The Protective Protein: A Multifunctional Lysosomal Enzyme

Niels Galjart

Aan Martine

THE PROTECTIVE PROTEIN: A MULTIFUNCTIONAL LYSOSOMAL ENZYME

HET 'PROTECTIVE PROTEIN': EEN MULTIFUNCTIONEEL LYSOSOMAAL ENZYM

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
vrijdag 29 november 1991 om 13:30 uur

door

Niels Jakob Galjart

geboren te Blaricum

Promotiecommissie

Promotor:	Prof. Dr. H. Galjaard
Overige leden:	Prof. Dr. D. Bootsma
	Prof. Dr. P. Borst
	Prof. Dr. A. Hasilik
Co-promotor:	Dr. A. d'Azzo

 Gedrukt door: Drukkerij Haveka B.V., Alblasserdam.

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 2.2

1. 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**: 755-64.

2. 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* **264**: 20655-63.

3. 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**: 4678-84.

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

5. Galjart, N. J., Willemsen, R., Gillemans, N., Zhou, X. Y., Morreau, H. and d'Azzo, A. "Analysis of the glycosylation, intracellular transport and structure of human lysosomal protective protein." *Submitted for publication*

6. Zhou, X. Y., Galjart, N. J., Willemsen, R., Gillemans, N., Galjaard, H. and d'Azzo, A. "A mutation in a mild form of galactosialidosis impairs dimerization of the protective protein and renders it unstable." *EMBO J* **(in press)**

7. Wiegant, J., Galjart, N. J., Raap, A. K. and d'Azzo, A. (1991). "The Gene Encoding Human Protective Protein (PPGB) Is on Chromosome-20." *Genomics* **10**: 345-49.

SCOPE OF THE THESIS

This thesis describes the characterization of a lysosomal protein, the 'protective protein', that has at least two functions. On the one hand it protects lysosomal β–galactosidase and neuraminidase from degradation within the lysosome, hence its name. On the other hand it has peptidase and deamidase activities, that could be involved in protein turnover in lysosomes and hormone (in)activation. Degradation is distinguished here from proteolytic processing, although both involve peptide hydrolysis. The first is, however, an aspecific random process, carried out at multiple sites, whereas the second is a highly specific (single) event. Given the putative function of the protective protein it seemed appropriate to start with an overview of intracellular sites of protein degradation, followed by a section on the biogenesis of lysosomes. The introduction ends with a summary on what is known about lysosomal storage disorders, a group of genetic diseases that are due to defects in lysosomal proteins. The protective protein itself is impaired in the rare disorder galactosialidosis and studies on this disease have been the basis for the discovery of the protein and analysis of its functions.

1. INTRODUCTION

1.1 Intracellular degradation

Intracellular catabolism of macromolecules ensures a counterbalance to the processes of biosynthesis and endocytosis and prevents the improper accumulation of products that would in turn impair the normal physiology of the cell. By means of degradation a cell can remove toxic or damaged components, generate energy, regulate cellular processes and, in times of starvation, provide the building blocks for the synthesis of new macromolecules. Degradation is carried out by a multitude of enzymes that are themselves subject to digestion for the same aforementioned reasons. It is widely agreed upon that intracellular breakdown can be divided into non-lysosomal and lysosomal degradation. The first is a collection of those degradative systems, cytosolic and compartmentalized, that are located outside lysosomes. The second includes breakdown carried out by the lysosomal hydrolases. Because in eukaryotic cells many macromolecules are degraded in a compartmentalized manner, unwanted hydrolysis elsewhere is prevented. At the same time the organized breakdown requires that either substrate or enzyme, or both, are tagged with signals that enable them to meet each other at the appropriate time and in the correct subcellular compartment.

Since all enzymes are endowed with a certain specificity the breakdown of macromolecules often occurs in a stepwise fashion. Complexes of enzymes catalyzing sequential steps speed up the degradation because the substrate travels a minimal distance, is passed on in a favourable conformation for the next hydrolytic step and the complexed enzymes may enhance each others activity when assembled. Examples of and variations on these general themes within a cell will be discussed in the next sections.

1.1.1 Non-lysosomal pathways of protein degradation

In mammalian cells all proteins are in a state of continuous turnover, with individual polypeptides being broken down at widely differing rates (for a recent review see Rivett, 1990). Degradation responds to changes in nutritional and hormonal conditions and to changes in the metabolic state of the cell (Ballard, 1987; Mortimore, 1987). Depending on their half lives proteins have been divided into short- and long-lived. The level of intracellular ATP influences the degradation rates of both classes (Gronostajski *et al,* 1985), although ATP-independent cytosolic proteolysis has also been reported (Woods and Lazarides, 1985; Fagan *et al,* 1986).

Short-lived cytosolic proteins are the prime targets for non-lysosomal pathways of protein degradation, as has been shown by the use of inhibitors of lysosomal function. One major system for selective cytosolic protein degradation is the ubiquitin pathway (reviewed by Ciechanover and Schwartz, 1989; Jentsch *et al,* 1990; 1991). Its selectivity depends largely on the ability of certain specialized enzymes, namely the ubiquitin-activating and -conjugating proteins E1, E2 and auxiliary factor E3, to recognize and tag proteins destined for degradation. This is achieved by the formation of a link between the C-terminal glycine of ubiquitin with the ε-amino group of an internal lysine residue of the substrate protein. Polyubiquitination can take place at defined lysine residue(s) within the 76 amino acid long ubiquitin polypeptide, resulting in branched chains (Chau *et al,* 1989). These structures are thought to be the real signal for degradation. Selective turnover of ubiquitinated substrate proteins is then carried out by an ATP- and ubiquitin-dependent protease complex of high molecular weight (Fagan *et al,* 1987; Hough *et al,* 1987). However, ubiquitin has also been found coupled to stable proteins and ubiquitin-conjugating enzymes have been implicated in other basic cellular functions such as DNA repair and cell cycle control (for review see Jentsch *et al,* 1991). Therefore, it is now understood that the marking of proteins for selective degradation is not the only role of the ubiquitin system.

The structural characteristics of a protein that determine its turnover rate are not well understood, but some rules have been proposed that relate protein primary structure to stability. One of these, the "N-end" rule, applies for ubiquitin-dependent degradation and relates the N-terminal amino acid of a protein to its intracellular stability (Bachmair *et al,* 1986). This signal, however, may be restricted to unfolded, nascent or highly flexible polypeptides (Rechsteiner, 1987) and/or it may operate on cleaved products of intracellular proteins (Dice, 1987). In addition, it depends on the correct location of a lysine residue, that could be the acceptor of a polyubiquitin chain (Bachmair and Varshavsky, 1989). Regions, rich in proline, glutamic acid, serine or threonine called PEST domains, have also been implicated in selective turnover, since they are common in primary structures of several short-lived proteins (Rogers *et al,* 1986). It has been demonstrated that deletion of such a PEST-domain from the cytosolic enzyme ornithine decarboxylase yields a truncated product that is considerably more stable (Ghoda *et al,* 1989). However, a second PEST region in this enzyme, still present in the truncated product, is apparently less influential, indicating that folding of a protein may mask determinants that otherwise would act as targets for degradation. In accordance with this is the finding that the rate of degradation of 35 proteins of known crystallographic struc-

ture, microinjected into HeLa cells, did not significantly conform to the aforementioned rules (Rogers and Rechsteiner, 1988a; 1988b; 1988c). Other events that may modify a polypeptide and render it abnormal and hence unstable are: damage (e.g. oxidation), phosphorylation or natural mutation(s).

