non)physiological stimuli or conditions.

After or even during transcription the precursor mRNA (pre-mRNA) is processed by constitutive or alternative splicing processes. Mature mRNA needs then to be transported out of the nucleus before it can be translated at ribosomal sites into proteins.

1.1.2 Promoter elements of genes encoding lysosomal proteins

The promoter regions of several genes encoding lysosomal hydrolases have been defined and characterized. In most cases, these promoters bear characteristics of housekeeping gene promoters: they are rich in GC residues, have features of CpG islands and lack the canonical TATA and CCAAT boxes. They usually contain one or more potential binding sites for the transcription factor Sp1 (GGGCGG). Due to these characteristics, heterogeneous transcription initiation sites are often encountered in mRNAs for lysosomal proteins. The latter feature has been described

for acid α-glucosidase (Hoefsloot *et al.*, 1990; Martiniuk *et al.*, 1990), acid phosphatase (Geier *et al.*, 1989), ß-galactosidase (Morreau *et al.*, 1991), ß-glucuronidase (Shipley *et al.*,1991), cathepsin D (Redecker *et al.*, 1991), arylsulphatase A (Kreysing *et al.*, 1990) and ß-hexosaminidase ß-chain (Proia, 1988; Neote *et al.*, 1988) genes. Furthermore, most of these promoters contain putative consensus sequences, that might interact with the repressor GCF (Kageyama and Pastan, 1989). However data on α-galactosidase A (Quinn *et al.*, 1987; Bishop *et al.*, 1988), ß-hexosaminidase α-chain (Proia & Soravia, 1987) and α-N-acetylgalactosaminidase (Wang and Desnick, 1991) are slightly different. These genes have still GC-rich promoter regions, but the α-galactosidase promoter includes a TATA and a CCAAT box as well as Sp1 binding sites. In the α-N-acetylgalactosaminidase promoter-area no HTF islands have been identified, although two CCAAT-like boxes and three Sp1 recognition sites are present. The putative promoter region of the human ß-hexosaminidase α-chain gene contains only a CCAAT and a TATA element.

The glucocerebrosidase promoter region is the only one sofar described with characteristics of promoters of genes with a tissue specific mode of expression. It has a normal GC content (≤50%), two typical TATA and CCAAT boxes and no Sp1 elements (Reiner *et al.*, 1988b; Horowitz *et al.*, 1989). The promoter of the glucocerebrosidase pseudogene has similar characteristics.

Within the 5' untranslated region (UTR) of the human ß-glucuronidase gene a consensus sequence similar to the 17 bp transcription "initiator" (Smale and Baltimore, 1989) has been located in the vicinity of multiple initiation sites (Shipley *et al.*, 1991). The latter are found around nucleotide position (-30), upstream of the ATG translation initiation codon. Sp1 binding sequences are also present 5' of these transcription initiation sites. A minor initiation site is encountered further upstream, at nucleotide position (-216), that was shown, however, to be unable to initiate basal transcription without the presence of other 5' sequences. From the characteristics of the 5' UTR of the ß-glucuronidase gene, it has been postulated that two promoters might control its transcription: one responsible for the constitutive expression and the other for the regulated expression in certain cell types. However, transcription initiation from the most upstream site has only been seen in transformed cells (Shipley et al., 1991).

In contrast to the human gene, the murine ß-glucuronidase structural gene (Gus-s) has a promoter which is rich in GC stretches, comprises a TATA box, but not a CCAAT box or Sp1 consensus sequences. In agreement with the presence of the TATA box, transcription of this gene is initiated at a single position. Furthermore, a TATA box containing promoter has been detected within intron 1 of the Gus-s gene (Wang *et al.*, 1988). This transcriptional unit directs initiation of transcription of the so-called ß-glucuronidase "intermingled" gene (Gig) in lever cells. Exon sequences of Gus-s and Gig transcripts appear not to overlap, although pre-mRNA of Gig and ß-glucuronidase must have sequence similarities. Selection of differential splice sites coupled to the use of different transcription initiation sites would generate either Gus-s or Gig mRNA. Whether transcription of Gig interferes with the regulation of ß-glucuronidase mRNA expression and what Gig mRNA encodes for is unknown.
The promoters of genes encoding human and mouse ß-galactosidase (Nanba and Suzuki, 1991; Morreau *et al.*, 1991) show similar characteristics (see Chapter II).

Other potential regulatory sequences encountered in 5' UTRs of genes encoding lysosomal proteins are binding sites for the transcription factor AP2 (Imagawa *et al.*, 1987). This is the case for the α-glucosidase, ß-glucosidase, ß-glucuronidase and cathepsin D promoter regions. Putative AP1 sites (Lee *et al.*, 1987) are also found in the ß-hexosaminidase ß-chain and α-glucosidase genes. Octamer sequences, "chorion box" enhancers, as well as a c-Fos enhancer element are present in the promoter region of the human α-galactosidase gene (Bishop et al., 1988 and references therein). A 60 nucleotide repeat, (GA)n, of unknown function, upstream of the promoter region of the latter gene, has been also identified (Quinn *et al.*, 1987).

1.1.3 Regulation of transcription of lysosomal genes

The expression pattern of genes encoding lysosomal enzymes reflects their ubiquitous distribution in different animal tissues and cell types that in turn relates to the relative amount and quality of substrates to be degraded. In this respect these genes can be compared to the aforementioned housekeeping genes. However, this does not imply that their expression is not regulated, for instance, at certain stages of embryonal development, tissue differentiation or under specific physiological conditions.

In recent years unexpected roles for lysosomal enzymes have emerged or have been postulated. A model of T and B lymphocytes interaction proposes that resting splenic B cells do not interact with a specific T cell population unless they are activated and do present antigens to T helper cells. This interaction process is mediated by membrane bound glycoprotein(s) on the surface of B cells that should be devoid of terminal sialic acid residues. It has been hypothesized that lysosomal neuraminidase could be involved in this process (Kearse *et al.*, 1988a;1988b). Cathepsin D and cathepsin B have been shown to participate in antigen processing and extracellular proteolysis (San Segundo *et al.*, 1986). The protective protein/cathepsin A (Jackman *et al.*, 1990; 1992; Galjart *et al.*, 1991) might play a role in specific degradation or regulation of bioactive peptides including oxytocin, substance P, angiotensin I and endothelin (discussed in Chapter 2). Cathepsin B or related enzymes might also be involved in tumor invasiveness (Mort and Recklies, 1985; Ryan *et al.*, 1985; Dufek *et al.*, 1984; Rochefort *et al.* 1987; Sloane *et al.*, 1987). Finally, an important role is known to be played by certain lysosomal enzymes in reproductive processes, like ovulation (Dimino and Elfont, 1980), acrosome reaction (McRorie and Williams, 1974; Wasserman, 1987) and egg implantation (Denker *et al.*, 1978). These diverse and specific lysosomal activities suggest that expression of these enzymes might be tightly controlled, at least in some cases. Transcription regulation might give rise to variable amounts of mRNA in different tissues in response to local stimili. Alternatively, regulation could occur at the level of mRNA stability, mRNA maturation or transport. Furthermore, efficiency of translation, posttranslational modifications, intracellular transport, protein degradation and lysosomal permeability could be envisaged as mechanisms to repress or activate lysosomal enzyme activities.

In general, it is believed that housekeeping genes are regulated foremost at the level of mRNA maturation and/or stability. This has been shown for rat glyceraldehyde 3-phosphate dehydrogenase (Piechaczyk *et al.*, 1984), mouse dihydrofolate reductase (Kaufman and Sharp, 1983), and chicken thymidine kinase (Groudine and Casimir, 1984). In the lysosomal system, a variation in amount of

mRNA in different tissues has sofar been observed only for murine protective protein/cathepsin A (Galjart *et al.*, 1990), murine ß-galactosidase (Morreau *et al.*, 1991) and rat cathepsin B (San Segundo *et al.*, 1986). All three enzymes show very high mRNA levels in kidney. Variable tissue distribution of rat cathepsin B was previously shown by the analysis of immunoreactive material (Katunuma and Kominami, 1983). Rat cathepsin D mRNA levels are rather uniform in the tissues analysed (San Segundo *et al.*, 1986). However, it has been reported that expression of the human counterpart of this enzyme varies in different tissues at the protein level (Dorn *et al.*, 1986; Reid *et al.*, 1986). A similar observation has been made for murine ß-glucuronidase, although differences in mRNA levels in this case cannot account for the observed differences at protein level (Bracey and Paigen, 1987). These authors have shown that the variation detected in the tissues tested are primarily due to differences in the translational yield of the ß-glucuronidase enzyme. Translational yield is here defined as the number of mature enzyme molecules formed per minute per mRNA molecule. In this study protein degradation is only significantly different in brain tissue.

Control of lysosomal gene expression was originally studied in different mouse strains using classic genetic approaches (for a review see Paigen, 1979). From this work it became apparent that a number of lysosomal and non lysosomal enzymatic activities can be induced in proximal tubulus cells of the kidney by androgens as well as high doses of progestational steroids. The activities of lysosomal ß-glucuronidase and to a lesser extent α-mannosidase, ß-galactosidase and ß-hexosaminidase were indeed enhanced. Several regulatory elements controlling this event have been suggested. For murine ß-glucuronidase next to the structural locus Gus-s, mentioned earlier, a cis-acting androgen inducible regulator (Gus-r) that is functional in kidney, a cis-acting systemic regulator (Gus-u) and a trans-acting temporal regulator Gus-t have been proposed. The latter would be involved in developmental programming. However, it appeared that the androgen-induced increase of ß-glucuronidase expression is absent in hypophysectomized mice. Thus, there is apparently a regulatory mechanism involving one of the pituitary hormones and androgen or androgen receptor. Further analysis has shown that androgen response in mouse kidney involves direct control of the amount of translatable ß-glucuronidase mRNA (Palmer *et al.*, 1983). The general response to androgen of many lysosomal enzymes might be related to the massive increase in glycosphingolipid content in kidneys of treated animals (Paigen, 1979).

Another example of a lysosomal enzyme, whose expression is influenced by hormonal treatment is human cathepsin D. Synthesis of this enzyme is regulated by estrogen in specific cell types (Cavailles *et al.*, 1983; Pietras *et al.*, 1979). In breast cancer cells, for instance, that have an estrogen receptor, a 52 kDa cathepsin D-like protease is induced, following estradiol treatment (or high dosis of androgens), and this induction is transcriptionally regulated. Also, in rat uterus cathepsin D has been reported to be induced by progresterone (Elangovan and Moulton, 1980), whereas in promonocytes the rate of both mRNA and protein expression of this enzyme is regulated by calcitriol (Redecker *et al.*, 1989). The activity of ß-hexosaminidase can be influenced by progesteron and estrogen (Rahi and Srivistava, 1983) and this conforms with the presence of a potential progesteron response element in the 5' UTR of the ß-hexosaminidase gene. (Neote *et al.*, 1988). In relation to the

aforementioned results, there is the interesting observation that hormones like estrogen, growth hormone and cortison are able to stabilize certain mRNAs via an as yet unknown mechanism (reviewed by Ross, 1989).

As mentioned before, glucocerebrosidase is sofar a unique example among lysosomal hydrolases. The gene is differentially expressed in a tissue specific manner consistent with the general features of its promoter (Reiner and Horowitz, 1988a). However, it has to be taken into account that these data are from cells that are either transformed or clearly abnormal. Intermediate mRNA expression is observed in skin fibroblasts, promyelocytes and macrophages, whereas high mRNA levels are found in epithelial cells (HeLa) and very low levels in B-cells. A primary expression in cells of the reticuloendothelial system would be consistent with the side of accumulation of glucocerebroside substrate in tissues of Gaucher patients, due to the deficiency of the enzyme. Particularly intruiging is the observation that in Gaucher cells glucocerebrosidase mRNA seems to be more abundantly expressed than in normal cells (Reiner and Horowitz, 1988). It is also interesting that the mRNA encoding the precursor of the glucocerebrosidase activator protein is also upregulated in these mutant cells (Reiner and Horowitz, 1988). A direct feedback regulation of transcription levels by the accumulated substrate intermediates has been postulated.

1.2 GENERAL FEATURES OF (ALTERNATIVE) SPLICING

The constitutive splicing of pre-mRNA molecules in the nucleus involves exact removal of intronic sequences that are demarcated by consensus 5' splice donor sites (EXON ... AG↓gtaagt ... intron) and 3' splice acceptor sites, flanked upstream by a polypyrimidine tract of about 6 residues (intron ... 6PyNcag↓ ... EXON). The splicing machinery is able to recognize the consensus sequences, splice out the intervening sequences and join the remaining exons (for reviews see Smith *et al.*, 1989; Leff *et al.*, 1986; Padgett *et al.*, 1986) The process occurs in a multistep manner. After the formation of a complex called the spliceosome, it begins with exact cleavage at the 5' donor site and continues with the joining of 5' intronic sequences to an adenosine residue located in a less conserved area, called the branch point sequence. At this stage a "lariat" form is generated of the intervening sequences. Cis-acting sequences within the intron as well as the adjacent exon are recognized by specific transacting factors that appear to insure that correct splicing takes place at a genuine splice site rather than at cryptic sites. In case of mutations at either the donor or the acceptor splice sites, cryptic sites might be recognized by the splicing machinery.

Alternative splicing is an elegant method employed by the cellular machinery to create protein diversity, thereby modulating or deleting specific protein functions. Furthermore it may alter the subcellular localization of a certain protein, although its function is maintained. The process may also serve as a direct mechanism of regulation of gene expression. Several modes of alternative splicing have been described (for reviews see Smith *et al.*, 1989; Andreadis *et al*, 1987).

A key issue is how alternative splicing is accomplished and regulated. Cis-regulation of splicing occurs because of the intrinsic characteristics of the mRNA template. For example multiple 5' or 3' splice sites may arise from the formation of diverse 5' or 3' ends of mRNA templates, due to initiation from differential promoters or termination at differential 3' sites (Smith *et al.*, 1989). Splice site selection is negatively regulated in a number of cases by the presence of specific

DNA sequences (Dabeva et al., 1986; Emeson *et al.*, 1989; Helfman *et al.*, 1990; Goux-Pelletan, 1990). Whether secondary structures in the mRNA template may play a role in the cis-regulation of alternative splicing is unknown.

A specifically regulated form of alternative splicing will be encountered in cells where variations in activities or levels of constitutive factors, like the splicing factor 2 (SF2; Krainer *et al.*, 1990) or the alternative splicing factor (ASF; Ge and Manley, 1990), could lead to alternate splicing patterns (for reviews see Smith *et al.*, 1989; Maniatis, 1991). Unique trans-acting factors might influence the splicing mechanism as it has been shown in studies on the sex determination pathway of Drosophila. Here, alternative splicing of the "transformer" pre-mRNA is regulated by the "Sex lethal" (Sxl) gene product functioning as transacting factor (for review see Maniatis, 1991). Until now, however, splice site selection by specific repressor proteins, like the aforementioned Sxl or the suppressor of the white apricot locus in Drosophila, is the only mechanism of trans-regulation that has been demonstrated (Smith *et al.*, 1989; Maniatis, 1991). Recent studies suggest that (alternate) splicing and polyadenylation might be coupled (Niwa *et al.*, 1990; Huang and Gorman, 1990; Pandey *et al.*, 1990; Hedley and Maniatis, 1991).

1.2.1 Alternative splicing of pre-mRNAs encoding lysosomal proteins

Among mRNAs for human lysosomal hydrolases alternatively spliced transcripts of ß-glucuronidase (Oshima *et al.*, 1987), sphingomyelinase (Quintern *et al.*, 1989; Schuchman *et al.*, 1991), ß-galactosidase (Morreau *et al.*, 1989; Yamamoto *et al.*, 1990), the sulfatide activator precursor protein (Nakano *et al.*, 1989; Zhang *et al.*, 1990; Rafi *et al.*, 1990; Zhang *et al.*, 1991; Holtschmidt *et al.*, 1991a), glucosylasparaginase (Fisher and Aronson, 1991) and α-N-acetylgalactosaminidase (Yamauchi *et al.*, 1990; Wang *et al.*, 1990) have been identified. Alternative splicing of placental ß-glucuronidase pre-mRNA leads to the skipping of a 153 bp exon. More complex alternatively spliced transcripts have been identified for acid sphingomyelinase and acid ß-galactosidase. In human placenta and fibroblasts 10% of the sphingomyelinase mRNA pool lacks an exon of 172 bp, but instead contains an in frame 40 bp intronic sequence, encoding 13 aminoacids not present in the normal protein (Quintern *et al.*, 1989). A cDNA derived from the latter transcript was also isolated from a retina cDNA library (Schuchmann *et al.*, 1991). A third type of sphingomyelinase mRNA, with a frequency of less than 1% lacks only the 172 bp exon (Schuchmann *et al.*, 1991). This third type of mRNA has been isolated from a human testis cDNA library and is also present in placenta. In human testis and cultured fibroblasts, alternative splicing of ß-galactosidase pre-mRNA occurs involving the exclusion of 3 different exons (see chapter 2).

In human brain only minor amounts of an alternate α-N-acetylgalactosaminidase mRNA is encountered, containing a 70 bp insertion (Yamauchi, 1990). Genomic analysis, however, could not identify the 70 bp stretch within the intervening sequences (Wang *et al.*, 1990). Curiously, a deletion of 45 bp in the normal α-N-acetylgalactosaminidase message, approximately at the same position of the 70 bp insertion, has been also detected in yet another alternatively spliced transcript. (Wang *et al.*, 1990).

A 9 bp insertion, normally occurring in mRNA encoding the sulfatide activator protein (SAP-I) has been identified in cDNA clones from hepatoma-, lung-

and skin fibroblast- libraries (Nakano *et al.*, 1989). Two cDNAs with a 9 bp and 6 bp insertion, respectively, overlapping for six nucleotides have been PCR-amplified from human fibroblasts. The 9 bp insertion is encoded by a separate exon (Holtschmidt *et al.*, 1990; Zhang *et al.*, 1991).

Transient expression in COS-I cells of ß-glucuronidase (Oshima *et al.*, 1987), sphingomyelinase (Schuchman *et al.*,1991) and ß-galactosidase (Morreau *et al.*, 1989) cDNAs derived for alternaltively spliced transcripts gives rise to proteins without catalytic activity towards standard synthetic substrates and at acidic buffer conditions. Whether translated products of the alternate transcripts are present in vivo and have any biological function is at the moment unknown.

In conclusion: the biological role of the alternative splicing among pre-mRNAs encoding lysosomal hydrolases is unknown. However, it is very tempting to speculate that generation of these mRNAs is a mode used by the cellular machinery to regulate expression of a specific enzyme.

Alternate 3' end formation of mRNAs may arise from alternative splicing of 3' exons that each contain a polyadenylation site (see for review Smith *et al.*, 1989). More frequently, however, different 3' ends emerge from the choice of differential polyadenylation sites present in a single 3' exon. Transcripts with poly A tails of variable length and specific features of the 3' UTR seem to have different stabilities (see for Jackson and Standart, 1990) Furthermore, translation efficiency might be influenced by alternate 3' ends. Among the lysosomal enzymes alternate 3' end formation was described for human arylsulphatase A (Kreysing *et al.*, 1990), ß-hexosaminidase α-chain (Proia and Soravia, 1987), glucocerebrosidase (Reiner *et al.*, 1988), the sulfatide activator precursor protein (Nakao *et al.*, 1989) and human α-N-acetylgalactosaminidase (Wang *et al.*, 1990).

1.3 REFERENCES

Andreadis, A., M.E. Gallego and B. Nadal-Ginard. 1987. Generation of protein isoform diversity by alternative splicing: mechanistic and biological implications. Annu Rev Cell Biol 3:207-42.

Banerji, J., S. Rusconi and W. Schaffner. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell 27:299-308.

Bishop, D.F., R. Kornreich and R.J. Desnick. 1988. Structural organization of the human alpha-galactosidase A gene: further evidence for the absence of a 3' untranslated region. Proc Natl Acad Sci U S A 85:3903-7.

Bracey, L.T. and K. Paigen. 1987. Changes in translational yield regulate tissue-specific expression of beta-glucuronidase. Proc Natl Acad Sci U S A 84:9020-4.

Breathnach, R. and P. Chambon. 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu Rev Biochem 50:349-83.

Cavailles, V., M. Garcia and H. Rochefort. 1989. Regulation of cathepsin-D and pS2 gene expression by growth factors in MCF7 human breast cancer cells. Mol Endocrinol 3:552-8.

Dabeva, M.D., M.A. Post-Beittenmiller and J.R. Warner. 1986. Autogenous regulation of splicing of the transcript of a yeast ribosomal protein gene. Proc Natl Acad Sci U S A 83:5854-7.

Denker, H.W., L.A. Eng and H. C.E. 1978. Studies on the early development and implantation in the cat. Implantation: proteinases. Anat. Embryol. 154:39-54.

Dimino, M.J. and E.A. Elfont. 1980. The role of lysosomes in ovarian physiology. 196-201.

Dorn, A., A. Muller, H. Kirschke, B. Wiederanders and H.G. Bernstein. 1986. Immunohistochemical analysis of the distribution of cathepsin D throughout human nervous system with reference to developmental aspects. Biomed Biochim Acta 45:1457-60.

Dufek, V., B. Matous, V. Kral and L. Bures. 1984. Characterization of cathepsin B-like proteinase from ascitic fluid of patients with primary liver cancer. Neoplasma 31:581-90.

Dynan, W.S. and R. Tjian. 1985. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature 316:774-8.

Emeson, R.B., F. Hedjran, J.M. Yeakley, J.W. Guise and M.G. Rosenfeld. 1989. Alternative production of calcitonin and CGRP mRNA is regulated at the calcitonin-specific splice acceptor. Nature 341:76-80.

Fisher, K.J. and N.J. Aronson. 1991. Deletion of exon 8 causes glycosylasparaginase deficiency in an African American aspartylglucosaminuria (AGU) patient. Febs Lett 288:173-8.

Galjart, N.J., N. Gillemans, D. Meijer and A. d'Azzo. 1990. Mouse "protective protein". cDNA cloning, sequence comparison, and expression. J Biol Chem 265:4678-84.

Galjart, N.J., H. Morreau, R. Willemsen, N. Gillemans, E.J. Bonten and A. d'Azzo. 1991. Human lysosomal protective protein has cathepsin A-like activity distinct from its protective function. J Biol Chem 266:14754-62.

Garcia, A.D., A.M. O'Connell and S.J. Sharp. 1987. Formation of an active transcription complex in the Drosophila melanogaster 5S RNA gene is dependent on an upstream region. Mol Cell Biol 7:2046-51.

Ge, H. and J.L. Manley. 1990. A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. Cell 62:25-34.

Geier, C., K. von Figura and R. Pohlmann. 1989. Structure of the human lysosomal acid phosphatase gene. Eur J Biochem 183:611-6.

Groudine, M. and C. Casimir. 1984. Post-transcriptional regulation of the chicken thymidine kinase gene. Nucleic Acids Res 12:1427-46.

Hedley, M.L. and T. Maniatis. 1991. Sex-specific splicing and polyadenylation of dsx pre-mRNA requires a sequence that binds specifically to tra-2 protein in vitro. Cell 65:579-86.

Helfman, D.M., R.F. Roscigno, G.J. Mulligan, L.A. Finn and K.S. Weber. 1990. Identification of two distinct intron elements involved in alternative splicing of beta-tropomyosin pre-mRNA. Genes Dev 4:98-110.

Hoefsloot, L.H., M. Hoogeveen-Westerveld, A.J. Reuser and B.A. Oostra. 1990. Characterization of the human lysosomal alpha-glucosidase gene. Biochem J 272:493-7.

Holtschmidt, H., K. Sandhoff, W. Furst, H.Y. Kwon, D. Schnabel and K. Suzuki. 1991. The organization of the gene for the human cerebroside sulfate activator protein. Febs Lett 280:267-70.

Holtschmidt, H., K. Sandhoff, H.Y. Kwon, K. Harzer, T. Nakano and K. Suzuki. 1991. Sulfatide activator protein. Alternative splicing that generates three mRNAs and a newly found mutation responsible for a

clinical disease. J Biol Chem 266:7556-60.
Horowitz, M., S. Wilder, Z. Horowitz, O. Reiner, T. Gelbart and E. Beutler. 1989. The human glucocerebrosidase gene and pseudogene: structure and evolution. Genomics 4:87-96.
Huang, M.T. and C.M. Gorman. 1990. The simian virus 40 small-t intron, present in many common expression vectors, leads to aberrant splicing. Mol Cell Biol 10:1805-10.
Imagawa, M., R. Chiu and M. Karin. 1987. Transcription factor AP-2 mediates induction by two different signal-transduction pathways: protein kinase C and cAMP. Cell 51:251-60.
Jackman, H.L., P.W. Morris, P.A. Deddish, R.A. Skidgel and E.G. Erdos. 1992. Inactivation of endothelin I by deamidase (lysosomal protein). J Biol Chem 267:2872-2875.
Jackman, H.L., F.L. Tan, H. Tamei, C. Beurling-Harbury, X.Y. Li, R.A. Skidgel and E.G. Erdos. 1990. A peptidase in human platelets that deamidates tachykinins. Probable identity with the lysosomal "protective protein". J Biol Chem 265:11265-72.
Jackson, R.J. and N. Standart. 1990. Do the poly(A) tail and 3' untranslated region control mRNA translation? Cell 62:15-24.
Kageyama, R. and I. Pastan. 1989. Molecular cloning and characterization of a human DNA binding factor that represses transcription. Cell 59:815-25.
Katunuma, N. and E. Kominami. 1983. Structures and functions of lysosomal thiol proteinases and their endogenous inhibitor. Curr Top Cell Regul 22:71-101.
Kaufman, R.J. and P.A. Sharp. 1983. Growth-dependent expression of dihydrofolate reductase mRNA from modular cDNA genes. Mol Cell Biol 3:1598-608.
Kearse, K.P., D.R. Cassatt, A.M. Kaplan and D.A. Cohen. 1988a. Characterization of cell surface molecules involved in the recognition of antigen-presenting cells by cloned helper T-cell lines. Cell Immunol 115:334-51.
Kearse, K.P., D.R. Cassatt, A.M. Kaplan and D.A. Cohen. 1988b. The requirement for surface Ig signaling as a prerequisite for T cell:B cell interactions. A possible role for desialylation. J Immunol 140:1770-8.
Krainer, A.R., G.C. Conway and D. Kozak. 1990. The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. Cell 62:35-42.
Kreysing, J., K. von Figura and V. Gieselmann. 1990. Structure of the arylsulfatase A gene. Eur J Biochem 191:627-31.
Lee, W., P. Mitchell and R. Tjian. 1987. Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. Cell 49:741-52.
Leff, S.E., M.G. Rosenfeld and R.M. Evans. 1986. Complex transcriptional units: diversity in gene expression by alternative RNA processing. Annu Rev Biochem 55:1091-117.
Maniatis, T. 1991. Mechanisms of alternative pre-mRNA splicing. Science 251:33-4.
Martiniuk, F., M. Mehler, S. Tzall, G. Meredith and R. Hirschhorn. 1990. Sequence of the cDNA and 5'-flanking region for human acid alpha-glucosidase, detection of an intron in the 5' untranslated leader sequence, definition of 18-bp polymorphisms, and differences with previous cDNA and amino acid sequences. Dna Cell Biol 9:85-94.
Mitchell, P.J. and R. Tjian. 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245:371-8.
McKnight, S. and R. Tjian. 1986. Transcriptional selectivity of viral genes in mammalian cells. Cell 46:795-805.
McRorie, R.A. and W.L. Williams. 1974. Biochemistry of mammalian fertilization. Ann. Rev. Biochem. 43:777-803.
Moreau, P., R. Hen, B. Wasylyk, R. Everett, M.P. Gaub and P. Chambon. 1981. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res 9:6047-68.
Morreau, H., E. Bonten, X.Y. Zhou and A. D'Azzo. 1991. Organization of the gene encoding human lysosomal beta-galactosidase. Dna Cell Biol 10:495-504.
Morreau, H., N.J. Galjart, N. Gillemans, R. Willemsen, G.T. van der Horst and A. d'Azzo. 1989. Alternative splicing of beta-galactosidase mRNA generates the classic lysosomal enzyme and a beta-galactosidase-related protein. J Biol Chem 264:20655-63.
Mort, J.S., A.D. Recklies and A.R. Poole. 1985. Release of cathepsin B precursors from human and murine tumours. Prog Clin Biol Res 180:243-5.

Nakano, T., K. Sandhoff, J. Stumper, H. Christomanou and K. Suzuki. 1989. Structure of full-length cDNA coding for sulfatide activator, a Co-beta-glucosidase and two other homologous proteins: two alternate forms of the sulfatide activator. J Biochem (Tokyo) 105:152-4.
Nanba, E. and K. Suzuki. 1991. Organization of the mouse acid beta-galactosidase gene. Biochem Biophys Res Commun 178:158-64.
Neote, K., B. Bapat, A. Dumbrille Ross, C. Troxel, S.M. Schuster, D.J. Mahuran and R.A. Gravel. 1988. Characterization of the human HEXB gene encoding lysosomal beta-hexosaminidase. Genomics 3:279-86.
Niwa, M., S.D. Rose and S.M. Berget. 1990. In vitro polyadenylation is stimulated by the presence of an upstream intron. Genes Dev 4:1552-9.
Oshima, A., J.W. Kyle, R.D. Miller, J.W. Hoffmann, P.P. Powell, J.H. Grubb, W.S. Sly, M. Tropak, K.S. Guise and R.A. Gravel. 1987. Cloning, sequencing, and expression of cDNA for human beta-glucuronidase. Proc Natl Acad Sci U S A 84:685-9.
Padgett, R.A., P.J. Grabowski, M.M. Konarska, S. Seiler and P.A. Sharp. 1986. Splicing of messenger RNA precursors. Annu Rev Biochem 55:1119-50.
Paigen. 1979. Acid hydrolases as models of genetic control . Ann.Rev.Genet. 13:417-466.
Palmer, R., P.M. Gallagher, W.L. Boyko and R.E. Ganschow. 1983. Genetic control of levels of murine kidney glucuronidase mRNA in response to androgen. Proc Natl Acad Sci U S A 80:7596-600.
Pandey, N.B., N. Chodchoy, T.J. Liu and W.F. Marzluff. 1990. Introns in histone genes alter the distribution of 3' ends. Nucleic Acids Res 18:3161-70.
Piechaczyk, M., J.M. Blanchard, L. Marty, C. Dani, F. Panabieres, S. El Sabouty, P. Fort and P. Jeanteur. 1984. Post-transcriptional regulation of glyceraldehyde-3-phosphate-dehydrogenase gene expression in rat tissues. Nucleic Acids Res 12:6951-63.
Proia, R.L. 1988. Gene encoding the human beta-hexosaminidase beta chain: extensive homology of intron placement in the alpha- and beta-chain genes. Proc Natl Acad Sci U S A 85:1883-7.
Proia, R.L. and E. Soravia. 1987. Organization of the gene encoding the human beta-hexosaminidase alpha-chain [published erratum appears in J Biol Chem 1987 Nov 5;262(31):15322]. J Biol Chem 262:5677-81.
Quinn, M., P. Hantzopoulos, V. Fidanza and D.H. Calhoun. 1987. A genomic clone containing the promoter for the gene encoding the human lysosomal enzyme, alpha-galactosidase A. Gene 58:177-88.
Quintern, L.E., E.H. Schuchman, O. Levran, M. Suchi, K. Ferlinz, H. Reinke, K. Sandhoff and R.J. Desnick. 1989. Isolation of cDNA clones encoding human acid sphingomyelinase: occurrence of alternatively processed transcripts. Embo J 8:2469-73.
Rafi, M.A., X.L. Zhang, G. DeGala and D.A. Wenger. 1990. Detection of a point mutation in sphingolipid activator protein-1 mRNA in patients with a variant form of metachromatic leukodystrophy. Biochem Biophys Res Commun 166:1017-23.
Rahi, H. and P.N. Srivistava. 1983. J. Reprod. Fert. 67:447-455.
Redecker, B., B. Heckendorf, H.W. Grosch, G. Mersmann and A. Hasilik. 1991. Molecular organization of the human cathepsin D gene. Dna Cell Biol 10:423-31.
Redecker, B., M. Horst and A. Hasilik. 1989. Calcitriol enhances transcriptional activity of lysozyme and cathepsin D genes in U937 promonocytes. Biochem J 262:843-7.
Reid, W.A., M.J. Valler and J. Kay. 1986. Immunolocalization of cathepsin D in normal and neoplastic human tissues. J Clin Pathol 39:1323-30.
Reiner, O. and M. Horowitz. 1988. Differential expression of the human glucocerebrosidase-coding gene. Gene 73:469-78.
Reiner, O., M. Wigderson and M. Horowitz. 1988. Structural analysis of the human glucocerebrosidase genes. Dna 7:107-16.
Rochefort, H., F. Capony, M. Garcia, V. Cavailles, G. Freiss, M. Chambon, M. Morisset and F. Vignon. 1987. Estrogen-induced lysosomal proteases secreted by breast cancer cells: a role in carcinogenesis? J Cell Biochem 35:17-29.
Ross, J. 1989. The turnover of messenger RNA. Scientific American 3:28-35.
Ryan, R.E., J.D. Crissman, K.V. Honn and B.F. Sloane. 1985. Cathepsin B-like activity in viable tumor cells isolated from rodent tumors. Cancer Res 45:3636-41.
San Segundo, B., S.J. Chan and D.F. Steiner. 1986. Differences in cathepsin B mRNA levels in rat tissues suggest specialized functions. Febs Lett 201:251-6.
Schuchman, E.H., M. Suchi, T. Takahashi, K. Sandhoff and R.J. Desnick. 1991. Human acid

sphingomyelinase. Isolation, nucleotide sequence and expression of the full-length and alternatively spliced cDNAs. J Biol Chem 266:8531-9.
Sen, R. and D. Baltimore. 1986. Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell 46:705-16.
Shipley, J.M., R.D. Miller, B.M. Wu, J.H. Grubb, S.G. Christensen, J.W. Kyle and W.S. Sly. 1991. Analysis of the 5' flanking region of the human beta-glucuronidase gene. Genomics 10:1009-18.
Sloane, B.F., J. Rozhin, J.S. Hatfield, J.D. Crissman and K.V. Honn. 1987. Plasma membrane-associated cysteine proteinases in human and animal tumors. Exp Cell Biol 55:209-24.
Smale, S.T. and D. Baltimore. 1989. The "initiator" as a transcription control element. Cell 57:103-13.
Smith, C.W., J.G. Patton and B. Nadal-Ginard. 1989. Alternative splicing in the control of gene expression. Annu Rev Genet 23:527-77.
Wang, A.M., D.F. Bishop and R.J. Desnick. 1990. Human alpha-N-acetylgalactosaminidase-molecular cloning, nucleotide sequence, and expression of a full-length cDNA. Homology with human alpha-galactosidase A suggests evolution from a common ancestral gene. J Biol Chem 265:21859-66.
Wang, A.M. and R.J. Desnick. 1991. Structural organization and complete sequence of the human alpha-N-acetylgalactosaminidase gene: homology with the alpha-galactosidase A gene provides evidence for evolution from a common ancestral gene. Genomics 10:133-42.
Wang, E., T.R. Korfhagen, P.M. Gallagher, M.A. D'Amore, J. McNeish, S.S. Potter and R.E. Ganschow. 1988. Overlapping transcriptional units on the same strand within the murine beta-glucuronidase gene complex. J Biol Chem 263:15841-4.
Wassarman, P.M. 1987. The biology and chemistry of fertilization. Science 235:553-60.
Yamamoto, Y., C.A. Hake, B.M. Martin, K.A. Kretz, A.J. Ahern-Rindell, S.L. Naylor, M. Mudd and J.S. O'Brien. 1990. Isolation, characterization, and mapping of a human acid beta-galactosidase cDNA. Dna Cell Biol 9:119-27.
Yamauchi, T., M. Hiraiwa, H. Kobayashi, Y. Uda, T. Miyatake and S. Tsuji. 1990. Molecular cloning of two species of cDNAs for human alpha-N-acetylgalactosaminidase and expression in mammalian cells. Biochem Biophys Res Commun 170:231-7.
Zhang, X.L., M.A. Rafi, G. DeGala and D.A. Wenger. 1990. Insertion in the mRNA of a metachromatic leukodystrophy patient with sphingolipid activator protein-1 deficiency. Proc Natl Acad Sci U S A 87:1426-30.
Zhang, X.L., M.A. Rafi, G. DeGala and D.A. Wenger. 1991. The mechanism for a 33-nucleotide insertion in mRNA causing sphingolipid activator protein (SAP-1)-deficient metachromatic leukodystrophy. Hum Genet 87:211-5.

CHAPTER II

2.1 INTRODUCTION TO THE EXPERIMENTAL WORK

2.1.1 ß-galactosidase: the enzyme and its deficiency

From 1967 on several patients have been described with excessive accumulation of G_{M1}-ganglioside in neuronal tissues as well as glycosaminoglycans and glycopeptides in visceral organs and other tissues. This syndrome, called G_{M1}-gangliosidosis, is inherited as an autosomal recessive trait and is characterized by an isolated deficiency of acid ß-D-galactosidase (Sacrez *et al.*, 1967; Okada and O'Brien, 1968). Depending on the clinical symptomatology, patients with this disease have been classified as having either the severe infantile, or the mild juvenile and adult form of the disease (O'Brien, 1989). Residual activity of ß-galactosidase among these clinically distinct patients ranges from less than 1% to 15% of normal levels. Patients with the infantile form show a severe neurodegenerative disorder with psychomotor deterioration, macular cherry-red spots, bone abnormalities and hepatosplenomegaly. These features lead to an early death. A milder clinical course with slow progressive mental deterioration, only slight bone abnormalities or visceromegaly is seen in patients with the late-infantile/juvenile form. Features such as dysarthria, choreoathetosis, mild or absent bone abnormalities, only little or no intellectual impairment and a prolonged life expectancy are characteristic of adult cases.