The delineation of non-lysosomal pathways of protein degradation has shifted scientific interest towards the isolation of the cytosolic and nuclear proteases responsible for intracellular protein turnover. Two high molecular weight complexes have been identified, which could fulfill this function (for review see Rivett, 1989a). One is a 1500 kDa (26S) particle, capable of degrading ubiquitinated proteins *in vitro* in an ATP-dependent fashion (Fagan *et al,* 1987; Hough *et al,* 1987). The other is smaller (~600 kDa) and has been given many names, the most common being multicatalytic proteinase complex, MCPC (reviewed by Rivett, 1989b). The latter can be part of a 26S proteolytic complex (Driscoll and Goldberg, 1990) and it is present also in yeast, where it is called proteinase yscE (Achstetter *et al,* 1984). Recently, it has been shown that certain subunits of the yscE complex are also important in the degradation of ubiquitinated proteins (Heinemeyer *et al,* 1991) and are essential for life (Fujiwara *et al,* 1990; Heinemeyer *et al,* 1991). Several hypotheses could explain why these proteases are of such a high molecular weight: the dissociation and association of key components, such as inhibitors or activators, allows a careful regulation of the proteolytic activity; the multifunctional and multicatalytic nature of the complexes involves many subunits; the formation of a channeling mechanism which binds and then cleaves ubiquitinated proteins sequentially, thereby avoiding diffusion of possible poisonous peptide intermediates, requires a multicomponent complex (Heinemeyer *et al,* 1991; Rivett, 1989a;1989b).

Besides the aforementioned example of cytosolic protein degradation, membrane enclosed non-lysosomal turnover has also been documented. In mammalian cells another selective degradation system has only recently been recognized in the endoplasmic reticulum, ER (for review see Klausner and Sitia, 1990). Its place so early in the secretory route is on one hand logical since here it can remove redundant or abnormal proteins before they become harmful to a cell, or accumulate to toxic levels within the ER. On the other hand it seems odd that another set of non-lysosomal proteases exist, thus far undiscovered, that exercise a function similar to the lysosome yet in an earlier biosynthetic compartment. Nevertheless, ER degradation is insensitive to lysosomotropic agents such as NH_4Cl and chloroquine, and it is not prevented by cycloheximide or other inhibitors of autophagosome formation such as methyladenine or colchicine (Lippincott-Schwartz

et al, 1988). Furthermore, it is accelerated by depletion of cellular calcium (Wileman *et al,* 1991), a cation thought to be important for the maintenance of ER structure (for review see Koch, 1990). The finding of a lag-time between completion of protein synthesis and start of ER degradation as well as a partial block in degradation at 16 ^{o}C suggest that it takes place at a site physically separated from the earliest biosynthetic compartment (Lippincott-Schwartz *et al,* 1988). The energy requirement of this process is contradictory in two reports of the same group (Lippincott-Schwartz *et al,* 1988; Klausner and Sitia, 1990).

Thus far a specific signal for ER-retention and -degradation has only been defined in the α–subunit of the T-cell receptor, TCR (Bonifacino *et al,* 1990a; 1990b). The domain is located in the transmembrane region of the α-chain and it is characterized by positively charged amino acid residues. Association with the CD3 δ subunit of the TCR complex is also mediated by this region. This assembly masks the retention/degradation signal and allows further complex formation and exit out of the ER. Other examples of proteins that might have putative signals for ER degradation are HMG-CoA reductase and apolipoprotein B-100 (Klausner and Sitia, 1990). An alternatively spliced subunit (H2A) of the asialoglycoprotein receptor is also degraded in the ER, the determinant for turnover being 5 amino acids immediately next to the transmembrane domain (Amara *et al,* 1989; Lederkremer and Lodish, 1991). Notably, like the TCR α-chain, these are all examples of normally occurring membrane associated proteins. In contrast, no specific determinant has yet been identified for the degradation of normal soluble ER proteins. Instead, only aberrant (mutated, chimaeric) or incompletely assembled soluble proteins have been shown to follow this degradative route with widely differing half lifes (e.g. Lau and Neufeld, 1989; Sitia *et al,* 1990; Stoller and Shields, 1989; our own results in Chapter 2). This could be related to the tendency of some proteins to form aggregates, thereby altering their proteolytic susceptibility. Alternatively, multiple pathways of ER-degradation may exist resulting in different rates of turnover. It remains to be determined whether proteins like the TCR α-chain are selectively degraded by proteases recognizing a specific tag, or selectively targeted to a novel proteolytic compartment (Klausner and Sitia, 1990).

The two non-lysosomal pathways of protein turnover depicted above function in protein depletion but their contribution to the production of essential amino acids for biosynthetic purposes is limited. The latter function is assigned to the lysosomal-vacuolar system, which will be described below.

1.1.2 Routes to lysosomal degradation

Lysosomes can be defined as a group of heterogeneous acidic vacuoles, surrounded by a single membrane with as a main function the digestion of macromolecules. They can be viewed as the terminal degradative compartment (reviewed by Kornfeld and Mellman, 1989). In yeast and plants vacuoles fulfill a similar function but they are involved also in metabolic storage and cytosolic ion and pH homeostasis (Klionsky *et al,* 1990). The pH of the lysosome is estimated to be ~4.7, it is maintained by an H^+-ATPase pump and it is essential for organelle function (for review see Ohkuma, 1987). In fact, weak bases (e.g. ammonia, chloroquine) can diffuse into lysosomes and become protonated and trapped, with consequent increase of intralysosomal pH and dysfunction. As far as their contribution to protein turnover is concerned, some authors state that lysosomes mainly degrade long-lived proteins (see reviews by Mortimore and Khurana, 1990; Rivett, 1990).

The routes by which macromolecules that need to be degraded reach the lysosomes are summarized in Figure 1. Extracellular material is taken up either selectively by receptor-mediated endocytosis (see section 1.2.3) or non-selectively by pinocytosis and phagocytosis. Intracellularly, micro- and macroautophagy account for most of the non-selective breakdown of cytosolic proteins. Microautophagy is the term coined to describe the invagination of the lysosomal membrane, followed by formation of intralysosomal vesicles containing cytoplasmic material. Macroautophagy is the process by which preexisting ER membranes engulf portions of the cytosol and form autophagosomes (for review see Marzella and Glaumann, 1987). As studied in the perfused rat liver, these vesicles are initially surrounded by two membranes, derived from preexisting smooth ER, with the intermembrane space equivalent to the lumen of the ER (Dunn, 1990a). They then fuse with primary lysosomes, i.e. lysosomes that do not contain degraded material, to form a secondary lysosome or autolysosome (Dunn, 1990b). Macroautophagy is in part a regulated, reversible process, which can be enhanced among others by stress. Starvation of liver cells by depletion of certain amino acids induces it after a lag-time of ~8 min. The effect is an increased non-selective protein turnover, which provides the cell with new amino acids. Recently it has been shown in cultured mouse mammary carcinoma cells that the ubiquitin-activating enzyme E1 is necessary for the heat-induced increase of lysosomal protein breakdown (Gropper *et al,* 1991). E1 may act either on the formation of the autophagic vacuoles or on the targeting of cytosolic proteins (Gropper *et al,* 1991). This finding links the cytosolic ubiquitin-pathway to the lysosomal pathway of protein degradation.

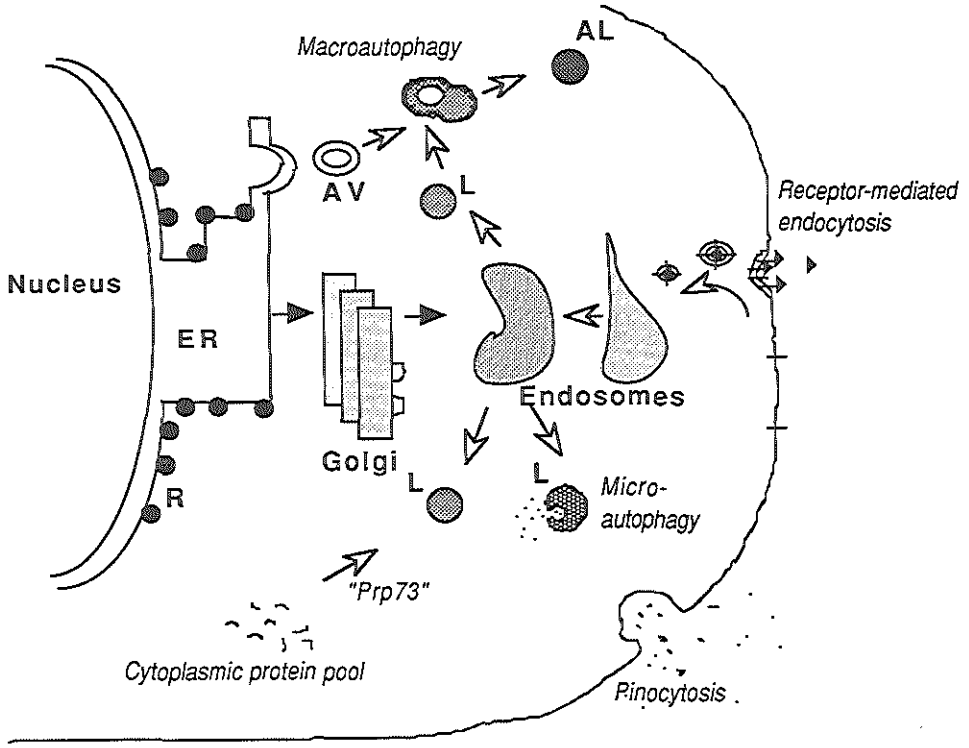


Figure 1. *Routes to lysosomal degradation.*
Shown is a eukaryotic cell in which the various pathways to lysosomal (protein) breakdown (their names are italicized), that are discussed in the text, are active. The directions of these routes to lysosomes are indicated with open arrows. Newly synthesized lysosomal enzymes are delivered to endosomes along the pathway marked with filled arrows and are subsequently transported to lysosomes. The different intracellular compartments and vesicles are not drawn proportional to each other. AV: autophagic vacuole, L: lysosome, AL: autophagolysosome, R: ribosome, ER: endoplasmic reticulum.

Dice and coworkers have demonstrated that cytosolic proteins can also be selectively broken down in lysosomes under stress conditions, like serum depletion (reviewed by Dice, 1990). This induces certain proteins to expose an amino acid stretch either identical or similar to Lys-Phe-Glu-Arg-Gln (KFERQ), which allows

their recognition by a member of the heat-shock protein (hsp) 70 family, termed prp73, that in turn targets them to the lysosome. The actual mechanism by which these substrate proteins are subsequently translocated through the lysosomal membrane is unknown. The rationale behind this selective targeting is that upon starvation the liver responds first by non-selectively turning over proteins through enhanced macroautophagy and then shifts to the selective degradation of dispensable proteins containing KFERQ-like sequences (Dice, 1990).

Another example of targeted vacuolar turnover of a cytosolic protein was recently found in yeast. Fructose 1,6 biphosphatase (FBPase), a key enzyme in gluconeogenesis, is highly expressed when cells are grown in poor glucose medium. When cells are then switched to rich medium the enzyme is targeted to the vacuole and selectively degraded. It is noteworthy that FBPase has a remote KFERQ-sequence (Chiang and Schekman, 1991).

1.1.3 Protein constituents of the lysosome

The degradative power of the lysosome is mastered by a set of acidic hydrolases and other supporting constituents, some 70 of which have now been described. Each enzyme recognizes a specific bond and sometimes additional features on a substrate and is classified accordingly. Lysosomal hydrolases have a low pH optimum and, except for the proteinases, many form higher order structures.

Because of the focus of this thesis some consideration will be given first to the lysosomal peptide hydrolases, collectively called cathepsins (a corruption of a Greek term meaning "to digest"), although they are not the biggest group (reviewed by Kirschke and Barrett, 1987). Like all other proteases cathepsins are subcategorized depending on whether they cleave within or at the extremes of a polypeptide chain, i.e. whether they are endoproteases (proteinases) or exoproteases (carboxy- or amino peptidases). Based on the four known catalytic mechanisms utilized by proteases to hydrolyze a peptide bond, these enzymes are further classified as serine-, cysteine-, aspartic acid- and metallo-proteases.

The serine proteases are specified by the socalled catalytic triad of Asp, His, and Ser amino acid side chains, that form a "charge relay" system in which electron density is "pushed" towards the serine-oxygen. In the three-dimensional conformation of active serine proteases these amino acids are neatly arranged next to each other in the order Asp-His-Ser, with the latter residue in the active center. In their primary structures instead these residues can be located far apart. The development of class specific inhibitors such as the compound diisopropylfluorophosphate

(DFP), that only reacts with the active site serine, has made assignment to a certain group of proteases more easy.

So far the best characterized lysosomal endoproteases are cathepsins B, H, L (cysteine type) and D (aspartic acid type) for which the corresponding cDNAs have been cloned. Remarkably, these proteinases are isolated as monomeric enzymes and it has been calculated for cathepsins B and D that their intralysosomal concentration could be as high as 25-45 mg/ml (~1 mM) (Kirschke and Barrett, 1987). This finding should be kept in mind when interpreting data obtained *in vitro* since such experiments are mostly performed using lysosomal enzyme concentrations that are a few orders of magnitude lower. Contrary to the proteinases the exoproteases are mostly isolated in higher order structures. Both amino- and carboxypeptidases have been characterized but, except for the protective protein/cathepsin A, no other primary structure through cDNA cloning is known. Given their high intralysosomal concentration and aspecificity it is assumed that lysosomal proteases have overlapping proteolytic activities and they do not need to act sequentially on a substrate.

The most diverse group of lysosomal hydrolases are the glycosidases. These are mostly exo-enzymes, specific for a glycosyl unit and its anomeric linkage. Hence they have to work in concert to sequentially remove monosaccharides from a variety of natural substrates. In this case enzyme complex formation would definitely add to the rate of substrate hydrolysis. Figure 2 shows two examples of macromolecules that are broken down in lysosomes in an ordered fashion and the enzymes involved in this event (for review see Aronson and Kuranda, 1989). Steps in the sequential degradation of macromolecules could be identified because of the presence of accumulated substrate(s) in tissues and urine of patients with a lysosomal enzyme deficiency. The latter can also be induced using inhibitors of lysosomal enzymes. Such studies have indicated that in glycoprotein breakdown peptide hydrolysis enhances the speed of oligosaccharide hydrolysis, probably because of the relieve of steric hindrance (Kuranda and Aronson, 1987).

Some hydrolases need an additional non-enzymic factor (activator) for full deployment of activity. Two of such activators have been characterized biochemically and purified some time ago. One cofactor was termed sulfatide activator protein, or SAP, because of its *in vivo* action on the substrates of arylsulfatase A, i.e. the sulfatides (Fischer and Jatzkewitz, 1975).

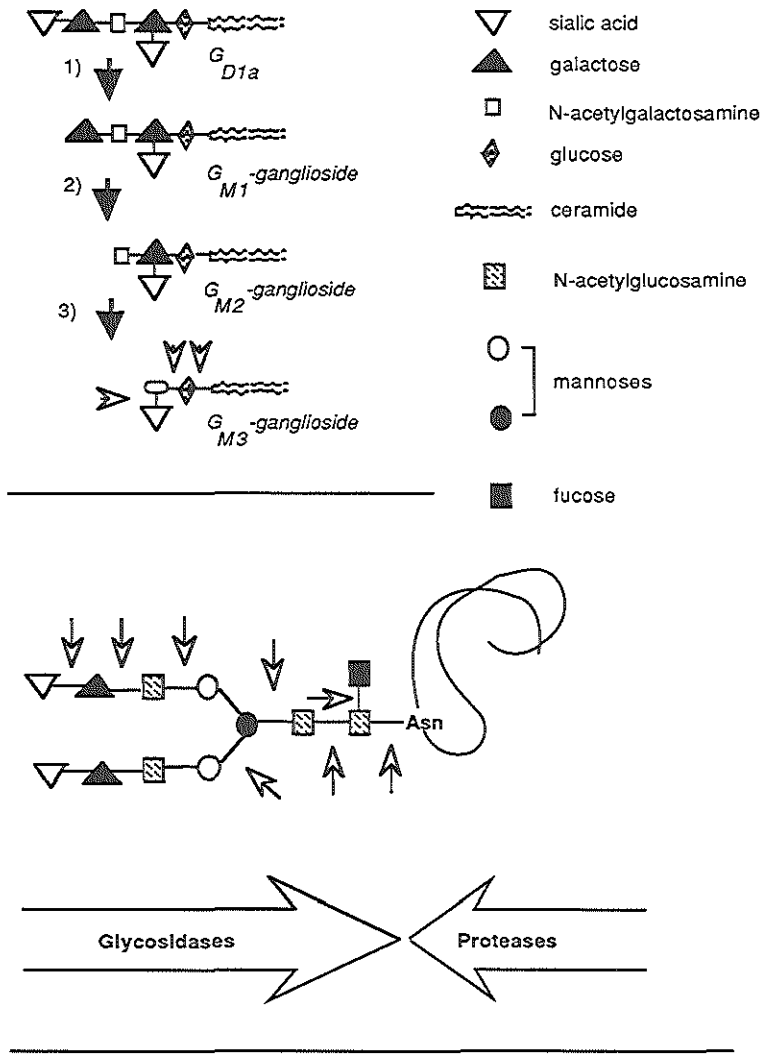


Figure 2 *Macromolecular substrates for lysosomal enzymes.*
Upper half: *the stepwise degradation of G_{D1a}-ganglioside to ceramide, reaction 1) requiring neuraminidase, 2) ß-galactosidase, and 3) ß-hexosaminidase A, respectively. If one of these enzymes or those acting later (arrowheads) is deficient storage products arise.* Lower half: *breakdown of Asn-linked glycoproteins in lysosomes. Reactions catalyzed by the glycosidases are indicated with the open arrows, again these have two work in concert. The protein backbone can presumably be broken down at random.*