A second lysosomal storage disorder, Morquio B syndrome, classified as one of the mucopolysaccharidoses, is also due to a single deficiency of acid ß-galactosidase (O'Brien *et al.*, 1976; Groebe *et al.* 1980). The residual enzyme activity in these cases is about 10% of normal levels. The patients present with a progressive skeletal dysplasia within the first decade of life, but, in contrast to those with G_{M1}-gangliosidosis, they have no neurological impairment since they do not accumulate G_{M1}-ganglioside in neuronal tissues. In fact, the residual ß-galactosidase in Morquio B syndrome has been shown to have an altered Km, affecting its affinity for keratansulfate and oligosaccharide substrates but not for G_{M1}-ganglioside (Paschke and Kresse, 1982; van der Horst, 1983).

Lysosomal ß-D-galactosidase hydrolazes non-reducing terminal galactose moieties from a wide variety of substrates, like G_{M1}-gangliosidosis, asialo-G_{M1}-gangliosides, asialofetuin, keratansulfate, lactosylceramide and glycopeptides (O'Brien, 1989; Conzelmann and Sandhoff, 1987). The enzyme has a pH optimum of ~4.2, is heat labile and is activated by chloride ions. The halflife of ß-galactosidase ranges from 2.0-3.5 days in mouse macrophages (Skudlarek and Swank, 1981; Tropea *et al.*, 1988) to 10 days in human skin fibroblasts (van Diggelen *et al.*, 1981). In vitro degradation of lipid substrates, like G_{M1}-ganglioside, by ß-galactosidase is accelerated in the presence of the activator protein, SAP-1 (Fischer and Jatzkewitz, 1975; Li and Li, 1976). The latter probably functions as a natural detergent, extracting the water insoluble substrate from membranes to make it available for hydrolysis. The enzyme also hydrolyzes artificial substrates, among which is the fluorescent compound 4-methylumbelliferyl-ß-D-galactopyranoside, used in most studies.

In a genetically distinct lysosomal storage disease, named galactosialidosis (Andria *et al.*, 1981), a coexistent deficiency of ß-D-galactosidase and N-acetyl-α-neuraminidase is caused by a primary defect of the so called protective protein (d'Azzo *et al.*, 1982). After the first report of a patient with such a combined ß-

galactosidase and neuraminidase deficiency (Wenger *et al.*, 1978), a number of cases have emerged, originally diagnosed as variants of G_{M1}-gangliosidosis. This disorder is rare and is transmitted with an autosomal recessive mode of inheritance. Based on the age of onset of the disease three phenotypes are again distinguished, ranging from severe early infantile forms resulting in early death or fetal hydrops to late infantile and juvenile/adult cases with milder manifestations and slower progression of the disease (see for reviews, Andria *et al.*, 1981; Galjaard *et al.*, 1984; Suzuki *et al.*, 1984; O'Brien, 1989). The accumulated products in placenta and urine from galactosialidosis patients are mainly sialylated oligosaccharides, as in the disorder sialidosis, due to an isolated neuraminidase deficiency (Okada *et al.*, 1978; van Pelt *et al.*, 1988a; 1988b; 1989). Thus, the major cause of the pathogenesis in these patients seems to be the very severe deficiency of neuraminidase activity (<1%). The 10-15% residual ß-galactosidase activity is apparently sufficient to remove terminal galactosyl residues.

2.1.2 ß-galactosidase/protective protein/neuraminidase complex

The enzymatic properties of ß-galactosidase have been extensively studied in a number of tissues and species. Depending upon the extraction procedure and source of material, ß-galactosidase activity is recovered in mono-, di- tetra- and multimeric forms (e.g. Hoeksema *et al.*, 1979; Norden *et al.*, 1974; Verheijen *et al.*, 1982; 1985; Hoogeveen *et al.*, 1983; Yamamoto *et al.*, 1982; Yamamoto and Nishimura, 1987; Potier *et al.* 1990; Scheibe *et al.*, 1990). The mature enzyme can be isolated using the affinity matrix p-aminophenylthiogalactoside-CH-Sepharose (van Diggelen et al., 1981). Using this purification procedure, ß-galactosidase was shown to copurify in complex with the protective protein and neuraminidase from bovine and porcine testis as well as human placenta glycoprotein preparations (Verheijen *et al.*, 1982; 1985; Yamamoto and Nishimura, 1987). A number of studies have pointed to the fact that ß-galactosidase, protective protein and neuraminidase form a complex because they depend on each other for the acquisition of full enzymatic activity and stability in lysosomes (d'Azzo *et al.* 1982; Hoogeveen *et al.*, 1983; Verheijen *et al.*, 1985; van der Horst *et al.*, 1989). In particular, the protective protein has been shown, to promote multimerization of ß-galactosidase monomers, thereby preventing their rapid intralysosomal degradation (Hoogeveen *et al.*, 1983). From studies on lysosomal neuraminidase, it has become clear that the latter dissociated from the rest of the complex is inactive (Verheijen *et al.*, 1982; 1985). It has further been determined by in vitro reconstitution experiments that neuraminidase needs to associate with the protective protein to be enzymatically active. Additional interaction with ß-galactosidase confers stability to this enzyme (van der Horst *et al.*, 1989). The stoichometry of the complex is not yet understood, althought Potier and coworkers (1990) have attempted to determine the molecular distribution of the three enzymes after gel filtration and HPLC of a purified placental preparation. These authors have detected three distinct forms with ß-galactosidase activity, of which only the largest species, with a molecular mass of about 680 kDa, expressed neuraminidase activity.

2.1.3 Biosynthesis and processing of ß-galactosidase

Immunoprecipitation experiments, using anti-ß-galactosidase antibodies have revealed that in normal cultured fibroblasts ß-galactosidase is synthesized as a precursor of 85 kDa which is posttranslationally processed through a number of steps into a mature 64-kDa enzyme (d'Azzo *et al.*, 1982). A partially processed intermediate of 66 kDa is occasionally observed in normal and galactosialidosis fibroblasts, that have been treated with the protease inhibitor leupeptin (d'Azzo *et al.*, 1982; Palmeri *et al.*, 1986). A similar processing patters has been observed for murine ß-galactosidase in macrophages (Skudlarek and Swank, 1979). The same authors have more recently studied the biosynthesis and maturation of the enzyme in mouse kidney after incorporation of either ^{3}[H]fucose or ^{35}[S]methionine, followed by immunoprecipitation with monospecific antibodies (Skudlarek and Swank, 1988). Forms of 85, 65 and 58 kDa are labeled with ^{3}[H]fucose. However, after labeling with ^{35}[S]methionine, polypeptides of 83, 63 and 56 kDa are immunoprecipitated. They have concluded that fucose-containing ß-galactosidase must represent a very small percentage of the total precursor pool. In contrast to the total pool of ß-galactosidase labeled with ^{35}[S]methionine, the ^{3}[H]fucose labeled molecules are very rapidly processed, after acquiring this sugar residue in the Golgi complex. Furthermore, it has been shown for macrophages labeled with ^{3}[H]mannose and for mouse kidney labeled with ^{3}[H]fucose that proteolytic maturation of ß-galactosidase releases peptides, which are heavily glycosylated (Skudlarek and Swank, 1981; 1988). It has been calculated that the mature enzyme purified from human liver and mouse macrophages contains only three carbohydrate chains per molecule (Overdijk *et al.*, 1986; Tropea *et al.*, 1988).

Studies on the biosynthesis and processing of ß-galactosidase in mutant cells have shown that in fibroblasts of an infantile and an adult G_{M1}-gangliosidosis patient a normally sized ß-galactosidase precursor protein is synthesized in a reduced amount and no mature form is detected. However, in a Morquio B cell strain synthesis and processing of the enzyme occur normally (Hoogeveen *et al.*, 1984). In cells from clinically different galactosialidosis patients ß-galactosidase is synthesized as a precursor molecule in normal amounts, only partially processed and rapidly degraded (d'Azzo *et al.*, 1982; Palmeri *et al.*, 1986). This is in agreement with the calculated turnover time of the enzyme in these mutant cells which is about ten-fold higher than that of the normal enzyme due to the deficiency of the protective protein (van Diggelen *et al.*, 1982).

2.1.4 ß-Galactosidase genes and cDNAs

The cDNAs and genes encoding human and mouse ß-galactosidase have been isolated and characterized (see publication 1 and 2; Nanba and Suzuki, 1990; 1991; Oshima *et al.*, 1988; Yamamoto *et al.*, 1990). The primary structure and expression of the human lysosomal enzyme as well as of a ß-galactosidase-related protein derived from an alternatively spliced transcript are described in publication 1. The latter protein is transported to the Golgi complex but does not reach the lysosomes and is inactive towards the artificial substrates used (publication 1). The amino acid sequence of murine ß-galactosidase, deduced from its cloned cDNA, is about 80% identical to that of the human enzyme (Nanba and Suzuki, 1990). Similarity is significantly lower at the C-termini of the two proteins, where 30 additional residues are encountered only

in the human molecule. Of the seven potential N-linked glycosylation sites present in human ß-galactosidase (publication 1), five are found at conserved positions in the murine enzyme, that carries eight in total. All cysteine residues are conserved between the two species, only one extra is present in the human protein.

The analysis of the gene encoding human ß-galactosidase and the partial characterization of its promoter region are described in publication 2. The mouse gene spans more than 80 kb of genomic DNA and consists, like the human gene, of 16 exons (Nanba and Suzuki, 1991). Exon/intron boundaries are completely conserved between the two species. The only difference is the length of exon I, XIII and XVI. Analysis of the murine promoter region has revealed several potential binding sites for the transcription factor SP1 within a GC-rich region, as it is the case for the human promoter. Furthermore, no TATA or CCAAT boxes are found at typical positions. Thus, the murine as well as the human promoter regions have characteristics of housekeeping gene promoters (see publication 2; Nanba and Suzuki, 1991).

The gene encoding human ß-galactosidase has been localized on the short arm of chromosome 3 (Shows *et al.*, 1979; Naylor *et al.*, 1982). Recent mapping of somatic cell hybrids with a partial ß-galactosidase cDNA has located the gene in the region 3p21-3pter (Yamamoto *et al.*, 1990). The murine gene is on chromosome 9 (Shows *et al.*, 1979; Naylor *et al.*, 1982).

The characterization of the human ß-galactosidase cDNA has enabled the identification of several mutations underlying different clinical forms of G_{M1}-gangliosidosis and Morquio B syndrome. The analysis of one genetic lesion present in two siblings with the severe infantile form of G_{M1}-gangliosidosis is described in publication 3. The mutations identified in clinically distinct Japanese patients are summarized in the table on page 28 (see References therein). In this table nucleotide and resulting amino acid changes are correlated with mRNA and protein expression. Five of the six G_{M1}-gangliosidosis patients with the infantile form carry different mutations, although in most of the cases only one allele has been identified. The ß-galactosidase activity measured after expression of the single allelic mutation in either COS-1 cells or transformed G_{M1}-gangliosidosis fibroblasts seem to match with an infantile phenotype. The C to T base substitution at position 1369 of the cDNA creates a stop codon and might render the mRNA unstable, since only trace amount of ß-galactosidase transcript is detected on Northern blots. Among 5 juvenile G_{M1}-gangliosidosis patients a common mutation is found, where Arg^{201} is changed into a Cys. Unfortunately, expression of this mutation in either COS-1 cells or transformed G_{M1}-gangliosidosis fibroblasts gives different results in two laboratories. The substitution of Ile^{51} to Thr is found in all adult patients. Ten of them are homozygotes for this mutation, whereas one is a compound heterozygote carrying a Gln instead of an Arg at position 457 in the other mutant allele. Expression data are in agreement with the adult phenotype. One common mutation, consisting of a substitution of Trp^{273} to Leu, is detected in mRNA samples of three different Morquio B patients.

2.1.5 ß-Galactosidase and protective protein

As mentioned earlier, ß-galactosidase requires the presence of the protective protein to resist rapid degradation in lysosomes. In fact, in normal cells at least part of the total protective protein pool is associated with ß-galactosidase and neuraminidase. This is reflected in the existence in galactosialidosis patients of a combined deficiency

of the two glycosidases due to a primary defect of the protective protein (see 3.1.1; 3.1.2 and References therein).

Little was known about the structural and enzymatic properties of the protective protein before the isolation and characterization of its cDNA. The biogenesis of the protective protein and its multifunctional activities have been recently reviewed in extent (Galjart, 1991).

In normal tissues and cultured cells the protective protein is synthesized as a 54-kDa precursor that is modified in the lysosomal/endosomal compartment into a two-chain mature product of 32- and 20- kDa polypeptides, held together by disulfide bridges (Galjart *et al.*, 1988). The predicted amino acid sequences of human, mouse and chicken protective proteins have a high degree of identity among each other and are homologous to the serine carboxypeptidase family of enzymes (Galjart *et al.*, 1988; 1990 and publication 4). Furthermore, the human protein is likely to be identical to a deamidase/carboxypeptidase isolated from human platelets that is involved in the degradation at least in vitro of bioactive peptides, like oxytocin, substance P, angiotensin I and endothelin I (Jackman *et al.*, 1990; 1992). We have established that the protective protein can function as a carboxypeptidase and that its catalytic activity overlaps with that of lysosomal cathepsin A (see publication 4 and references therein). The generation of active site mutants of the human protective protein brought us to the understanding that its catalytic activity is distinct from its protective function towards ß-galactosidase and neuraminidase. These findings are discussed in publication 4. Finally, we have analyzed a set of mutated protective proteins, carrying targeted amino acid substitutions, in order to define signals in the human protein important for its intracellular transport, and to determine the site of its association with ß-galactosidase. These results and major conclusions are reported in publication 5.

Clinical subtypes	Genotype/ Allele		Amino acid changes		mRNA	Expression ß-gal activity*		No of	
G_{M1}-gangliosidosis	I	II				Allele I	Allele II	cases	References
Infantile	$947^{A\to G}$	-	$Tyr^{315}\to Cys$	-	+	4x		1	Yoshida et al., 1991
	Duplication 1069-1233	-	Frameshift	-	+large			2	Yoshida et al., 1991
	$367^{G\to A}$	-	$Glyc^{123}\to Arg$	-	+	5x		1	Yoshida et al., 1991
	$145^{C\to T}$	-	$Arg^{49}\to Cys$	-	+/-	ND°		1	Nishimoto et al., 1991
	$1369^{C\to T}$	$1369^{C\to T}$	$Arg^{457}\to Ter$	$Arg^{457}\to Ter$	-	ND°	ND°	1	Nishimoto et al., 1991
Late infantile/ juvenile	$601^{C\to T}$	-	$Arg^{201}\to Cys$	-	+	102x ND°		1 3	Yoshida et al., 1991 Nishimoto et al., 1991
	$601^{C\to T}$	$601^{C\to T}$	$Arg^{201}\to Cys$	$Arg^{201}\to Cys$	+	ND°	ND°	1	Nishimoto et al., 1991
Adult/chronic	$152^{T\to C}$	$1370^{G\to A}$	$Ile^{51}\to Thr$	$Arg^{457}\to Gln$	+	325x	8x	1	Yoshida et al., 1991
	$152^{T\to C}$	$152^{T\to C}$	$Ile^{51}\to Thr$	$Ile^{51}\to Thr$	+	325x ND°	325x ND°	4 6	Yoshida et al., 1991 Nishimoto et al., 1991
Morquio B	$817\text{-}818^{TG\to CT}$	$1527^{G\to T}$	$Trp^{273}\to Leu$	$Trp^{509}\to Cys$	+			1	Oshima et al., 1991
	$817\text{-}818^{TG\to CT}$	$1445^{G\to A}$	$Trp^{273}\to Leu$	$Arg^{482}\to His$	+			2	Oshima et al., 1991

* = nmol/mg protein/hour;
x = expression in G_{M1}-gangliosidosis fibroblasts
° = expression in COS-1 cells.
ND = Not Detected

2.2 REFERENCES

Andria, G., P. Strisciuglio, G. Pontarelli, W.S. Sly and W.E. Dodson. 1981. Infantile neuraminidase and beta-galactosidase deficiencies (galactosialidosis) with mild clinical courses. p. 379-95. *In* Tettamanti, G., P. Durand, S. Didonato (ed.), Sialidases and Sialidosis. Edi. Ermes, Milan.

Conzelmann, E. and K. Sandhoff. 1987. Glycolipid and glycoprotein degradation. Adv Enzymol Relat Areas Mol Biol 60:89-216

D'Azzo, A., A. Hoogeveen, A.J. Reuser, D. Robinson and H. Galjaard. 1982. Molecular defect in combined beta-galactosidase and neuraminidase deficiency in man. Proc Natl Acad Sci U S A 79:4535-9.

Fischer, G. and H. Jatzkewitz. 1975. The activator of cerebroside sulfatase. Purification from human liver and identification as protein. Hoppe-Zeyler Z Physiol Chem 356:605-13.

Galjaard, H., A. d'Azzo, A.T. Hoogeveen and F.W. Verheijen. 1984. Combined beta-galactosidase-sialidase deficiency in man: genetic defect of "a protective protein". p. 113-32. *In* Barranger, J.A., O. Brady (ed.), Molecular Basis of Lysosomal Storage Disorders. Academic Press, New York.

Galjart, N.J. 1991. The protective protein: A multifunctional lysosomal enzyme. Thesis, Erasmus University, Rotterdam

Galjart, N.J., N. Gillemans, A. Harris, G.T. van der Horst, F.W. Verheijen, H. Galjaard and A. d'Azzo. 1988. Expression of cDNA encoding the human "protective protein" associated with lysosomal beta-galactosidase and neuraminidase: homology to yeast proteases. Cell 54:755-64.

Groebe, H., M. Krins, H. Schmidberger, K. von Figura, K. Harzer, H. Kresse, E. Paschke, A. Sewell and K. Ullrich. 1980. Morquio syndrome (mucopolysaccharidosis IV-B) associated with beta-galactosidase deficiency. Report of two cases. Am J Hum Genet 32:258-72.

Hoeksema, H.L., O.P. van Diggelen and H. Galjaard. 1979. Intergenic complementation after fusion of fibroblasts from different patients with beta-galactosidase deficiency. Biochim Biophys Acta 566:72-9.

Hoogeveen, A.T., H. Graham Kawashima, A. d'Azzo and H. Galjaard. 1984. Processing of human beta-galactosidase in GM1-gangliosidosis and Morquio B syndrome. J Biol Chem 259:1974-7.

Hoogeveen, A.T., F.W. Verheijen and H. Galjaard. 1983. The relation between human lysosomal beta-galactosidase and its protective protein. J Biol Chem 258:12143-6.

Jackman, H.L., P.W. Morris, P.A. Deddish, R.A. Skidgel and E.G. Erdos. 1992. Inactivation of endothelin I by deamidase (lysosomal protein). J Biol Chem 267:2872-2875.

Jackman, H.L., F.L. Tan, H. Tamei, C. Beurling-Harbury, X.Y. Li, R.A. Skidgel and E.G. Erdos. 1990. A peptidase in human platelets that deamidates tachykinins. Probable identity with the lysosomal "protective protein". J Biol Chem 265:11265-72.

Li, S.C. and Y.T. Li. 1976. An activator stimulating the enzymatic hydrolysis of sphingoglycolipids. J Biol Chem 251:1159-63.

Nanba, E. and K. Suzuki. 1990. Molecular cloning of mouse acid beta-galactosidase cDNA: sequence, expression of catalytic activity and comparison with the human enzyme. Biochem Biophys Res Commun 173:141-8.

Nanba, E. and K. Suzuki. 1991. Organization of the mouse acid beta-galactosidase gene. Biochem Biophys Res Commun 178:158-64.

Naylor, S.L., R.W. Elliott, J.A. Brown and T.B. Shows. 1982. Mapping of aminoacylase-1 and beta-galactosidase-A to homologous regions of human chromosome 3 and mouse chromosome 9 suggests location of additional genes. Am J Hum Genet 34:235-44.

Nishimoto, J., E. Nanba, K. Inui, S. Okada and K. Suzuki. 1991. GM1-gangliosidosis (genetic beta-galactosidase deficiency): identification of four mutations in different clinical phenotypes among Japanese patients. Am J Hum Genet 49:566-74.

Norden, A., L. Tennant and J.S. O'Brien. 1974. Ganglioside GM1 beta-galactosidase A: purification and studies of the enzyme from human liver. J Biol Chem 249:7969-76.

O'Brien, J.S. 1989. Beta-galactosidase deficiencies (GM1-gangliosidosis, galactosialidosis, and Morquio syndrome type B); ganglioside sialidase deficiency (mucolipidosis type IV). p. 1797-1806. *In* Scriver, A., A.L. Beaudet, W.S. Sly, D. Valle (ed.), The Metabolic Basis of Inherited Disease. McGraw-Hill, New York.

O'Brien, J.S., E. Gugler, A. Giedion, U. Wiessmann, N. Herschkowitz, C. Meier and J. Leroy. 1976. Spondyloepiphyseal dysplasia, corneal clouding, normal intelligence and acid beta-galactosidase

deficiency. Clin Genet 9:495-504.
Okada, S., T. Kato, S. Miura, H. Yabuuchi, M. Nishigaki, A. Kobata, H. Chiyo and J. Furuyama. 1978. "Hypersialyloligosaccharidurie in mucolipidosis: a method for diagnosis". Clin Chim Acta 86:159-67.
Okada, S. and J.S. O'Brien. 1968. Generalized gangliosidosis: Beta-galactosidase deficiency. Science 160:1002-04.
Oshima, A., K. Yoshida, M. Shimmoto, Y. Fukuhara, H. Sukuraba, Y. Suzuki. 1991. Human ß-galactosidase gene mutation in Morquio B disease Am J Hum Genet 49:1091-3.
Oshima, A., A. Tsuji, Y. Nagao, H. Sakuraba and Y. Suzuki. 1988. Cloning, sequencing, and expression of cDNA for human beta-galactosidase. Biochem Biophys Res Commun 157:238-44.
Overdijk, B., E.F.J. Hiensch-Goormachtig, E.F.J. Beem and G.J. Steijn. 1986. The carbohydrate structures of beta-galactosidase from human liver. Glyconjugate J 3:339-50.
Palmeri, S., A.T. Hoogeveen, F.W. Verheijen and H. Galjaard. 1986. Galactosialidosis: molecular heterogeneity among distinct clinical phenotypes. Am J Hum Genet 38:137-48.
Paschke, E. and H. Kresse. 1982. Morquio disease, type B: activation of GM1-beta-galactosidase by GM1-activator protein. Biochem Biophys Res Commun 109:568-75.
Potier, M., L. Michaud, J. Tranchemontagne and L. Thauvette. 1990. Structure of the lysosomal neuraminidase-beta-galactosidase-carboxypeptidase multienzymic complex. Biochem J 267:197-202.
Sacrez, R., J.G. Juif, J.M. Gigonnet and J.E. Gruner. 1967. La Maladie de Landing, idiote amaurotique infantile precoce avec gangliosidose generalizee de type GM1. Pediatrie 22:143-62.
Scheibe, R., K. Hein and K.W. Wenzel. 1990. Lysosomal beta-galactosidase from rat liver: purification, molecular forms and association with neuraminidase. Biomed Biochim Acta 49:547-56.
Shows, T.B., L.R. Scrafford-Wolff, J.A. Brawn and M. Meisler. 1979. GM1-gangliosidosis: Chromosome 3 assignment of the beta-galactosidase A gene (Beta-GAL A). Somat Cell Genet 5:147-58.
Skudlarek, M.D. and R.T. Swank. 1979. Biosynthesis of two lysosomal enzymes in macrophages. Evidence for a precursor of beta-galactosidase. J Biol Chem 254:9939-42.
Skudlarek, M.D. and R.T. Swank. 1981. Turnover of two lysosomal enzymes in macrophages. J Biol Chem 256:10137-44.
Skudlarek, M.D. and R.T. Swank. 1988. Synthesis and processing of lysosomal enzymes of mouse kidney after treatment with [3H]fucose. J Biol Chem 263:11302-5.
Suzuki, Y., H. Sakuraba, T. Yamanaka, Y.M. Ko, Y. Limori, Y. Okurama and A.T. Hoogeveen. 1984. Galactosialidosis: a comparative study of clinical and biochemical data on 22 patients. p. 161-75. *In* Arina, M., Y. Suzuki, H. Yabuuchi (ed.), The Developing Brain and its Disorders. University of Tokyo Press, Tokyo.
Tropea, J.E., R.T. Swank and H.L. Segal. 1988. Effect of swainsonine on the processing and turnover of lysosomal beta-galactosidase and beta-glucuronidase from mouse peritoneal macrophages. J Biol Chem 263:4309-17.
van der Horst, G.T., N.J. Galjart, A. d'Azzo, H. Galjaard and F.W. Verheijen. 1989. Identification and in vitro reconstitution of lysosomal neuraminidase from human placenta. J Biol Chem 264:1317-22.
van der Horst, G.T., W.J. Kleijer, A.T. Hoogeveen, J.G. Huijmans, W. Blom and O.P. van Diggelen. 1983. Morquio B syndrome: a primary defect in beta-galactosidase. Am J Med Genet 16:261-75.
van Diggelen, O.P., A.T. Hoogeveen, P.J. Smith, A.J. Reuser and H. Galjaard. 1982. Enhanced proteolytic degradation of normal beta-galactosidase in the lysosomal storage disease with combined beta-galactosidase and neuraminidase deficiency. Biochim Biophys Acta 703:69-76.
Van Diggelen, O.P., A.W. Schram, M.L. Sinnott, P.J. Smith, D. Robinson and H. Galjaard. 1981. Turnover of beta-galactosidase in fibroblasts from patients with genetically different types of beta-galactosidase deficiency. Biochem J 200:143-51.
van Pelt, J., K. Hard, J.P. Kamerling, J.F. Vliegenthart, A.J. Reuser and H. Galjaard. 1989. Isolation and structural characterization of twenty-one sialyloligosaccharides from galactosialidosis urine. An intact N,N'-diacetylchitobiose unit at the reducing end of a diantennary structure. Biol Chem Hoppe Seyler 370:191-203.
Van Pelt, J., D.G. Van Bilsen, J.P. Kamerling and J.F. Vliegenthart. 1988b. Structural analysis of O-glycosidic type of sialyloligosaccharide-alditols derived from urinary glycopeptides of a sialidosis patient. Eur J Biochem 174:183-7.
Van Pelt, J., J.A. Van Kuik, J.P. Kamerling, J.F. Vliegenthart, O.P. Van Diggelen and H. Galjaard. 1988a. Storage of sialic acid-containing carbohydrates in the placenta of a human galactosialidosis fetus.

Isolation and structural characterization of 16 sialyloligosaccharides. Eur J Biochem 177:327-38.
Verheijen, F., R. Brossmer and H. Galjaard. 1982. Purification of acid beta-galactosidase and acid neuraminidase from bovine testis: evidence for an enzyme complex. Biochem Biophys Res Commun 108:868-75.
Verheijen, F.W., S. Palmeri, A.T. Hoogeveen and H. Galjaard. 1985. Human placental neuraminidase. Activation, stabilization and association with beta-galactosidase and its protective protein. Eur J Biochem 149:315-21.
Wenger, D.A., T.J. Tarby and C. Wharton. 1978. Macular cherry-red spots and myoclonus with dementia: coexistent neuraminidase and beta-galactosidase deficiencies. Biochem Biophys Res Commun 82:589-95.
Yamamoto, Y., M. Fujie and K. Nishimura. 1982. The interrelation between high- and low-molecular-weight forms of GM1-beta-galactosidase purified from porcine spleen. J Biochem (Tokyo) 92:13-21.
Yamamoto, Y., C.A. Hake, B.M. Martin, K.A. Kretz, A.J. Ahern-Rindell, S.L. Naylor, M. Mudd and J.S. O'Brien. 1990. Isolation, characterization, and mapping of a human acid beta-galactosidase cDNA. Dna Cell Biol 9:119-27.
Yamamoto, Y. and K. Nishimura. 1987. Copurification and separation of beta-galactosidase and sialidase from porcine testis. Int J Biochem 19:435-42.
Yoshida, K., A. Oshima, M. Shimmoto, Y. Fukuhara, H. Sakuraba, N. Yanagisawa and Y. Suzuki. 1991. Human beta-galactosidase gene mutations in GM1-gangliosidosis: a common mutation among Japanese adult/chronic cases. Am J Hum Genet 49:435-42.

2.3 PUBLICATIONS

Publication 1

J. Biol. Chem. **264** (1989), 20655-20663

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1989 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 264, No. 34, Issue of December 5, pp. 20655–20663, 1989
Printed in U.S.A.

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

(Received for publication, July 20, 1989)

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 G_{M1}-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 G_{M1}-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 G_{M1}-gangliosidosis fibroblasts. The subcellular localization of cDNA-encoded β-galactosidase and β-galactosidase-related proteins is different.

Acid β-D-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 G_{M1}[1]-gangliosidosis and Morquio B syndrome (4–6). Among G_{M1}-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 an 85-kDa precursor, which is posttranslationally processed to the mature lysosomal form of 64 kDa (8). In cells of an infantile and an adult G_{M1}-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, pTZ18 and pTZ19 plasmid vectors were obtained from Pharmacia LKB Biotechnology Inc. T4 DNA ligase was from BRL. The sequenase and the sequencing kit were purchased from United States Biochemical Corp. *Taq* 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.: [α-^{32}P]dATP and [α-^{32}P]dCTP, 3000 Ci/mmol; [γ-^{32}P]dATP, 6000 Ci/mmol;

* 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.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J05124.

‡ To whom correspondence and reprint requests should be sent: Dept. of Cell Biology and Genetics, Erasmus University, P. O. Box 1738, 3000 DR Rotterdam, The Netherlands. Fax: (0)10-4087212.

[1] The abbreviations used are: G_{M1}, IIINeuAc-GgOse$_4$Cer; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; HPLC, high pressure liquid chromatography; bp, base pairs, kb, kilobases, PCR, polymerase chain reaction.

[α-^{35}S]dATPαS, >1000 Ci/mmol; [^{35}S]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 G_{M1}-gangliosidosis, and one obligated heterozygote were obtained from the European Human Cell Bank, Rotterdam (Dr. W. J. Kleyer). Fibroblasts were cultured in Dulbecco's modified Eagle's medium/Ham's F10 medium (1:1 v/v) supplemented with antibiotics and 10% fetal bovine serum. COS-1 cells were grown in the same medium supplemented with 5% fetal bovine serum.

Protein 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 against Immobilon PVDV transfer 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 1×10^6 independent clones with insert sizes ranging from 0.7 to 3.3 kb, was plated out at a density of 5×10^4 plaque-forming units per 90-mm plate and screened with anti-β-galactosidase antibodies as described previously (21). Antibody-positive clones were rescreened with oligonucleotide probes labeled at the 5′ end with ^{32}P using γ-^{32}P 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—HβGa(S) and HβGa(L) cDNAs and their restriction fragments were subcloned into plasmid vectors pTZ18 and pTZ19. Nucleotide sequences on both strands were obtained by the dideoxy chain termination methods of Sanger *et al.* (24), for single-stranded DNA, and of Murphy and Kavanagh (25), for double-stranded DNA. M13 universal reverse primer and a synthetic oligonucleotide were used. Sequence data were analyzed using the program of Staden (26).

Isolation and Sequencing of Genomic β-Galactosidase λ Clones—A human EMBL-3 λ library (kindly provided by Dr. G. Grosveld, Erasmus University, Rotterdam), derived from DNA of leukocytes of a chronic myeloid leukemia patient, was screened with the 5′ 850-bp *Eco*RI fragment of cDNA clone HβGa(L). The inserts of three overlapping λ clones were subcloned into the plasmid vector pTZ18. Sequences of genomic exons were determined by the chain termination method on double-stranded DNA, using synthetic oligonucleotide primers derived from the β-galactosidase cDNA sequence.

RNA Isolation and Northern Blot Hybridization—Total RNA was isolated from cultured fibroblasts as described (27). Polysomal mRNA, immunoselected using antibodies raised against purified placental complex, was obtained following the procedure of Myerowitz and Proia (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 polysomal 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 type-specific oligonucleotide probes or a 90-bp *Pst*I DNA fragment. These probes were labeled as mentioned above.

Transient Expression of β-Galactosidase cDNAs in COS-1 Cells—Subcloning of the two cDNAs into a derivative of the mammalian expression vector pCD-X and conditions of transfections of pCDHβGa constructs to COS-1 cells were as described previously (15). Labeling with [^{35}S]methionine was carried out in the presence or absence of NH_4Cl (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 artificial 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 (15). Human recipient cells were from an infantile G_{M1}-gangliosidosis patient (Fig. 5, patient II). 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 pCDHβGa 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 IgG) 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. 1*A*). 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. 1*A*, *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. 1*B*). 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 rescreened with both oligonucleotide probes. One clone, λHβGa39, with a total insert size of 1.7 kb, carried an internal *Eco*RI site which released, upon digestion with *Eco*RI, two fragments of 500 and 1200 bp, hybridizing with probe 2 and 1, 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 preproform.[2] Therefore, λHβGa39 could not contain the entire coding region of β-galactosidase precursor. Rescreening of the library with this cDNA probe yielded a clone, λHβGa(L), that consisted of a 5′ *Eco*RI 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.

A

Sample	Aminoacid sequence
T1	D* E A V A X X L Y D I L A R
T2	F A Y G K
T3	A Y V A V D G I P Q G V L E R
N-ter	Q R M F E I D Y S R D S F L K D G Q

B

Sample	Oligonucleotide probes	
T3	A Y V A V D G I .P GCC TAC GTC GCC GTG GAC GGT ATC CCC	(1)
	Q G V L E R CAG GGC GTG CTC GAG GGC	
N-ter	M F E I D ATG TTT GAA ATT GAT GA C G C C T A	(2)

FIG. 1. **Partial amino acid sequences of placental β-galactosidase and oligonucleotide probes.** *A, T1–T3* are the amino acid sequences of three tryptic peptides derived from purified human placental β-galactosidase. *N-ter* indicates the amino terminal sequence of the mature protein. *Asterisk* refers to a discrepancy between chemically derived and predicted amino acid sequences; unassigned residues are indicated by *X*. Amino acids are identified by the single-letter code. *B, T3* and part of *N-ter* were used to synthesize oligonucleotide probes 1 and 2. Mismatches to the actual cDNA sequence are *underlined*.

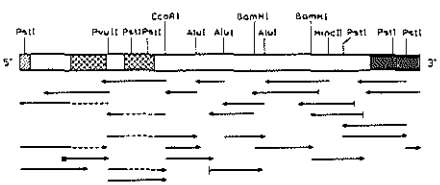


FIG. 2. **Composite restriction maps and sequencing strategy of HβGa cDNA clones.** All restriction enzyme sites in the cDNAs used for subcloning are shown. *Arrows* indicate the direction and extent of sequencing reactions. *Cross-hatched boxes* represent the two protein-coding regions not present in the short β-galactosidase cDNA. *Broken arrows* indicate sequencing reactions used for the short clone. *Arrows* starting with a *vertical line* represent 5′ or 3′ sequences of independent cDNAs. The *arrow* starting with a *solid square box* was a sequence reaction primed with a synthetic oligonucleotide. The *hatched* and *solid bars* are the 5′- and 3′-untranslated regions, respectively.

sequenced using the dideoxy 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 strategy used. The complete sequences of HβGa39 and HβGa(L) are combined in Fig. 3.

A common ATG translation initiation codon is found at the 5′ end of both cDNAs (Fig. 3, position 51). This ATG represents the beginning of an open reading frame for HβGa(L) of 2031 nucleotides, which is interrupted by three consecutive stop codons, and it is flanked at the 3′ end by a 318-nucleotide untranslated region. A putative polyadenylation signal (AATAAA) is present at position 2379. The sequences of the two cDNAs from their internal *Eco*RI site toward the 3′ end are identical, except that HβGa39 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 HβGa39 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 *Eco*RI fragment encompassing the 5′ end of HβGa39 is 393 nucleotides shorter than the corresponding fragment of HβGa(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 mutation in the open reading frame of HβGa39 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 *Eco*RI fragment of HβGa39 for the HβGa(L). The resulting cDNA construct, HβGa(S), has an open reading frame of 1638 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 HβGa(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 185–245), 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 *83–114*), 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

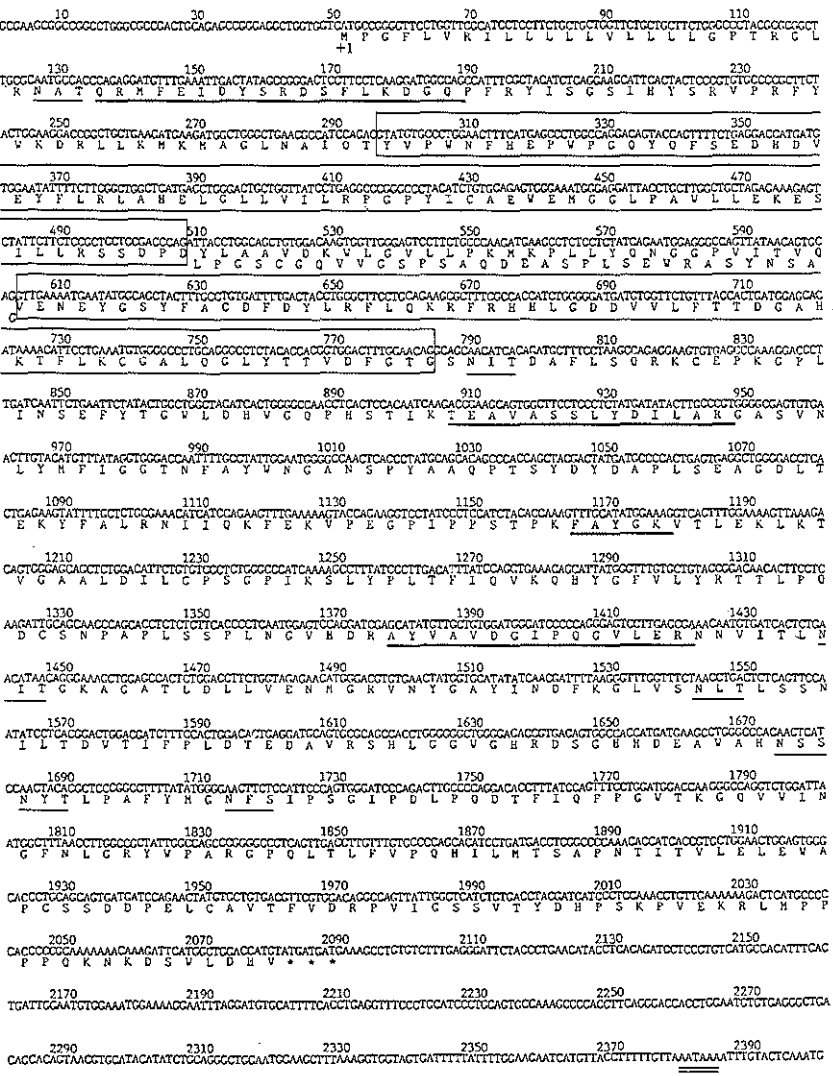
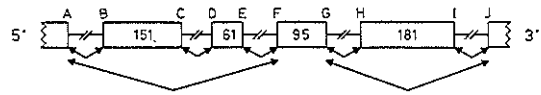


FIG. 3. **Combined nucleotide and predicted amino acid sequences of β-galactosidase cDNAs.** The numbers *above* the nucleotide sequence refer to H,βGa(L) cDNA. Numbers on the *right* and on the *left* specify the amino acids of the long and short β-galactosidase-predicted sequences, respectively. Amino 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*.