In vitro this cofactor also stimulates the degradation of GM1-ganglioside and globotriaosylceramide by β–galactosidase and α-galactosidase, respectively (Li and Li, 1976). Its function is to extract the lipid substrates from membranes to make them available for hydrolysis.The second cofactor is essential for the activation of membrane associated lysosomal glucocerebrosidase (Ho and O'Brien, 1971), but it acts on the enzyme rather than on the substrate. It has recently been demonstrated that both the aforementioned cofactors as well as two other heat stable activator proteins, now collectively called saposins, are derived from a single high molecular weight precursor by proteolytic processing (Furst *et al,* 1988; O'Brien *et al,* 1988). Interestingly, this uncleaved proform was shown to be similar to a sulfated glycoprotein secreted by rat sertoli cells, which may have a function in spermatogenesis (Collard *et al,* 1988). Thus, multiple functions seem to be gathered within one amino acid sequence. A 22 kDa cofactor of β–hexosaminidase A, needed for the hydrolysis of G_{M2}-ganglioside but not related to the saposins, has also been isolated and sequenced (Meier *et al,* 1991; Xie *et al,* 1991). The lysosomal protective protein, although needed for the stabilization and activation of β–galactosidase and neuraminidase respectively, differs from the other cofactors in that it has a distinct enzymatic activity (see chapter 2).

After complete hydrolysis of a macromolecule, low molecular weight building blocks need to be transported accross the lysosomal membrane for reutilisation and to clear the lysosomes. This might occur in some cases by simple diffusion. However, an increasing number of carrier-mediated transport systems have been demonstrated to exist. Thus, metabolites like cystine (Gahl *et al,* 1982), acidic monosaccharides such as sialic acid and glucuronic acid (Mancini *et al,* 1989), neutral hexoses (Mancini *et al,* 1990), phosphate (Pisoni, 1991), calcium (Lemons and Thoene, 1991) and Vitamin B12 (Idriss and Jonas, 1991) are specifically transported.

In addition to the enzymes, activators and transport proteins lysosomes contain structural components which are the lysosomal integral membrane proteins (LIMP), also called lysosome-associated membrane proteins (LAMP) or lysosomal membrane glycoproteins (lgp). As deduced from the available primary structures this is a family of glycoproteins with a short cytosolic tail (10-11 amino acids), a single transmembrane segment and a heavily glycosylated (sialylated) luminal portion (Granger *et al,* 1990). At present the function of these different membrane components is not known. It is assumed that their abundant glycosylation may serve to protect these proteins (Barriocanal *et al,* 1986) as well as the lipid bilayer (Schauer, 1985) from degradation by lysosomal hydrolases. Normally the LAMPs

are detected almost exclusively on the lysosomal membrane, but certain cellular processes, like differentiation, may modulate their glycosylation state and cell surface expresssion, suggesting a role for these proteins in cell adhesion (Amos and Lotan, 1990). Upon stimulation with thrombin, activated platelets expose LAMP-1 at their surface and in this case it was reasoned that the protein could aid in platelet aggregation (Febbraio and Silverstein, 1990). This provides another example of an extralysosomal role for a lysosomal protein.

Protein sequencing and deduction of primary structures from cloned cDNAs encoding various lysosomal proteins have failed to show any linear stretch of amino acids common to this class of proteins. This was somewhat surprising since many of these enzymes are targeted in a similar manner to lysosomes. However, these studies did make clear that many lysosomal proteins are homologous to other proteins present either within or outside the lysosome and that there may even exist cognate proteins in lower eukaryotic and/or prokaryotic organisms. Table 1 lists the different protein families, with the various lysosomal and non-lysosomal members. It is apparent that a similar enzymatic activity is often the basis for a partial sequence identity among different proteins. For example the lysosomal enzymes α- and β-hexosaminidase are homologous, as are α-galactosidase and α-N-acetylgalactosaminidase and the family of sulfatases. The lysosomal enzyme involved in the degradation of glycogen, α-glucosidase, is homologous to the non-lysosomal intestinal brush border sucrase/isomaltase. The protective protein shares sequence identity with the yeast vacuolar carboxypeptidase Y as well as with the Golgi complex located KEX1 gene product and several plant serine carboxypeptidases. Here, the finding of sequence homology has revealed that the protective protein bears protease activity.

Table 1. *Related lysosomal and other proteins (the table is adapted from Neufeld 1991)*

related lysosomal proteins	protein homologues	refs[a)]
β-hexosaminidase α- and β-subunits	β-N-acetylglucosaminidase (*Dictyostelium discoideum*)	1-7
β-glucuronidase	β-glucuronidase (*E. coli*)	8-12
α-galactosidase α-N-acetylgalactosaminidase	α-galactosidase (yeast, *E. coli*)	13-17
α-glucosidase	intestinal brush border sucrase/isomaltase	18
acid phosphatase	prostatic acid phosphatase	19-21
arylsulfatases A and B N-acetylglucosamine 6-sulfatase iduronate sulfatase	microsomal steroid sulfatase arylsulfatase (sea urchin)	22-29
lysosomal integral membrane proteins		30
cathepsins B, H, L, S	papain family of thiol proteases	31, 32
cathepsin D	renin/pepsinogen family of aspartyl proteases	33
protective protein/cathepsin A	serine carboxypeptidase family (yeast, plants)	34

a) 1: Myerowitz *et al,* 1985; 2: Korneluk *et al,* 1986; 3: Proia *et al,* 1987; 4: Bapat *et al,* 1988;
5: Graham *et al,* 1988; 6: Neote *et al,* 1988; 7: Proia ,1988;
8: Nishimura *et al,* 1986; 9: Oshima *et al,* 1987; 10: d'Amore *et al,* 1988; 11: Powell *et al,* 1988; 12: Miller *et al,* 1990;
13: Bishop *et al,* 1986; 14: Kornreich *et al,* 1989; 15: Tsuji *et al,* 1989; 16: Wang *et al,* 1990; 17: Wang *et al,* 1991;
18: Hoefsloot *et al,* 1988;
19: Yen *et al,* 1987; 20: Pohlmann *et al,* 1988; 21: Robertson *et al,* 1988;
22: Geier *et al,*1989; 23: Peters *et al,* 1989; 24: Stein *et al,* 1989a; 25: Stein *et al,* 1989b;
26: Kreysing *et al,* 1990; 27: Peters *et al,* 1990; 28: Schuchman *et al,* 1990; 29: Wilson *et al,* 1990;
30: see references in Zot *et al,* 1990;
31: see references in Kirschke and Barrett, 1987; 32: Ritonja *et al,* 1991;
33: Faust *et al,* 1985;
34: see chapter 2.