Exon/Intron Boundaries:

```
A/B  CCATCCACAC/gtaagt.....tgtgtgtcttgccag/CTATGTCCCC
C/D  CTGCCAAAATG/gtgagt.....ttttcccttaaactag/CCAGCATTAC
E/F  TCCCACCCAG/gtaggt.....ttttttatttttcag/ATTACCTGGC
G/H  AACAGTGCAG/gtaagc.....tttttcactcacag/GTTGAAAATG
I/J  TTTGCAAACAG/gtcggt.....cctcttacaatttttcag/GCAGCAACAT
```

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 *J*) 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.

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 in the predicted primary sequence (Fig. 3, *thin line*). 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 terminus of the purified placental enzyme. The predicted M_r 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 G_{M1}-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 G_{M1}-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 G_{M1}-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. 6*B*. 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

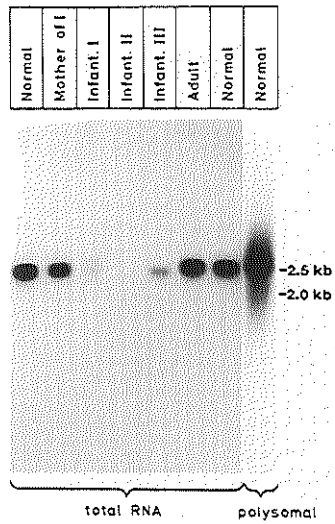


FIG. 5. **Northern blot analysis of fibroblast RNA.** Total and polysomal fibroblast RNA from three normal individuals and total fibroblast RNA from an adult and three infantile G_{M1}-gangliosidosis patients as well as one heterozygote were fractionated on a formaldehyde-agarose gel and probed with the HβGa(L) cDNA. The sizes of the two β-galactosidase transcripts are indicated. Exposure time was 2 days.

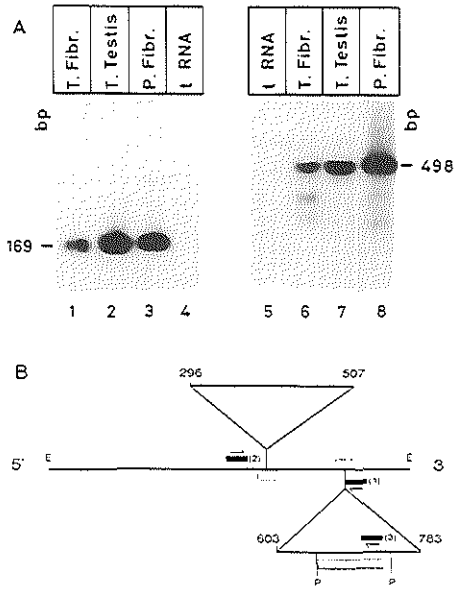


FIG. 6. **Detection of two mRNAs for β-galactosidase by PCR amplification.** *A*, total (*T. Fibr.*) and polysomal fibroblast (*P. Fibr.*) RNA and total testis (*T. Testis*) RNA were used to synthesize single-stranded cDNAs that were subsequently subjected to 24 rounds of amplification. Amplified products 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 *Eco*RI 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′-AAGCATCTGTGATGTTGCTG-3′; of the antisense primer *3* is 5′-ACATTTCAGGAATGTTTTATGTGCT-3′; of the sense primer *2* is 5′-TGGAAGGACCGGCTGCTGAA-3′. *Cross-hatched bars* designate a 90-bp *Pst*I probe and two 20-mer oligonucleotide probes. The sequence of the 5′-oligonucleotide probe is 5′-CCATCCAGAC/ATTACCTGGC-3′; of the 3′ probe is 5′-AACAGTGCAG/GCAGCAACAT-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. 6*B*, *cross-hatched bars*). Two 20-mers were synthesized on the basis of sequences of the HβGa(S) cDNA, which are colinear with the 10 nucleotides flanking each end of regions 1 and 2 of HβGa(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 *Pst*I probe present in region 2. As shown in Fig. 6*A*, 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.

Transient Expression of β-Galactosidase cDNAs in COS-1 Cells—HβGa(S) and HβGa(L) cDNAs were cloned in sense and antisense orientations into a derivative of the mammalian expression vector pCD-X and transfected separately to COS-1 cells. After 48 h, normal and transfected cells were incubated for an additional 16 h, with [^{35}S]methionine. In some instances the labeling step was done in presence of NH_4Cl to induce maximal secretion of lysosomal protein precursors (19). Radiolabeled proteins from cells and media were immunoprecipitated with anti-β-galactosidase antibodies. The results are shown in Fig. 7. A β-galactosidase polypeptide of 85 kDa is detected intra- and extracellularly after transfection of COS-1 cells with pCDHβGa(L)-sense construct (*lanes 1, 2,* and *7, 8*). A protein of 68 kDa is synthesized and secreted upon transfection with the pCDHβGa(S)-sense plasmid (*lanes 3* and *9*). Treatment with NH_4Cl does not have any detectable influence. The estimated molecular mass of the large molecule (85 kDa) correlates with that observed for the glycosylated β-galactosidase precursor immunoprecipitated in human cells. The 68-kDa polypeptide is a form that was not noticed previously. These cDNA-derived proteins are not present in mock-transfected cells or in cells transfected with an anti-

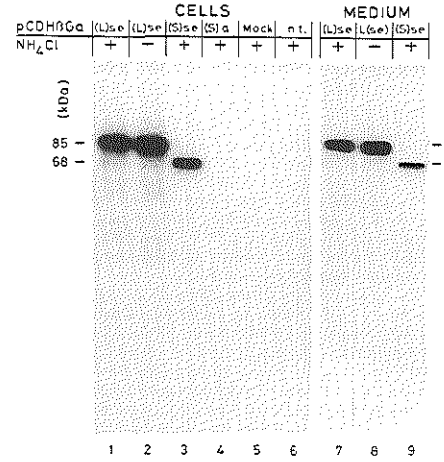


FIG. 7. **Transient expression of pCDHβGa constructs in COS-1 cells.** The pCDHβGa(L) and pCDHβGa(S) cDNA constructs in the sense (*se*, *lanes 1–3* and *7–9*) and antisense (*a*, *lane 4*) orientations were transfected to COS-1 cells. A mock transfection was carried out without the addition of DNA (*lane 5*). *n.t.* indicates not transfected (*lane 6*). After 48 h cells were incubated with [^{35}S]methionine for an additional 16 h with and without NH_4Cl (+ 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.

sense construct (*lanes 4* and *5*). It appears, therefore, that the antibody preparation used in these experiments hardly recognizes COS-1 endogenous β-galactosidase, since untransfected cells also do not show any cross-reactive bands (*lane 6*). As seen in lanes 1 and 2, the cDNA-derived 85-kDa β-galactosidase precursor is poorly processed into the mature 64-kDa form in transfected cells. This is due to the transfection procedure, as observed before (15). A 5-fold increase in β-galactosidase activity above the endogenous COS-1 values is measured only in cells transfected with the pCDHβGa(L)-sense construct (Table I). Using the same assay conditions, the β-galactosidase-related molecule is apparently not active.

We also tested whether the cDNA-encoded proteins were able to correct β-galactosidase activity in G_{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. Aliquots of the different conditioned media were added to the culture medium of fibroblasts from an infantile G_{M1}-gangliosidosis patient (patient II in Fig. 5). After 2 days of uptake, activities were measured in cell homogenates using 6-methylumbelliferyl substrate. As shown in Table II, COS-1 cell-derived 85-kDa precursor taken up by G_{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 typical lysosomal distribution as well as uniformly diffuse perinuclear labeling of β-galactosidase is observed in COS-1 cells transfected with the pCDHβGa(L)-sense construct (Fig. 8*A*). However, a strong fluorescent labeling restricted to the perinuclear region is present in cells transfected with the short construct (Fig. 8*B*). Adjacent untransfected cells react poorly with the human antibodies. Taken together, these results demonstrate that the long and short cDNAs direct the synthesis of two proteins, one of which behaves as the classic lysosomal β-galactosidase, whereas the other is not enzymatically active at the pH value and substrate concentration used. This β-galactosidase-related protein also has a different subcellular localization.

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	1.6

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

DISCUSSION

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 G_{M1}-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 G_{M1}-gangliosidosis patients, not yet analyzed at the molecular level, do synthesize β-galactosidase precursor (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 Hoogeveen *et al.* (37) that all G_{M1}-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_{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-

TABLE II

Correction of β-galactosidase activity in G_{M1}-gangliosidosis fibroblasts after uptake of COS-1 cell-derived β-galactosidase precursor

Transfection in COS-1 cells	Addition of COS-1 cell-derived proteins	Activity in G_{M1}-gangliosidosis fibroblasts	
		β-Galactosidase	β-Glucuronidase
		microunit/mg protein	*milliunit/mg protein*
pCDHβGa (L) sense	+	234	2.48
pCDHβGa (S) sense	+	6.6	2.12
pCDHβGa (S) sense	+	7.6	2.01
Mock	+	12.3	3.58
	−	7.5	1.88

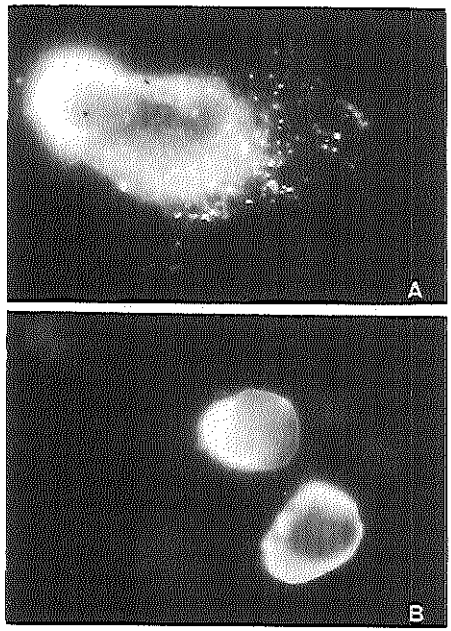


FIG. 8. **Immunocytochemical localization of β-galactosidase proteins in transiently transfected COS-1 cells.** *A*, transfection with pCDHβGa(L) cDNA construct; *B*, transfection with pCDHβGa(S) cDNA construct.

bly, are read in different frames only in the 95-nucleotide stretch between regions 1 and 2. To our knowledge this is the first example of such a configuration in a mammalian gene.

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 this phenomenon has been observed for human β-glucuronidase mRNA (40). The genomic data also rule out the possibility that the short cDNA is the product of a cloning artifact. The amount of the short mRNA, however, must be less than 1/10 of the long one, if we consider the signal obtained on Northern blots. Therefore, the existence of the two β-galactosidase transcripts was further proven by PCR amplification of partial cDNA fragments specifying the two mRNAs. The short transcript does not seem to be testis-specific, since it is also detected in fibroblast total and polysomal mRNA, indicating that this transcript is actively translated in fibroblasts. It is not excluded, however, that the two mRNAs may be expressed in differential amounts in other tissues.

The open reading frames of the long and short β-galactosidase 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 G_{M1}-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 NH_4Cl. 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 Grosveld 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; Sjozef 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., Scrafford-Wolff, L. R., Brown, J. A., and Meisler, M. (1979) *Somatic Cell Genet.* **5**, 147–158
4. Okada, S., and O'Brien, J. S. (1968) *Science* **160**, 1002–1004
5. O'Brien, J. S., Gugler, E., Giedion, A., Wiesmann, U., Herschkowitz, N., Meijer, C., and Leroy, J. (1976) *Clin. Genet.* **9**, 495–504
6. Groebe, H., Krins, M., Schmidberger, H., von Figura, K., Harzer, K., Kresse, H., Paschke, E., Sewell, A., and Ullrich, K. (1980)

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., Fujie, M., and Nishimura, K. (1982) *J. Biochem. (Tokyo)* **92**, 13–21
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. (1985) *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. (1980) *J. Biol. Chem.* **255**, 4937–4945
20. Aebersold, R. H., Teplow, 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. 1, 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., Hagemeijer, A., Wiedemann, L. M., and Grosveld, G. (1988) *Leukemia* **2**, 628–633
32. Proia, R. L., d'Azzo, A., and Neufeld, E. F. (1984) *J. Biol. Chem.* **259**, 3350–3354
33. Van Dongen, J. M., Barneveld, R. A., Geuze, H. J., and Galjaard, H. (1984) *Histochem. J.* **16**, 941–954
34. von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690
35. Benham, F. J., Hodgkinson, S., and Davies, K. E. (1984) *EMBO J.* **3**, 2635–2640
36. Nanba, E., Tsuji, A., Omura, K., and Suzuki, Y. (1988) *Biochem. Biophys. Res. Commun.* **152**, 794–800
37. Hoogeveen, A. T., Reuser, A. J. J., Kroos, M., and Galjaard, H. (1986) *J. Biol. Chem.* **261**, 5702–5704
38. Brenthnach, R., and Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349–383
39. Andreadis, A., Gallego, M. E., and Nadal-Ginard, B. (1987) *Annu. Rev. Cell. Biol.* **3**, 207–242
40. Oshima, A., Kyle, J. W., Miller, R. D., Hoffmann, J. W., Powell, P. P., Grubb, J. H., Sly, W. S., Tropak, M., Guise K. S., and Gravel, R. A. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 685–689
41. Tropea, J. E., Swank, R. T., and Segal, H. L. (1988) *J. Biol. Chem.* **263**, 4309–4317
42. Skudlarek, M. D., and Swank, R. T. (1988) *J. Biol. Chem.* **263**, 11302–11305

Publication 2

DNA Cell Biol. 7 (1991), 495-504

DNA AND CELL BIOLOGY
Volume 10, Number 7, 1991
Mary Ann Liebert, Inc., Publishers
Pp. 495–504

Organization of the Gene Encoding Human Lysosomal β-Galactosidase

HANS MORREAU, ERIK BONTEN, XIAO YAN ZHOU, and ALESSANDRA D'AZZO

ABSTRACT

Human β-galactosidase precursor mRNA is alternatively spliced into an abundant 2.5-kb transcript and a minor 2.0-kb species. These templates direct the synthesis of the classic lysosomal β-D-galactosidase enzyme and of a β-galactosidase-related protein with no enzymatic activity. Mutations in the β-galactosidase gene result in the lysosomal storage disorders G_{M1}-gangliosidosis and Morquio B syndrome. To analyze the genetic lesions underlying these syndromes we have isolated the human β-galactosidase gene and determined its organization. The gene spans >62.5 kb and contains 16 exons. Promoter activity is located on a 236-bp *Pst* I fragment which works in a direction-independent manner. A second *Pst* I fragment of 851 bp located upstream from the first negatively regulates initiation of transcription. The promoter has characteristics of a housekeeping gene with GC-rich stretches and five potential SP1 transcription elements on two strands. We identified multiple cap sites of the mRNA, the major of which maps 53 bp upstream from the translation initiation codon.

The portion of the human pre-mRNA undergoing alternative splicing is encoded by exons II–VII. Sequence analysis of equivalent mouse exons showed an identical genomic organization. However, translation of the corresponding differentially spliced murine transcript is interrupted in its reading frame. Thus, the mouse gene cannot encode a β-galactosidase-related protein in a manner similar to the human counterpart. Differential expression of the murine β-galactosidase transcript is observed in different mouse tissues.

INTRODUCTION

ACID β-D-GALACTOSIDASE (EC 3.2.1.23) catalyzes the hydrolysis of β-linked terminal galactosyl residues from a variety of natural and synthetic substrates (Conzelmann and Sandhoff, 1987; O'Brien, 1989). The human β-galactosidase gene on chromosome 3 (Shows *et al.*, 1979) encodes a glycosylated precursor protein of 85 kD that is proteolytically processed in lysosomes into a mature form of 64 kD (d'Azzo *et al.*, 1982). The majority of the active enzyme is found in a complex with lysosomal neuraminidase (sialidase, EC 3.2.1.18) and the protective protein/cathepsin A (EC 3.4.16.1) (Verheijen *et al.*, 1982; Yamamoto *et al.*, 1982; Galjart *et al.*;, 1988, 1991; van der Horst *et al.*, 1989). The latter is essential for the intralysosomal activation and stabilization of the other two glycosidases (d'Azzo *et al.*, 1982; Hoogeveen *et al.*, 1983; Yamamoto and Nishimura, 1987). Mutations in the β-galactosidase locus result in the metabolic storage diseases G_{M1}-gangliosidosis and Morquio B syndrome (Okada and O'Brien, 1968; O'Brien *et al.*, 1976; Groebe *et al.*, 1980; O'Brien, 1989). Among patients with G_{M1}-gangliosidosis, a wide spectrum of clinical manifestations have been described, ranging from severe infantile forms to milder juvenile and adult variants (O'Brien, 1989).

Three groups have independently isolated and sequenced the cDNA encoding human lysosomal β-galactosidase (Oshima *et al.*, 1988; Morreau *et al.*, 1989; Yamamoto *et al.*, 1990). Recently, the cDNA encoding the murine enzyme has also been characterized (Nanba and Suzuki, 1990). Using the human cDNA as probe, a major transcript of 2.5 kb is detected in total RNA preparations from normal fibroblasts. This mRNA is either not detectable or present in reduced amount in fibroblasts from clinically different G_{M1}-gangliosidosis patients (Morreau *et al.*, 1989). We and O'Brien's group demonstrated the existence in different human tissues of a minor 2.0-kb transcript derived from alternative splicing of β-galactosidase precursor mRNA (Morreau *et al.*, 1989; Yamamoto *et al.*, 1990).

To study the structure and function of β-galactosidase

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

and to understand fully the molecular nature of the mutations underlying different clinical phenotypes, the complete human β-galactosidase gene has been isolated, and its exon/intron organization has been characterized. Furthermore, we have investigated 5′ regulatory sequences that contain promoter elements and identified the cap site of the mRNA. To elucidate the significance of the alternatively spliced human transcript, a partial genomic fragment spanning 5′ exons of the mouse β-galactosidase gene has been isolated and its exon/intron organization compared to the one of the human gene. Using the murine β-galactosidase cDNA as probe, we have analyzed the expression of the β-galactosidase transcript in different mouse tissues.

EXPERIMENTAL PROCEDURES

Materials

Restriction endonucleases were from Boehringer Mannheim, Bethesda Research Laboratories (BRL), New England Biolabs, Pharmacia LKB Biotechnology Inc., and Promega Biotec. DNA polymerase (Klenow) was from Promega Biotec. T4 polynucleotide kinase, M13 reverse sequencing primers, deoxy and dideoxy nucleotides, pTZ18 and pTZ19 plasmid vectors were purchased from BRL. A T7 polymerase sequencing kit was purchased from Pharmacia LKB Biotechnology Inc. *Taq* polymerase was from Cetus Corp. Synthetic oligonucleotides primers were synthesized on an Applied Biosystems 381A oligonucleotide synthesizer. Radionucleotides were obtained from Amersham Corp: [α-^{32}P]dATP and [α-^{32}P]dCTP, 3,000 Ci/mmole; [γ-^{32}P]ATP, 6,000 Ci/mmole; [α-^{35}S]dATP >1,000 Ci/mmole. All other reagents were from standard commercial suppliers, if not specified otherwise.

Isolation of human and mouse genomic β-galactosidase clones and DNA sequencing

Both mouse and human genomic libraries (complexity 3 × 10^6 and 2 × 10^6, respectively) were kindly provided by Dr. G. Grosveld (Erasmus University, Rotterdam). They were constructed by cloning human or mouse genomic DNA fragments, partially digested with *Mbo* II into the *Bam* HI restriction site of the λ EMBL-3 vector. Human and mouse total DNAs were prepared from leukocytes of a chronic myeloid leukemia patient and from mouse embryonic stem cells, respectively. The human library (2 × 10^6 plaques) was screened with two *Eco* RI cDNA fragments (Morreau *et al.*, 1989), spanning the entire human β-galactosidase cDNA, using standard protocols (Sambrook *et al.*, 1989). Probes were labeled according to the procedure of Feinburg and Vogelstein (1983). The mouse library (2 × 10^6 plaques) was screened only with the most 5′ *Eco* RI fragment derived from the human β-galactosidase cDNA. Hybridization and washing of the mouse library were performed under nonstringent conditions (58°C). Several positive phages from both libraries were isolated, digested with different restriction enzymes, and mapped to each other. Complete or partial inserts of overlapping clones were subcloned in pTZ18 or 19 plasmid vectors using suitable restriction sites. Nucleotide sequencing of genomic fragments was performed by the chain-termination method on double-stranded DNA (Murphy and Kavanagh, 1988) using M13 universal or reverse primers and synthetic oligonucleotide primers derived from the HβGa(L) cDNA sequence. The positions of splice sites were obtained by comparing genomic sequences to the known cDNA sequence of HβGa(L). Sequence data were analyzed using the program of Staden (1986).

The polymerase chain reaction

To determine the length of the different introns, genomic fragments subcloned in pTZ-vectors were subjected to the polymerase chain reaction (PCR). The primers used in amplification reactions were derived from adjacent exons present in one genomic subfragment. Alternatively, plasmid-derived universal or reverse M13 primer was used in combination with a primer derived from exonic sequences. Intronic fragments were amplified with *Taq* polymerase as described (Hermans *et al.*, 1988), using a programmable DNA incubator (Cetus, Emeryville, CA). Amplified material was separated on 0.75% agarose gels in the presence of standard DNA markers. Intron lengths were either directly determined by comparison with marker fragments or calculated combining genomic mapping data and PCR results.

Promoter analysis

Pst I restriction fragments from the 5′ end of the human β-galactosidase gene were subcloned in the *Pst* I cloning site of the pCATEnh reporter plasmid (Promega) in sense and antisense orientations. Independent constructs were transiently expressed in HeLa cells. Cells were grown in Dulbecco's modified Eagles medium-Ham's F10 medium (1:1 vol/vol) (GIBCO BRL, Grand Island, NY), supplemented with antibiotics and 8% fetal bovine serum. These cells were transfected with different pCAT-constructs together with a plasmid containing the LacZ gene driven by the RSV promoter. The latter was used as a control of transfection efficiency. Conditions of transfection were as described previously (Graham and Van der Eb, 1973). Sixty hours after transfection HeLa cells were harvested in 250 m*M* Tris pH 7.8, freeze/thawed three times, and centrifuged to eliminate cell debris. Supernatants were assayed for CAT activity as described by Gorman *et al.* (1982), and analyzed on thin-layer chromatography (TLC) for the separation of the [^{14}C]chloramphenicol substrate from its products.

Primer extension

Two polysomal mRNA preparations, immunoselected using antibodies raised against purified placental β-galactosidase, were obtained following the procedure of Myerowitz and Proia (1984). One pmol of a 27-mer synthetic oligonucleotide primer was end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase.

The radiolabeled 27-mer (10^5 cpm) was annealed to 0.2

μg of polysomal mRNA for 12 hr at 30°C. Primer extension was performed using the conditions described by Sambrook *et al.* (1989). Products were analyzed on a 6% acrylamide/7 *M* urea sequencing gel.

Northern blot analysis

Total RNA was extracted from tissues of adult BCBA mice and fetuses (at days 13, 16, or 19 of gestation) using the method of Auffray and Rougeon (1980). RNA samples were electrophoresed on a 0.8% agarose gel containing 0.66 *M* formaldehyde (Fourney *et al.*, 1988), and blotted onto Zeta-Probe membranes (Bio-Rad). Standard hybridization and washing conditions were applied.

RESULTS

Isolation of β-galactosidase genomic clones and genomic organization

We have previously described the isolation and characterization of recombinant λ phages spanning the most 5′ region of the β-galactosidase gene (Morreau *et al.*, 1989). These recombinants were derived from a human EMBL-3 genomic library, screened with a 5′ β-galactosidase cDNA fragment of 852 bp (HβGa(L), nucleotides 1–852). We have rescreened the same library with a 1.5-kb *Eco* RI fragment encompassing the entire 3′ region of HβGa(L) cDNA (nucleotides 853–2,399). Several positive phages were isolated that were subjected to restriction endonuclease mapping and Southern blot analysis using radiolabeled cDNA fragments covering the entire cDNA. As shown in Fig. 1, exonic sequences were mapped on seven independent λ phage inserts, all of them overlapping except for λ 9.6/λ 7.1 and λ 7.3/λ 3, that extend into two large intervening sequences. These phages together cover a genomic area >70 kb. The *Eco* RI restriction pattern of cloned genomic DNA containing exon sequences is equivalent to the one obtained after Southern blot analysis of *Eco* RI-digested human genomic DNA probed with the complete HβGa(L) cDNA (data not shown).

The β-galactosidase coding region is divided into 16 exons (Fig. 1), distributed over >62.5 kb of genomic DNA. All exons were sequenced on both strands from selected phage fragments. The only discrepancies between genomic and cDNA sequences were a single nucleotide change in exon I [C instead of a T at position 79 of HβGa(L)] and an additional G following a stretch of three in exon XVI (at position 2,277 of HβGa(L)) encoding the 3′ untranslated region (UTR). The sequences of the exon/intron boundaries are compiled in Table 1. All splice junctions conform to the GT/AG rule formulated by Breathnach and Chambon (1981). The sizes of β-galactosidase exons range from 41 to >615 bp. The 15 introns vary in length from 84 bp to 15 kb. The sizes of introns 1 and 10 could not be calculated since no overlapping clones including these genomic areas were available. This could be due to the fact that these genomic areas are not represented in the library. We estimated the length of these introns to be larger than 10 kb.

Characterization of the human β-galactosidase promoter region and identification of the cap site

A 108-bp *Eco* RI–*Apa* I cDNA probe (nucleotides 1–108), derived from the most 5′ end of HβGa(L), hybridized to a 2.9-kb *Eco* RI genomic fragment present in λ 9.6. This fragment was subcloned into pTZ-plasmid vectors and sequenced. Within the 2.9-kb genomic stretch, which contains the 5′ UTR and the ATG start codon of β-galactosidase cDNA, two adjacent *Pst* I fragments are included. The first is 236 bp long (Fig. 2A, fragment 1) and contains nucleotides 1–30 of the cDNA 5′ UTR; the second is 851 bp long and flanks the first upstream (Fig. 2A, fragment 2). To assess whether this genomic region contained promoter activity, the 851- and 236-bp *Pst* I fragments as well

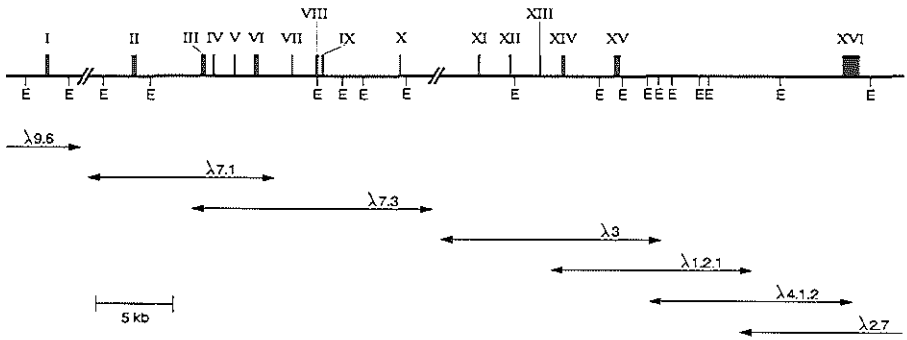


FIG. 1. *Eco* restriction map of the human β-galactosidase gene. *Eco* RI restriction sites are shown with a capital E. Arrows indicate the relative position and overlap of seven independent genomic λ phages. Interrupted lines represent intronic sequences of unknown size. Exons are denoted as black vertical bars with roman numbers. The bar underneath indicates the size in kilobases.

TABLE 1. NUCLEOTIDE SEQUENCES OF THE EXON/INTRON BOUNDARIES PRESENT IN THE HUMAN β-GALACTOSIDASE GENE

EXON	SIZE (bp)	cDNA position EXON	5' SPLICE DONOR	INTRON NO	SIZE (kb)	3' SPLICE ACCEPTOR
I	128/130/131	→ 125	GGCTTGCGC/gtaagtctgcgggacccgggtacggggaggcattgctagg	1	>10	aaggtggaagttctcatgtgaccccatccttgctctgtag/AATGCCACC
II	170	126→ 295	CATCCAGAC/gtaagtaggagggcatgtggctctcacctaggcctgaaca	2	4.2	ctgccaagtatccaggaaggggctgtgtgtgtcttggcag/GTATGTGCC
III	151	296→ 446	TGGGAAATG/gtgagtgtggctagactgggctctacctggtgtgtgtgag	3	0.55	acggaacagacattataatggcttgttttccctgaactag/GGAGGATTA
IV	61	447→ 507	CCGACCCAG/gtaggttgttacagatgtcttgagaagatttataaaagct	4	1.3	gggttggacaagcttgtgttaaacttttcttgttttccag/ATTACCTGG
V	95	508→ 602	ACAGTGCAG/gtaaccctggagttgagcatggatggattcccggaggggt	5	1.4	aggatctcctcatttttccctgctctttttttcactcacag/GTTGAAAAT
VI	181	603→ 783	TTGGAACAG/gttggtgtttgtagtaacagaaattgcagcttaaaatcag	6	2.4	ttccacgactgacctaccttctccctcttacaattttcag/GCAGCAACA
VII	59	784→ 842	GGACCCTTG/gtaagaatcgagggtgaggtgtgtgtttgcagctgttaag	7	2.0	ataagtctatcctaagggctacgttctttttttgttttag/ATCAATTCT
VIII	122	843→ 964	TGTGAACTT/gtgagtgttcattgcctgaggggagggtgccctcagtagc	8	84 bp	ggctcacagtgaagtaccgtgttcttttgtttccttgtag/GTACATGTT
IX	41	965→1005	ATTGGAATG/gtaagagcactttaatatctgtagaacttatgctagaatg	9	5.3	caccatcagtaataactagtggcacattttgtcttctag/GGGCCAACT
X	113	1006→1118	ATCCAGAAG/gtgggtgctttgtatgagctgtggtgggtgctttgtatga	10	>10	actttttttcttttctctctttttctactctggtttatag/TTTGAAAAA
XI	75	1119→1193	TTGGAAAAG/gtaagaaaaaaaaataacagaaactttgaaagtgagagac	11	2.3	aggggtgcagtgtgtgaatgctgcttttgtttcttttcag/TTAAACACA
XII	90	1194→1283	GTGAAACAG/gtaggccttcaaaggtaatgcctcgccttttttctacttg	12	1.9	gcaatagtgatgtatttttcttaaatccttttctaaccag/CATTATGGG
XIII	114	1284→1397	GTGGATGGG/gtaagaatcgtctctgaactgtgcgttttgtcccagtgag	13	1.4	taccctcaattcaattgtcatttggtcttgaaccccacag/ATCCCCCAG
XIV	132	1398→1529	GATTTTAAG/gtaggacaaaccacactgtcaagactaaggtttaaggggt	14	2.5	ttccctccccagctcactgtgctctgtttctgctttgcag/GGTTTGGTT
XV	255	1530→1784	TGGACCAAG/gtacgtgtcttcatgggaagggtttgaattc	15	15	tgcatctggggttgatggttctctgtctctctctctccag/GGCCAGGTC
XVI	≥615	1785→2399				

Exon sequences are in upper case letters, intron sequences in lowercase.

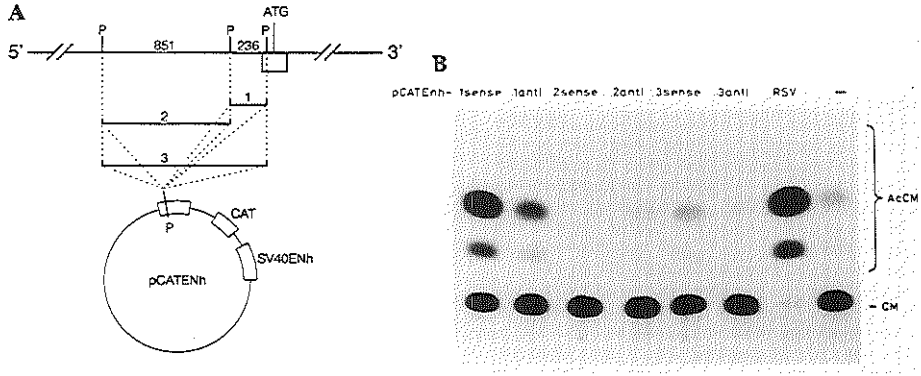


FIG. 2. A. Schematic representation of the β-galactosidase promoter region and of the different constructs used in the CAT assay. Direction of transcription is indicated. Exon I is shown as a hatched box including the ATG translation initiation codon of β-galactosidase cDNA. P is *Pst* I. Numbers 1, 2, and 3 designate the 236-bp, the 851-bp, and the 1.1-kb *Pst* I fragments, respectively. These fragments were cloned into the *Pst* I site of the reporter plasmid pCATEnh in front of the bacterial CAT gene. The position of the SV40 enhancer is also indicated. B. Identification of the human β-galactosidase promoter. Lysates from HeLa cells transfected with the various pCATEnh-constructs mentioned above were used in CAT assays. "Sense" and "anti" indicate the orientation of the fragments in the reported plasmid. CM represents the unprocessed substrate and AcCM marks the position of the acetylated products after substrate conversion. RSV is the promoter of the Rous sarcoma virus, used as a control in the assay.

as their combined 1.1-kb partial fragment (Fig. 2A, fragment 3) were subcloned separately in sense and antisense orientations 5′ of the chloramphenicolacetyl transferase (CAT) gene into the pCATEnh reporter plasmid (Fig. 2A). This plasmid contains a SV40 enhancer but lacks a promoter element. The different constructs were transfected into HeLa cells together with a plasmid expressing *E. coli* β-galactosidase used as a reference for transfection efficiency. Cell lysates were then tested for promoter activity using a CAT assay. The results are shown in Fig. 2B. The 236-bp fragment (pCAT1 sense and antisense) has clear promoter activity in both orientations, although the antisense construct stimulates transcription to a lesser extent than the sense construct. By contrast, the upstream 851-bp fragment (pCAT2 sense and antisense) fails to stimulate CAT transcription. Furthermore, it might contain negative regulatory sequences, because endogenous background activity of the promoterless plasmid is totally shut off when the fragment is used in the sense orientation. This observation is further substantiated by the low degree of transcription stimulation detected after transfection of the construct that carries the entire 1.1-kb *Pst* I partial fragment including the putative promoter element (pCAT3 sense and antisense). These results demonstrate that the β-galactosidase promoter is confined within a 236-bp *Pst* I fragment and suggest the presence of a putative silencer in the immediately flanking 5′ sequences.

The sequence of the entire 1.1-kb genomic fragment is given in Fig. 3. The 206-bp stretch (nucleotides −256 to −51) upstream from the known sequence of HβGa(L) cDNA (indicated with an arrowhead in Fig. 3) has a G + C content of 71% and a ratio of observed/expected CpG dinucleotide of 0.76 (Gardiner-Garden and Frommer,

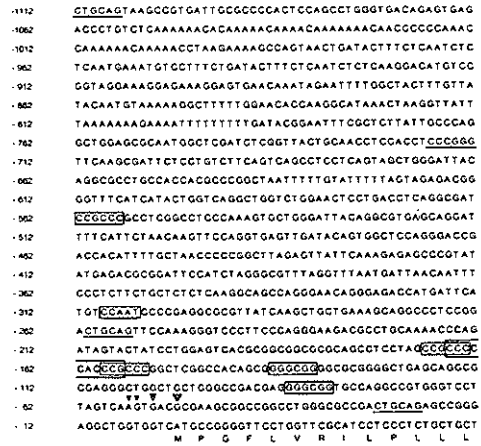


FIG. 3. The 5′ nucleotide sequence of the human β-galactosidase gene. The sequence includes 1,062 nucleotides of 5′-flanking DNA, 50 nucleotides of 5′ UTR present in the cDNA (indicated as a big open arrowhead), and 38 nucleotides of coding sequence. The first 13 amino-terminal amino acids of the β-galactosidase protein are shown. *Pst* I and *Sma* I restriction sites are underlined. SP1 binding sites are boxed. The only potential AP-2 motif is included in a larger box and overlaps with two SP1 sites. The big closed arrowhead indicates the major transcription start sites; the two minor sites are indicated with small close arrowheads.