1.2 Biogenesis of lysosomes

1.2.1 Transcription, translation and translocation

More than ten genes encoding different lysosomal proteins have been isolated and characterized. These studies have shown that with the prominent exception of the gene encoding β–glucocerebrosidase (Reiner *et al,* 1988) the respective promoters are very GC-rich and bear characteristics of promoters present in housekeeping genes, which are ubiquitously expressed. However, tissue specific expression of mRNAs encoding lysosomal proteins has also been observed (e.g. cathepsin B: San Segundo *et al,* 1986; see also chapter 2). This might indicate a differential need for these proteins in various tissues. Remarkably, the cathepsins B, H, L, and D are overexpressed and secreted in increased amounts in certain tumours (Rochefort *et al,* 1987; Sloane *et al,* 1987). This process may contribute to the metastatic potential of the transformed cells, because of the degradation of extracellular matrix by the proteases.

Some of the genes encoding lysosomal proteins give rise to transcripts, that undergo alternative splicing. These include the genes for β–glucuronidase (Oshima *et al,* 1987), β–galactosidase (Morreau *et al,* 1989; Yamamoto *et al,* 1990), sphingomyelinase (Quintern *et al,* 1989; Schuchman *et al,* 1991), α-N-acetylgalactosaminidase (Wang *et al,* 1990; Yamauchi *et al,* 1990), the sulfatide activator protein, or saposin 1 (Holtschmidt *et al,* 1991; Nakano *et al,* 1989; Zhang *et al,* 1990; Zhang *et al,* 1991) and aspartylglucosaminidase (Fisher and Aronson, 1991). Only in the case of β–galactosidase, β–glucuronidase and sphingomyelinase the cDNAs, derived from the alternatively spliced transcript, were expressed. In all instances, however, the protein products were enzymatically inactive. Thus far none of the alternatively spliced transcripts has been shown to encode a physiologically functional protein.

The segregation of secretory proteins, including lysosomal, from other polypeptides begins at the level of translation and is mediated by a discrete topogenic signal, called the signal sequence (Blobel, 1980). Statistical analysis of several known signal sequences revealed that they do not share any amino acid homology and vary considerably in length. Still they maintain characteristic features that include an N-terminal region with a positive net charge, a hydrophobic central core of 7-16 amino acids and a C-terminal domain consisting of 4-6 relatively polar residues (von Heijne, 1986). When the signal sequence is exposed on a nascent chain it is recognized and bound by the signal recognition particle, SRP (Walter and Blobel, 1980; 1982), which causes a transient arrest of translation. It is

the SRP that directs the nascent polypeptide-ribosome complex to a "docking" protein (SRP receptor) on the ER-membrane (Gilmore *et al,* 1982a; 1982b; Meyer *et al,* 1982). This event permits association of the ribosome to its receptor (Savitz and Meyer, 1990), GTP-dependent release of the SRP from the signal sequence, cessation of translation arrest (Connolly and Gilmore, 1989; Gilmore and Blobel, 1983; 1985) and translocation of the signal sequence through the ER-membrane. Finally, GTP-hydrolysis is probably needed for the subsequent release of the SRP from the "docking" protein, which enables the various components to mediate another round of signal sequence binding and translocation. The latter two processes can be biochemically uncoupled, indicating that they are sequential reactions (Nicchitta *et al,* 1991) . Recently, Simon and Blobel (1991) have demonstrated the existence of a proteinaceous protein-conducting channel within the ER membrane that mediates translocation of nascent polypeptides and is kept open by attached ribosomes. These studies show that only protein-protein interactions determine the fate of a translocating polypeptide chain and that the lipid bilayer of the ER membrane is not involved in this process. At the luminal side of the ER a signal peptidase cleaves off the signal sequence to release the proform of a secretory (lysosomal) protein.

Inside the ER a prominent modification of precursor proteins that takes place co- or posttranslationally is glycosylation on specific Asn residues in the primary sequence context Asn-X-Ser/Asn-X-Thr (X being any amino acid except Pro). The reaction is accomplished by transferring a triantennary structure of 14 sugar molecules (9 mannose, 3 glucose, 2 N-acetylglucosamine) *en bloc* from a lipid carrier, dolichol-pyrophosphate, to the protein (reviewed by Kornfeld and Kornfeld, 1985). In cultured cells glycosylation can be prevented by addition of drugs like tunicamycin (for review on inhibition of the biosynthesis and processing of N-linked oligosaccharide chains see Elbein, 1987). ER localized glucosidases and mannosidases remove posttranslationally the terminal glucose residues and one mannose from the oligosaccharide side chains.

Translocation and glycosylation are accompanied and/or followed by the folding of the precursor polypeptide, an event that includes formation of correct disulfide bridges. Previously, the concept of self-catalyzed or spontaneous protein folding and assembly was generally accepted, the information for proper folding being completely enclosed within the primary structure of a given protein. More recently, however, this notion has been abandoned in favor of the theory of "protein-catalyzed" protein folding (for review see Rothman, 1989). The process is now thought to occur in a number of discrete steps under the guidance of a specialized

class of molecules, called "molecular chaperones" (for reviews see (Ellis *et al,* 1989; Ellis and van der Vies, 1991). These are members of a family of "heat shock" proteins (hsp), termed as such because they were first detected in increased amounts after heat treatment. It is now established that they are also constitutively expressed proteins. Chaperonins bind to partially unfolded proteins, thereby promoting a proper conformation and preventing the aggregation of polypeptides. Cycles of binding and release are possible at the expense of ATP. Exposed hydrophobic domains in a partially unfolded precursor could be viewed, analogously to the signal sequence, as the targeting domains to which the chaperonins could bind (Landry and Gierasch, 1991). Protein folding within the ER was shown to be mediated by a a member of the hsp70 family (Munro and Pelham, 1986), a protein that turned out to be identical to the immunoglobulin heavy chain binding protein, BiP (Haas and Wabl, 1983). Not only does BiP aid in the folding of newly synthesized luminal proteins, it also mediates the assembly of multicomponent protein structures (reviewed by Hurtley and Helenius, 1989; Pelham, 1989). In yeast, the homologue of BiP is also required for the translocation of secretory proteins into the ER (Nguyen *et al,* 1991). Another ER enzyme, protein disulfide isomerase or PDI (Edman *et al,* 1985), catalyzes thiol-disulphide interchange *in vitro* and aids in the formation of correct disulfide bridges in newly synthesized precursors *in vivo* (Freedman, 1989).

Attaining a proper conformation is a prerequisite for exit out of the ER and this has been termed the quality-control system of this compartment (Hurtley and Helenius, 1989). It prevents unassembled, denatured, mutated or otherwise aberrant proteins to cause damage further down the secretory pathway. Perhaps as a consequence of their prolonged lifespan in the ER these faultily folded precursors are cleared by the aforementioned ER-degradation system.

Soluble proteins such as BiP and PDI carry out essential functions and must be constitutively present within the ER. This compartment is filled with these and other resident proteins, collectively termed reticuloplasmins (Koch, 1987). How is the segregation between resident and traversing proteins accomplished ? It has been proposed that the lumen of the ER has a highly ordered supramolecular protein structure, maintained among others by calcium (for review see Koch, 1990). This scaffold could be the major determinant in the retention of soluble reticuloplasmins. Perturbation of cellular calcium levels indeed causes the secretion of several resident ER proteins (Booth and Koch, 1989). However, a small proportion of proteins escaping from this structure would have to be selectively retrieved. BiP and PDI are retained because they contain a specific signal consisting of a C-ter-

minal extension of the 4 amino acid residues Lys-Asp-Glu-Leu, or KDEL (Munro and Pelham, 1987). In higher eukaryotes the KDEL-sequence requirement is not strict , as a limited number of other residues have similar effects (Andres *et al,* 1990). A variation of this retention motif is present in yeast, where the signal is HDEL (Pelham *et al,* 1988). It has been demonstrated that the KDEL-sequence is responsible for retrieval of resident proteins, presumably from a compartment between the ER and Golgi which has been called "salvage compartment". Soluble proteins that lack this signal leave the ER. These observations have been substantiated by the identification in yeast of a receptor that recognizes the "HDEL" signal (Lewis *et al,* 1990b; Semenza *et al,* 1990). Also in man two putative receptors have been identified, one of which is homologous to the yeast receptor (Lewis and Pelham, 1990a; Vaux *et al,* 1990).

Some transmembrane ER proteins, like the adenoviral E3/19K glycoprotein, contain retention motifs different from the KDEL sequence on their cytoplasmically exposed tails, indicating they are retrieved by different factors (Jackson *et al,* 1990; Nilsson *et al,* 1989; Paabo *et al,* 1987).