1987). The G + C frequency falls back to 40–50% in the immediately upstream 851-bp *Pst* I fragment (nucleotides −1,107 to −257). Several putative transcription regulatory elements are found in the 1.1-kb genomic region, the most frequent being the promoter specific SP1 binding motif GGGCGG (Dynan, 1986; Mitchell and Tjian, 1989). Five such sites are present in both sense and antisense orientation, four of which are clustered within the small 236-bp fragment. A potential binding site for the transcription factor AP-2 (TCCCCANGGC) is located at nucleotides −166 to −157 and overlaps with two SP1 binding sites (Imagawa *et al.*, 1987). One CCAAT box (Breathnach and Chambon, 1981) is found at positions −309 to −305.

A primer extension experiment was performed to determine the cap site of human β-galactosidase mRNA (Fig. 4). Immunoselected polysomal mRNA was reverse-transcribed using an end-labeled antisense 27-mer oligonucleotide primer complementary to nucleotides −34 to −7 of exon I (Fig. 3). A nucleotide sequence of the corresponding genomic region was performed using the same primer and included as a reference. Multiple cap sites were identified, the major of which maps 53 nucleotides upstream from the translation initiation codon. Two minor extension products are 55 bp and 56 bp upstream from this ATG (Fig. 3). These results demonstrate that the mRNA encoding human β-galactosidase is only 3–6 bp longer than the longest cDNA isolated (Morreau *et al.*, 1989).

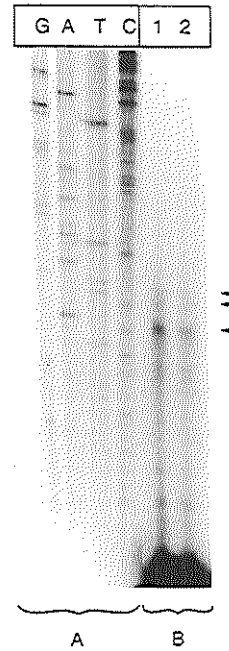


FIG. 4. Primer extension of the 5′ end of human β-galactosidase mRNA. A 27-mer spanning nucleotides −34 to −7 of exon I was used as a primer. B. Extended products obtained using two different polysomal mRNA preparations. The thick arrow indicates the major transcription initiation site (position −53 in Fig. 3); the small arrows (position −55/−56 in Fig. 3) indicate two minor extension products. A. Nucleotide sequence of the corresponding genomic area obtained using the same oligonucleotide primer.

Expression of mouse β-galactosidase mRNA and mouse genomic organization

A full-length cDNA encoding mouse β-galactosidase was recently isolated and characterized (Nanba and Suzuki, 1990). The nucleotide and amino acid sequences are highly conserved between human and mouse species. We have independently isolated a 2.0-kb partial cDNA missing the 5′-most 400 bp. To determine the level of expression of mouse β-galactosidase transcript, we analyzed total RNA from different mouse tissues on a Northern blot. Samples were applied in equal amounts and the blot was probed with the 2.0-kb cDNA fragment. As shown in Fig. 5, a 2.5-kb β-galactosidase mRNA is seen in all tissues tested, but its amount varies in a tissue-specific manner. It is very high in kidney, rather high in brain and spleen, and comparatively lower in other tissues. The overall tissue distribution of mouse β-galactosidase mRNA would conform with the "housekeeping" features of the human gene (Dynan, 1986). Still, transcription regulation might be necessary to modulate the level of mRNA depending upon its need in particular cell types.

We have shown earlier than in human tissues a minor 2.0-kb β-galactosidase transcript is present together with the abundant 2.5-kb species. These different transcripts derive from alternatively spliced pre-mRNA and encode the classic lysosomal β-galactosidase and a minor β-galactosidase-related protein (Morreau *et al.*, 1989). In none of the mouse total RNA samples analyzed was the small 2.0-kb species detected. To verify whether mouse precursor mRNA can undergo a similar splicing event and give rise to two translation products, we isolated mouse genomic phages spanning the area of interest. The exons of the mouse β-galactosidase gene equivalent to human exons III/IV/V/VI and part of exon II were sequenced and their exon/intron boundaries determined. Figure 6 shows that the partial genomic organization of the mouse gene is identical to the one of its human counterpart. All exon/intron splice junctions conform to the GT/AG rule. We have previously demonstrated that alternative splicing of human pre-mRNA involves exons III, IV, and VI. As a consequence of this differential splicing, sequences comprised within exon V are translated in two different reading frames, one giving rise to lysosomal β-galactosidase and the other to the β-galactosidase-related protein. In Fig. 7, the mouse counterpart of human exon V is depicted together with its splice sites. Translation of this exon is only possible in the reading frame (Fig. 7, reading frame 3) that

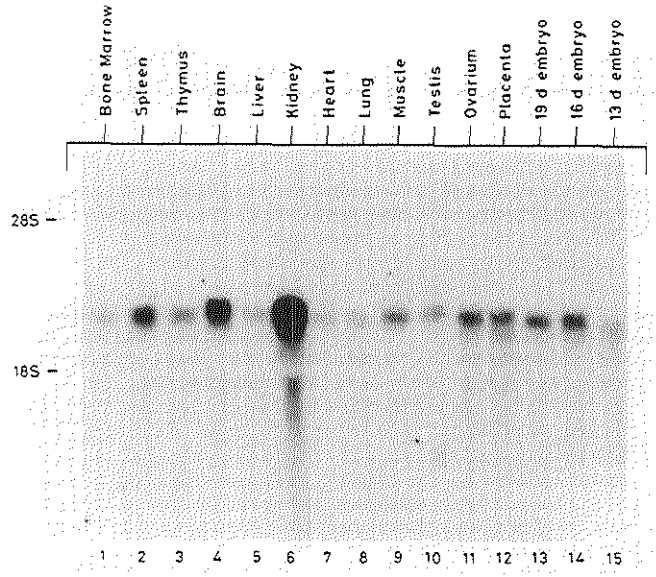


FIG. 5. Expression of β-galactosidase mRNA in different mouse tissues. RNA samples of different mouse tissues and embryos at indicated days of gestation were fractionated on gel, blotted onto Zeta-Probe membranes, and probed with a 2.0-kb partial mouse β-galactosidase cDNA. Ribosomal RNA markers are shown. Exposure time was 1 week.

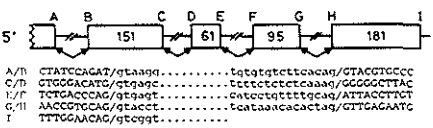


FIG. 6. Genomic organization of the 5′ region of the mouse β-galactosidase gene. Boxes represent separate exons; numbers specify their length in base pairs. The 5′ end sequence of the left-most exon was not determined as indicated by vertical zigzag lines. The sequences of exon/intron boundaries are marked with sequential letters (A–I).

codes for the classic lysosomal β-galactosidase. This implies that a mouse β-galactosidase-related protein, if present at all, cannot derive from a minor transcript that is alternatively spliced from pre-mRNA in a similar fashion as the human transcript.

DISCUSSION

We have previously characterized a partial 5′ region of the human lysosomal β-galactosidase gene undergoing alternative splicing and giving rise to two distinct transcripts. Except for two large introns, we now report the isolation and analysis of the entire gene encoding β-galactosidase. It consists of 16 exons that vary in size from 41 bp to ≥615 bp and are dispersed over a genomic region of more than 62.5 kb. Intron lengths range from 84 bp to ≥15 kb. All exon/intron splice junctions follow the CT/AG rule defined by Breathnach and Chambon (1981). It has been proposed that intervening sequences demarcate important functional or structural domains of proteins that are encoded by different exons (Gilbert, 1985). As it is the case for several secretory and membrane proteins (Gilbert, 1985), the signal sequence of pre-pro-β-galactosidase is confined to exon I. Exon II encodes the first putative glycosylation site present in the precursor polypeptide and the amino-terminal sequence of the mature 64-kD enzyme (Morreau *et al.*, 1989).

The site of carboxy-terminal proteolytic processing of the β-galactosidase precursor was suggested to be around amino acid 530 (Yamamoto *et al.*, 1990) whose coding sequence falls in the middle of exon XV. This postulation was made on the basis of the hydrophilic profile of the protein in this region. Although the amino acid sequence between residues 500–550 is the least homologous between human and mouse β-galactosidases (Nanba and Suzuki, 1990), we have verified that the highly hydrophilic charac-

FIG. 7. Complete sequence of the intermediate 95-bp mouse exon. Exonic sequences are shown as uppercase letters, intronic as lowercase letters. Predicted amino acid sequences translated from reading frames 1, 2, or 3 are indicated. Asterisks refer to stop codons.

teristic of this part of the protein has been conserved (H. Morreau, unpublished observation).

Functional analysis of subfragments of a 1.1-kb genomic region, spanning part of exon I and the 5'-flanking sequences, allowed us to identify promoter activity for the β-galactosidase gene to a 236-bp *Pst* I fragment. The latter is rich in GC residues (71%) and bears characteristics of a CpG island (Bird, 1986; Gardiner-Garden and Frommer, 1987). It lacks a conventional TATA box and contains four of the five consensus sequences for the promoter specific transcription factor SP1 (Dynan, 1986; Mitchell and Tjian, 1989), identified in the entire 1.1-kb fragment. Additional analysis by primer extension showed that transcription initiation sites are clustered in a 5-nucleotide stretch and are located within the 236-bp *Pst* I fragment. Promoters of other genes encoding lysosomal enzymes have been characterized. Some appeared to be typical "housekeeping" promoters, like β-galactosidase (Neote *et al.*, 1988; Geier *et al.*, 1989; Hoefsloot *et al.*, 1990; Martiniuk *et al.*, 1990), while others have GC-rich promoters containing a TATA box (Proia and Soravia, 1987; Bishop *et al.*, 1988; D'Amore *et al.*, 1988). The only exception is the gene encoding glucocerebrosidase whose promoter lacks GC-rich stretches and comprises conventional CAAT and TATA boxes (Reiner *et al.*, 1988; Reiner and Horowitz, 1988). The 236-bp fragment of β-galactosidase directs transcriptional initiation in an orientation-independent manner, as observed for other GC-rich promoter elements carrying SP1 binding sites (Hata *et al.*, 1989; Baarends *et al.*, 1990). A putative recognition site for transcription factor AP-2 (Imagawa *et al.*, 1987) is also found at positions −166 to −156 and overlaps with two SP1 sequences. Whether this is of functional importance remains to be determined.

Surprisingly, if the 236-bp promoter element is placed downstream from its authentic 851-bp upstream sequences, promoter activity is drastically diminished. Furthermore, antisense transcription is completely abolished by the presence of the 851-bp fragment which makes transcription orientation unidirectional. This fragment by itself also abolishes the endogenous background transcription activity of the pCATEnh reporter plasmid. It is conceivable that *cis*-acting "silencer" sequences are contained within the 851-bp region, as has been reported for other mammalian genes including rat α-fetoprotein, insulin I, and growth hormone genes (Laimins *et al.*, 1986; Muglia and Rothman-Denes, 1986; Wight *et al.*, 1987) and human interferon-β (IFN-β), apolipoprotein CIII, and calpain genes (Goodbourn *et al.*, 1986; Reue *et al.*, 1988; Hata *et al.*, 1989). However, we could not identify within the 851-bp fragment any conserved sequences similar to those present in negative regulatory elements of human IFN-β and apolipoprotein CIII gene promoters. Sequential deletions of this genomic region coupled to footprinting and band-shift assays will enable us to determine whether specific transcription factors bind to consensus sequences in this area.

Cis-acting regulatory elements might modulate transcription of the β-galactosidase gene in a tissue-specific manner. The level of expression of β-galactosidase mRNA varies in different mouse tissues in spite of its ubiquitous distribution. However, we do not know if these differences reflect variation in transcription of the gene. A differential pattern of expression has also been observed for mouse protective protein/cathepsin A mRNA (Galjart *et al.*, 1990). Although the two enzymes mutually cooperate at the protein level, the amount of their corresponding transcripts in different tissues need not necessarily overlap. For example, placental β-galactosidase mRNA is rather low compared to a highly expressed protective protein mRNA. This might reflect either specific regulation of transcription or conversely different rate of degradation of the aforementioned transcripts.

We have determined the partial genomic organization of the mouse β-galactosidase gene to define whether 5' exons corresponding to the ones alternatively spliced in the human gene can undergo a similar event. We have shown that exon/intron splice junctions are completely conserved between mouse and human β-galactosidase genes. Nevertheless, if alternative splicing of mouse exons corresponding to human exons III, IV, and VI takes place this leads to interrupted translation of the resulting transcript. Therefore, a mouse counterpart of the human β-galactosidase-related protein, if present, cannot arise through a similar splicing process. It is still possible that alternative splicing of human β-galactosidase pre-mRNA is a means to suppress translation of an active lysosomal enzyme or to change the subcellular fate of a translated product. Such mechanisms of down regulation has been postulated for differentially spliced transcripts encoding specific proteins

in *Drosophila* (Bingham *et al.*, 1988) and the H_2 subunit of human asialoglycoprotein receptor (Lederkremer and Lodish, 1991).

ACKNOWLEDGMENTS

We wish to acknowledge the constant support of Professor Hans Galjaard. We would like to thank Dr. Gerard Grosveld and Dies Meijer for providing us with the genomic libraries and for their experimental suggestions and discussions. We also thank Pim Visser and Mirko Kuit for excellent graphic and photographic work and Jeannette Lokker for typing and editing the manuscript. At last we like to thank Sjozèf van Baal for his essential help in handling computer programs.

REFERENCES

AUFFRAY, C., and ROUGEON, F. (1980). Purification of mouse immunoglobulin heavy chain messenger RNAs from total myeloma tumor RNA. Eur. J. Biochem. **107**, 303–314.

BAARENDS, W.M., THEMMEN, A.P.N., BLOK, L.J., MACKENBACH, P., BRINKMANN, R.O., MEIJER, D., FABER, P.W., TRAPMAN, J., and GROOTEGOED, J.A. (1990). The rat androgen receptor gene promoter. Mol. Cell. Endocrinol. **74**, 75–84.

BINGHAM, P.M., CHOU, T.-B., MIMS, I., and ZACHAR, Z. (1988). On/off regulation of gene expression at the level of splicing. Trends Genet. **4**, 134–138.

BIRD, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature **321**, 209–213.

BISHOP, K.F., KORNREICH, R., and DESNICK, R.J. (1988). Structural organization of the human α-galactosidase A gene: Further evidence for the absence of a 3′ untranslated region. Proc. Natl. Acad. Sci. USA **85**, 3903–3907.

BREATHNACH, R., and CHAMBON, P. (1981). Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**, 349–383.

CONZELMANN, E., and SANDHOFF, K. (1987). Glycolipid and glycoprotein degradation. Adv. Enzymol. **60**, 89–217.

D'AMORE, M.A., GALLAGHER, P.M., KORFHAGEN, T.R., and GANSCHOW, R.E. (1988). Complete sequence and organization of the murine β-glucuronidase gene. Biochemistry **27**, 7131–7140.

D'AZZO, A., HOOGEVEEN, A.T., REUSER, A.J.J., ROBINSON, D., and GALJAARD, H. (1982). Molecular defect in combined β-galactosidase and neuraminidase deficiency in man. Proc. Natl. Acad. Sci. USA **79**, 4535–4539.

DYNAN, W.S. (1986). Promoters for housekeeping genes. Trends Genet. **2**, 196–197.

FEINBURG, A.P., and VOGELSTEIN, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**, 6–13.

FOURNEY, R.M., MIYAKOSHI, J., DAY, R.S., and PATTERSON, M.C. (1988). Northern blotting: Efficient RNA staining and transfer. Focus **10**, 5–7.

GALJART, N.J., GILLEMANS, N., HARRIS, A., VAN DER HORST, G.T.J., VERHEIJEN, F.W., GALJAARD, H., and D'AZZO, A. (1988). Expression of cDNA encoding the human "protective protein" associated with lysosomal β-galactosidase and neuraminidase: Homology to yeast proteases. Cell **54**, 755–764.

GALJART, N.J., GILLEMANS, N., MEIJER, D., and D'AZZO, A. (1990). Mouse "protective protein": cDNA cloning sequence comparison and expression. J. Biol. Chem. **265**, 4678–4684.

GALJART, N.J., MORREAU, H., WILLEMSEN, R., GILLEMANS, N., BONTEN, E.J., and D'AZZO, A. (1991). Lysososomal protective protein is cathepsin A and its catalytic activity is distinct from its protective function. J. Biol. Chem. (in press).

GARDINER-GARDEN, M., and FROMMER, M. (1987). CpG Islands in vertebrate genomes. J. Mol. Biol. **196**, 261–282.

GEIER, C., VON FIGURA, K., and POHLMANN, R. (1989). Structure of the human lysosomal acid phosphatase gene. Eur. J. Biochem. **183**, 611–616.

GILBERT, W. (1985). Genes-in-pieces revisited. Science **228**, 823–824.

GOODBOURN, S., BURSTEIN, H., and MANIATIS, T. (1986). The human β-interferon gene enhancer is under negative control. Cell **45**, 601–610.

GORMAN, C.M., MOFFAT, L.F., and HOWARD, B.H. (1982). Recombinant genomes which express chloramphenicol acetyl-transferase in mammalian cells. Mol. Cell Biol. **2**, 1044.

GRAHAM, F.L., and VAN DER EB, A.J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**, 456.

GROEBE, H., KRINS, M., SCHMIDBERGER, H., VON FIGURA, K., HARZER, K., KRESSE, H., PASCHKE, E., SEWELL, A., and ULLRICH, K. (1980). Morquio syndrome (mucopolysaccharidosis IV-B) associated with β-galactosidase deficiency. Report of two cases. Am. J. Hum. Genet. **32**, 258–272.

HATA, A., OHNO, S., AKITA, Y., and SUZUKI, K. (1989). Tandemly reiterated negative enhancer-like elements regulate transcription of a human gene for the large subunit of calcium-dependent protease. J. Biol. Chem. **264**, 6404–6411.

HERMANS, A., GOW, J., SELLERI, L., VON LINDERN, M., HAGEMEIJER, A., WIEDEMANN, L.M., and GROSVELD, G. (1988). Ber-abl oncogene activation in Philadelphia chromosome-positive acute lymphoblastic leukemia. Leukemia **2**, 628–633.

HOEFSLOOT, L.H., HOOGEVEEN-WESTERVELD, M., REUSER, A.J.J., and OOSTRA, B.A. (1990). Characterization of the human lysosomal α-glucosidase gene. Biochem. J. **272**, 493–497.

HOOGEVEEN, A.T., VERHEIJEN, F.W., and GALJAARD, H. (1983). The relation between human lysosomal β-galactosidase and its protective protein. J. Biol. Chem. **258**, 12143–12146.

IMAGAWA, M., CHIU, R., and KARIN, M. (1987). Transcription factor AP-2 mediates induction by two different signal-transduction pathways: Protein kinase C and cAMP. Cell **51**, 251–260.

LAIMINS, S., HOLMGERN-KÖNIG, M., and KHOURY, G. (1986). Transcriptional "silencer" element in rat repetitive sequences associated with the rat insulin 1 gene locus. Proc. Natl. Acad. Sci. USA **83**, 3151–3155.

LEDERKREMER, G.Z., and LODISH, H.F. (1991). An alternatively spliced miniexon alters the subcellular fate of the human asialoglycoprotein receptor H2 subunit. J. Biol. Chem. **266**, 1237–1244.

MARTINIUK, F., MEHLER, M., TZALL, S., MEREDITH, G., and HIRSCHHORN, R. (1990). Sequence of the cDNA and 5′-flanking region for human acid α-glucosidase, detection of an intron in the 5′ untranslated leader sequence, definition of 18 bp polymorphisms, and differences with previous cDNA and amino acid sequences. DNA Cell Biol. **9**, 85–94.

MITCHELL, P.J., and TJIAN, R. (1989). Transcriptional regulation of mammalian cells by sequence-specific DNA binding proteins. Science **245**, 371–378.

MORREAU, H., GALJART, N.J., GILLEMANS, N., WILLEMSEN, R., VAN DER HORST, G.T.J., and D'AZZO, A. (1989). Alternative splicing of β-galactosidase mRNA generates the classic lysosomal enzyme and a β-galactosidase-related protein. J. Biol. Chem. **34**, 20655–20663.

MUGLIA, L., and ROTHMAN-DENES, L.B. (1986). Cell type-specific negative regulatory element in the control region of the rat α-fetoprotein gene. Proc. Natl. Acad. Sci. USA **83**, 7653–7657.

MURPHY, G., and KAVANAGH, T. (1988). Speeding-up the sequencing of double-stranded DNA. Nucleic Acids Res. **16**, 5198.

MYEROWITZ, R., and PROIA, R.L. (1984). cDNA clone for the α-chain of human β-hexosaminidase: Deficiency of α-chain mRNA in Ashkenazi Tay-Sachs fibroblasts. Proc. Natl. Acad. Sci. USA **81**, 5394–5398.

NANBA, E., and SUZUKI, K. (1990). Molecular cloning of mouse acid β-galactosidase cDNA, expression of catalytic activity and comparison with the human enzyme. Biochem. Biophys. Res. Commun. **173**, 141–148.

NEOTE, K., BAPAT, B., DUMBRILLE-ROSS, A., TROXEL, C., SCHUSTER, S.M., MAHURAN, D.J., and GRAVEL, R.A. (1988). Characterization of the human hex B gene encoding lysosomal β-hexosaminidase. Genomics **3**, 279–286.

OKADA, S., and O'BRIEN, J.S. (1968). Generalized gangliosidosis: Beta-galactosidase deficiency. Science **160**, 1002–1004.

O'BRIEN, J.S. (1989). The gangliosidosis. In *The Metabolic Basis of Inherited Disease*. C.R. Scriver, A.L. Beadet, W.S. Sly, and D. Valle, eds. (McGraw-Hill, New York) pp. 1797–1806.

O'BRIEN, J.S., GUGLER, E., GIEDION, A., WIESSMAN, U., HERSCHKOWITZ, N., MEIER, C., and LEROY, J. (1976). Spondyloepiphyseal dysplasia, corneal clouding, normal intelligence and acid β-galactosidase deficiency. Clin. Genet. **9**, 495–504.

OSHIMA, A., TSUJI, A., NAGAO, Y., SAKURABA, H., and SUZUKI, Y. (1988). Cloning, sequencing, and expression of cDNA for human β-galactosidase. Biochem. Biophys. Res. Commun. **157**, 238–244.

PROIA, R.L., and SORAVIA, E. (1987). Organization of the gene encoding the human β-hexosaminidase α-chain. J. Biol. Chem. **262**, 5677–5681.

REINER, O., and HOROWITZ, M. (1988). Differential expression of the human glucocerebrosidase-coding gene. Gene **73**, 469–478.

REINER, O., WIGDERSON, M., and HOROWITZ, M. (1988). Structural analysis of the human glucocerebrosidase genes. DNA **7**, 102–116.

REUE, K., LEFF, T., and BRESLOW, J.L. (1988). Human apolipoprotein CIII gene expression is regulated by positive and negative cis-acting elements and tissue-specific protein factors. J. Biol. Chem. **263**, 6857–6864.

SAMBROOK, J., FRITSCH, E.F., and MANIATIS, T. (1989). *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

SHOWS, T.B., SCRAFFORD,-WOLFF, L.R., BROWN, J.A., and MEISLER, M. (1979). G_{M1}-gangliosidosis: Chromosome 3 assignment of the β-galactosidase A gene (βGAL_A). Somat. Cell Genet. **5**, 147–158.

STADEN, R. (1986). The current status and portability of our sequence handling software. Nucleic Acids Res. **14**, 217–231.

VAN DER HORST, G.T.J., GALJART, N.J., D'AZZO, A., and VERHEIJEN, F.W. (1989). Identification and *in vitro* reconstitution of lysosomal neuraminidase from human placenta. J. Biol. Chem. **264**, 1317–1322.

VERHEIJEN, F.W., BROSSMER, R., and GALJAARD, H. (1982). Purification of acid β-galactosidase and acid neuraminidase from bovine testis: Evidence for an enzyme complex. Biochem. Biophys. Res. Commun. **108**, 868–875.

WIGHT, P.A., CREW, M.D., and SPINDLER, S.R. (1987). Discrete positive and negative thyroid hormone-responsive transcription regulatory elements of the rat growth hormone gene. J. Biol. Chem. **262**, 5659–5663.

YAMAMOTO, Y., and NISHIMURA, K. (1987). Copurification and separation of β-galactosidase and sialidase from porcine testis. Int. J. Biochem. **5**, 435–442.

YAMAMOTO, Y., FUJIE, M., and NISHIMURA, K. (1982). The interaction between high- and low-molecular-weight forms of G_{M1}-β-galactosidase purified from porcine spleen. J. Biochem. **92**, 13–21.

YAMAMOTO, Y., HAKE, C.A., MARTIN, B.M., KRETZ, K.A., AHERN-RINDELL, A.J., NAYLOR, S.L., MUDD, M., and O'BRIEN, J.S. (1990). Isolation, characterization, and mapping of human acid β-galactosidase cDNA. DNA Cell Biol. **9**, 119–127.

Address reprint requests to:
Dr. Alessandra d'Azzo
Department of Cell Biology and Genetics
Erasmus University
POB 1738
3000 DR Rotterdam
The Netherlands

Received for publication March 5, 1991, and in revised form May 24, 1991.

Publication 3

(Submitted for publication)

ABERRANTLY SPLICED ß-GALACTOSIDASE mRNA WITH A TWENTY NUCLEOTIDE INTRONIC INSERTION CAUSES AN INFANTILE FORM OF G_{M1}-GANGLIOSIDOSIS

Amelia Morrone§, Hans Morreau⊙, Xiao Yan Zhou⊙, Enrico Zammarchi§, Wim J. Kleyer♦, Hans Galjaard⊙ and Alessandra d'Azzo⊙

From The ⊙Department of Cell Biology and Genetics, The ♦Department of Clinical Genetics, Erasmus University, Rotterdam, The Netherlands; and §The Department of Pediatrics, University of Florence, Florence, Italy

Summary

The lysosomal storage disorders G_{M1}-gangliosidosis and Morquio B syndrome are caused by a complete or partial deficiency of acid ß-galactosidase. We characterized the mutation segregating in a family with two sibs having the severe infantile form of G_{M1}-gangliosidosis. In total mRNA preparations derived from the patient's fibroblasts two aberrant ß-galactosidase transcripts (1 and 2) have been identified. Both transcripts contain a 20 bp insertion derived from the 5' end of intron 1 of the ß-galactosidase gene due to aberrant splicing of pre-mRNA molecules. Furthermore in transcript 2 sequences encoded by exon II are deleted during the splicing process. Comparison of the 20 nucleotide insertion with wildtype intronic sequences showed that in genomic DNA of the patients an extra T is inserted immediately downstream the conserved GT splice donor dinucleotide. Both patients are homozygotes for the T nucleotide insertion. We speculate that this single base substitution is responsible for the preferential usage of a cryptic splice donor site during the splicing process.

Mutations in the human ß-galactosiadase gene on chromosome 3 (Shows et al., 1979) lead to absence or abrogation of the normal function of ß-galactosidase, which is the hydrolysis of ß-linked terminal galactosyl-residues from a variety of natural substrates, like glycolipids, glycopeptides and oligosaccharides (Conzelmann and Sandhoff, 1987; O'Brien, 1989). The majority of active ß-galactosidase is found in a complex with lysosomal neuraminidase and the protective protein/ cathepsin A (Verheijen et al., 1982; 1985; Yamamoto and Nishimura, 1987). It has been demonstrated that the latter is essential for intralysosomal activation and stabilization of the other two glycosidases (d'Azzo et al., 1982; Hoogeveen et al., 1983; Verheijen et al., 1985; Van der Horst et al., 1989).

An isolated ß-galactosidase deficiency results in the metabolic storage disorders G_{M1}-gangliosidosis and Morquio-B syndrome (Okada and O'Brien, 1968; O'Brien et al., 1976; Groebe et al., 1980; for review see O'Brien, 1989). Depending on the clinical symptomatology G_{M1}-gangliosidosis patients are classified as having either the severe infantile or the mild juvenile and adult form of the disease. The infantile G_{M1}-gangliosidosis patients suffer from a severe neurodegenerative disorder, dysmorphic features and hepatosplenomegaly, leading to early death. The residual activity of ß-galactosidase in these patients is less than 1%. Pathologic examination shows excessive accumulation of G_{M1}-ganglioside in neuronal tissues as well as glycosaminoglycans and glycopeptides in visceral organs and other tissues. The juvenile and adult variants have milder symptoms, a prolonged life expectancy and higher (10-15%) residual ß-galactosidase activity. In Morquio-B syndrome, the mutant

enzyme can normally degrade G_{M1}-ganglioside, but has an altered affinity for keratansulphate and oligosaccharides, which are the major substrates accumulated in patient's tissues (Paschke and Kresse, 1982; van der Horst et al., 1983). Therefore, no obvious neurologic impairment is detected in the latter syndrome, in contrast to G_{M1}-gangliosidosis.

In human cultered fibroblasts the newly synthesized ß-galactosidase is a glycosylated precursor molecule of 85 kDa that is processed through intermediate forms to a 64-kDa mature enzyme (d'Azzo et al., 1982).

The cDNAs and genes encoding human and mouse ß-galactosidase have been isolated and characterized (Oshima et al., 1988); Morreau et al., 1989; 1991; Yamamoto et al., 1990; Nanba and Suzuki, 1990; 1991). The primary structures of the human and murine enzyme, as deduced from the cDNAs, share 80% identity (Morreau et al., 1989; Nanba et al., 1990). Exon/ intron boundaries are also completely conserved between the two species, the only difference being the length of exon I, XIII and XVI. Analysis of the promoter regions of both genes has revealed several binding sites for the transcription factor SP1, within a GC rich sequence and no conventional TATA and CCAAT elements (Morreau et al., 1991; Nanba et al., 1991). It has been shown that the human gene can give rise to at least two alternatively spliced mRNAs: one major transcript of 2.5 kb encodes the classic, catalytically active ß-galactosidase protein, and one minor of about 2.0 kb gives rise to a non-lysosomal ß-galactosidase-related protein which is inactive towards the artificial substrate used (Morreau et al., 1989; Yamamoto et al., 1990).

The characterization of the human ß-galactosidase cDNA has led to the identification of several mutations underlying distinct clinical forms of G_{M1}-gangliosidosis and Morquio-B syndrome (Yoshiba et al, 1991; Nishimoto et al., 1991; Oshima et al., 1991). All G_{M1}-gangliosidosis patients sofar analyzed are of Japanese origin. Most of the identified mutations are single base substitutions leading to single aminoacid changes. The majority of the mutations are different in clinically distinct patients, although in most of the cases only one allele has been identified. In this report we describe the genetic lesion present in two siblings affected with the severe infantile form of G_{M1}-gangliosidosis. We have found that the insertion of an extra T nucleotide at the donor splice site leads to aberrant splicing of ß-galactosidase pre-mRNA.

Experimental Procedures

Cell culture

Human skin fibroblasts from the index patient, the fetus and their parents, were obtained from the European Cell Repository, Rotterdam (Dr W.J. Kleijer). Fibroblasts were cultered in Dulbecco's modified Eagles - Hams F10 medium (1:1 vol/vol) with antibiotics and 10 % fetal bovine serum.

Oligonucleotides

For cDNA synthesis, PCR amplification, nucleotide sequencing and hybridization studies several oligonucleotide primers were constructed on the basis of either HßGaL cDNA or genomic sequences (Morreau *et al.*, 1989; 1991). The oligonucleotide primers were synthesized on an Applied Biosystems 381A oligonucleotide synthesizer and purified as recommended by the manufacturer. The following primers were made:

(a) 5'CGAATTCATGCCGGGGTTCCTGGTTCGC 3' (sense)
(b) 5'CGAATTCCTCCCATTTCCCACTCTGCACAG 3' (antisense)
(c) 5' GGCTTGCGCAATGCCACC 3'(sense)
(d) 5' CTTGCGCGTTAAGTCTGC 3'(sense)
(e) 5' GGGACCCGGGTATGTGCC 3'(sense)

(f) 5' ggttccccgccagcctgt 3'(antisense intronic)
(g) 5' CTTGCGCGTAAGTCTGC 3'(sense)
(h) 5' AGCTCATGAGCCAGCCGAAG 3'(antisense)

cDNA synthesis; PCR amplification of cDNA and genomic DNA

Total RNA was isolated from cultured fibroblasts as described by Auffray and Rougeon (1980). The entire coding sequence of the ß-galactosidase mRNA was reverse transcribed into six overlapping single stranded cDNA fragments using specific antisense primers (Hermans *et al.* 1988). Fragments were subsequently amplified on a Perkin Elmer Cetus thermocycler using sense and antisense primers in the presence of Taq polymerase. Genomic DNA was amplified similarly (Saiki *et al.*, 1988). Single stranded cDNA or genomic DNA samples were synthesized for direct sequencing by the asymmetric polymerase chain reaction method, as described by Gyllenstein and Erlich, (1988). Nucleotide sequencing was performed by the dideoxy chain termination method of Sanger *et al.* (1977), with internal primers or primers used for the first strand cDNA synthesis. In few occasions, PCR products were digested with EcoRI and subcloned into pTZ18 or 19 plasmid vectors (BRL). Nucleotide sequencing of these subclones was carried out with the dideoxy-chain termination method on double stranded DNA (Murphy and Kavanagh, 1988) using either M13 reverse or universal primers (BRL). The sequencing kit was purchased from United States Biochemical Corp.

For hybridization of Southern blots containing amplified DNA fragments, oligonucleotide probes were labeled at their 5' end with γ $^{[32]}$P and T_4 polynucleotide kinase. Hybridization and washing conditions were as previously described (Wood *et al.*, 1985). cDNA probes were labeled according to the procedure of Feinburg and Vogelstein (1983).

Results

Two altered cDNAs are derived from aberrantly spliced ß-galactosidase pre-mRNAs

A male infant, born in 1978 from healthy unrelated parents, was diagnosed at 14 months of age to suffer from severe infantile G_{M1}-gangliosidosis. The subsequent pregnancy of the mother resulted in early termination since the fetus was found to be also affected.

In order to identify the mutation(s) underlying the G_{M1}-gangliosidosis in this family, total RNA was isolated from skin fibroblasts of the index patients, the fetus and the parents. As seen in Fig.1, in total RNA preparations from both patients a very reduced amount of the 2.5 kb ß-galactosidase transcript is detected in the Northern blot, that is also resolved in a broad fuzzy band. The parents show about half of the amount of the 2.5 kb ß-galactosidase mRNA present in the normal control.

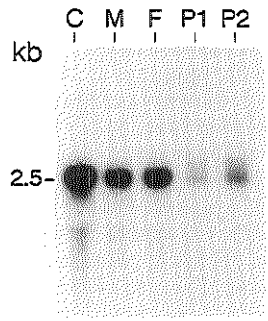


Fig. 1. Northern blot analysis of fibroblast RNA.
Total fibroblasts RNA (20 µg) from a control (C), two G_{M1}-gangliosidosis patients (P1 , P2) and their parents (M, F) were fractionated on a formaldehyde agarose gel and probed with the full length HßGaL cDNA.

Biosynthetic labeling experiments carried out on fibroblasts from the different family members, showed in both parents a reduced synthesis of ß-galactosidase precursor that was correctly processed into the 64-kDa lysosomal form. In contrast, the patients were found to be completely devoid of precursor as well as mature proteins (not shown).

To identify the mutation(s), we first analyzed the sequence of the ß-galactosidase mRNA isolated from the family members. For this purpose, six overlapping cDNA fragments encompassing the entire coding region of the ß-galactosidase mRNA were amplified from total RNA preparations, using first strand cDNA synthesis and the polymerase chain reaction (PCR). Amplified fragments were either sequenced directly or subcloned into pTZ18 or 19 plasmids and then sequenced. Only in the region spanning nucleotides 49-452 of wild type ß-galactosidase cDNA, amplified with oligonucleotide primers (a) and (b) (sequences are given in the Experimental Procedures), we detected differences from the normal sequence. In both patients two abnormal cDNAs were identified carrying both a 20 nucleotide (nt) insertion at position 126 of HßGaL cDNA. The sequences of these aberrant transcripts, 1 and 2, including the insertion, are shown in Fig 2A (panel 2 and 3). In addition to the 20 nt insertion, transcript 2 is also devoid of sequences encoded by exon II (Fig 2A, panel 3). The cDNA sequences spanning the normal exon I/II and exon II/III boundaries are also shown (Fig. 2A, panels 1 and 4).

Comparison with the corresponding genomic area demonstrated that the 20 nt insertion resembles the splice donor site and 3' flanking sequences of intron 1 of the human ß-galactosidase gene (Morreau et al., 1991). However, an additional T nucleotide is present in transcripts 1 and 2 of the patients after the conventional GT dinucleotide of the splice donor site. These abnormal mRNAs must be therefore be derived from aberrantly spliced pre-mRNA molecules, as depicted schematically in Fig. 2B. Apparently, the normal splice donor site of intron 1 is not longer recognized, but instead a cryptic site 21 nucleotide downstream is preferentially chosen. The aberrant transcripts are not amplified from mRNA preparations of both parents, indicating that they probably represent a minor pool. The 20 nt insertion causes a frameshift in the open reading frames (ORFs) of both aberrant transcripts. The reading frame of transcript 1 terminates in an early stopcodon. Consequently, only a small aberrant peptide could be potentially translated from transcript 1. In contrast, due to the deletion of exon II-encoded sequences in transcript 2, the normal ß-galactosidase ORF is resumed in transcript 2, and the latter could give rise to a protein missing only the exon II encoded aminoacid stretch. However, as mentioned above, immunoprecipitation studies have failed to reveal the presence of any ß-galactosidase protein in patient's fibroblasts.

To confirm the cDNA sequence findings we performed the following experiment, taking advantage of a unique PpuM1 recognition site present within the 20 nucleotide insertion. The cDNA fragment spanning nucleotide 49-452 was again amplified from total RNAs of all family members using the same oligonucleotide primers mentioned in Fig. 2. After digestion of the PCR products with the restriction enzyme PpuM1, the resulting cDNA fragments were separated on a 1% agarose gel and subsequently transferred onto nylon membranes. The lengths of undigested and digested cDNA fragments are shown in Fig. 3B. The blots were hybridized with different ^{32}P labeled oligonucleotide probes, indicated as blocks in the scheme of Fig. 3A (sequences are given in the Experimental Procedures). Oligo (c) is specific for the

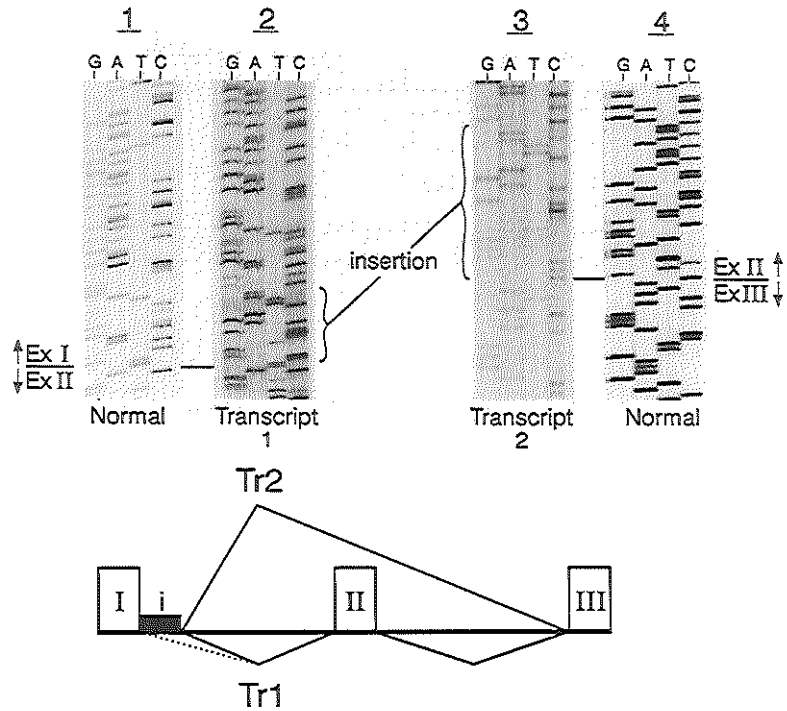
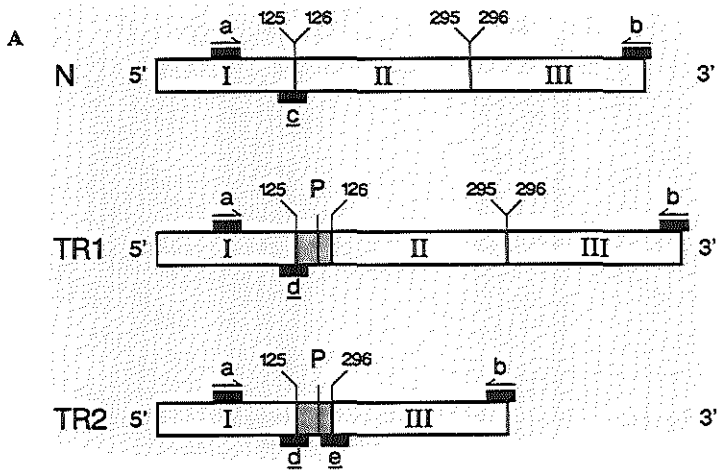


Fig. 2. Partial nucleotide sequence of aberrant transcripts 1 and 2.
A. Total RNA was isolated from fibroblasts of a normal individual, the two G_{M1}-gangliosidosis patients and their parents. The mRNA samples were reverse transcribed into cDNA with antisense primers located in different parts of the ß-galactosidase mRNA molecule. These cDNAs were subsequently amplified with sense and antisense primers, subcloned and sequenced (panel 1, 2 and 4) or sequenced directly after an asymmetric PCR reaction (panel 3). A portion of the normal antisense sequence of ß-galactosidase mRNA spanning ExI/ExII and ExII/ExIII boundaries is shown. Furthermore sequences of the aberrant transcripts 1 and 2 are given which include the 20 nucleotide insertion (indicated with brackets).
B. *Schematic representation of the splicing mechanism leading to aberrant transcripts 1 and 2.* Exon sequences are indicated by roman numbers I/II and III. The 20 nucleotide intronic sequence (i) present in transcript 1 and 2 is shown as a black bar. Normal splicing is indicated by underbroken lines.

normal transcript and is an exon I/II overlapping primer, spanning nt 117-134 of HßGaL, oligo (d) recognizes both abnormal transcripts, whereas oligo (e) is unique for transcript 2 and overlaps between insertion and exon III sequences. Furthermore, we used a ^{32}P labeled cDNA probe specific for exon II, spanning nt 150-219 of HßGaL cDNA. The results are shown in Fig.4. After hybridization with oligonucleotide (c) only the normal 413 bp fragment is seen in samples of controls and parents, although in the latter it is present in clearly reduced amounts. In amplified samples from both patients the 413 bp fragment is completely absent. These results implied that either both ß-galactosidase alleles carry the same mutation or



B

	Fragment length	
	- PpuM1 (bp)	+ PpuM1 (bp)
N	413	413
TR1	433	95 + 338
TR2	263	95 + 168

Fig.3. PpuM1 restriction analysis of partiel cDNAs derived from transcript 1 and 2.
A. Schematic representation of ß-galactosidase normal (N) and mutant (TR1, TR2) cDNA fragments amplified with oligonucleotide primers (a) and (b), indicated as black bars. Regions encoded by specific exons are marked by I,II or III. The numbers above the open bar refer to the nucleotide position in the HßGaL cDNA. The restriction enzyme PpuM1 is indicated with a capital P.
B. Amplified cDNA fragments were subjected to restriction analysis with the restriction enzyme PpuM1. The length of digested and undigested cDNA fragments is given in base pairs (bp).

only one allele is transcribed. With oligonucleotide (d) a 95 bp fragment, that proves the presence of abnormal transcripts 1 and 2, is visible only in samples from the two patients. No such fragment is detected among the amplified products derived from mRNA of the parents. A transcript 2-specific 168 nucleotide fragment hybridizes in both patients with oligo (e). Furthermore, a 338 bp fragment specific for the aberrant transcript 1 is also detected with the cDNA probe comprising exon II sequences and is present in samples from both patients. With the same probe, the 413 bp normal fragment is also seen in samples from controls and parents. These results confirm the direct sequencing data.

Analysis of exon I/intron 1 ß-galactosidase genomic region.
A 20 nucleotide intronic insertion is found in the ß-galactosidase mRNA of both affected children. In order to confirm the presence of the additional T within this nucleotide stretch at genomic level, we amplified a 150 bp genomic region with the exonic primer (a) in the sense orientation and the intronic primer (f) in the antisense orientation (sequences are given in the Experimental Procedures). The outline of the experiment is shown in Fig.5 (upper panel). The sequences of amplified fragments from the patient's DNA are compared with the normal sequence of this genomic

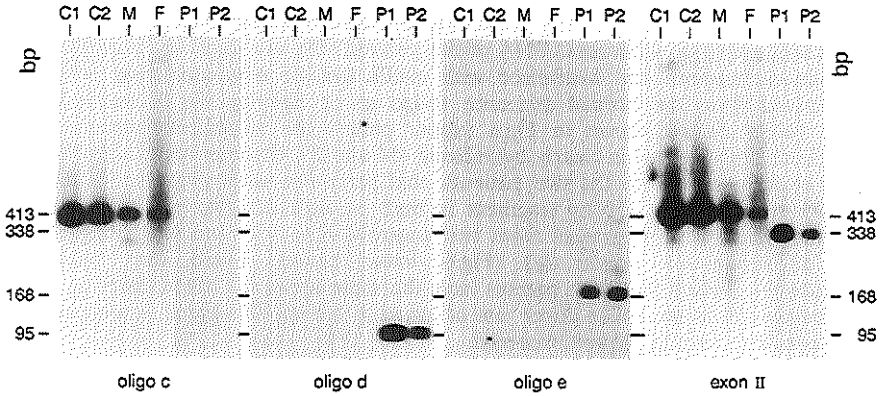


Fig.4. Southern blot analysis of PpuM1 restricted cDNA fragments.
The cDNA fragments, mentioned in Fig.3, were separated on agarose gels and subjected to Southern blot analyses using as probes either oligo (c) (specific for the normal transcript), oligo (d) (specific for transcript 1 and 2) or oligo (e) (specific for transcript 2). Furthermore a cDNA probe derived from exon II of the human ß-galactosidase gene was also tested. The length of the hybridizing cDNA fragments is given in base pairs (bp). Amplified samples from controls are indicated as C1/C2, from the two sibs as P1/P2 and from their parents by F/M.

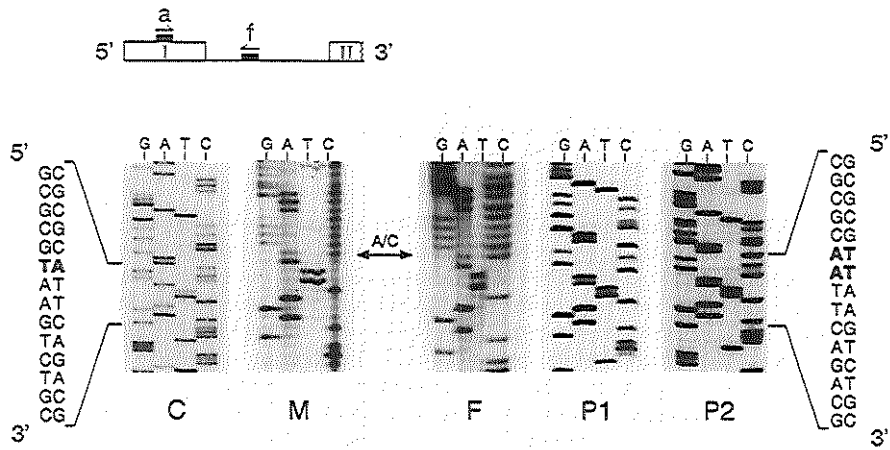


Fig.5. Partial nucleotide sequence of the ß-galactosidase genes from two GM1-gangliosidosis patients (P1, P2), their parents (M, F) and a control (C).
Genomic DNA was isolated and subjected to asymmetric PCR of the mutated genomic region using exon derived oligo (a) and intronic oligo (f), giving fragments of 150 nucleotides long. The antisense sequence is shown.

region and are shown in Fig.5 (lower panel). The results demonstrate that in genomic

DNA from both parents a double sequence starts right after the insertion of the extra T nucleotide, indicating that both parents carry the same mutation. The genomic sequences of the patients indeed confirm that they are homozygotes for the T nucleotide insertion. Finally, a 150 nucleotide genomic fragment encompassing the mutation was amplified from total DNA of the parents, sibs, and controls, transferred onto nylon membranes and hybridized with allelic specific oligonucleotide probes either derived from the normal sequence (oligo g) or carrying the T nucleotide insertion (oligo d). As shown in Fig.6 (upper panel) the amplified fragment in samples from both parents is clearly hybridizing (M and F), however with half of the intensity of fragments from the two patients, that carry the same mutation in both alleles (P1 and P2). The 150 bp fragment in both parent's samples hybridizes with half of the intensity of fragments from two control-DNAs (C1 and C2) using the normal oligonucleotide primer (g) (Fig.6, lower panel). These data further substantiate the finding that both parents have the same abnormal allele, carrying an extra T nucleotide at the splice donor site of intron 1. No specific signal is seen with oligonucleotide (g) in the patients material.

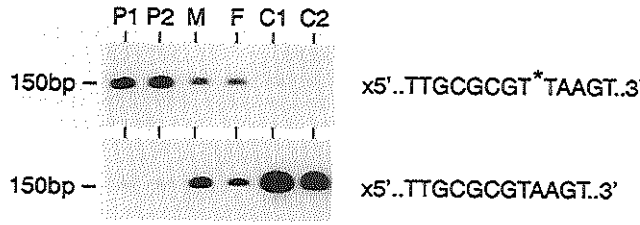


Fig.6. Hybridization analysis of β-galactosidase genomic fragments.
A 150 nucleotide genomic fragment spanning the mutation was amplified using oligo (a) and (f), from total DNA of two normal individuals (C1, C2), the patients (P1, P2) and their parents (M, F). The fragments were subjected to Southern blot analysis using oligonucleotide probes either specific for the normal (lower panel) or abnormal (upper panel) allele.

Discussion

We have shown that homozygosity for a genomic mutation causes a severe infantile form of G_{M1}-gangliosidosis in a family with two affected sibs. A single T nucleotide insertion, immediately after the conserved dinucleotide GT of the splice donor site of intron 1, is likely to interfere with the normal splicing process, leading to a preferential use of a more downstream cryptic splice site. In total RNA preparations from fibroblasts of both patients we have identified two aberrant transcripts, however we can not rule out the possibility that more alternatively spliced forms may exist.

The role of conserved structural elements in splicing of higher eukaryotic pre-mRNAs have been well documented (for a review see Breathnach and Chambon, 1981). These elements consists of a 5' splice donor site, a 3' splice acceptor site and a less conserved branch point sequence. Mutations at the GT/AG dinucleotides of 5' and 3' splice sites were shown to interfere with the normal splicing process both in site directed mutagenized genes and naturally occuring mutants (Padgett *et al.*, 1986). It has been empirically calculated that a T nucleotide in the third position of a

normal splice donor site is found only in 5% of the cases, whereas an A nucleotide is most commonly present (70%; Padgett *et al.*, 1986). Indeed an A is the base normally encountered in the third position of the ß-galactosidase splice donor site of intron 1 (Morreau *et al.*, 1991). At the moment it is unclear whether the A to T substitution at the third position of the donor site, due to the additional T nucleotide, is the cause of the abnormal splicing or it is the insertion per se. Eventually, both alterations might either modify the recognition site for the splicing machinery or interfere with the normal lariat formation of intron 1. It is noteworthy that the more downstream splice donor site of intron 1, that is recognized as splice donor site during splicing of the ß-galactosidase pre-mRNA of the patients, has the conventional 5'GTA 3' nucleotide triplet. Preferential usage of this donor site may cause the generation of a completely different lariat form and subsequent exon skipping during the splicing event. Such a mechanism might explain the presence of transcript 2 in these patients. We have shown on Northern blots that the amount of ß-galactosidase mRNA in patient's fibroblasts is severely reduced. Apparently the aberrantly spliced transcripts are highly unstable. Only from transcript 2 a ß-galactosidase protein could potentially be translated, missing the sequence encoded by exon II. However, this product can not be immunoprecipitated from fibroblast extracts of both patients, using anti-ß-galactosidase antibodies.

Several different mutations leading to the infantile type of G_{M1}-gangliosidosis in Japanese patients have been documented (Yoshiba *et al.*, 1991; Nishimoto *et al.*, 1991; Oshima *et al.*, 1991). In five of these patients only one of the alleles has been analyzed. Three of them carry point mutations resulting in a single amino acid changes, the other two have a duplication of both exon XI and XII that leads to a frameshift mutation. Only one patient was found to be homozygote for a single base substitution that creates a stop codon.

Acknowledgments

We would like to thank Professor H. Galjaard for continuous support, Dr Gerard Grosveld for the many experimental suggestions, Mirko Kuit for excellent graphic and photographic work. Furthermore we would like to thank Jeanette Lokker for editing the manuscript.

References

Breathnach, R., Chambon, P. (1981) *Annu. Rev. Biochem.* , **50**, 349-383

Conzelmann, E., Sandhoff, K. (1987) *Adv. Enzymol. Relat. Areas Mol. Biol.*, **60**, 89-216

D'Azzo, A., Hoogeveen, A., Reuser, A.J., Robinson, D., and Galjaard, H. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79**, 4535-4539

Feinberg, A.P., Vogelstein B., (1983) *Anal. Biochem.* **132**, 6-13

Groebe, H., Krins, M., Schmidberger, H., von Figura K., Harzer K., Kresse H., Paschke, A. Sewell, A., and Ullrich K. (1980) *Am. J. Hum. Genet.* **32**, 258-272

Gyllensten, U.B., Erlich, H.A. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 7652-7656

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

Morreau, H., Galjart N.J., Gillemans N., Willemsen R., van der Horst, G.J., and d'Azzo, A. (1989) *J. Biol. Chem.* **264**, 20655-20663

Morreau, H., Bonten, E., Zhou, X.Y., and D'Azzo, A. (1991) *Dna Cell. Biol.* **10**, 495-504

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

Nanba, E., Suzuki, K. (1990) *Biochem. Biophys. Res. Commun.* **173**, 141-148

Nanba, E., Suzuki, K. (1991) *Biochem. Biophys. Res. Commun.* **178**, 158-164

Nishimoto, J., Nanba, E., Inui, K., Okada, S., and Suzuki, K. (1991) *Am. J. Hum. Genet.***49**, 566-74

O'Brien, J.S., in *The Metabolic Basis of Inherited Disease* Scriver, A, Beaudet, A.L., Sly, W.S., Valle, D., Eds. (McGraw-Hill, New York, 1989), vol. 6, pp. 1797-1806

O'Brien, J.S., Gugler, E., Giedion, A., Wiessmann, U., Herschkowitz, N., Meier, C., and Leroy J. (1976) *Clin. Genet.* **9**, 495-504
Okada, S., O'Brien, J.S. (1968) *Science* **160**, 1002-1004
Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, and Sharp, P.A. (1986) *Annu. Rev. Biochem.* **55**, 1119-1150
Oshima, A., Tsuji, A., Nagao, Y., Sakuraba, H., and Suzuki, Y. (1988) *Biochem. Biophys. Res. Commun.* **157**, 238-244
Oshima, A., Yoshida, K., Shimmoto, M., Fukuhara, Y., Sakuraba, H., and Suzuki, Y. (1991) *Am. J. Hum. Genet.* **49**, 1091-1093
Paschke, E., Kresse, H. (1982) *Biochem. Biophys. Res. Commun.* **109**, 568-575
Saiki, R.K. , Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich H.A. (1988) *Science* **239**, 487-491
Sanger, F.G., Nicklen, S., and Coulsen, A. (1977) *Proc. Natl. Acad. Sci.* **74**, 5463-5467
Shows, T.B., Scrafford-Wolff, L.R., Brawn, J.A., and Meisler, M. (1979) *Somat. Cell Genet.* **5**, 147-158
van der Horst, G.T., Kleijer, W.J., Hoogeveen A.T., Huijmans, J.G., Blom, W., and van Diggelen O.P. (1983) *Am. J. Med. Genet.* **16**, 261-275
van der Horst, G.T., Galjart, N.J. d'Azzo, A., Galjaard, H., and Verheijen, F.W. (1989) *J. Biol. Chem.* **264**, 1317-1322
Verheijen, F., Brossmer, R., and Galjaard, H. (1982) *Biochem. Biophys. Res. Commun.* **108**, 868-875
Verheijen, F.W., Palmeri, S., Hoogeveen, A.T., and Galjaard, H. (1985) *Eur. J. Biochem.* **149**, 315-321
Wood, W.I., Gitschier, J., Lasky, L.A., and Lawn, R.M. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 1585-1588
Yamamoto, Y, Nishimura, K. (1987) *Int. J. Biochem.* **19**, 435-442
Yamamoto, Y., Hake, C.A., Martin, B.M., Kretz, K.A., Ahern-Rindell, A.J., Naylor, S.L., Mudd, M., and O'Brien J.S.(1990) *Dna Cell. Biol.* **9**, 119-127
Yoshida, K., Oshima, A., Shimmoto, M., Fukuhara, Y., Sakuraba, H., Yanagisawa, N., and Suzuki Y. (1991) *Am. J. Hum. Genet.* **49**, 435-442

Publication 4

J. Biol. Chem. **266** (1991), 14754-14762

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1991 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 266, No. 22, Issue of August 5, pp. 14754–14762, 1991
Printed in U.S.A.

Human Lysosomal Protective Protein Has Cathepsin A-like Activity Distinct from Its Protective Function*

(Received for publication, January 28, 1991)

Niels J. Galjart, Hans Morreau, Rob Willemsen, Nynke Gillemans, Erik J. Bonten, and Alessandra d'Azzo‡

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

The protective protein was first discovered because of its deficiency in the metabolic storage disorder galactosialidosis. It associates with lysosomal β-galactosidase and neuraminidase, toward which it exerts a protective function necessary for their stability and activity. Human and mouse protective proteins are homologous to yeast and plant serine carboxypeptidases. Here, we provide evidence that this protein has enzymatic activity similar to that of lysosomal cathepsin A: 1) overexpression of human and mouse protective proteins in COS-1 cells induces a 3–4-fold increase of cathepsin A-like activity; 2) this activity is reduced to ~1% in three galactosialidosis patients with different clinical phenotypes; 3) monospecific antibodies raised against human protective protein precipitate virtually all cathepsin A-like activity in normal human fibroblast extracts. Mutagenesis of the serine and histidine active site residues abolishes the enzymatic activity of the respective mutant protective proteins. These mutants, however, behave as the wild-type protein with regard to intracellular routing, processing, and secretion. In contrast, modification of the very conserved Cys^{90} residue interferes with the correct folding of the precursor polypeptide and, hence, its intracellular transport and processing. The secreted active site mutant precursors, endocytosed by galactosialidosis fibroblasts, restore β-galactosidase and neuraminidase activities as effectively as wild-type protective protein. These findings indicate that the catalytic activity and protective function of the protective protein are distinct.

Intralysosomal degradation is a composite process that is largely controlled by a battery of acidic hydrolases. The majority of these glycoproteins are synthesized on membrane-bound polysomes as high molecular weight precursors and routed to the lysosomes via a series of compartment-dependent posttranslational modifications. For the stepwise catabolism of different macromolecules to occur efficiently, a number of these hydrolases must work in concert and might, therefore, reside in a multienzymic complex. An example of such a complex could be the one consisting of lysosomal β-galactosidase (EC 3.2.1.23), *N*-acetyl-α-neuraminidase (sialidase, EC 3.2.1.18), and the protective protein (1–3). In human placenta (3), bovine testis (2), and porcine spleen and testis (4, 5) these three glycoproteins copurify through an affinity matrix for β-galactosidase.

The association of the protective protein with β-galactosidase and neuraminidase is essential for the stability and activity of these two glycosidases within the lysosomes (3, 6, 7). This is reflected by the existence of the metabolic storage disorder galactosialidosis (8), in which a primary defect of the protective protein results in a combined β-galactosidase/neuraminidase deficiency (1, 9). Among galactosialidosis patients distinct clinical phenotypes exist, ranging from severe early infantile forms in which visceromegaly with nephrotic syndrome, heart failure, and other abnormalities lead to early death or fetal hydrops, to milder late infantile and juvenile/adult variants (8, 10). Biochemical heterogeneity within these recognized phenotypes has also been observed (11, 12).

In human cultured fibroblasts the protective protein is synthesized as a precursor of 54 kDa which is proteolytically processed into a mature two-chain form of 32- and 20-kDa polypeptides linked together by disulfide bridges (1, 12). The predicted amino acid sequences of human as well as mouse protective proteins are homologous to yeast and plant serine carboxypeptidases (12, 13). Both protective proteins react with the serine protease inhibitor DFP,[1] but only in their mature state (13). Together these findings allowed us to predict a serine carboxypeptidase activity for the protective protein that is apparently synthesized and transported to the lysosomes as a zymogen. Some of its characteristics correlate well with those of a previously identified carboxypeptidase, cathepsin A (EC 3.4.16.1). This enzyme has been partially purified from different sources (14) and was shown to exist in small and large aggregate forms (15). In the native small aggregate, subunits with molecular masses of 20, 25 and 55 kDa are present, of which the 25-kDa polypeptide reacts with DFP (16). Besides its carboxypeptidase activity, optimal at acidic pH, cathepsin A can also function as a peptidyl aminoacylamidase (14, 17). Recently, a deamidase/carboxypeptidase purified from human platelets was shown to have sequence identity to the NH_2 termini of the protective protein chains (18). Enzymatic characterization of this deamidase with a variety of substrates and inhibitors also suggested a similarity to cathepsin A.

Here we provide direct evidence that the protective protein maintains cathepsin A-like activity. Galactosialidosis is therefore the first example of a lysosomal storage disorder associated with a protease deficiency. We also demonstrate by site-

* 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.

‡ To whom correspondence should be addressed: Dept. of Cell Biology and Genetics, Erasmus University, P. O. Box 1738, 3000 DR Rotterdam, The Netherlands. Tel.: 31-10-4087173-8140; Fax: 31-10-4087212.

[1] The abbreviations used are: DFP, diisopropylfluorophosphate; MES, 2-(*N*-morpholino)ethanesulfonic acid; Z, benzyloxycarbonyl; bp, basepair(s); G_{M1}, II^3NeuAc-GgOse^{4}Cer; anti-54, antibodies raised against recombinant human protective protein; anti-32, antibodies raised against the mature denatured human 32-kDa protective protein subunit; ER, endoplasmic reticulum.

directed mutagenesis of the human protective protein that its cathepsin A-like activity can be separated from its protective function toward β-galactosidase and neuraminidase.

EXPERIMENTAL PROCEDURES

Cell Culture—Human skin fibroblasts from normal individuals, patients with the early infantile (19) and juvenile/adult (20) forms of galactosialidosis, and a G_{M1}-gangliosidosis patient were obtained from the European Cell Bank, Rotterdam (Dr. W. J. Kleijer). Cells from the late infantile galactosialidosis patient (21) and both parents were provided by Dr. G. Andria, Dept. of Pediatrics, University of Napels, Italy. Fibroblasts were maintained in Dulbecco's modified Eagle's medium, Ham's F10 medium (1:1 v/v) supplemented with antibiotics and 10% fetal bovine serum. COS-1 cells (22) were grown in the same medium, supplemented with 5% fetal bovine serum.

Enzyme Assays—For enzyme activity assays and immunotitration experiments cells were harvested by trypsin treatment and homogenized in double-distilled water. When necessary, cell lysates were subsequently diluted in 20 mM sodium phosphate, pH 6.9, containing 1 mg/ml bovine serum albumin. Cathepsin A activity was measured in cell homogenates using a modification of the method of Taylor and Tappel (23). Briefly, 5 μl (2–10 μg of protein) of cell homogenates were incubated for 30 min at 37 °C in 100 μl of 50 mM MES, pH 5.5, 1 mM EDTA, in the absence or presence of 1.5 mM N-blocked dipeptides Z-Phe-Ala, Z-Phe-Leu, or Z-Glu-Tyr (Bachem). Reactions were stopped by addition of an equal volume of 10% (v/v) trichloroacetic acid. Precipitates were removed by centrifugation, and a fraction (5%) of the supernatants was taken to measure the concentration of released amino acid by the fluorimetric method outlined by Roth (24). The activities of β-galactosidase, neuraminidase, and β-hexosaminidase were measured with artificial 4-methylumbelliferyl substrates (25). Total protein concentrations were determined as described previously (26).

Antibodies and Immunotitration of Cathepsin A-like Activity—We have previously described the preparation of antibodies raised against a denatured form of the 32-kDa subunit of human protective protein (12). These "anti-32" antibodies recognize under reducing and denaturing conditions the 54-kDa precursor form as well as the 32-kDa mature component of the protective protein. To obtain a monospecific antiserum that immunoprecipitates human protective protein under native conditions, the latter was overexpressed in *Spodoptera frugiperda* (Sf9) insect cells, that had been infected with recombinant baculovirus containing human protective protein cDNA.[1] Protective protein was purified from the culture medium of infected cells using a concanavalin A-Sepharose column (Pharmacia), as described earlier (2). "Anti-54" antibodies were raised in rabbits against this purified protein preparation. An IgG fraction (2.2 mg of protein/ml) was prepared from anti-54 antiserum using a protein A-Sepharose column (Pharmacia).

Immunotitration of cathepsin A with anti-54 antibodies was performed essentially as described before (3). Formalin-fixed *Staphylococcus aureus* cells (Immunoprecipitin, Bethesda Research Laboratories) were added to the samples to remove antigen-antibody complexes.

Isolation of cDNA Clones and DNA Sequence Analysis—A chicken embryo λgt11 cDNA library (Clontech, Palo Alto, CA) (27), consisting of 1×10^6 independent clones, was plated out as described before (13) and screened using the heterologous human protective protein cDNA, Hu54, as a probe (12, 28). The longest cDNA insert was subcloned into pTZ18 and 19 (Pharmacia) (29) and sequenced on both strands (30, 31). Comparison to the human sequence showed that the chicken cDNA lacks the ATG start codon and part of the signal peptide. Sequence data were analyzed with the programs of the University of Wisconsin Genetics Computer Group (32). Protein alignments were also done with the latter software package.

Plasmid Constructs—*In vitro* mutagenesis of human protective protein cDNA was carried out using the method described by Higuchi *et al.* (33). Polymerase chain reaction-amplified DNA fragments, containing the desired mutations giving rise to single amino acid substitutions, were introduced in the normal human cDNA by suitable restriction enzyme sites. Using the same procedure the deletion construct 32(Δ20) was generated by introducing a stop codon in one of the amplification primers. The stop codon follows immediately the Arg^{277} residue. A 365-bp *Bam*HI fragment, with the point mutation that gives rise to the Ser^{150} to Ala^{150} amino acid change, was subsequently substituted for the wild-type fragment into the 32(Δ20) construct, using standard cloning procedures (29). This resulted in the 32SA(Δ20). The 20(Δ32) construct encodes the human 20-kDa subunit tagged with the signal sequence (residues 299 to 452 and −28 to +1 in the human protein, respectively). The cDNA stretches encoding these two parts of the protective protein preproform were amplified by polymerase chain reaction and afterwards ligated together using an *Nco*I restriction enzyme site introduced in two of the amplification primers. This site does not alter the amino acid sequences of the 20-kDa subunit or signal peptide. All DNA fragments resulting from polymerase chain reaction amplification were verified by sequencing as described above. The oligonucleotides needed for site-directed mutagenesis were synthesized on an Applied Biosystems 381A oligonucleotide synthesizer.

Human/chicken cDNA, HCh1, was made by exchange of 5′-end chicken with human cDNA sequences at a conserved *Pst*I restriction enzyme site. All constructs were cloned into a derivative of the mammalian expression vector pCD-X (34) as described previously (12).

Transfection in COS-1 Cells—COS-1 cells were seeded out in 30-mm dishes 1–2 days prior to transfection and grown to 30% confluency. Transfection in COS-1 cells, metabolic labeling of transfected cells, and preparation of cell extracts and media were carried out as described before (12, 35). Cells were labeled with [^{3}H]leucine (143 Ci/mmol, Amersham Corp.). Immunoprecipitation of radiolabeled proteins from cell lysates and media was performed using anti-32 or anti-54 antibodies, as reported earlier (35). Radioactive proteins were resolved on 12.5% polyacrylamide gels under reducing and denaturing conditions and visualized by fluorography of gels impregnated with Amplify (Amersham Corp.). For the DFP-binding assay and direct measurement of cathepsin A activity, COS-1 cells were transfected with various pCD constructs and maintained afterwards for 72 h in normal culture medium. Subsequently, cells were harvested by treatment with trypsin. Cell lysates were either incubated with [^{3}H]DFP (Du Pont-New England Nuclear, 3.0 Ci/mmol) (13) or used as such for detection of cathepsin A activity as described above.

Uptake Studies in Human Fibroblasts—COS-1 cell-derived protective protein precursors were obtained from the medium of unlabeled COS-1 cells, transfected in 100-mm Petri dishes. Media were concentrated as described previously (35) and half of the concentrated material was added to the medium of recipient early infantile galactosialidosis fibroblasts (12). After 5 days of uptake the medium was replaced with fresh medium containing the other half of concentrated material. 2 days later cells were harvested by trypsin treatment, and cell lysates were partly used for enzyme activity assays. The remainder of these homogenates was diluted 7-fold in 10 mM sodium phosphate buffer, pH 6.0, containing 100 mM NaCl and 1 mg/ml bovine serum albumin. After centrifugation to remove insoluble material the cell lysates were divided into three aliquots of 25 μl each and incubated for 1.5 h with 1.5 μl of preimmune serum, anti-54 antibodies, or antinative human β-galactosidase antibodies. Immunoprecipitin, extensively washed in the aforementioned buffer, was subsequently added to the samples, and after 30 min antibody-antigen complexes were removed by centrifugation. All steps were performed on ice or at 4 °C. The supernatants were assayed for β-galactosidase activity.

Indirect Immunofluorescence—COS-1 cells, transfected with selected pCD constructs, were treated mildly with trypsin 48 h after transfection and subsequently reseeded at low density on coverslips. 16 h later cells were fixed and incubated with anti-32 antibodies and in a second step with goat anti-(rabbit IgG) conjugated with fluorescein (36).

Immunoelectron Microscopy—Transfected COS-1 cells were fixed in 0.1 M phosphate buffer, pH 7.3, containing 1% acrolein and 0.4% glutaraldehyde. Further embedding in gelatin, preparation for ultracryotomy, and the methods for immunoelectron microscopy were as reported earlier (37).

RESULTS

Evidence That the Protective Protein Is Similar to Cathepsin A—We first ascertained whether the protective protein maintains carboxypeptidase activity aside from its protective function. The choice of the synthetic substrate to use in the assay was dictated by the similarity of the protein to cathepsin A (14–18). The latter hydrolyzes preferentially at acidic pH acylated dipeptides having a hydrophobic residue in the pe-

[1] E. J. Bonten and A. d'Azzo, manuscript in preparation.

nultimate (P1) position (38). Of these N-blocked dipeptides Z-Phe-Ala was reported to be the most specific substrate for cathepsin A (38).

In total cell homogenates from human cultured fibroblasts we have measured the hydrolysis of Z-Phe-Ala as well as Z-Glu-Tyr and Z-Phe-Leu. The rate of hydrolysis is maximal for Z-Phe-Ala (Table I), 3-fold lower for Z-Phe-Leu (normal fibroblasts: 1, 163 milliunits/mg protein; 2, 93 milliunits/mg protein) and barely detectable for Z-Glu-Tyr (not shown). In order to prove that the protective protein is the enzyme responsible for the cleavage of Z-Phe-Ala, we raised monospecific polyclonal antibodies in rabbits against a human native protective protein preparation (anti-54 antibodies). As shown in Fig. 1, virtually all carboxypeptidase activity toward this substrate is precipitated at increasing antibody concentrations. Since the purified preparation used for immunization of the rabbits was obtained from the culture medium of Sf9 insect cells infected with a recombinant baculovirus expression vector (39),[2] it is unlikely that proteins of human origin, other than the protective protein, are directly precipitated by the antibodies. From these results we conclude that lysosomal protective protein has a substrate specificity overlapping with that of cathepsin A. We have also tested whether β-galactosidase activity is coprecipitated with cathepsin A by virtue of the association of these two proteins. Indeed, about one-third of total β-galactosidase activity is brought down at maximal antibody concentration. The values for β-hexosaminidase, measured in the fibroblast homogenates as a reference enzyme, remained unchanged throughout the experiment.

TABLE I

Cathepsin A-like activity in normal and mutant human fibroblasts

Lysates of different human cultured fibroblasts were incubated for 30 min at 37 °C in 50 mM MES, pH 5.5, containing 1.5 mM Z-Phe-Ala. Cathepsin A-like activity was determined by indirect fluorimetric quantitation of liberated alanine.

Cell strain		Cathepsin A-like activity
		milliunits/mg protein*
Normal fibroblasts	1	439
	2	266
G_{M1} gangliosidosis		407
Early infantile galactosialidosis		1.3
Late infantile galactosialidosis		3.0
Juvenile/adult galactosialidosis		4.0
Parents late infantile patient	M	139
	F	117

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

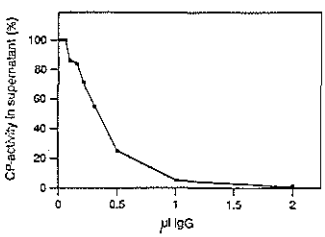


FIG. 1. **Immunotitration of cathepsin A-like activity.** Increasing amounts of an IgG antibody fraction, raised against human native protective protein precursor, were added to a cell extract of normal human fibroblasts. Antibody-antigen complexes were removed by addition of *S. aureus* cells and the remaining carboxypeptidase (*CP*) activity toward the acylated dipeptide Z-Phe-Ala was measured in the supernatant.

Carboxypeptidase Deficiency in Galactosialidosis Fibroblasts—The same dipeptide, Z-Phe-Ala, was used as substrate to measure cathepsin A-like activity in fibroblast homogenates from normal individuals, a G_{M1}- gangliosidosis patient with an isolated β-galactosidase deficiency, different galactosialidosis patients, and carriers (Table I). In contrast to normal as well as G_{M1}-gangliosidosis fibroblasts the galactosialidosis cell strains tested have minute activities toward the substrate. Clear heterozygote values are measured in the carrier samples. The normal hydrolysis of Z-Phe-Ala measured in G_{M1}-gangliosidosis cell extract indicates that an isolated β-galactosidase deficiency does not influence the carboxypeptidase activity of the protective protein.