1.2.2 Targeting to lysosomes

The major pathway for segregation of soluble lysosomal enzymes starts in the pre-Golgi/Golgi compartment (Lazzarino and Gabel, 1988; for review see Kornfeld and Mellman, 1989). At this site lysosomal proteins are specifically recognized by the enzyme UDP-N-acetylglucosamine:lysosomal enzyme N-acetylglucosamine-1-phosphotransferase (phosphotransferase), that transfers N-acetylglucosamine-1-phosphate to specific mannose residues on selected Asn-linked oligosaccharide moieties. This process is followed by the removal of terminal N-acetylglucosamine, catalyzed by a second enzyme, N-acetylglucosamine-1-phosphodiester α-N-acetylglucosaminidase, that leaves free mannose-6-phosphate (M6P) residues, one or two per oligosaccharide chain. In spite of the specificity of the reaction the protein determinant on lysosomal precursors that is recognized by the phosphotransferase is not a linear amino acid sequence, but rather a combination of non-contiguous stretches brought together in three dimensional space (Baranski *et al,* 1990). The structure must also be oriented properly with respect to the substrate oligosaccharide chain. In the case of lysosomal cathepsin D a lysine residue is a critical component of the protein determinant (Baranski *et al,* 1990).

Transit through the three defined regions of the Golgi stack, i.e. cis-, medial- and trans-Golgi, brings about a further modification of the oligosaccharide chains on the heterogeneously phosphorylated lysosomal proteins. This is achieved by

Golgi resident glycosidases and glycosyltransferases, of which some 100 different types are presumed to exist (Paulson and Colley, 1989). Modification by a glycosyltransferase is taken as evidence that a protein has traveled through the compartment where the transferase is located. The stepwise addition of sugars like N-acetylglucosamine, galactose and sialic acid builds "hybrid" or "complex" types of oligosaccharide chains on lysosomal proteins (see Figure 2). Instead, unmodified side chains are said to be of "high mannose" type. Only the latter and the "hybrid" chains contain M6P residues. At the trans-side of the Golgi complex, in a reticular structure called the trans-Golgi network, TGN (Griffiths and Simons, 1986), lysosomal protein precursors with a fully modified M6P recognition marker are recognized by and bind to a specific receptor. The ligand-receptor complexes are at this point competent for targeting to lysosomes.

The discovery of the M6P recognition marker (Kaplan *et al,* 1977) led to the identification of two distinct M6P receptors, MPR (Hoflack and Kornfeld, 1985; Sahagian *et al,* 1981), for which the corresponding cDNAs have been cloned from various species and the primary structures determined (for review see Kornfeld and Mellman, 1989, and references therein; Ma *et al,* 1991). One is about 300 kDa and binds ligand in the absence of divalent cations (cation independent- or CI-MPR). Its preproform is composed of a signal sequence, followed by a large extracellular domain, containing 15 contiguous repeating elements 16-38 % identical to each other, a transmembrane domain and a cytoplasmic tail. The smaller receptor is 46 kDa and exhibits enhanced ligand binding affinity in the presence of divalent cations (cation dependent- or CD-MPR). It is also an integral membrane glycoprotein, oriented in the same way as the CI-MPR. Moreover, its extracellular domain is homologous to each of the repeating units of the CI-MPR, indicating that the two receptors are derived from a common ancestor. Both receptors exhibit different affinities for ligands and have different pH optima. The CI-MPR assumes a monomeric conformation, whereas the small receptor is presumably dimeric, although mono- and tetrameric forms have also been reported (Waheed *et al,* 1990a; Waheed and von Figura, 1990b). As monomer and dimer, respectively, the CI- and CD-MPR can bind one mole of diphosphorylated ligand per mole of "native" receptor. From the ligand side efficient binding to the receptor depends on the degree of phosphorylation, the type of oligosaccharide chain (Faust and Kornfeld, 1989) and in some cases protein determinants. The latter parameter was found to play a role in the segregation of cathepsin L in transformed mouse fibroblasts, where it is the major excreted protein (MEP). In these cells MEP displays a reduced binding to the CI-MPR in comparison with other lysosomal enzymes, which results in the enhanced

secretion of this protease (Dong *et al,* 1989). This is in spite of the presence of an oligosaccharide chain with high binding capacity for the same receptor (Lazzarino and Gabel, 1990) .

As mentioned above newly synthesized lysosomal enzymes are segregated at the level of the TGN. Ligand-receptor complexes following the biosynthetic route are clustered in clathrin-coated pits, that bud off and are transported to a prelysosomal compartment (see also section 1.2.3). The low pH of the latter causes dissociation of lysosomal proteins from their receptors, which can then recycle to the TGN. Agents that raise intra(pre)lysosomal pH impair the dissociation of the ligands from their receptors, which are in turn both recycled. In the absence of unoccupied receptors newly synthesized lysosomal precursors cannot cluster into coated pits and take a non-selective or default pathway, resulting in their enhanced secretion. In the biosynthetic route the CI-MPR is more effective in targeting lysosomal precursors. Evidence for this is the observation that overexpression of murine CD-MPR in a murine cell line, deficient in the CI-MPR, never completely restores the efficient sorting of lysosomal enzymes (Ma *et al,* 1991). Overexpression of the CD-MPR in cells that normally express both types of receptor actually enhances secretion of lysosomal proteins (Chao *et al,* 1990). An explanation for this phenomenon is that the CD-MPR releases its ligands too early, i.e. in a compartment where secretion can still occur. These recent studies raise the question whether the CD-MPR is also involved in receptor-mediated secretion of lysosomal precursor proteins.

A receptor-mediated targeting to lysosomes takes place also at the plasma membrane, where it is called endocytosis. Secreted lysosomal precursor proteins, carrying the M6P recognition marker, can be taken up and delivered to the lysosome. About 10 % of the lysosomal enzymes are delivered via this endocytic route. Surprisingly the CD-MPR is not involved in endocytosis even though it does reach the plasma membrane (Stein *et al,* 1987). Probably at this site it can not efficiently bind the ligand (Watanabe *et al,* 1990).

Deletion mutagenesis of the CI-MPR has shown that the cytoplasmic tail contains the relevant and distinct signals for targeting in the biosynthetic as well as endocytic routes (Canfield *et al,* 1991; Lobel *et al,* 1989). The signal for endocytosis is a general motif rather than a specific sequence. It consists of an aromatic side chain (e.g. Tyr, but in the CI-MPR this residue can be replaced by Phe), separated from a bulky hydrophobic amino acid side chain by two amino acids, one of which positively charged. This signal is present in primary structures of other transmembrane proteins (Canfield *et al,* 1991). Examples are the CD-MPR (Johnson *et al,*

1990), the low density lipoprotein receptor, LDLR (Chen *et al,* 1990), h-LAMP-1 (Williams and Fukuda, 1990) and lysosomal acid phosphatase, LAP (Peters *et al,* 1990).

Remarkably, the CI-MPR was found to be identical to the insulin-like growth factor II receptor, IGFIIR (Morgan *et al,* 1987). Although the binding sites for lysosomal proteins and IGFII are located at different positions, these ligands compete for a single receptor molecule. The biological significance of the dual function of the receptor is not yet clear. Interestingly, chicken CI-MPR lacks affinity for IGFII (Canfield and Kornfeld, 1989). It has been proposed that the CI-MPR evolved its IGF-binding function to target the hormone for degradation before it can transmit a growth signal (e.g. Haig and Graham, 1991). However, there is also evidence that the receptor is involved in signal transduction through IGFII binding. In addition to the M6P-dependent route other pathways for lysosomal enzyme targeting exist. A well documented example of a protein that makes use of such an alternative mechanism is LAP. This protein is synthesized as a membrane bound precursor with a short cytoplasmic tail (Pohlmann *et al,* 1988; Waheed *et al,* 1988) and it is transported to the lysosome via the plasma membrane (Braun *et al,* 1989). Recycling between plasma membrane and endosomes can occur before LAP ends up in lysosomes. Also in this case the endocytic signal invokes an essential Tyr residue in the cytoplasmic tail, but contrary to the CI-MPR in LAP this residue cannot be replaced by Phe (Peters *et al,* 1990). In lysosomes LAP is converted to a soluble form by the sequential action of a cytoplasmic thiol protease and a lysosomal aspartyl protease (Gottschalk *et al,* 1989). Aside from LAP other enzymes utilize a M6P-independent mechanism for lysosomal enzyme targeting. These include β–glucocerebrosidase (Aerts *et al,* 1986; 1988) and the LAMPs (Kornfeld and Mellman, 1989; Williams and Fukuda, 1990). Moreover, some tissues like liver, or cultured cells like HepG2, have alternative systems for targeting lysosomal enzymes, even those that contain the M6P recognition marker (Rijnboutt *et al,* 1991). In conclusion, while it is clear that the M6P sorting system is the most prevalent and best characterized for lysosomal proteins, other routes also exist, which will presumably be better defined in the near future.