Analysis of Conserved Domains in Protective Proteins of Different Species—Amino acid sequence comparison with other well defined serine carboxypeptidases (40–42) revealed that the protective protein/cathepsin A is a member of this family of enzymes (12, 13). Similarly, comparison of the primary structures of protective proteins from different species could disclose domains in the human protein important for its association with β-galactosidase/neuraminidase and, hence, for its protective function. However, the previously characterized mouse protective protein appeared to be almost identical to its human counterpart (13). We therefore isolated the cDNA encoding chicken protective protein. Its predicted amino acid sequence is shown in Fig. 2, aligned with those of the human and mouse proteins. The chicken sequence lacks the first methionine residue and part of the signal peptide.

Identity between the different proteins is 67% (chicken/human), 66% (chicken/mouse), and 87% (mouse/human). The serine, histidine, and aspartic acid residues that are known to form the catalytic triad of serine carboxypeptidases (43) are found in the chicken protective protein/cathepsin A at positions 150, 431, and 375, respectively. Ser^{150} and His^{431} are included in two of the three highly conserved regions (*boxed* in Fig. 2) in this family of enzymes. Remarkably, however, the chicken enzyme has a glycine for alanine substitution at position 152 that occurs within the Gly-Glu-Ser-Tyr-Ala-Gly domain, containing the active site serine. All three protective proteins have 9 conserved cysteine residues, probably crucial for their tertiary structure as well as function. Both chicken and mouse homologues have two additional cysteines, one on each subunit of their respective two-chain forms, but at different positions.

Additional essential residues and domains emerge from the sequence alignment. Amino acids surrounding the two proteolytic cleavage sites (Fig. 2, *vertical arrows*) are largely identical. An internal repeating motif (*underlined* in Fig. 2), characterized by 2 recurring Trp residues 16 amino acids apart, is present in each subunit of all three protective proteins, suggesting an ancient intragenic duplication. Notably, this "repeat" within the 32-kDa polypeptide includes the 10-amino acid domain (residues 53–62, *boxed* in Fig. 2), conserved in all serine carboxypeptidases. Four potential *N*-linked glycosylation sites are found in the chicken protective protein (Fig. 2, *hatched boxes*), two of which are in identical positions in the three sequences.

Mutagenesis of Human and Chicken Protective Proteins—To investigate whether the cathepsin A-like activity of the protective protein is essential for the activation and stabilization of β-galactosidase and neuraminidase we used a genetic approach. As summarized in Fig. 3, mutant and hybrid protective proteins were obtained by site-directed mutagenesis of wild-type cDNAs, encoding either the human or chicken forms. The first series of mutants (Fig. 3, *upper bar*) carried single amino acid substitutions in the human protein. Of the

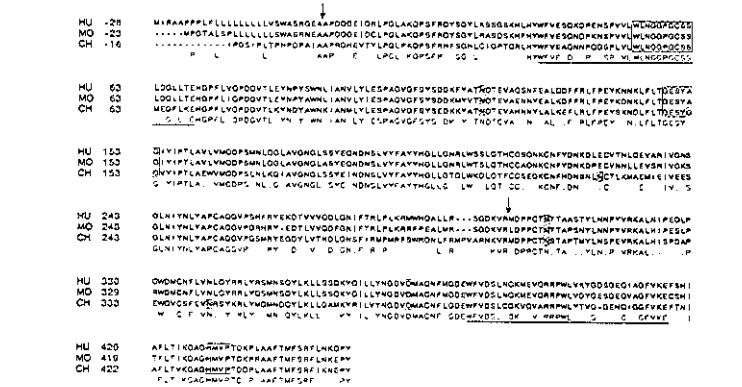


FIG. 2. **Alignment of predicted amino acid sequences of human, mouse, and chicken protective proteins.** Amino acid sequences of previously characterized human (*HU*) and mouse (*MO*) protective proteins are shown aligned with the chicken (*CH*) homologue. Residues on the *fourth line* (consensus sequence) denote identity at those positions in the three protective proteins. The three domains, characteristic of serine carboxypeptidases, and the aspartic acid residue, conserved throughout this family of enzymes, are *boxed*. Potential *N*-linked glycosylation sites are indicated with *hatched boxes*. *Vertical arrows* denote proteolytic cleavage sites. The internal repeating motif in 32- and 20-kDa subunits is *underlined*. *Numbers* on the *left* refer to positions of the amino acids within the sequences.

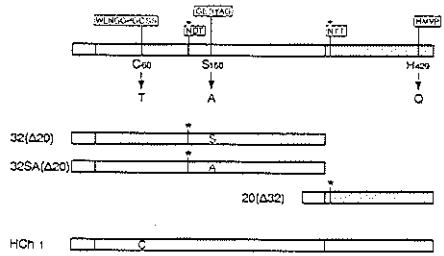


FIG. 3. *In vitro* **mutagenized and hybrid protective proteins.** The preprofrom of human protective protein is represented by the *upper bar*. *Hatched part*, signal sequence; *stippled part*, 32-kDa subunit; *cross-hatched part*, 20-kDa subunit. Single amino acid substitutions are indicated. The *three middle bars* represent the deletion mutants. In the 32SA(Δ20) mutant the Ser^{150} to Ala^{150} substitution is also present. The 20(Δ32) mutant is tagged with the human signal sequence (*hatched part*). The *lower bar* represents the human/chicken hybrid protein, having the signal sequence (*hatched part*) and the most NH_2-terminal 60 amino acids (*stippled part*) of the human protective protein.

catalytic triad, Ser^{150} was replaced by alanine and His^{429} was mutated into glutamine, in order to abolish cathepsin A-like activity. The third point mutation was introduced at residue Cys^{60} to study the effect of this alteration on the correct folding, transport, and processing of the protective protein. The mutants 32(Δ20) and 20(Δ32) were deleted of either of the two subunits in order to check their reciprocal influence within the two-chain form on the cathepsin A and/or protective activities. They encode truncated 32- and 20-kDa polypeptides, respectively. The latter was tagged with the human signal sequence to allow its translocation into the endoplasmic reticulum (ER). The 32SA(Δ20) mutant encodes a truncated 32-kDa protein carrying the Ser^{150} to Ala^{150} amino acid substitution (Fig. 3, *middle bars*).

The human-chicken hybrid construct, HCh1 (Fig. 3, *lower bar*), was made to identify determinants on the human protective protein, important for its association to β-galactosidase/neuraminidase. Furthermore, the replacement of chicken with human 5′ end cDNA sequences provided HCh1 with the correct translation initiation codon. As a result, the NH_2 terminus of the HCh1 hybrid precursor contains 60 amino acids of human origin.

Transient Expression of Mutant and Hybrid Protective Proteins in COS-1 Cells—To follow the intracellular transport and processing of mutant and hybrid protective proteins, their corresponding cDNAs were subcloned into a derivative of the expression vector pCD-X and transfected into COS-1 cells. Human and mouse protective protein constructs (pCDHu54 and pCDMo54) were included in the experiments as controls. Transiently expressed proteins were detected 2 days after transfection by metabolic labeling with [^{3}H]leucine followed by immunoprecipitation from cell lysates and media (Fig. 4). Immunoprecipitations were carried out using two different antibodies: those raised against the human denatured 32-kDa polypeptide (anti-32 antibodies), that recognize under reducing conditions the 54-kDa precursor as well as the mature 32-kDa component, or the anti-54 antibodies.

Fig. 4 shows that singly transfected pCD constructs direct the synthesis of mutated or hybrid protective proteins that are stable under the conditions used. The SA^{150} and HQ^{429} mutant precursors behave as wild-type protective protein in that they undergo normal proteolytic processing and the unprocessed precursors are secreted into the medium (Fig. 4, *lanes 1*, *3*, and *4*). In contrast, the cysteine to threonine substitution apparently interferes with the maturation and secretion of the precursor molecule (Fig. 4, *lane 2*). However, it is also possible that an aberrantly folded mature mutant protein evades recognition by the anti-32 antibodies.

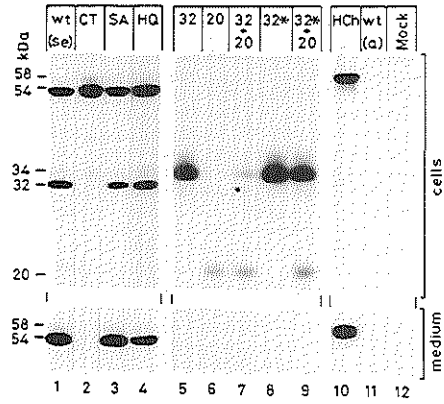


FIG. 4. **Transient expression of normal human, mutant, and hybrid protective proteins in COS-1 cells.** COS-1 cells were transfected with various pCD constructs: *wt(se)* and *(a)* represent human protective protein cDNA in the sense and antisense orientation, respectively; *CT*, *SA*, and *HQ* are the single amino acid substitutions; *32* and *20* represent the 32(Δ20) and 20(Δ32) deletion mutants, and *HCh* is the human-chicken hybrid protective protein. All abbreviations are explained in Fig. 3. 2 days after transfection newly synthesized proteins were labeled with [^{3}H]leucine for an additional 10 h. Labeled proteins were immunoprecipitated from cells and media using anti-32 antibodies (*lanes 1–4* and *10–12*) or anti-54 antibodies (*lanes 5–9*). Proteins were separated by gelelectrophoresis under reducing and denaturing conditions and visualized by fluorography. Molecular sizes of precursors, mature subunits, and truncated polypeptides are indicated at *left*. For the medium samples only the 54-kDa part of the gel is shown since the uncleaved 54-kDa precursor is the only form of the protective protein detected in the medium of transfected COS-1 cells. Exposure time for *lanes 1–4* and *10–12* was 2 days; for *lanes 5–9*, 3 days.

The 32(Δ20) deletion mutant construct encodes a 34-kDa polypeptide that is ~2 kDa larger than the corresponding wild-type subunit (Fig. 4, *lane 5*). A possible explanation for this size difference is that additional carboxyl-terminal processing of the latter takes place, an event that is impaired in the mutant 32(Δ20). This processing step may normally occur after endoproteolytic cleavage of the 54-kDa precursor either in an endosomal or lysosomal compartment. In addition, altered glycosylation of the 32(Δ20) mutant in comparison with the wild-type subunit may also contribute to this observed difference. The 20(Δ32) mutant is very similar in size to the 20-kDa component of the mature protective protein (Fig. 4, *lane 6*). Neither of the two independently synthesized polypeptides are secreted into the medium. To test their capacity to associate and to analyze the influence of this event on their intracellular transport and secretion, COS-1 cells were cotransfected with both pCD32(Δ20) and pCD20(Δ32). Binding of the two truncated proteins was proven by their coprecipitation with monospecific antibodies against either the 32- or the 20-kDa denatured subunit (not shown). Surprisingly, however, their interaction seems to cause a severe reduction in the amount of immunoprecipitable 34-kDa polypeptide, as compared to the single transfections (Fig. 4, *lanes 5–7*). We envisaged that formation of an "active" two-chain cathepsin A soon after synthesis could underly this effect. This hypothesis was supported by the observation that assembly of a truncated 34-kDa polypeptide, carrying the Ser^{174} to Ala^{176} active site mutation, with the 20-kDa polypeptide does not lead to reduced immunoprecipitable material (Fig. 4, *lanes 8* and *9*). The results further indicate that association of the different subunits does not induce their secretion.

The HCh1 hybrid precursor is about 4 kDa larger than the human proform (Fig. 4, *lanes 1* and *10*). This is due to the presence of two extra sugar chains in the chicken protective protein, since tunicamycin treatment prior to and during labeling leads to the synthesis of precursor molecules of identical size (not shown). The hybrid precursor is secreted into the culture medium, but no mature form can be precipitated intracellulary (Fig. 4, *lane 10*). A likely explanation is that proteolytic processing to the mature hybrid two-chain protein does occur but this form is not brought down by the antibodies under the experimental conditions used.

Localization of Mutant and Hybrid Protective Proteins—Given the differential behavior of mutant and hybrid protective proteins in transfected COS-1 cells, we analyzed their subcellular distribution by indirect immunofluorescence and immunogold labeling techniques. At light microscopy a typical lysosomal labeling pattern and a diffuse staining of the perinuclear region are observed in cells expressing the wild-type human, the HCh1 hybrid, and the SA^{150} mutant protective proteins (Fig. 5, *A–C*). The HQ^{420} mutant protein behaves similarly (not shown). In contrast, the CT^{40} precursor as well as the deletion mutants all seem to accumulate in the perinuclear region (Fig. 5, *D–F*).

For a more refined localization, ultrathin sections of transfected cells were probed with anti-32 antibodies followed by

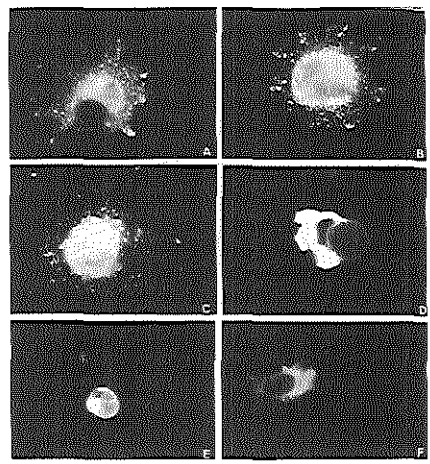


FIG. 5. **Immunocytochemical localization of normal, mutant, and hybrid protective proteins in transfected COS-1 cells.** COS-1 cells were treated with trypsin 48 h after transfection and reseeded on coverslips. 16 h later, cells were fixed and incubated with anti-32 antibodies. The intracellular distribution of normal human protective protein (*A*), HCh1 hybrid (*B*), and mutants SA^{150} (*C*), CT^{40} (*D*), 32(Δ20) (*E*), and 32(Δ20)/20(Δ32) (*F*) is shown. Abbreviations used to define mutant and hybrid protective proteins are explained in Fig. 3. Magnification: *A*, *B*, *C*, 16500×; *D*, *E*, *F*, 2000×.

goat anti-(rabbit IgG)-gold and analyzed by electron microscopy. As shown in Fig. 6, overexpressed wild-type protective protein is compartmentalized in lysosomes and is detected in large amounts in the Golgi complex and rough ER (Fig. 6, *A* and *B*). A similar pattern is seen in cells transfected with the HCh1 hybrid protein, the SA^{150} mutant (Fig. 6, *C* and *D*) and the HQ^{429} protein (not shown). Immunogold labeling is restricted exclusively to the rough ER in cells expressing either the CT^{60} mutant or the truncated subunits (not shown). COS-1 cells transfected with an antisense cDNA were used to estimate background labeling due to endogenous protective protein. The number of gold particles in lysosomes, Golgi complex, and rough ER was on average one or two. All together these results identify two types of modified protective proteins: those whose intracellular transport and processing overlap with wild-type protective protein (SA^{150}, HQ^{429}, HCh1) and those that accumulate in the ER and are neither processed nor secreted (CT^{60}, deletion mutants).

Protective Protein Active Site Mutants Lack Cathepsin A-like Activity—The similar characteristics observed thus far for the active site mutants with respect to wild-type protective protein imply that their tertiary structures are not grossly modified by the amino acid substitutions. We next ascertained whether cathepsin A-like activity was measurable in cells expressing these two mutant proteins compared to cells transfected with Hu54, Mo54, HCh1, and 32(Δ20) constructs. Two independent assays were used. First, COS-1 cell extracts were incubated with [^{3}H]DFP, followed by immunoprecipitation with anti-54 antibodies (Fig. 7, *upper panel*). As we have shown before, human and mouse protective proteins are able to react with the inhibitor after proteolytic cleavage of their zymogens (Fig. 7, *lanes 1* and *2*). Only the large subunit, carrying the serine active site, is detectable. The mouse form is slightly bigger in size and reacts poorly with the antibodies. In contrast, neither the SA^{150}, containing a modified active site serine, nor the HQ^{429} mutants show any binding capacity (Fig. 7, *lanes 4* and *5*). Likewise, the 32(Δ20) deletion mutant, missing the 20-kDa subunit, does not react with the inhibitor (Fig. 7, *lane 6*). Mature HCh1 hybrid molecules, if present, again are not immunoprecipitable (Fig. 7, *lane 3*). Cells transfected with an antisense wild-type construct were included in the experiment as estimate of the level of endogenous COS-1

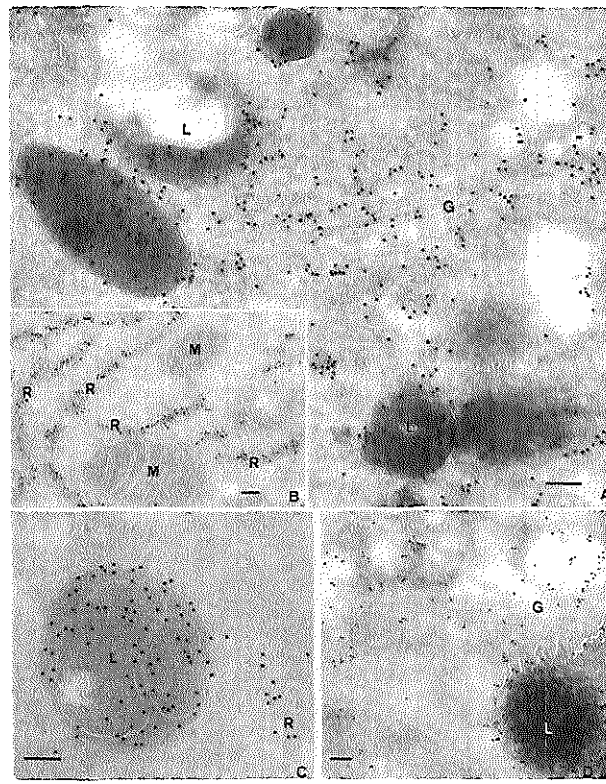


FIG. 6. **Cryosections of COS-1 cells, transfected with normal human (*A* and *B*), HCh1 hybrid (*C*) or SA^{150} mutant (*D*) protective proteins and labeled with anti-32 antibodies and goat anti-(rabbit IgG)-gold.** *A*, shows extensive labeling of the Golgi complex (*G*) and lysosomes (*L*), in cells expressing the normal human protein. A low magnification of the perinuclear region is shown in *B* with extensive labeling of rough endoplasmic reticulum structures (*R*), but not of a mitochondrion (*M*). An identical labeling pattern is observed in cells expressing the HCh1 hybrid or SA^{150} mutant protective proteins. Correct lysosomal targeting of these modified proteins is shown in (*C*) and (*D*). *Bars*, 0.1 μm.

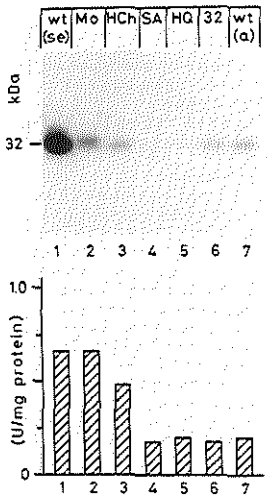


FIG. 7. [^{3}H]DFP labeling and cathepsin A activity of normal, mutant, and hybrid protective proteins. Cell extracts of COS-1 cells, transfected with selected constructs, were incubated with [^{3}H]DFP, followed by immunoprecipitation using anti-54 antibodies (*upper panel*). Molecular size is indicated at *left*. Exposure time was 6 days. The pCD constructs used are abbreviated as before with the addition of pCDMo54 sense (*Mo*). The same set of constructs was transfected separately into COS-1 cells to directly test cathepsin A activity in the different cell extracts, using Z-Phe-Ala as substrate (*lower panel*). *Hatched vertical bars* represent rates of hydrolysis. One unit is defined as the enzyme activity that releases 1 μmol of alanine per min.

protective protein (Fig. 7, *lane 7*). Since the SA^{150} and HQ^{429} mutants resemble most the wild-type protective protein, it is conceivable that all antibodies are efficiently competed out by unlabeled molecules only in these cell extracts. This explains the lack of signal in *lanes 4* and *5* compared to *lane 7*.

The results obtained with DFP inhibitor are well supported by direct measurements of Z-Phe-Ala hydrolysis in lysates of cells transfected independently with the same set of constructs (Fig. 7, *lower panel*). Equal increase in cathepsin A-like activity above endogenous COS-1 levels is measured in cells expressing the wild-type human and mouse protective proteins. These results demonstrate directly that both proteins have cathepsin A-like activity and that only a small proportion of the mouse protective protein is immunoprecipitated. The SA^{150} and HQ^{429} mutants are completely inactive, as is the case for the deletion mutant (Fig. 7) and the CT^{60} mutant (not shown). The ~2.5-fold increase in activity, detected in HCh1-expressing cells, confirms that this hybrid protein must be present in its mature two-chain form. Surprisingly, however, it has an altered substrate specificity, as compared to human and mouse wild-type proteins, since it preferentially cleaves Z-Phe-Leu over Z-Phe-Ala (not shown). This effect could be due to the glycine for alanine substitution at position 152 in the chicken protective protein.

Dual Function of the Protective Protein—The two active site mutations have been shown by different criteria to abolish the carboxypeptidase activity of the protective protein without disturbing its tertiary structure. Therefore, these mutants are excellent candidates to test, in uptake studies, whether loss of cathepsin A-like activity influences the protective function. For this purpose secreted modified precursors from transfected COS-1 cells were added to the medium of early infantile galactosialidosis fibroblasts deficient in protective protein mRNA. Similarly, the secreted HCh1 hybrid as well as mouse precursor proteins were also tested. After uptake, cells were harvested and β-galactosidase and neuraminidase activities measured. As shown in Table II, the SA^{150} and HQ^{429} mutant precursors, endocytosed and processed by the deficient cells, restore β-galactosidase and neuraminidase activities as efficiently as the wild-type protein. Thus, the protective protein has catalytic activity clearly distinct from its protective function. Surprisingly, the HCh1 hybrid molecule has maintained the capacity to bind and activate β-galactosidase and neuraminidase, although the overall identity between chicken and human sequences is only 67%. It seems that the modified active site domain (Gly-Glu-Ser-Tyr-Gly-Gly) and altered substrate specificity have no influence on the protective function of HCh1. As observed earlier (13), the mouse homologue, 87% similar to human protective protein, adopts a configuration not entirely suitable for the activation of human neuraminidase.

Confirmation that the SA^{150} and HQ^{429} mutants exert their protective function via physical association with β-galactosidase was obtained by examining the coprecipitation of this enzyme with different endocytosed protective proteins. Galactosialidosis cell lysates used to measure correction of β-galactosidase/neuraminidase activities were incubated with anti-54 antibodies and, as control, with preimmune serum or anti-native human β-galactosidase antibodies. Fig. 8 shows that 22, 25, and 34% of β-galactosidase activity is coprecipitated with anti-54 antibodies in cells that have taken up the wild-type human, the SA^{150}, and the HQ^{429} mutant protective proteins, respectively. A comparable percentage of activity is coprecipitated in a normal human fibroblast homogenate. These results demonstrate that a proportion of active β-galactosidase is indeed associated with the SA^{150} and HQ^{429} mutant proteins, and that the enzyme has equal affinity for the active site mutants and wild-type protective protein. Furthermore, the values are specific since β-galactosidase activity is either not at all or to a lesser extent coprecipitated in cells treated with the HCh1 hybrid or mouse protective proteins.

TABLE II

Restoration of β-galactosidase and neuraminidase activities in galactosialidosis fibroblasts after uptake of various COS-1 cell-derived protective protein precursors

Equivalent amounts of secreted precursor proteins from the medium of transfected COS-1 cells were added to the culture medium of early infantile galactosialidosis fibroblasts. After 7 days of uptake, cells were harvested and enzyme activities measured using 4-methylumbelliferyl substrates. The results are representative of experiments carried out several times.

Cell strain	Addition of protective proteins	Activity	
		β-Galactosidase	Neuraminidase
		milliunits/mg protein*	*microunits/mg protein*
Early infantile galactosialidosis	wt (se)	5.18	550
	SA	5.18	533
	HQ	4.83	700
	Mo54	4.17	80
	HCh1	5.27	500
	wt (a)	0.58	12
		0.72	5.1
Normal fibroblasts		7.33	900

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

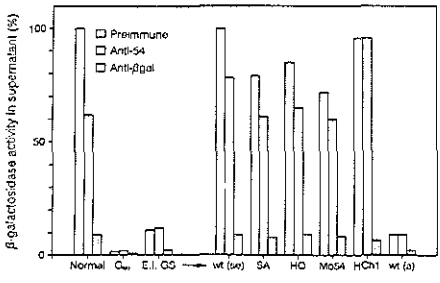


FIG. 8. **Precipitation of β-galactosidase activity in galactosialidosis fibroblasts after uptake of various COS-1 cell-derived protective protein precursors.** Early infantile galactosialidosis (*E.I.GS*) cell homogenates from the uptake experiment described in Table II were each divided in three aliquots and incubated with either preimmune serum, the anti-54 antibodies, or anti-native human β-galactosidase antibodies. As controls, normal human fibroblasts and cells from a G_{M1}-gangliosidosis patient were treated in the same manner. After precipitation of antibody-antigen complexes, the remaining β-galactosidase activity was measured in the supernatants (*vertical bars*). The value obtained in preimmune serum treated cells after endocytosis of wild-type human protective protein is taken as 100% activity. Except for normal and G_{M1}-gangliosidosis fibroblasts, all β-galactosidase activities are expressed as a percentage of this value. Cell homogenates to the *right* of the *arrow* represent E.I.GS fibroblasts that have taken up the different COS-1 cell-derived protective protein precursors (abbreviated as in Fig. 4).

respectively. This, in turn, is not surprising if we take into account the lower affinity of anti-54 antibodies for chicken and mouse mature protective proteins. On the other hand, a complex formed by human β-galactosidase and protective proteins from other species may be more susceptible to dissociation. In human fibroblasts neuraminidase is inactivated upon dilution or freeze/thawing, therefore its coprecipitation with the protective protein could not be examined.

DISCUSSION

The primary structure of human protective protein has suggested a putative carboxypeptidase activity by virtue of its homology with yeast carboxypeptidase Y and the KEX1 gene product. We now directly demonstrate that the protective protein at acid pH cleaves the acylated dipeptides Z-Phe-Ala, Z-Phe-Phe, and Z-Phe-Leu with clear preference for the first named substrate. This chymotrypsin-like activity closely resembles that of lysosomal cathepsin A (14). Several lines of evidence confirm this similarity: 1) monospecific antibodies against native human protective protein precursor precipitate virtually all carboxypeptidase activity toward Z-Phe-Ala; 2) overexpression of protective protein in COS-1 cells leads to increased cathepsin A-like activity; 3) cells from a galactosialidosis patient deficient in protective protein mRNA have less than 1% residual cathepsin A-like activity.

Considering the highly specific binding of the protein to lysosomal β-galactosidase and neuraminidase, it was reasonable to assume that terminal processing of these two enzymes would be the principal role of the carboxypeptidase. Our genetic analysis, however, provides evidence that the catalytic and protective functions of the protective protein are distinct, since loss of its cathepsin A activity does not influence its ability to stabilize and activate the other two enzymes. These separable functions could relate to the existence of free and associated pools of protective protein and β-galactosidase in human tissues. A number of indications support this notion. Preliminary studies by gel filtration suggest that precursor and mature protective protein/cathepsin A can form homodimers of ~95 kDa free of β-galactosidase/neuraminidase. Conversely, the immunotitration experiments presented here have shown that not all β-galactosidase activity is coprecipitated with cathepsin A using anti-54 antibodies. Earlier data agree with these results since a fraction of β-galactosidase was found unassociated in crude glycoprotein preparations of human placenta (3).

The reason for maintenance of these different pools of enzymes could be the need to catabolize a broad spectrum of substrates in different metabolic pathways. A recent report by Jackman *et al.* (18) emphasizes this hypothesis. These authors, in an effort to characterize a deamidase released from human platelets, came to the unexpected finding that their purified enzyme is probably identical to the protective protein. They further demonstrate that *in vitro* this platelet enzyme has deamidase as well as carboxypeptidase activity on biologically important peptides, like substance P, bradykinin, angiotensin I, and oxytocin. The deamidase activity is optimal at neutral pH, whereas the carboxypeptidase works best at pH 5.5. The purified two-chain enzyme forms homodimers of 95 kDa at this pH (18). In view of the characteristics of the enzyme, they also came to the conclusion that it is similar to cathepsin A. We can deduce from our mutagenesis studies that the deamidase activity of the protective protein is also separable from its protective function. This does not exclude, however, that in lysosomes the cathepsin A/deamidase works in cooperation with β-galactosidase and neuraminidase. For example, an exopeptidase might be required after endoproteolytic cleavage of glycoprotein substrates, to trigger the efficient hydrolysis of their sugar side chains by the associated glycosidases. On the other hand, complex formation may modulate cathepsin A/deamidase activity. A better understanding of the functions of the protective protein requires the identification of substrates that are targets of the enzyme *in vivo*. It is noteworthy that protective protein mRNA expression is high in mouse kidney, brain, and placenta (13), suggesting the need of a cathepsin A/deamidase activity in these tissues, *e.g.* for the inactivation of bioactive peptides such as oxytocin and kinins.

Extended knowledge of the protective protein could arise from the analysis of individual galactosialidosis patients, done in light of the results reported here. These patients have so far been identified and diagnosed on the basis of their reduced β-galactosidase/neuraminidase activities. Only recently, a carboxypeptidase deficiency was reported for the first time in three late infantile/juvenile patients (44), although the less specific Z-Phe-Leu substrate was used in these studies. The ability to directly detect residual cathepsin A activity in patients will allow the identification of individuals having an isolated cathepsin A/deamidase deficiency but normal protective protein function. The creation of animal models having targeted cathepsin A/deamidase active site mutations could prove instructive in this context.

Except for the active site mutants the other modified human protective proteins are all retained in the ER. In the case of the mutant precursor with a cysteine to threonine substitution at position 60 this is likely due to improper folding of the precursor polypeptide (45). This cysteine residue is embedded within the 10-residue region that is most conserved among all carboxypeptidases and must be important for their three-dimensional structures (46). Moreover, this domain in the protective proteins is part of an internal repeat occurring once in the 32- and 20-kDa chains. Since this motif is characterized by tryptophan residues it could be engaged in intra-

or intermolecular hydrophobic interactions. As deduced from the cotransfection experiments, the truncated 32/20 and 32SA/20 polypeptides can spontaneously associate in the ER, but are subsequently retained. This could imply that a single chain precursor is essential for correct transfer of the protective protein to the Golgi complex. Alternatively, aberrant assembly of the two chains could also cause retention, although we have indications that coexpression of the separate subunits in insect cells leads to a 3–4-fold increase in cathepsin A activity. The possibility that formation of an active dimer in the ER has a deleterious effect on *de novo* synthesized proteins awaits further investigations.

The crystallization of wheat serine carboxypeptidase II has recently revealed remarkable structural homology of this enzyme to zinc carboxypeptidase A (43). It was speculated that these two proteases share a common ancestor, perhaps a binding protein that had divergently acquired greater catalytic activity by two different mechanisms. In this scenario binding to other proteins comes before catalytic activity. The protective protein is about 30% identical to wheat serine carboxypeptidase II. Interesting questions that arise are those of how the catalytic/protective activities of this pleiotropic member of the serine carboxypeptidase family have evolved and what came first.

Acknowledgments—We wish to thank Professor Hans Galjaard for continuous support and we are grateful to Dr. Gerard Grosveld for stimulating discussions and useful suggestions. We are indebted to Dr. Just Vlak (Dept. of Virology, Agricultural University, Wageningen, The Netherlands) for introducing us to the use of the baculovirus expression system and providing us with some of his improved vectors. We also thank Dr. Martine Jaeglé for critical reading of the manuscript, Sjozef van Baal for his help with the computer data analysis, Pim Visser for the graphic work, Mirko Kuit for the excellent photography, and Jeannette Lokker for typing and editing the manuscript.

REFERENCES

1. 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
2. Verheijen, F., Brossmer, R., and Galjaard, H. (1982) *Biochem. Biophys. Res. Commun.* **108**, 868–875
3. Verheijen, F. W., Palmeri, S., Hoogeveen, A. T., and Galjaard, H. (1985) *Eur. J. Biochem.* **149**, 315–321
4. Yamamoto, Y., Fujie, M., and Nishimura, K. (1982) *J. Biochem.* **92**, 13–21
5. Yamamoto, Y., and Nishimura, K. (1987) *Int. J. Biochem.* **5**, 435–442
6. Hoogeveen, A. T., Verheijen, F. W., and Galjaard, H. (1983) *J. Biol. Chem.* **258**, 12143–12146
7. 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
8. Andria, G., Strisciuglio, P., Pontarelli, G., Sly, W. S., and Dodson, W. E. (1981) in *Sialidases and Sialidoses* (Tettamanti, G., Durand, P., and DiDonato, S., eds) pp. 379–395, Edizioni Ermes, Milano, Italy
9. Wenger, D. A., Tarby, T. J., and Wharton, C. (1978) *Biochem. Biophys. Res. Commun.* **82**, 589–595
10. Suzuki, Y., Sakuraba, H., Yamanaka, T., Ko, Y. M., Iimori, Y., Okamura, Y., and Hoogeveen, A. T. (1984) in *The Developing Brain and Its Disorders* (Arima, M., Suzuki, Y., and Yabuuchi, H., eds) pp. 161–175, University of Tokyo Press, Tokyo
11. Palmeri, S., Hoogeveen, A. T., Verheijen, F. W., and Galjaard, H. (1986) *Am. J. Hum. Genet.* **38**, 137–148
12. 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
13. Galjart, N. J., Gillemans, N., Meijer, D., and d'Azzo, A. (1990) *J. Biol. Chem.* **265**, 4678–4684
14. McDonald, J. K., and Barrett, A. J. (1986) in *Mammalian Proteases: A Glossary and Bibliography*, Vol. 2, pp. 186–191, Academic Press, New York
15. Doi, E., Kawamura, Y., Matoba, T., and Hata, T. (1974) *J. Biochem.* **75**, 889–894
16. Kawamura, Y., Matoba, T., and Doi, E. (1980) *J. Biochem. (Tokyo)* **88**, 1559–1561
17. Matsuda, K. (1976) *J. Biochem. (Tokyo)* **80**, 659–669
18. Jackman, H. L., Tan, F., Tamei, H., Beurling-Harbury, C., Li, X.-Y., Skidgel, R. A., and Erdos, E. G. (1990) *J. Biol. Chem.* **265**, 11265–11272
19. Kleijer, W. J., Hoogeveen, A. T., Verheijen, F. W., Niermeijer, M. F., Galjaard, H., O'Brien, J. S., and Warner, T. G. (1979) *Clin. Genet.* **16**, 60–61
20. Loonen, M. C. B., van der Lugt, L., and Franke, L. C. (1974) *Lancet* **2**, 785
21. Andria, G., Del Giudice, E., and Reuser, A. J. J. (1978) *Clin. Genet.* **14**, 16–23
22. Gluzman, Y. (1981) *Cell* **23**, 175–182
23. Taylor, S., and Tappel, A. L. (1973) *Anal. Biochem.* **56**, 140–148
24. Roth, M. (1971) *Anal. Chem.* **43**, 880–882
25. Galjaard, H. (1980) *Genetic Metabolic Disease: Early Diagnosis and Prenatal Analysis*, Elsevier Science Publishers B. V., Amsterdam
26. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) *Anal. Biochem.* **150**, 76–85
27. Young, R. A., and Davis, R. W. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 1194–1198
28. Feinberg, A. P., and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13
29. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
30. Sanger, F. G., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
31. Murphy, G., and Kavanagh, T. (1988) *Nucleic Acids Res.* **16**, 5198
32. Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395
33. Higuchi, R., Krummel, B., and Saiki, R. K. (1988) *Nucleic Acids Res.* **16**, 7351–7367
34. Okayama, H., and Berg, P. (1983) *Mol. Cell. Biol.* **3**, 280–289
35. Proia, R. L., d'Azzo, A., and Neufeld, E. F. (1984) *J. Biol. Chem.* **259**, 3350–3354
36. van Dongen, J. M., Barneveld, R. A., Geuze, H. J., and Galjaard, H. (1984) *Histochem. J.* **16**, 941–954
37. Willemsen, R., Hoogeveen, A. T., Sips, H. J., van Dongen, J. M., and Galjaard, H. (1986) *Eur. J. Cell Biol.* **40**, 9–15
38. Kawamura, Y., Matoba, T., Hata, T., and Doi, E. (1977) *J. Biochem. (Tokyo)* **81**, 435–441
39. Luckow, V. A., and Summers, M. D. (1988) *Biotechnology* **6**, 47–55
40. Dmochowska, A., Dignard, D., Henning, D., Thomas, D. Y., and Bussey, H. (1987) *Cell* **50**, 573–584
41. Sørensen, S. B., Svendsen, I., and Breddam, K. (1987) *Carlsberg Res. Commun.* **52**, 285–295
42. Valls, L. A., Hunter, C. P., Rothman, J. H., and Stevens, T. H. (1987) *Cell* **48**, 887–897
43. Liao, D.-I., and Remington, S. J. (1990) *J. Biol. Chem.* **265**, 6528–6531
44. Tranchemontagne, J., Michaud, L., and Potier, M. (1990) *Biochem. Biophys. Res. Commun.* **168**, 22–29
45. Hurtley, S. M., and Helenius, A. (1989) *Annu. Rev. Cell Biol.* **5**, 277–307
46. Sørensen, S. B., Svendsen, I., and Breddam, K. (1989) *Carlsberg Res. Commun.* **54**, 193–202

Publication 5

J. Biol. Chem. (1992), *in press*

HUMAN LYSOSOMAL PROTECTIVE PROTEIN: GLYCOSYLATION, INTRACELLULAR TRANSPORT AND ASSOCIATION WITH ß-GALACTOSIDASE IN THE ENDOPLASMIC RETICULUM

Hans Morreau*, Niels J. Galjart*, Rob Willemsen†, Nynke Gillemans, Xiao Yan Zhou and Alessandra d'Azzo¶.

From the MGC Department of Cell Biology and Genetics and the Department of Clinical Genetics†, Erasmus University, Rotterdam, The Netherlands.

*These authors contributed equally to the results presented in this report.

SUMMARY

In lysosomes ß-galactosidase and neuraminidase acquire a stable and active conformation through their association with the protective protein. The latter is homologous to serine carboxypeptidases and has cathepsin A-like activity which is distinct from its protective function towards the two glycosidases. To define signals in the human protective protein important for its intracellular transport, and to determine the site of its association with ß-galactosidase, we have generated a set of mutated protective protein cDNAs carrying targeted base substitutions. These mutants were either singly transfected into COS-1 cells or co-transfected together with wild type human ß-galactosidase. We show that all point mutations cause either a complete or partial retention of the protective protein precursor in the endoplasmic reticulum (ER). This abnormal accumulation leads to degradation of the mutant proteins probably in this compartment. Only the oligosaccharide chain on the 32-kDa subunit acquires the mannose-6-phosphate recognition marker, the one on the 20-kDa subunit seems to be merely essential for the stability of the mature protein. In cotransfection experiments, wild type ß-galactosidase and protective protein appear to assemble already as precursors, soon after synthesis, in the ER. Mutated protective protein precursors that are retained in the ER or pre-Golgi complex interact with and withhold normal ß-galactosidase molecules in the same compartments, thereby preventing their normal routing.

Lysosomal protective protein has two so far identified modes of action: a catalytic activity overlapping with that of cathepsin A (Jackman *et al.*, 1990; Galjart *et al.*, 1991), and a protective function towards lysosomal ß-D-galactosidase (EC 3.2.1.23) and N-acetyl-α neuraminidase (EC 3.2.1.18, sialidase), whose stability and activity depend on their interaction with the protective protein (d'Azzo *et al.*, 1982; Hoogeveen *et al.*, 1983; van der Horst *et al.*, 1989). Using ß-galactosidase affinity columns the three hydrolases copurify from different tissues of various species as complexes of variable molecular mass (Verheijen *et al.*, 1982, 1985; Yamamoto *et al.*, 1982, 1987). The latter have been analyzed mainly by determining the distribution of the two glycosidase activities over high and low molecular weight forms (Hoogeveen *et al.*, 1983; Potier *et al.*, 1990; Scheibe *et al.*, 1990; Verheijen *et al.*, 1982, 1985; Yamamoto *et al.*, 1982, 1987). Immunochemical detection of the protective protein in one of these studies revealed that the majority of the mature two-chain form is not resolved in the complex (Hoogeveen *et al.*, 1983). Recently, we have demonstrated that protective protein precursor by itself can form oligomers, probably homodimers, of ~85 kDa at neutral pH (Zhou *et al.*, 1991). Together these data suggest that different pools of protective protein either associated or not associated with the two glycosidases may exist (Galjart *et al.*, 1991).

The primary structures of human, mouse and chicken protective proteins as well as human and mouse ß-galactosidases have been determined from their cloned cDNAs (Galjart *et al.*, 1988, 1990, 1991; Oshima *et al.*, 1988; Morreau *et al.*, 1989;

Yamamoto *et al.*, 1990, Nanba and Suzuki, 1990). The protective protein is homologous to yeast and plant serine carboxypeptidases (for review see Breddam, 1986) and could be identical to a deamidase/carboxypeptidase released from human platelets that has been purified as a homodimer of ~94 kDa and may function in the local (in)activation of bioactive peptides (Jackman *et al.*, 1990; 1992). The carboxypeptidase activity of this enzyme has been compared with that of cathepsin A (Jackman *et al.*, 1990). We have given direct evidence that the protective protein has cathepsin A-like activity but this catalytic activity is separable from the protective function towards ß-galactosidase and neuraminidase, hence these two roles are distinct (Galjart *et al.*, 1991).

In cultured human fibroblasts biosynthetic labeling studies have shown that the first immunoprecipitable forms of human protective protein and ß-galactosidase are glycosylated precursors of 54 kDa and 85 kDa, respectively (d'Azzo *et al.*, 1982). The 54-kDa precursor is converted within one hour after synthesis into a mature two-chain product of disulfide linked 32- and 20-kDa subunits (d'Azzo *et al.*, 1982, Galjart *et al.*, 1988; Palmeri *et al.*, 1986). The two-chain form binds the serine protease inhibitor DFP demonstrating that maturation serves to release the carboxypeptidase activity (Galjart *et al.*, 1990, 1991). The 85-kDa ß-galactosidase precursor also undergoes extensive proteolysis giving rise to a mature form of 64 kDa. During these steps, a 66-kDa intermediate occasionally accumulates in cultured fibroblasts that have been treated with the protease inhibitor leupeptin (d'Azzo *et al.*, 1982; Palmeri *et al.*, 1986). Curiously, human ß-galactosidase cDNA, overexpressed in COS-1 cells, directs the synthesis of the 85 kDa precursor that is secreted but hardly converted into the mature 64-kDa lysosomal form (Oshima *et al.* 1988; Morreau *et al.*, 1989).

In view of the apparently aberrant processing of human ß-galactosidase in COS-1 cells, we wanted to assess the role played by the protective protein in the maturation and stability of the enzyme in this system. Furthermore, we were interested in extending the structural analysis of the protective protein to define domains and signals that govern its intracellular transport, processing and quaternary structure. For these studies a series of *in vitro* mutagenized protective proteins with targeted amino acid substitutions were overexpressed in COS-1 cells either after single transfections or after cotransfections with wild type human ß-galactosidase. The results show that the majority of the site-directed mutations impede the correct routing of the protective protein to the lysosomes. Surprisingly, mutated precursors that are withheld in the ER are able to coretain normal ß-galactosidase molecules in the same compartment. Therefore, the association of ß-galactosidase and protective protein soon after synthesis seems to interfere with efficient transport of ß-galactosidase out of the ER and in turn with its maturation and stability in lysosomes.

Experimental Procedures

Plasmid constructs

In vitro mutagenesis of human protective protein cDNA was carried out using the method of Higuchi *et al* (1988). DNA fragments with the desired point mutations were amplified by PCR and exchanged for wild type cDNA segments using suitable restriction enzyme sites. In the final construct the integrity of the DNA derived from PCR amplification was verified by dou0xe strand plasmid sequencing (Murphy and Kavanagh, 1988). Oligonucleotides for site-directed mutagenesis were synthesized on an Applied Biosystems 381A oligonucleotide synthesizer. All constructs were cloned into a derivative of the mammalian expression vector pCD-X (Galjart *et al*, 1988; Okayama and Berg, 1982), using standard procedures (Sambrook *et al*, 1989).

COS-1 cells (Gluzman, 1981) were maintained in Dulbecco's modified Eagle's medium- Ham's F10 medium (1:1, v/v), supplemented with antibiotics and 5% (v/v) fetal bovine serum. For biosynthetic labeling studies cells were seeded in 30 mm dishes and grown to 30% confluency. Transfection of COS-1 cells, metabolic labeling and preparation of cell extracts and media have been carried out as described previously (Galjart *et al*, 1988; Proia *et al*, 1984). In both single and cotransfections the amounts of plasmid DNA used were 1 μg for the various protective protein constructs and 2 μg for the ß-galactosidase construct. The ß-hexosaminidase cDNA construct, ßSG, was kindly provided by Dr. R. Proia (NIDDK, NIH, Bethesda, MD, U.S.A.). Labeling was carried out with 60 μCi [^{3}H]leucine per ml labeling medium (143 Ci/mmol, Amersham Corp.) or 100 μCi [^{32}P]phosphate per ml labeling medium (carrier free, Amersham Corp.), for the time periods indicated in the legends to the figures. In pulse-chase experiments 0.3 mg of unlabeled leucine per ml labeling medium was added to the dishes after the 30 min pulse. The immunoprecipitation conditions using fixed *Staphylococcus aureus* cells (Immunoprecipitin, BRL) have been described earlier (Proia *et al*, 1984). The antibody preparations used in immunoprecipitation experiments were: the anti-protective protein antibodies, anti-54 (Galjart *et al.*, 1991), that recognizes all forms (54, 32, 20 kDa) of the protective protein; the anti ß-galactosidase antibodies, αd64 (Morreau *et al.*, 1989), that primarily precipitate the precursor polypeptide, and αn64 that was raised against a purified placental preparation of ß-galactosidase and precipitate both mature and precursor forms of the enzyme. After immunoprecipitation, radioactive proteins were resolved on 12.5% polyacrylamide gels under reducing and denaturing conditions (Hasilik and Neufeld, 1980), fixed and visualized by autoradiography ([^{32}P]labeled samples) or fluorography ([^{3}H]leucine labeled material). In the latter case Amplify (Amersham Corp.) was used to enhance the signals.
ß-galactosidase and ß-hexosaminidase activities in single and cotransfected COS-1 cells were measured with artificial 4-methylumbelliferyl substrates using standard assay conditions (Galjaard, 1980).

Limited proteolysis with trypsin
In order to obtain large quantities of secreted proteins COS-1 cells were transfected in 100 mm dishes. 48 hr after transfection, cells were labeled in medium without fetal bovine serum. Medium samples were collected 16 hr later, and centrifuged at 1000 rpm for 5 min at room temperature, in order to remove detached cells. Afterwards, bovine serum albumin (BSA, Boehringer Mannheim) was added in a final concentration of 1 mg/ml. Media were concentrated and desalted as described previously (Proia *et al*, 1984). Aliquots of 60 μl, corresponding to about 12.5% of the original volume, were taken and brought to 200 μl with 20 mM sodium phosphate pH 6.8. One sample was left as such on ice, the remainder was incubated with 1.5 μg trypsin (Sigma) for 0, 2, 5, 10, 30 min at 37 °C. Trypsin was inactivated by the addition of 3 μg bovine pancreas trypsin inhibitor (Sigma). In the 0 min time point the inhibitor was added before the trypsin. The procedure used here is a modification of the one described by Frisch and Neufeld (1981). After proteolysis 10 μl samples were taken for detection of cathepsin A-like activity with the N-blocked dipeptide benzyloxycarbonyl-phenylalanyl-alanine (Z-Phe-Ala) as substrate. The assay was carried out using a modified procedure (Galjart *et al.*, 1991) of the method of Taylor and Tappel (1973). Liberated alanine was measured by the fluorimetric method outlined by Roth (1971). From the remainder of the aliquots (about 150 μl) radiolabeled proteins were immunoprecipitated using anti-54 antibodies. Proteins were resolved and visualized as described above.

Immunoelectron microscopy
Transfected COS-1 cells in 100 mm dishes were fixed in 0.1 M phosphate buffer pH 7.3, containing 1% acrolein and 0.4% glutaraldehyde. Further embedding in gelatin, preparation for ultracryotomy and other methods for immunoelectron microscopy were as reported (Willemsen *et al*, 1986). Immunolabeling was performed using αd64, anti-ß-galactosidase antibodies mentioned above and α32, anti-protective protein antibodies. The latter were raised against the 32-kDa denatured chain of the human protective protein and were shown in immunoprecipitation experiments to recognize under denaturing and reducing conditions the 54-kDa precursor as well as the 32-kDa mature subunit of the protective protein (Galjart *et al.*, 1988).

Results

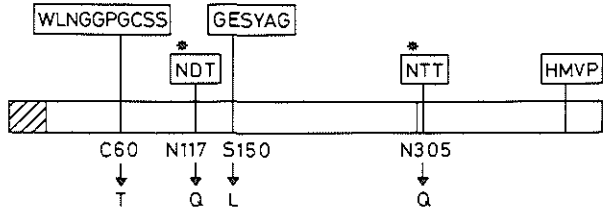
Mutagenesis of human protective protein

We have recently described point mutations in the human protective protein cDNA, that alter key amino acid residues within the three domains highly conserved among the members of the serine carboxypeptidase family (Galjart *et al*, 1991). Two of these amino acid substitutions, Ser_{150} to Ala and His_{429} to Gln (henceforth called SA_{150} and HQ_{429}), abolish cathepsin A-like activity without affecting the protective function. The third, Cys_{60} to Thr (CT_{60}), impairs the transport of human protective protein out of the ER. To investigate the role of other residues, crucial for structure and function of the protein, additional amino acid substitutions were introduced *in vitro*. As summarized in Fig. 1 (upper part) the acquisition of either one or both oligosaccharide chains present on the human protein (Galjart *et al*, 1988) was prevented by modification of Asn_{117} to Gln (NQ_{117} or NQ1) and/or Asn_{305} to Gln (NQ_{305} or NQ2). Furthermore the active site serine residue at position 150 was changed into leucine (SL_{150}) to determine the influence of a bulky amino acid in that position on the transport and activity of the protective protein.

Turnover of normal and mutant protective proteins in transfected COS-1 cells.

To investigate the rate of synthesis and degradation of mutant protective proteins pulse-chase experiments were carried out in transfected COS-1 cells. After transfections, cells were labeled with [^{3}H]leucine for 30 min, and chased for a maximal period of 6 h. Radiolabeled proteins from cell lysates and media were immunoprecipitated with anti-54 antibodies. As shown in Fig. 1 (lower part), processing of the 54 kDa wildtype precursor into the 32/20-kDa two-chain form is detected after 30-60 min chase (Fig. 1, wt(se), lanes 1-5). The majority of the overexpressed precursor, however, is secreted into the culture medium, where it is visible already at 30 min chase. Extracellular accumulation of radioactive material reaches a maximum level at 3 h and declines after 6 h, indicating that either some degradation or partial reinternalisation of the precursor has taken place.

The three glycosylation mutants, NQ_{117}, NQ_{305} and $NQ_{117,305}$, are synthesized as precursors of the expected reduced size with respect to the normal precursor (Fig. 1, panels NQ1, NQ2, NQ1&2). Tunicamycin treatment of transfected COS-1 cells prior to and during labeling revealed that the protein moieties of wild type precursor and glycosylation mutants have an identical electrophoretic mobility (not shown). Deletion of the first oligosaccharide chain on residue 117 prevents proteolytic processing of the mutant precursor, but does not impair its secretion (Fig. 1, NQ1, lanes 1-5). Release into the medium of the NQ_{117}-precursor seems to be more gradual than that of wildtype precursor and no decline is visible after 6 h chase. This suggests that, once secreted, NQ_{117} is stable and not endocytosed. The NQ_{305}-mutant instead is processed into a 32/18-kDa two-chain form and is apparently cleaved at the correct site. In this case maturation is followed by degradation of the aberrant mature protein (Fig. 1, NQ2, lanes 1-5). Thus, either the absence of the oligosaccharide chain on the 20-kDa subunit is deleterious for the stability of the mature form or it is the amino acid substitution per se that causes this instability. Secretion of NQ_{305} is delayed in comparison to wildtype and NQ_{117}-precursors and the amount of immunoprecipitable material is reduced. Kinetics and degree of secretion of NQ_{117}- and NQ_{305}-mutants suggest that both precursors are in part withheld in an early biosynthetic compartment. This effect is apparently additive since the double glycosylation mutant



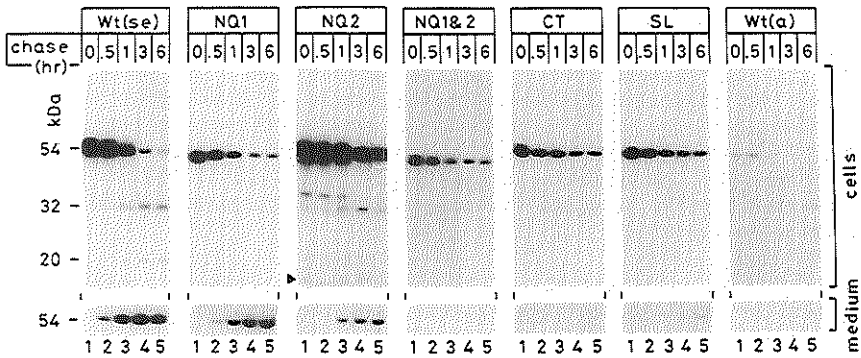


Fig. 1. Pulse-chase analysis of overexpressed wildtype and mutant protective proteins. **Upper part:** the preproform of human protective protein is represented by the bar with the hatched part being the signal sequence. The position and sequence of three domains, highly conserved in serine carboxypeptidases, are shown above the bar. The asterisks indicate the two glycosylation sites. Single amino acid substitutions are depicted with numbers referring to the position of these residues within the protective protein. **Lower part:** COS-1 cells were transfected with pCD-constructs containing the various (mutant) cDNAs: wt(se) and (a) represent human protective protein cDNA cloned in the sense and antisense orientation respectively. Transfected cells were labeled for 30 min with [^{3}H]leucine and chased for the periods of time indicated above the lanes. Labeled proteins were immunoprecipitated from cell extracts and media with anti-54 antibodies, separated by SDS-PAGE under reducing conditions and visualized by fluorography. Molecular sizes of precursor and mature polypeptides are at left. The arrowhead indicates the position of the 18 subunit of the NQ_{305} mutant protective protein. Exposure times of the fluorographs were: wt(se), 8 days; NQ1, NQ1&2 CT, SL, 4 days; NQ2, wt(a), 20 days, all media samples, 2 days.

is not detected in the culture medium. Instead it is slowly degraded after synthesis without undergoing any processing (Fig. 1, NQ1&2, lanes 1-5).

A similar fate is observed for the CT_{60}- and SL_{150}- proteins, yet in the latter case small quantities of the mutant precursor polypeptide escape to the culture medium after 3-6 h chase (Fig. 1, CT and SL, lanes 1-5). Processing of endogenous COS-1 protective protein can be visualized after a prolonged exposure of the fluorograph (Fig. 1, wt(a), lanes 1-5). Intracellularly, its maturation pattern is similar to that observed for wildtype human protective protein, but no immunoprecipitable form is recovered from the culture medium.

Phosphorylation

To determine which of the two oligosaccharide chains on human protective protein acquires the mannose-6-phosphate (M6P) recognition marker transfected COS-1 cells were labeled with [^{32}P]phosphate. Radiolabeled proteins from cell homogenates and media were immunoprecipitated as above. For comparison cDNAs encoding mouse protective protein (Galjart *et al*, 1990) and a human-chicken hybrid protective protein (Galjart *et al*, 1991) were also transfected. As shown in Fig. 2, both the precursor polypeptide as well as the 32-kDa chain of human and mouse protective proteins incorporate the label (Fig. 2, lanes 1 and 2). The anti-54 antibodies immunoprecipitate mouse protective protein less efficiently. The same holds true for the human-chicken hybrid, of which only the 58-kDa precursor form is brought down by the antibodies under the experimental conditions used (Fig. 2, lane 3). Of the three glycosylation mutants only NQ_{305} is phosphorylated and in part processed (Fig. 2, lanes 7-9), indicating that loss of the sugar moiety on the 20-kDa polypeptide does not influence the phosphorylation process. A notable difference is now detected between CT_{60}- and SL_{150}-mutant precursors: the latter is phosphorylated and secreted in tiny quantities into the culture medium (Fig. 2, lanes 4 and 5). The faint 54- and 32-kDa bands visible throughout the autoradiograph represent endogenous COS-1 protective protein.

Limited proteolysis of secreted wildtype and mutant protective proteins with trypsin.

We have previously demonstrated that the normal endoproteolytic processing of the wild type precursor, secreted by transfected COS-1 cells, can be mimicked by trypsin and yields a mature two-chain enzyme (Zhou *et al.*, 1991), whose big subunit binds radiolabeled DFP (not shown). We have now extended these experiments to determine the influence of the deletion of either of the two oligosaccharide chains on the stability and/or activity of the protective protein/cathepsin A. Radiolabeled COS-1 cell-derived precursor samples were aliquotted and each aliquot incubated with a fixed amount of trypsin for increasing periods of time. Reactions were stopped by the addition of trypsin inhibitor. Afterwards cathepsin A-like activity was measured in each aliquot followed by immunoprecipitation with anti-54 antibodies. As shown in Fig. 3, trypsin digestion of wildtype human protective protein precursor gives rise to a two-chain product with constituent polypeptides of about 32 and 20 kDa. This protein is resistant to further trypsin digestion up to 30 min at 37 ^{0}C (Fig. 3, lanes 1-6). Concomitant with the appearance of the two-chain product there is a sharp increase in cathepsin A-like activity, which declines only after 30 min. The first cleaved bond(s) is highly trypsin-sensitive, since even at t=0 small amounts of precursor protein are converted (Fig. 3, lane 2). The residual cathepsin A-like activity in the

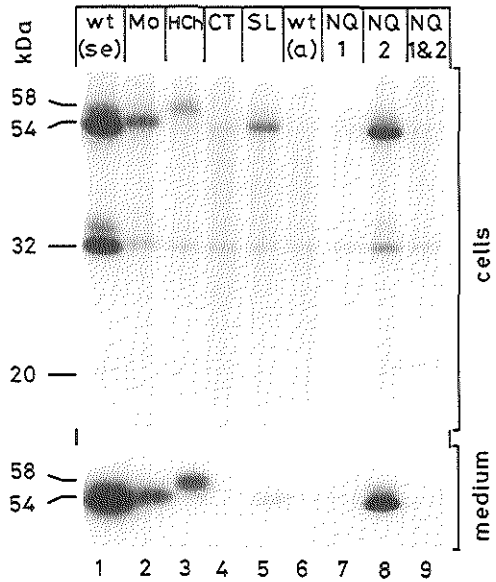


Fig. 2. Phosphorylation of transiently expressed protective proteins. COS-1 cells were transfected with the pCD-constructs described in Fig. 1 as well as with two expression plasmids containing the cDNAs encoding mouse protective protein (Mo) and a human-chicken hybrid protective protein (HCh). Two days after transfection cells were labeled with [^{32}P] phosphate and radiolabeled proteins were immunoprecipitated from cell lysates and media using anti-54 antibodies. Molecular sizes of precursor and mature polypeptides are indicated at left. Exposure time of the autoradiograph was one day.

aliquot that was not treated with trypsin (Fig. 3, lane 1) is due to the presence of some processed protective protein, inadvertently generated while manipulating the culture medium.

Surprisingly, the secreted and cleaved NQ_{117}-mutant is as resistant to further trypsin digestion as wildtype precursor (Fig. 3, lanes 7-12). The generated two-chain product with a large subunit of about 30 kDa and a normally sized 20-kDa subunit, is active but cathepsin A-like activity is about one third compared to wild type cleaved precursor aliquots. Deletion of the second oligosaccharide chain instead has a drastic effect on protease resistance (Fig. 3, lanes 13-18). Cleavage of the NQ_{305}-mutant precursor with trypsin initially gives rise to a two-chain product with subunits of about 32 and 18 kDa. Along with the appearance of the two-chain product a very low cathepsin A-like activity can be detected, but both immunoprecipitable material as well as enzymatic activity disappear upon prolonged trypsin treatment. These results correlate well with the pulse-chase experiments presented earlier. Both types of study indicate that once an NQ_{305}-mature protein is formed it is unstable.

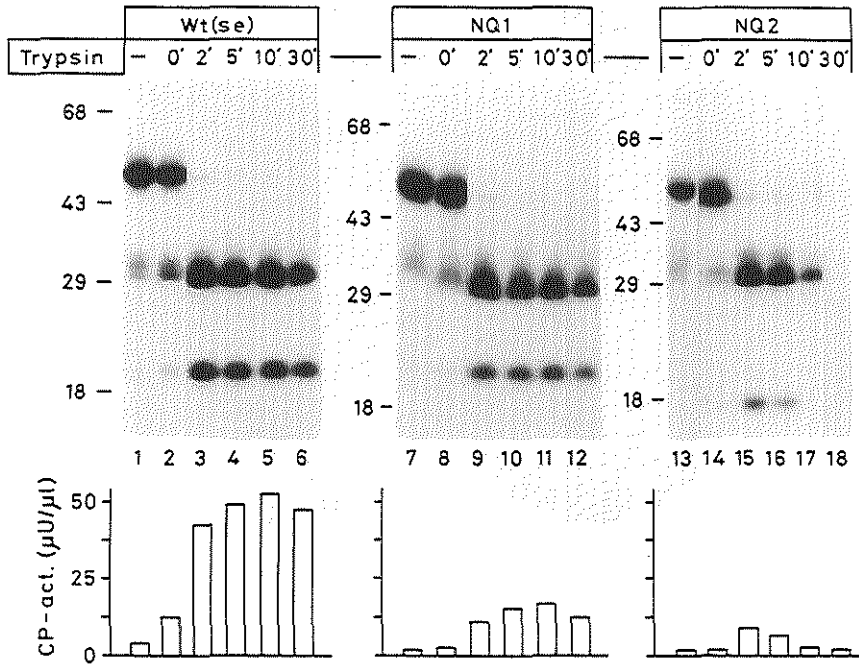


Fig. 3. Limited proteolysis with trypsin or normal and selected mutant human protective protein precursors. COS-1 cells were transfected with pCD-constructs encoding wildtype human, NQ1- or NQ2- mutant protective proteins. After transfections, radiolabeled secreted proteins were concentrated, using 1 mg/ml BSA as carrier. Aliquots of the concentrated preparations were incubated at 37°C with 2 mg trypsin for the indicated periods of time. Reactions were stopped with trypsin inhibitor. Cathepsin A-like activity towards the acylated dipeptide Z-Phe-Ala was measured in a part of each aliquot. One milliunit of activity is defined as the enzyme activity that releases one picomole of alanine per min. The remainder of the samples were used for immunoprecipitation using anti-54 antibodies. Molecular sizes of prestained protein standards are indicated to the left of each panel. Exposures were: left panel 4 days; middle panel and right panel 28 days.

Localization of mutant human protective proteins.

In order to correlate the turnover of the different mutant protective proteins with a subcellular compartment, immunoelectron microscopy was carried out on transfected COS-1 cells. Ultrathin sections of these cells were probed with antibodies against the denatured 32 kDa subunit of human protective protein (anti-32 antibodies), followed by an incubation with goat anti-(rabbit IgG)-gold. Previously we have shown that in COS-1 cells wildtype human protective protein compartmentalizes in structures corresponding to rough ER (RER), Golgi complex and lysosomes. Fig. 4 shows that

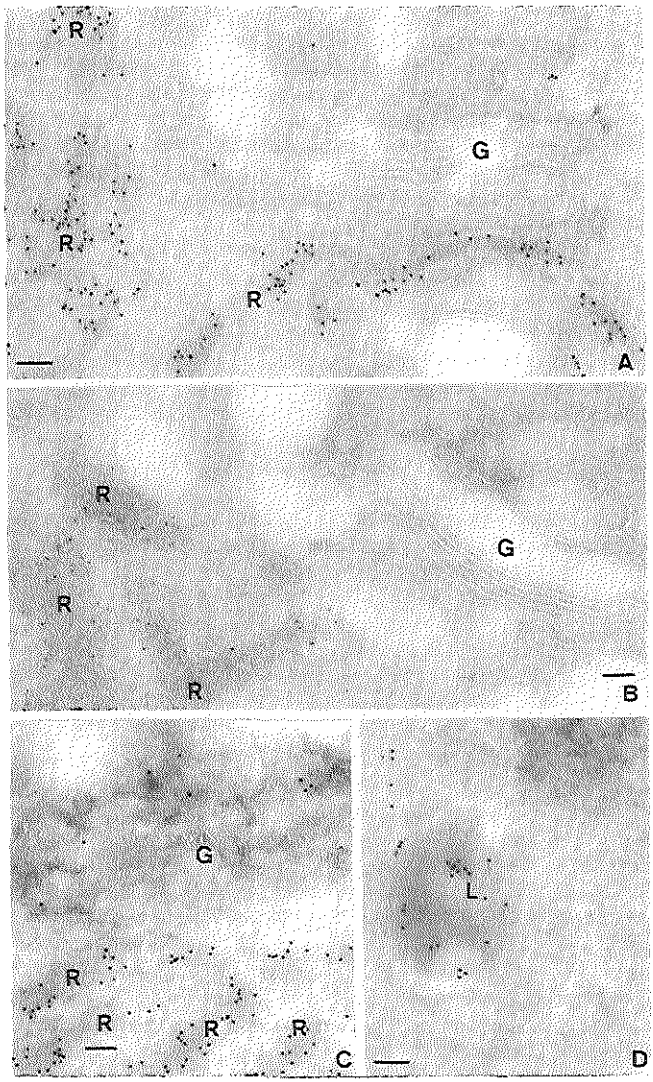


Fig. 4. Localization of overexpressed mutant protective proteins. COS-1 cells, transiently transfected with pCD-SL_{150} (A), -NQ_1(B), or -NQ_2(C, D), were fixed 72 hr after transfection, embedded and prepared to immunoelectron microscopy. Cryosections were incubated with anti-32 antibodies followed by goat anti-(rabbit IgG)-gold labeling. The SL_{150}- and NQ_1- mutant proteins are only detected in rough endoplasmic reticulum structures (R), whereas the NQ_2-mutant is also detected in low amounts in the Golgi complex (G) and lysosomes (L), Bars, 0.1 μm.

COS-1 cells transfected with SL_{150}- or NQ_{117}-mutant cDNAs are labeled almost exclusively in RER structures and not in the Golgi complex or lysosomes (Fig. 4A and B respectively). The same results were obtained with cells overexpressing the double glycosylation mutant (not shown). Thus, degradation of, SL_{150}- and $NQ_{117-, 305}$-precursors is likely to occur within the ER. In cells overexpressing the NQ_{305}-mutant protein instead some labeling above background is observed in Golgi complex as well as in lysosomes (Fig. 4C, D). This indicates that the NQ_{305}-mutant protein reaches the lysosome, where it is subsequently processed and degraded.

Maturation of human ß-galactosidase precursor in cotransfected COS-1 cells

It was shown previously that in COS-1 cells, transiently expressing human ß-galactosidase, the 85-kDa precursor was poorly processed into the mature 64-kDa polypeptide (Oshima *et al.*, 1988; Morreau *et al.*, 1989). The lack of processing was thought to be due to the transfection procedure. However, it is also conceivable to assume that the amount of endogenous protective protein available in COS-1 cells is not sufficient to stabilize the bulk of overexpressed ß-galactosidase molecules. To further investigate the role played by the protective protein in the maturation and stability of ß-galactosidase in COS-1 cells, single and cotransfections were performed using both human ß-galactosidase (HßGaL) and protective protein (Hu54) cDNA constructs. In the same experiment we also tested whether any of the site directed mutagenized protective proteins with an altered subcellular distribution (i.e. CT_{60}; SL_{150}; 32ь20) or catalytic activity (i.e. SA_{150}; HQ_{429}), would influence ß-galactosidase intracellular behaviour. The biochemical characteristics of mutant SA_{150}, HQ_{429}, CT_{60} and 32ь20 have been described in an earlier report (Galjart *et al.*, 1991).

Single and cotransfected cells were labeled with [^{3}H]leucine and radiolabeled proteins from cell lysates were immunoprecipitated using two different anti-ß-galactosidase antibodies: αd64 and αn64 (see Experimental Procedures). As seen in Fig. 5 (lanes 1 and 11), in cells transfected with only ß-galactosidase cDNA αd64 preferentially recognize the precursor polypeptide, whereas αn64 have a broader specificity and precipitate both precursor and mature forms of the enzyme. However, the latter is detected in relatively little amounts and is only partially processed into an intermediate 66-kDa form. A similar immunoprecipitation pattern is observed in cells cotransfected with HßGaL cDNA and an antisense Hu54 construct (Fig. 5, lanes 8 and 18). In contrast, cells cotransfected with both ß-galactosidase and wildtype protective protein cDNAs (Fig. 5, lanes 2 and 12) contain a sharply increased amount of mature and fully processed 64-kDa polypeptide. This is regardless of the type of antibodies used, although the αn64 preparation shows a better affinity for the mature protein. Thus, coexpression of the two wildtype human proteins in COS-1 cells seems to facilitate the maturation of the ß-galactosidase precursor and/or to stabilize the mature enzyme in lysosomes.

Early association of ß-galactosidase precursor with wildtype and mutant protective proteins

Another interesting finding emerged from the experiment discussed above. In cotransfected COS-1 cells the protective protein precursor appears to coprecipitate with ß-galactosidase using both αd64 and αn64 antibody preparations (Fig. 5, lanes 2 and 12). Given the monospecificity of the antibodies, coprecipitation of the two proteins must be the result of their association that apparently takes place at

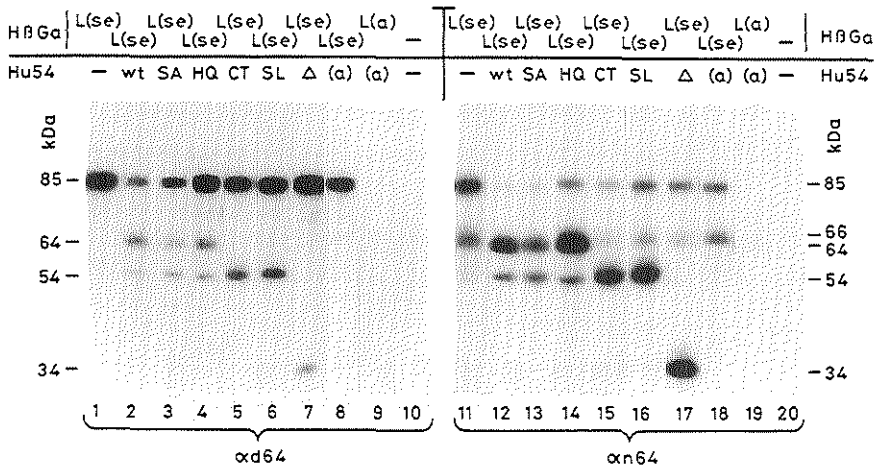


Fig. 5. Early association of human ß-galactosidase with various mutant protective proteins. COS-1 cells were either singly or cotransfected with HßGaL and different Hu54 cDNA constructs (lanes 1-20) in sense(se) or antisense(a) orientations. The protective protein constructs indicated are wildtype Hu54(wt), SA_{150}(SA), HQ_{429}(HQ), CT_{60}(CT), SL_{150}(SL), and 32(Δ20)(Δ), all cloned in the sense orientation. Cell lysates were immunoprecipitated with either αd64 or αn64 anti ß-galactosidase antibodies. Molecular weights are indicated. Exposure time was 7 days.

precursor level. Under the stringent immunoprecipitation conditions used, the mature two-chain protective protein does not coprecipitate with ß-galactosidase, although it is normally present, as confirmed by a subsequent round of immunoprecipitation of the same samples with anti-54 antibodies (not shown). Similar results were obtained with COS-1 cells coexpressing ß-galactosidase, and the SA_{150}- or HQ_{429}- protective protein mutants (Fig. 5, lanes 3, 4 and 13, 14). The latter are active site mutants that, expressed in COS-1 cells, were shown to be transported intracellularly and secreted in a mode identical to the wildtype protective protein (Galjart *et al.*, 1991). As clearly seen in Fig. 5, the two mutant precursors are still able to associate with ß-galactosidase precursor and are coprecipitated with both antibody preparations (lanes 3, 4 and 13, 14). Furthermore, in these cotransfected samples ß-galactosidase precursor is fully processed into the mature 64-kDa form (lanes 13, 14).

Coexpression of ß-galactosidase with the CT_{60}-, SL_{150}- and 32Δ20 protective protein mutants also leads to coprecipitation of these mutant proteins with both anti ß-galactosidase antibodies (Fig. 5, lanes 5-7 and 15-17). However, in these cotransfections maturation of ß-galactosidase appears greatly reduced, and only a small amount of the intermediate 66-kDa polypeptide is immunoprecipitated with αn64 antibodies (lanes 15-17). A notable difference between these three protective protein mutants and the active site mutants is that the former are retained in the ER. Therefore, these results suggest that a correct intracellular compartmentalization of the protective protein is a prerequisite for ß-galactosidase to be present in its mature state. Furthermore, the association between the two proteins that occurs already at precursor level does not seem to be hampered by the various amino acid substitutions

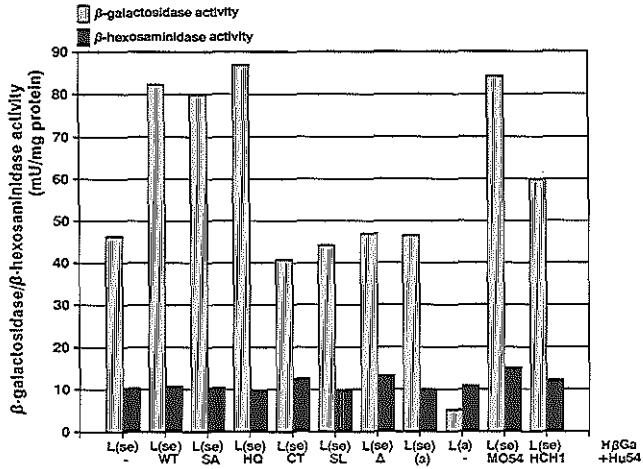


Fig. 6. β-galactosidase and β-hexosaminidase activities in single and cotransfected COS-1 cells. COS-1 cells were either transfected only with HßGaL constructs (sense, L(se), or antisense, L(a)) or cotransfected with HßGaL(se) and various protective protein encoding cDNAs. The different protective proteins used in the cotransfections are wildtype Hu54 (wt), SA_{150}(SA), HQ_{429}(HQ), CT_{60}(CT), SL_{150}(SL), 32(Δ20), Hu54 antisense (a), murine Hu54 (Mo54), and a human/chicken hybrid protective protein (HCh1). ß-Galactosidase and ß-hexosaminidase activities were measured with the appropriate 4-methylembelliferyl substrates. The activities are given in mU/mg protein. One milliunit of enzyme activity is defined as the activity that releases 1 nmol of 4-methylumbelliferyl/min.