1.2.3. Selective vesicular transport

Protein transport between compartments of the secretory route is carried out by vesicles. Receptor-mediated endocytosis has contributed largely to the current understanding of the mechanisms of vesicular transport. The process is initiated by the formation of clathrin networks underlying the plasma membrane. Clathrin, the

name means "lattice-like" (Pearse, 1987), is a protein complex with a three-legged structure (triskelion) that consists of three heavy and three light chains (for reviews see Brodsky, 1988; Keen, 1990; Pearse, 1987). Of the latter two groups exist, LCa and LCb, that can bind in any combination to the heavy chains (see review by (Brodsky *et al,* 1991). Alternative splicing of mRNAs encoding the light chains gives rise to tissue specific forms of LCa and LCb. Clathrin spontaneously forms cage-like structures *in vitro*. Clathrin networks induce the local invagination of the plasma membrane, thereby sequestering clustered receptor-ligand complexes. The so-formed "clathrin-coated" pit buds off and becomes a coated vesicle. Uncoating of the vesicle commences soon after. The process requires ATP and is catalyzed by a member of the hsp70 family, which is identical to the prp73 protein. Selective, clathrin-controlled transport operates not only in the endocytic route, but also in the segregation of newly synthesized lysosomal proteins at the level of the TGN. Regulated secretion through the formation of secretory vesicles requires clathrin as well, whereas constitutive secretion does not (Orci *et al,* 1987).

A group of characteristic proteins, called "adaptins", have been isolated that link clathrin to receptor-ligand complexes, via the cytoplasmic tails of the receptors. Adaptins are assembled into complexes called AP-1 and -2, or HAI and -II (reviewed by Keen, 1990), which promote coated pit and -vesicle formation. Within the vesicles adaptins are located between the membrane and the clathrin-cage (Vigers *et al,* 1986a; 1986b). AP-1 resides exclusively at the TGN, whereas AP-2 is localized to the plasma membrane. The latter recognizes the aforementioned "tyrosine signal" (Canfield *et al,* 1991) in the cytoplasmic tail of the CI-MPR (Glickman *et al,* 1989), whereas another determinant is important for coated pit formation at the TGN (Lobel *et al,* 1989). Both complexes contain proteins of 100-115 kDa, α- and β–adaptin in AP-2, ß'- and γ-adaptin in AP-1, and smaller proteins of about 50 and 17 kDa. (see Keen, 1990).

1.2.4 Endosomes/lysosomes

After uncoating, an endocytic vesicle packed with receptor-ligand complexes or empty receptors (e.g. CD-MPR) can fuse with a compartment at the periphery of the cell, called the early endosome (reviewed by Gruenberg and Howell, 1989b). The latter is a collection of vesicles with tubulovesicular extensions, that can fuse with each other. Fusion requires cytosolic factors, among which are a ras-like GTP-binding protein, rab5 (Gorvel *et al,* 1991), and the N-ethylmaleimide (NEM)-sensitive fusion protein, NSF (Diaz *et al,* 1989). The slightly acidic pH of the endosome may cause release of certain ligands from their receptors. Recycling receptors es-

cape the degradative pathway, perhaps via the tubular extensions, whereas the majority of the ligands remains confined to the multivesicular main body of the endosome and is subsequently transported to late endosomes or prelysosomes. Early and late endosomes are different entities. They can be distinguished by various biochemical and morphological criteria (Schmid *et al,* 1988), and appear to have differential need for microtubules (for review see Kelly, 1990). It is, however not clear how transport between these compartments is organized. One model states that late endosomes derive from early ones simply by maturation (Stoorvogel *et al,* 1991). Another view assumes that they are preexisting compartments that communicate via vesicular traffic (Gruenberg *et al,* 1989a). A third model envisages an interconnected network of early and late endosomes (Hopkins *et al,* 1990). Late endosomes are presumably the site where newly synthesized lysosomal enzymes arrive (reviewed by Kornfeld and Mellman, 1989). The pH is low enough for dissociation of ligands from the CD/CI-MPR. The next step is the formation of a fully equipped lysosome. No recycling of ligands is possible from the lysosome and its membrane is devoid of MPR (Griffiths *et al,* 1988).

Many lysosomal proteins have been shown to undergo discrete proteolytic processing, called "maturation", prior to or upon arrival in lysosomes. Various functions could rationalize this process. First of all, a precursor and not a mature protein may contain targeting information necessary to shuttle it to lysosomes (see chapter 2). For some proteins maturation could simply be a consequence of the proteolytic environment of the lysosome, something that could apply for the enzyme β–hexosaminidase that is as active in its precursor form as in its mature state (Hasilik and Neufeld, 1980a; 1980b). However, in the case of the protective protein/cathepsin A proteolytic processing converts an inactive form (zymogen) to an active one (see chapter 2). In this case maturation becomes a functional event and ensures the release of the peptidase activity only within the endosomal/lysosomal compartment. Various functions may underly the maturation of cathepsins B, H, L and D, which occurs in several steps (for review see Erickson 1989). A first endoproteolytic event causes the loss of a propeptide segment and is followed by further proteolytic processing as well as N- and/or C-terminal amino acid trimming. Previously, it was thought that the propeptide also served to prevent early proteolytic activity. While this may hold true intracellularly, it was recently found that secreted precursor forms of these cathepsins are catalytically active under certain conditions (Erickson 1989). The function(s) of the further intralysosomal processing and trimming of these cathepsins is not exactly known, but in the case of cathepsin D it might render the enzyme less stable (Horst and Hasilik, 1991). Therefore, another function of

processing could indeed be to alter stability and/or conformation of lysosomal enzymes.

1.2.5 Non-selective vesicular transport

Non-clathrin-coated vesicular transport, is a non-selective process dealing with "bulk" flow of proteins from ER to Golgi and between the different Golgi stacks (reviewed by Hicke and Schekman, 1990; Rothman, 1991). Several steps in the formation and fusion of non-clathrin-coated vesicles have been defined. Vesicle budding is catalyzed by cytosolic factors, among which is a complex, called the "coatomer", that consists of at least seven components (Waters *et al,* 1991). One of these is a 110 kDa protein, β–COP, that is homologous to β–adaptin of the AP-2 complex, indicating that the molecular mechanisms involved in selective and non-selective vesicle formation have some similar features (Duden *et al,* 1991; Serafini *et al,* 1991). After budding the "coatomer" is removed in a reaction that requires a ras-like GTP-binding protein. The vesicle then fuses with an acceptor membrane, by means of the same factor, NSF, that is involved in fusion of early endosomes. Fusion is inhibited by NEM and causes the accumulation of uncoated vesicles (Malhotra *et al,* 1988). NSF has been purified from Chinese hamster ovary cells (Block *et al,* 1988) and its corresponding cDNA cloned (Wilson *et al,* 1989). Surprisingly this factor is the mammalian homologue of the yeast SEC18 gene product, that is essential in one of the early steps in the transport of secretory proteins (Eakle *et al,* 1988). In fact the yeast sec18 protein can functionally substitute for the mammalian NSF (Wilson *et al,* 1989). These data indicate that vesicle fusion is an evolutionary conserved process that is controlled by distinct GTP-binding proteins working in concert with NSF and other cytosolic components (Bourne, 1988).