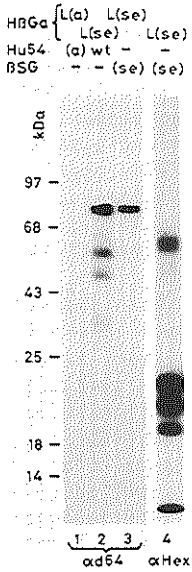


Fig. 7. Absence of coprecipitation of β-galactosidase and β-hexosaminidase. COS-1 cells were cotransfected with HßGaL and wildtype Hu54 (wt) in sense (se) or antisense (a) orientations (lanes 1/2). In addition COS-1 cells were cotransfected with HßGaL(se) and ßSG sense (se) constructs. Immunoprecipitations were carried out using anti-ß-galactosidase antibodies (αd64) or anti hexosaminidase antibodies (αHex). Molecular weight markers are indicated. Exposure time was 3 days.

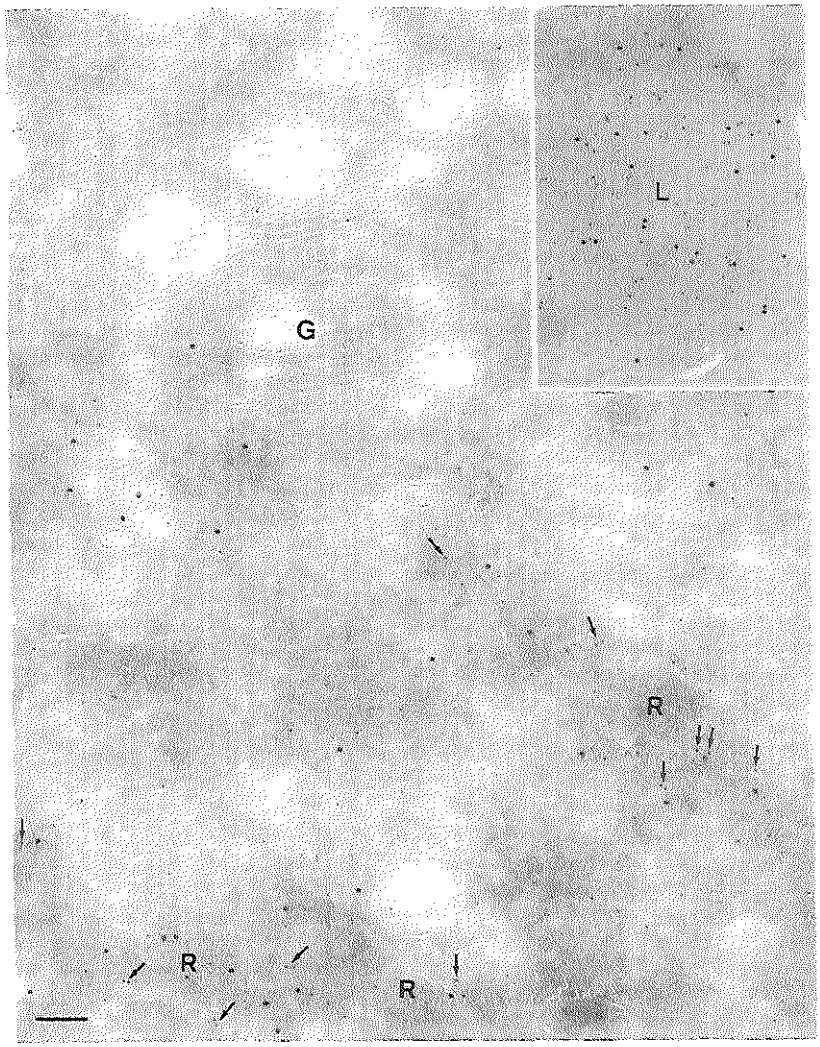


Fig. 8. Immunoelectron microscopy colocalization of ß-galactosidase and protective proteins in cotransfected COS-1 cells. COS-1 cells were cotransfected with HßGaL(se) cDNA and wildtype Hu54 (inset), or SL_{150} protective protein constructs. Ultrathin sections of cotransfected cells were probed with a combination of αd64 and α32 antibodies followed by incubation with goat anti-(rabbit IgG) coupled to gold particles of large size for ß-galactosidase antibodies and small size for protective protein antibodies. Cells coexpressing the two wildtype proteins show a colocalization of ß-galactosidase and protective protein in lysosomes: L (inset). In cells cotransfected with HßGaL and SL_{150}-constructs the gold particles are largely confined to the rough endoplasmic reticulum (R)/pre-Golgi complex (G). Bar, 0.1 µm.

generated within the protective protein molecule.

The improved maturation and/or stability of ß-galactosidase, coexpressed with wildtype protective protein, is reflected in the clear increase in enzymatic activity measured in cotransfected COS-1 cells (Fig. 6, L(se)/WT). Again the SA_{150}- and HQ_{429}- protective protein mutants have an effect on ß-galactosidase activity comparable to that of the wildtype protective protein. The same holds true for the mouse protective protein (Mo54) and the human/chicken hybrid protein (HCh1). However, in cells cotransfected with the CT_{60}-, SL_{150}- and 32o20 (o) protective protein mutants the level of enzymatic activity is similar to that detected in cells expressing only ß-galactosidase (L(se)). In the latter case the activity is clearly higher than the endogenous COS-1 level, measured in cells transfected with an antisense ß-galactosidase construct (L(a)). This activity is probably due to the small amount of mature ß-galactosidase present in cells transfected with only ß-galactosidase construct or cotransfected with the CT_{60}-, SL_{150}- and 32o20 protective protein mutants. In all samples ß-hexosaminidase activity was determined as a reference enzyme.

To exclude that coprecipitation of ß-galactosidase and protective protein precursors is not merely caused by overexpression of these proteins, COS-1 cells were cotransfected with HßGaL cDNA and a construct, ßSG, encoding the ß-subunit of the soluble lysosomal enzyme ß-hexosaminidase (Proia, 1988). As shown in Fig. 7, after immunoprecipitation with αd64 antibodies, ß-hexosaminidase is not coprecipitated with ß-galactosidase precursor (lanes 2 and 3), although it is expressed in a comparable amount (lane 4). Likewise, anti ß-hexosaminidase antibodies (αHex) do not coprecipitate ß-galactosidase. An increased amount of mature ß-galactosidase is again observed in COS-1 cells cotransfected with HßGaL and Hu54 cDNA constructs in comparison with COS-1 cells cotransfected with HßGaL and ßSG constructs (Fig. 7, lanes 2 and 3). No endogenous COS-1 proteins are immunoprecipitated from cells transfected with antisense constructs (Fig. 7, lane 1).

Localization of ß-galactosidase and protective protein in cotransfected cells.
In order to determine the intracellular distribution of ß-galactosidase precursor in cotransfected cells immunoelectron microscopy studies were performed (Fig. 8). Ultrathin sections of these cells were probed with anti-32 and αd64 antibodies followed by an incubation with goat anti-(rabbit IgG) coupled to gold particles of different sizes. In cells cotransfected with HßGaL and Hu54 cDNA constructs a clear colocalization in lysosomes is observed of both ß-galactosidase (large gold particles) and protective protein (small gold particles) (Fig. 8, inset). However, in cells cotransfected with HßGaL and SL_{150}-protective protein constructs ß-galactosidase is mainly retained together with the mutant protein in the ER/pre-Golgi compartment (Fig. 7). Only a few gold particles are present throughout the Golgi stacks, that could give rise to the small amount of partially processed ß-galactosidase seen in the aforementioned immunoprecipitation studies. These data confirm that ß-galactosidase and protective protein precursors interact soon after synthesis in the ER, and imply that this event might regulate the efficiency by which ß-galactosidase molecules are transported out of this compartment to the lysosomes.

Discussion

We have analyzed the intracellular transport, processing and secretion of protective proteins carrying targeted amino acid substitutions in order to get insight in the conformational characteristics of normal precursor and mature proteins. Furthermore, we have studied the interaction between these protective proteins and ß-galactosidase.

The results show that all amino acid substitutions interfere with or impair completely the exit of the different mutant proteins from the ER. These findings are in contrast to the behaviour of two protective protein active site mutants, SA_{150} and HQ_{429}, described earlier (Galjart *et al*, 1991), whose biosynthesis and processing resemble completely that of the wildtype protein. After synthesis the ER-retained mutant proteins are degraded. Their turnover is not as rapid as the degradation of the T cell receptor α-chain (Lippincott-Schwartz *et al*, 1988), but it compares well to the turnover rate of a natural mutant of the α-subunit of lysosomal ß-hexosaminidase (Lau *et al*, 1989). Although different amino acid substitutions could well result in differential folding of the various mutant precursors, the observed degradation rates seem quite similar. The SL_{150}-mutant is even phosphorylated, albeit poorly, indicating that part of the total pool of synthesized molecules reach a pre-Golgi site where the first phosphorylation step is thought to occur (Lazzarino *et al*, 1988). Some of the phosphorylated molecules are allowed to leave the ER, as judged by the detection of small quantities of SL_{150}-precursor in the culture medium. The other part could be specifically retrieved from a post-ER compartment (Pelham 1989).

A large amount of labeled human precursor protein is secreted instead of being transported to the lysosomes. This is not uncommon for overexpressed lysosomal enzymes; it has for example been found for the α-subunit of human ß-hexosaminidase in transfected COS-1 cells (Lau *et al.*, 1989), for human cathepsin D in baby-hamster kidney cells (Horst *et al.*, 1991; Isidori *et al.*, 1991) and for mouse cathepsin L, which is the major excreted protein (MEP) in transformed mouse fibroblasts (Dong *et al.*, 1989). It has been suggested that different protein sorting systems are present, that target lysosomal enzymes to their final destination with variable efficiency (Horst *et al.*, 1991; Isidoro *et al.*, 1991).

NQ_{117}- and NQ_{305}- glycosylation mutants are secreted in reduced amounts compared to wildtype protective protein, a fact that is explained by their partial intracellular retention. Assuming that an Asn to Gln amino acid substitution have little influence on the three dimensional structure of the protective protein, we can ascribe the following potential functions to the two oligosaccharide chains: 1) both are important for the timely exit of protective protein precursors from the ER; 2) the chain on the 32-kDa subunit is necessary and sufficient for the acquisition of the M6P-recognition marker; it is not required for the stability of the mature two-chain protective protein that is nonetheless less active than the wildtype protein; 3) the chain on the 20-kDa subunit is crucial for the stability of the mature form. The latter supposition is substantiated both by the pulse-chase experiments as well as by the limited proteolysis with trypsin. In addition to what we observe experimentally, the position of the oligosaccharide chain on the 20-kDa subunit, near the 32/20-kDa boundary and its location within a conserved stretch of amino acid residues, might also imply a specific role for this sugar chain. Combined the findings indicate that it could actually serve as an age marker for the protective protein (Segal *et al*, 1984; Winkler *et al*, 1984). In this model its stepwise trimming by glycosidases in the

lysosome would render the mature protective protein increasingly unstable.

Limited proteolysis of the wildtype protective protein precursor with trypsin readily gives rise to a mature two-chain molecule, that is rather resistant to further cleavage. From the size of the generated 20-kDa fragments it seems as if both the initial intracellular endoproteolytic processing as well as the trypsin cleavage take place within the same domain, perhaps even after the same residue. The results indicate that a region is present in the precursor molecule that is very sensitive to proteolysis and it will be of interest to determine whether other enzymes could cleave within the same domain. If so one could imagine that the conversion of protective protein to its mature form could occur in an extracellular environment by circulating proteases. This supposition might be relevant in view of an interesting finding made by Erdös and coworkers. These authors have identified a deaminidase in human platelets and smooth muscle cells, probably identical to the protective protein, that inactivates a number of bioactive peptides, including endothelin I, and is homodimeric in its native state (Jackman *et al*, 1990, 1992).

Recent studies have demonstrated that the protective protein also dimerizes at precursor level and neutral pH, suggesting that this event takes place in an early biosynthetic compartment (Zhou *et al*, 1991). Oligomerization could be a requirement for timely exit of the precursor out of the ER. This hypothesis is supported by the observation that a natural mutant of the protective protein fails to form dimers and is largely retained in the ER (Zhou *et al*, 1991). On the other hand human ß-galactosidase, the second component of the complex, overexpressed in COS-1 cells is only in part routed to the lysosomes and is poorly processed into the mature form (Oshima *et al*, 1988; Morreau *et al*, 1989). In relation to this observation it is noteworthy that in galactosialidosis fibroblasts, which are completely devoid of the protective protein, ß-galactosidase can reach the lysosomes. Its proteolytic processing in these cells is, however, clearly delayed compared to normal fibroblasts (d'Azzo *et al*, 1982). We have now evidence that coexpression of the protective protein and ß-galactosidase in COS-1 cells leads to accumulation of the mature and fully processed ß-galactosidase polypeptide. Unexpectedly, protective protein and ß-galactosidase precursors appear to associate at an early stage of biosynthesis, indicating that their assembly is not restricted to the lysosomal environment. Whether monomeric ß-galactosidase interacts with protective protein dimers is still unclear.

The advantage of the association event for ß-galactosidase could be similar to that for the protective protein: it might facilitate the exit of the precursor polypeptide out of the ER and in turn stabilize the mature enzyme in lysosomes. This assumption is substantiated by the finding that some *in vitro* mutagenized protective proteins, retained in the ER, prevents efficient lysosomal targeting of ß-galactosidase. These findings could have important consequences for the interpretation of the molecular events in galactosialidosis and perhaps even G_{M1}-gangliosidosis. Natural mutants of the protective protein could exist that retain a large part of the ß-galactosidase precursor pool in the ER, thereby causing an additional negative effect on its lysosomal activity. Viceversa, there might be mutations in the ß-galactosidase protein that influence the intracellular transport and processing of protective protein. Finally the early assembly of ß-galactosidase and protective protein in normal cells could create pools of free and assembled molecules that are committed to different functions already at an early stage and could potentially be targeted differentially according to the respective functions. It needs to be further investigated to what

extend and in which stoichiometry association occurs in normal cells and tissues.

Acknowledgements
We would like to thank Professor Hans Galjaard and the Rotterdam Foundation of Clinical Genetics for their support and Dr. Gerard Grosveld for stimulating discussions. We are grateful to Pim Visser for the graphic work, Mirko Kuit and Tom de Vries Lentsch for excellent photography and Jeannette Lokker for typing and editing this manuscript.

Abbreviations

DFP, diisopropylfluorophosphate; PCR, polymerase chain reaction; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis.

References

Breddam, K. (1986) *Carlsberg Res.Commun.* **51**, 83-128
d'Azzo, A., Hoogeveen, A., Reuser, A. J., Robinson, D. and Galjaard, H. (1982) *Proc.Natl.Acad.Sci.U.S.A.* **79**, 4535-4539
Dong, J. M., Prence, E. M. and Sahagian, G. G. (1989) *J.Biol.Chem.* **264**, 7377-7383
Frisch, A. and Neufeld, E. F. (1981) *J.Biol.Chem.* **256**, 8242-8246
Galjaard, H. (1980) *Genetic Metabolic Disease: Early diagnosis and prenatal analysis.* Elsevier Science Publishers B.V., Amsterdam
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
Galjart, N. J., Gillemans, N., Meijer, D. and d'Azzo, A. (1990) *J.Biol.Chem.* **265**, 4678-4684
Galjart, N. J., Morreau, H., Willemsen, R., Gillemans, N., Bonten, E. J. and d'Azzo, A. (1991) *J.Biol.Chem.* **266**, 14754-14762
Gluzman, Y. (1981) *Cell* **23**, 175-182
Hasilik, A. and Neufeld, E. F. (1980) *J.Biol.Chem.* **255**, 4937-4945
Higuchi, R., Krummel, B. and Saiki, R. K. (1988) *Nucleic Acids Res.* **16**, 7351-7367
Hoogeveen, A. T., Verheijen, F. W. and Galjaard, H. (1983) *J.Biol.Chem.* **258**, 12143-12146
Hoogeveen, A.T., Graham-Kawashima, H., d'Azzo, A. and Galjaard H. (1984) *J.Biol.Chem.* **259**, 1974- 1977
Horst, M. and Hasilik, A. (1991) *Biochem.J.* **273**, 355-361
Isidoro, C., Horst, M., Baccino, F. M. and Hasilik, A. (1991) *Biochem.J.* **273**, 363-367
Jackman, H.L., Tan, F.L., Tamei, H., Buerling-Harbury, C., Li, X.Y., Skidgel, R.A. and Erdos, E.G. (1990) *J.Biol.Chem.* **265**, 11265-11272
Jackman, H.L., Morris P.W., Deddish P.A., Skidgel R.A., Erdos E.G. (1992) *J. Biol. Chem.* **267**, 2872-2875
Lau, M. M. and Neufeld, E. F. (1989) *J. Biol. Chem.* **264**, 21376-21380
Lazzarino, D. A. and Gabel, C. A. (1988) *J. Biol. Chem.* **263**, 10118-10126
Lippincott-Schwartz, J., Bonifacino, J. S., Yuan, L. C. and Klausner, R. D. (1988) *Cell* **54**, 209-220
Morreau, H., Galjart, N. J., Gillemans, N., Willemsen, R., van der Horst, G. T. J. and d'Azzo, A. (1989) *J. Biol. Chem.* **264**, 20655-20663
Murphy, G. and Kavanagh, T. (1988) *Nucleic Acids Res.* **16**, 5198
Nanba, E. and Suzuki, K. (1990) *Biochem. Biophys. Res. Commun.* **173**, 141-148
Okayama, H. and Berg, P. (1982) *Mol. Cell. Biol.* **2**, 161-170
Oshima, A., Tsuji, A., Nagao, Y., Sakuraba, H. and Suzuki, Y. (1988) *Biochem. Biophys. Res. Commun.* **157**, 238-244
Palmeri, S., Hoogeveen, A.T., Verheijen, F.W. and Galjaard, H. (1986) *Am. J. Hum. Genet.* **38**, 137-148
Pelham, H. R. (1989) *Annu. Rev. Cell Biol.* **5**, 1-24
Potier, M., Michaud, L., Tranchemontagne, J. and Thauvette, L. (1990) *Biochem. J.* **267**, 197-202
Proia, R.L., d'Azzo, A. and Neufeld, E.F. (1984) *J. Biol. Chem.* **259**, 3350-3354
Roth, M. (1971) *Anal. Chem.* **43**, 880-882
Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual.* (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory)
Scheibe, R., Hein, K. and Wenzel, K.W. (1990) *Biomed. Biochim. Acta* **49**, 547-556

Segal, H.L. and Winkler, J.R. (1984) *Curr. Top. Cell Regul.*
Simmons, W.H. and Walter, R. (1980) *Biochemistry* **19**, 39-48
Taylor, S. and Tappel, A.L. (1973) *Anal Biochem* **56**, 140-148
Van der Horst, G., Galjart, N.J., d'Azzo, A., Galjaard, H. and Verheijen, F.W. (1989) *J. Biol. Chem.* **264**, 1317-1322
Verheijen, F., Brossmer, R. and Galjaard, H. (1982) *Biochem. Biophys. Res. Commun.* **108**, 868-875
Verheijen, F.W., Palmeri, S., Hoogeveen, A.T. and Galjaard, H. (1985) *Eur. J. Biochem.* **149**, 315-321
Willemsen, R., Hoogeveen, A.T., Sips, H.J., van Dongen, J.M. and Galjaard, H. (1986) *Eur J Cell Biol* **40**, 9-15
Winkler, J.R. and Segal, H.L. (1984) *J. Biol. Chem.* **259**, 15369-15372
Yamamoto, Y., Fujie, M. and Nishimura, K. (1982) *J. Biochem. (Tokyo)* **92**, 13-21
Yamamoto, Y. and Nishimura, K. (1987) *Int. J. Biochem.* **19**, 435-442
Yamamoto, Y., Hake, C.A., Martin, B.M., Kretz, K.A., Ahern-Rindell, A.J., Naylor, S.L., Mudd, M. and O'Brien, J.J. (1990) *DNA Cell Biol.* **9**, 119-127
Zhou, X.Y., Galjart, N.J., Willemsen, R., Gillemans, N., Galjaard, H. and d'Azzo, A. (1991) *EMBO J.* **10**, 4041-4048

SUMMARY

The isolation and characterization of the cDNAs as well as the genes encoding lysosomal enzymes have greatly facilitated studies on their primary structures, intracellular transport and enzymatic characteristics. Furthermore it allowed the identification of genetic lesions that affect either the synthesis or the biochemical properties of the enzymes and in turn lead to lysosomal storage disorders.

This thesis focusses on the structural and functional analysis of acid ß-galactosidase starting with the identification and characterization of its cDNA and gene, followed by new insights on the intracellular transport of this enzyme and its assembly with the protective protein. Chapter I gives an overview of the characteristics of transcriptional elements of genes encoding lysosomal enzymes. In accordance with their ubiquitous distribution in various tissues, the promoters of most lysosomal genes have characteristics of "housekeeping" gene promoters: they are rich in G and C nucleotide residues, have features of CpG-islands, contain SP1 binding sites and lack the conventional TATA and CCAAT boxes. In spite of these features, a number of lysosomal enzymes have been shown to function in tightly regulated processes which may call for a controlled gene expression. A few examples are given, where regulation of transcription of lysosomal genes might occur. Chapter I ends with a discussion on the occurrence of alternative splicing of pre-mRNAs encoding lysosomal proteins.

The experimental work discussed in chapter II is preceded by an introduction on the lysosomal storage disorders characterized by an isolated or combined ß-galactosidase deficiency and on the biochemical features of the enzyme itself. Mutations at the ß-galactosidase structural locus cause the lysosomal storage disorders G_{M1}-gangliosidosis and Morquio B syndrome; a combined deficiency of ß-galactosidase and neuraminidase due to a primary defect of the protective protein is found in the disorder galactosialidosis. The clinical manifestations in patients with these genetic diseases are heterogeneous, ranging from severe infantile to milder juvenile and adult forms. Lysosomal ß-galactosidase hydrolyses non-reducing terminal galactose moieties from glycoproteins, glycolipids and glycosaminoglycans. In normal cultured fibroblasts the enzyme is synthesized as a precursor protein of 85 kDa, that is subsequently cleaved in lysosomes into a 64-kDa mature form. Using an affinity chromotography matrix, ß-galactosidase can be purified together with the protective and under certain experimental conditions with N-acetyl-α-neuraminidase. It is apparent that these three glycoproteins associate, since they have been shown to depend on each other for their intralysosomal stability and full enzymatic activity. Although they probably form a functional complex in lysosomes, the stoichiometry of this complex is not yet understood.

Publication 1 describes the isolation of two different cDNA molecules. The first encoding the classic lysosomal ß-galactosidase recognizes a mRNA of 2.5 kb in total RNA preparations from normal fibroblasts. The expression of this transcript differs considerably in RNA preparations from clinically distinct G_{M1}-gangliosidosis patients. Analysis of the predicted amino acid sequence has revealed that the ß-galactosidase precursor contains seven potential N-linked glycosylation sites, five of which are present at the C-terminus. From the location of the chemically determined N-terminal sequence of the mature 64-kDa polypeptide we could deduce, that the substantial proteolytic processing of the 85 kDa ß-galactosidase precursor observed in human

fibroblasts as well as in mouse kidney cells and macrophages occurs exclusively at the C-terminus.

The second cDNA molecule derives from alternatively spliced ß-galactosidase pre-mRNA. This transcript of 2.0 kb is only visible in immunoselected polysomal RNA. The region of the human pre-mRNA undergoing alternative splicing is encoded by exon II-VII of the ß-galactosidase gene. The alternatively spliced mRNA encodes for a ß-galactosidase-related protein that has a different subcellular distribution than the classic enzyme and lacks enzymatic activity towards synthetic substrates. A similar ß-galactosidase-related protein can not be encoded by the mouse gene, since the reading frame of a putative alternatively spliced message would contain several stop codons and give rise to a truncated polypeptide.

Analysis of the gene encoding human lysosomal ß-galactosidase (publication 2) taught us that it contains 16 exons and spans over 62.5 kb of genomic sequences. Promoter activity has been narrowed to a 236 bp fragment containing GC-rich stretches and several SP1 transcription elements. These characteristics of the promoter as well as the multiple cap sites of the ß-galactosidase mRNA classifies it in the group of housekeeping gene promoters. Most strikingly A second fragment of 851 bp, located upstream from the first, negatively regulates initiation of transcription.

The identification of the mutation in the ß-galactosidase gene present in both alleles of two sibs with the severe infantile form of G_{M1}-gangliosidosis is presented in publication 3. The mutation consists of a T nucleotide insertion immediately after the GT dinucleotide of the splice donor site of intron 1. In total RNA samples of these patients two aberrant transcripts (1 and 2) have been identified containing a 20 nucleotide insertion derived from the 5' end of intron 1. Furthermore, in transcript 2 sequences encoded by exon II are deleted. Apparently, the T nucleotide insertion may lead to an aberrant splicing mechanism, due to the preference for a cryptic splice site present in intron 1.

The isolation and characterization of the cDNA encoding the protective protein have shown that this protein is homologous to yeast and plant serine carboxypeptidases. In fact, it closely resembles a previously characterized lysosomal enzyme cathepsin A, and is identical to a deamidase/carboxypeptidase isolated from human platelets, that is involved in the local (in)activation of bioactive peptides. In publication 4 we present evidence that the protective protein has indeed cathepsin A-like activity and this activity is severely deficient in fibroblasts from clinically different galactosidosis patients. Furthermore, using site-directed mutagenesis, we have demonstrated that the protective function towards ß-galactosidase and neuraminidase is distinct from the catalytic activity and the latter can be exerted outside the complex. A number of targeted base substitutions in the protective protein cDNA, described in publication 4 and 5 has allowed us to study the intracellular transport and biochemical characteristics of these mutant proteins in comparison with the wildtype protective protein. Furthermore, both mutants and wildtype protective proteins have been used in cotransfection experiments to study their site and mode of interaction with normal ß-galactosidase. Surprisingly, we have found that wildtype ß-galactosidase and protective protein associate as precursors already at the level of the endoplasmic reticulum. A remarkable finding was that most of the mutated protective proteins that are aberrantly retained in the endoplasmic reticulum or pre-Golgi complex are able to withhold ß-galactosidase molecules and prevent their normal intracellular routing.

SAMENVATTING

De isolatie en karakterisering van cDNA's die coderen voor lysosomale enzymen hebben de mogelijkheid tot onderzoek naar hun primaire structuur, intracellulair transport en enzymatische werking verbreed. Voorts kunnen de genetische mutaties in kaart te gebracht worden die de normale functie van deze eiwitten verstoren en dientengevolge leiden tot lysosomale stapelingsziekten. Het werk in dit proefschrift is gericht op structurele en functionele analyse van het lysosomale, zure ß-galactosidase enzym beginnend met de identificatie en karakterisering van het ß-galactosidase cDNA en gen. Het eindigt met nieuwe inzichten in het intracellulair transport en assemblage van dit eiwit met het protective protein.
Hoofdstuk I geeft een overzicht van de kenmerken van transcriptie elementen van genen die coderen voor lysosomale enzymen. In overeenstemming met de wijdverbreide distributie van deze enzymen in weefsels hebben de meeste promoters van de lysosomale genen de kenmerken die gevonden worden voor de expressie zgn. huishoudgenen: het frequent voorkomen van G and C nucleotiden in zogenaamde CpG-eilanden. Bovendien bevatten ze SP1-bindingsplaatsen en zijn de TATA en CCAAT elementen, kenmerkend voor weefsel specifiek geexpresseerde genen, veelal afwezig. Ondanks deze karakteristieken is een aantal lysosomale enzymen betrokken bij streng gereguleerde processen, hetgeen een gecontroleerde genexpressie zou kunnen vereisen. Hiervan worden verschillende voorbeelden gegeven. Hoofdstuk I eindigt met een discussie over het voorkomen van "alternatieve splicing" van voorloper mRNA moleculen, die coderen voor lysosomale enzymen.
In hoofdstuk II gaat aan het experimentele werk een introduktie betreffende de lysosomale stapelingsziekten vooraf, die worden gekennmerkt door een geïsoleerd of gecombineerd defect van het enzym ß-galactosidase. Mutaties in het genlocus, coderend voor ß-galactosidase zijn de oorzaak van de lysosomale stapelingsziektes G_{M1}-gangliosidose en Morquio B syndroom. Een gecombineerde deficiëntie van ß-galactosidase en neuraminidase is het gevolg van een primair defect van het protective protein, hetgeen gevonden wordt bij de ziekte galactosialidosis. De klinische verschijnselen bij patiënten met deze genetische aandoeningen zijn wisselend en variëren van ernstig infantiele tot milde juveniele en adulte vormen. Lysosomaal ß-galactosidase is verantwoordelijk voor de hydrolyse van niet reducerende terminale galactose groepen van glycoproteïnen, glycolipiden en glycosaminoglycans. In gekweekte normale fibroblasten wordt het enzym aangemaakt als een voorloper eiwit met een molekuul gewicht van 85 kDa wat wordt omgezet in een enzymatisch actieve eindvorm van 64 kDa. De enzymatische kant van het ß-galactosidase is uitgebreid bestudeerd in de afgelopen 10 jaar. Daarbij bleek o.a. dat het "protective protein" meezuivert met ß-galactosidase en onder specifieke experimentele condities ook met N-acetyl-α-neuraminidase. Deze drie glycoproteinen zijn voor hun stabiliteit en optimale enzymatische activiteit in het lysosoom van elkaar afhankelijk. Hoewel het duidelijk is dat zij een functioneel complex vormen in de lysosomen is de stochiometrie van dit complex nog niet bekend.

In publicatie 1 wordt de isolatie van twee verschillende cDNA-moleculen beschreven. Het eerste cDNA codeert voor het klassieke lysosomale ß-galactosidase en herkent een boodschapper RNA van 2.5 kb in RNA preparaties afkomstig van normale fibroblasten. De mate van expressie van dit transcript verschilt aanmerkelijk in fibroblasten van verschillende, klinisch heterogene G_{M1}-gangliosidose patiënten. De

analyse van de voorspelde aminozuur sequentie laat zien dat de voorloper van matuur ß-galactosidase zeven mogelijke "N-linked" glycosyleringsplaatsen heeft, waarvan er zich vijf bevinden in het C-terminale eind. Op basis van de chemisch bepaalde N-terminale aminozuur sequentie van het mature ß-galactosidase, kan worden geconcludeerd dat de uitgebreide maturatie van het 85 kDa ß-galactosidase voorloper eiwit zoals werd gezien in humane fibroblasten, muis niercellen en macrophagen, vrijwel uitsluitend plaatsvindt in het C-terminale gedeelte van het eiwit. Het tweede cDNA komt voort uit alternatief "gespliced" ß-galactosidase voorloper-RNA. Het gedeelte van dit molecuul dat betrokken is bij de alternatieve "splicing" wordt gecodeerd door exon II tot VII van het ß-galactosidase gen. Dit tweede transcript van 2.0 kilobase kan zichtbaar worden gemaakt na analyse van, met antilichaam geselecteerd polysomaal RNA en codeert voor een wat wij het ß-galactosidase gerelateerd eiwit noemen. Dit eiwit heeft een andere subcellulaire distributie dan het klassieke lysosomale ß-galactosidase en vertoont geen enzymatisch activiteit voor omzetting van synthetische ß-galactoside substraten. De analyse van het overeenkomstige muize ß-galactosidase gen leerde ons, dat een identieke alternatieve "splicing" van muis voorloper-RNA nooit zal kunnen leiden tot een muis ß-galactosidase-gerelateerd eiwit, aangezien het open leesraam in het RNA onderbroken is. De analyse van het gen dat codeert voor humaan lysosomaal ß-galactosidase leerde ons dat het 16 exonen bevat en zich uitstrekt over meer dan 62.5 kb genomisch DNA. De promoter bevindt zich op een fragment van 236 bp dat voorts gekenmerkt wordt door een GC-rijke sequentie die SP1 bindingsplaatsen bevat. In combinatie met de meervoudige "cap sites" die gevonden worden na analyse van boodschapper RNA zijn dit de klassieke kenmerken van promoters nodig voor transcriptie van huishoudgenen. Een 851 nucleotiden lang fragment wat voor het 236 bp fragment ligt is mogelijk betrokken bij negatieve regulatie van transcriptie van het gen.

In beide allelen van twee verwante patientjes met de ernstige infantiele vorm van G_{M1}-gangliosidose werd een insertie gevonden van een T-nucleotide onmiddellijk achter het GT-nucleotiden paar dat in de "splice donor site" van intron 1 (publicatie 3). In totaal RNA afkomstig van deze patiënten hebben wij twee aberrante transcripten (1 en 2) geïdentificeerd die een 20 nucleotiden insertie bevatten, afkomstig van het 5' eind van intron 1. Bovendien blijken in transcript 2 sequenties gedeleteerd te zijn die worden gecodeerd door exon 2. Blijkbaar wordt de T-nucleotide insertie tot een abnormaal "splicings"patroon veroorzaakt door de keuze van een cryptische "splice site" aanwezig in intron 1 (publicatie 3).

De isolatie en karakterisatie van het cDNA coderend voor het "protective eiwit" heeft laten zien dat het homologie vertoont met gist en plant serine-carboxypeptidasen. Het lijkt op een voorheen gekarakteriseerd lysosomaal enzym cathepsine A en is voorts identiek aan een deamidase/carboxypeptidase geïsoleerd uit humane bloedplaatjes. De laatste functie is mogelijk betrokken is bij de lokale (in)activatie van bio-actieve peptiden. In publicatie 4 wordt bewezen dat het protective protein inderdaad cathepsin A-achtige activiteit heeft en dat deze activiteit afwezig is in fibroblasten afkomstig van verschillende types galactosialidose patiënten. Bovendien werd, gebruikmakend van mutagenese op speciale plaatsen in het molekuul, aangetoond dat de beschermende werking van het "protective protein" opzichte van ß-galactosidase/neuraminidase te scheiden is van haar catalytische activiteit. Een aantal gerichte base substituties in het protective protein cDNA

beschreven in publicatie 4 en 5 heeft het mogelijk gemaakt het intracellulaire gedrag van deze mutant eiwitten te bestuderen en te vergelijken met het wildtype protective protein. Bovendien zijn zowel mutanten als normale protective protein cDNA konstructen gebruikt in cotransfectie experimenten met ß-galactosidase cDNA's om de plaats en type van de interactie tussen de beide eiwitten te bestuderen. Verrassend genoeg hebben we gevonden dat wildtype ß-galactosidase en protective eiwit reeds in het endoplasmatisch reticulum associëren. De meesten van de gemuteerde protective eiwitten blijven achter in het endoplasmatisch reticulum of pre-Golgi. Bovendien blijken deze "protective protein" mutanten in staat ß-galactosidase te weerhouden van zijn normale intracellulair transport.

TOT SLOT

Dit proefschrift is tot stand gekomen dankzij steun en toewijding van een aantal mensen. Natuurlijk betreft dit voornamelijk Sandra d'Azzo die de afgelopen jaren al haar talenten in de strijd heeft geworpen om dit werk te laten slagen. Soms was die betrokkenheid zo groot, dat ik bevreesd was dat het vaak gehoorde slogan: "to eat your own liver" ook daadwerkelijk in praktijk zou worden gebracht. Sandra, daarnaast wil ik je ook bedanken voor je vriendschap.

Professor dr. H. Galjaard wil ik bedanken voor al de geboden kansen en ik weet dat Arthur van der Kamp achter de schermen hierin altijd een belangrijke rol vervulde. Zowel Niels Galjart als Gerard Grosveld, bedankt voor de hulp en het plezier. Xiao Yan Zhou (Big Bad Egg), Amelia Morrone (La mattina da leone.....), Nynke Gillemans, Erik B. Bonten (-jongen !), Theresa van Amelsfoort, Arno van 't Hoog : bedankt. De nieuwlichters Robbert Rottier en Aarnoud van der Spoel: voor jullie de aanmoedigingsprijs. "Van de andere kant" wil ik ook graag wat mensen naar voren halen. Dies, bedankt voor je vriendschap. Sjozef, altijd die onzelfzuchtige bereidheid om te helpen. Maarten Fornerod, leve de subtiele humor. Marcel Koken, een koppige bok zonder streken. Geert Weeda, leve de diplomatie. Martine Jogel, voortaan de hele muis in een keer in de formaline. Verder iedereen bedankt die dat verdient. Piet, was ik een gewillig oor of niet? Reyer Hoogendoorn, Mirko Kuit, Jeannette Lokker, Pim Visser en Rob Willemsen hebben veel werk verricht voor onze groep, bedankt. De commissie aangesteld om dit proefschrift te beoordelen, heeft dat in een ongewoon korte tijd gedaan. Prof. dr. D. Bootsma, Prof. dr. A. Westerveld, Prof. dr. B. Wieringa en Prof. dr. H. Galjaard, mijn oprechte dank. Corine, al vind jij het heel eng jij staat voorop.

CURRICULUM VITAE

De schrijver van dit proefschrift werd geboren op 27 maart 1959. Na het behalen van het gymnasium-ß diploma aan het Gemeentelijk Gymnasium te Hilversum in 1977 werd aanvankelijk een jaar onderricht gevolgd in de studie economie. Dit pad werd verlaten en er werd vervolgens een aanvang gemaakt in 1978 met de studie geneeskunde aan de Universiteit van Amsterdam. In 1985 resulteerde dit in het behalen van het artsexamen. Op 1 maart 1986 trad hij in dienst van het instituut Celbiologie en Histologie I (hoofd Prof.Dr. H. Galjaard) als wetenschappelijk assistent en participeerde aanvankelijk in een onderzoek, dat werd verricht onder leiding van Dr. D. Halley naar (klinisch) genetische aspecten van cystische fibrose.
In september 1987 begon hij aan een onderwerp onder leiding van Dr.A.d'Azzo betreffend het humane zure ß-galactosidase, hetgeen beschreven wordt in dit proefschrift. Sinds februari 1992 is de auteur werkzaam als patholoog in opleiding, verbonden aan het Academisch Ziekenhuis Dijkzigt te Rotterdam.