The bulk flow of proteins from ER to Golgi in non-clathrin-coated vesicles is also called "anterograde" transport. This process concomitantly depletes the ER from a considerable amount of lipid. A counteracting mechanism was shown to exist that could regulate lipid flow and take care of the retrieval of ER resident proteins from a post-ER compartment. This "retrograde" transport system utilizes microtubules and it can be visualized under conditions that slow down anterograde transport (Lippincott-Schwartz *et al,* 1989; 1990). Recent experiments indicate that antero- and retro-grade transport systems might use a common set of components, that catalyze (a) crucial step(s) in the advancement of both processes (Orci *et al,* 1991). If true the balance between the two transport systems would simply be regulated by their competition for limiting factors.

1.3 Lysosomal storage disorders

Previous sections described the role of lysosomes in degradative processes and pathways along which lysosomal proteins are sorted. Faulty targeting or reduced stability and/or activity of a lysosomal enzyme, due to a mutation in its gene, are conceivable. This leads to a deficiency of such an enzyme, causing accumulation in lysosomes of non-degradable substrates and eventually cellular dysfunction. The latter forms the basis of a lysosomal storage disorder. The concept was developed by Hers (1965), who was the first to discover a lysosomal enzyme deficiency (acid α-glucosidase) in a disorder called glycogenosis type II, or Pompe disease. Since this description over thirty lysosomal storage disorders have been documented,which are commonly grouped according to the accumulated substrate(s) (see Neufeld, 1991 for a recent review). The mode of inheritance of these disorders is autosomal recessive, except for Fabry disease (α-galactosidase deficiency) and Hunter syndrome (iduronate sulfatase deficiency), which are both X-linked. Individual lysosomal storage disorders are generally very rare but within defined isolated populations the incidence of a disease may be much higher. This was for example the case in the Ashkenazi-Jewish and French-Canadian populations, where G_{M2}-gangliosidosis (hexosaminidase A deficiency) frequently occurred but voluntary carrier detection has almost eliminated the disorder in these groups. The carrier frequency for a mutated hexosaminidase A allele is estimated to be 1:30 within the first group (Petersen *et al,* 1983).

A characteristic feature of lysosomal storage diseases is that even within one disorder patients may show widely variable clinical symptoms. Several attempts have been made over the years to correlate clinical heterogeneity to biochemical parameters. One model proposes a "critical threshold" of residual lysosomal enzyme activity, above which the mutated enzyme can cope with incoming substrate and postpone the most severe symptoms of a disease (Conzelmann and Sandhoff, 1983). The model demonstrates the importance of substrate influx in addition to residual enzyme activity in the development of a disorder. More recently the biochemical characterisation of lysosomal proteins has been facilitated and extended by the isolation of their corresponding cDNAs, which has concomitantly enabled the identification of different gene mutations involved in different variants of a lysosomal storage disorder. The major conclusions that can be drawn from all these studies carried out in many laboratories will be briefly summarized here.

It has been shown that many different mutations may underly a given lysosomal storage disorder, even within isolated populations such as the Ashkenazi-

Jewish. In the latter group, however, enrichment of few mutated alleles was also found. Therefore, the clinical and biochemical heterogeneity detected in earlier studies could in part be caused by genetic variability. From the data gathered on the two most extensively studied lysosomal enzyme deficiencies, namely those of β–hexosaminidase and β–glucocerebrosidase, one can deduce that the majority of the patients with a given lysosomal storage disorder will be compound heterozygotes, i.e. they carry two different mutated alleles. The combination of both alleles has to be considered in predictions of clinical outcome from genotype assesment. Only in a few cases such a genotype-phenotype correlation could be made. For example, in Gaucher disease, which is caused by a deficiency of the lysosomal enzyme β–glucocerebrosidase, the "Asn370 to Ser" mutation has thus far only been found in patients with the milder non-neurologic form of the disease (Tsuji *et al,* 1988). However, other data indicate that genotype-phenotype correlations are not always perfect. Homozygosity for the "Leu444 to Pro" β–glucocerebrosidase allele in patients with Gaucher disease is normally associated with a severe neuronopathic form (Firon *et al,* 1990; Tsuji *et al,* 1987; Wigderson *et al,* 1989), but in patients of Japanese origin the same mutation is associated with the non-neuronopathic type (Masuno *et al,* 1990).

Mutational analysis has further demonstrated that apparently similar protein deficiencies can be caused by different gene mutations. For example lack of a lysosomal protein in patients' tissues can be the result of gene deletions, gene rearrangements, splicing errors, or otherwise unstable mRNA. Other mutations might result in amino acid substitutions that cause improper folding of a lysosomal enzyme precursor, followed by its degradation in the ER. Only few mutations will allow residual functioning of a lysosomal protein, whereas the majority of alterations will completely inhibit enzymatic activity. Alleles giving rise to less severe disease will, therefore, probably be limited. It remains to be determined whether their presence in a population confers some kind of selective advantage to heterozygotes.

The advent of techniques enabling genotype assigment in a lysosomal storage disorder have added DNA based methods, in addition to enzymatic detection, to the field of prenatal diagnosis. Hopes for improved therapeutic methods for treatment of this group of diseases in the near future have been raised now that many cDNAs encoding lysosomal enzymes are available. Since this allows most lysosomal proteins to be overproduced and purified using recombinant DNA technology the original idea of enzyme replacement therapy (Hers, 1965) has received renewed interest. It is based on the fact that lysosomal enzymes can be taken up by

receptor-mediated endocytosis. In patients with a lysosomal storage disorder one could envisage the administration of sufficient amounts of purified normal enzyme that would be targeted to lysosomes and overcome the deficiency. A weekly infusion with purified human placental β–glucocerebrosidase, modified for endocytosis by macrophages, in a child with type I Gaucher disease demonstrated clinical improvement (Barton *et al,* 1990). However, nothing is known yet about the long term effects of these infusions. Therefore, the next stage of this therapeutic approach should be the very careful examination whether the structure of recombinant enzymes permits a regular intake by patients.

Bone marrow transplantation and gene replacement are being mentioned as other feasible therapeutic approaches. In both cases the target tissue is the bone marrow that contains the hematopoietic stem cells. The idea is that once differentiated and circulating these cells would provide normal enzyme to deficient cells through secretion or direct cell-to-cell contact. Bone marrow transplantation has been carried out and biochemical and clinical benefit was shown in some patients (Krivit *et al,* 1990). For gene replacement the attention is focused on achieving retrovirus-mediated integration of recombinant cDNA into hematopoietic progenitor cells of the patient. Whatever the therapy, it is still the consensus that the blood brain barrier can not be crossed by exogenous enzyme, therefore those patients with neurological symptoms may not be cured of their severe complaints.

Studies on any lysosomal storage disorder and on the effect of a certain therapeutic strategy for this disease would greatly benefit from the existence of animals, carrying a disease-producing mutation in the gene encoding the lysosomal protein of interest. However, such model systems of human disease have thus far been limited to the rarely detected animals with a naturally occurring lysosomal storage disorder. As it is possible now to "knock out" a gene by homologous recombination in (murine) embryonic stem cells (Thomas and Capecchi, 1987), or even to insert a selected point mutation in the chosen gene of interest via the "hit and run" procedure (Hasty *et al,* 1991), more of these model systems will become available in the near future. It is obvious that this will be of great benefit to the understanding of the factors that underly this heterogeneous class of diseases.

1.4 References

Achstetter, T., Ehmann, C., Osaki, A. and Wolf, D. H. (1984). "Proteolysis in eukaryotic cells. Proteinase yscE, a new yeast peptidase." *J Biol Chem* **259**, 13344-8.

Aerts, J. M., Brul, S., Donker-Koopman, W. E., van Weely, S., Murray, G. J., Barranger, J. A., Tager, J. M. and Schram, A. W. (1986). "Efficient routing of glucocerebrosidase to lysosomes requires complex oligosaccharide chain formation." *Biochem Biophys Res Commun* **141**, 452-8.

Aerts, J. M., Schram, A. W., Strijland, A., van Weely, S., Jonsson, L. M., Tager, J. M., Sorrell, S. H., Ginns, E. I., Barranger, J. A. and Murray, G. J. (1988). "Glucocerebrosidase, a lysosomal enzyme that does not undergo oligosaccharide phosphorylation." *Biochim Biophys Acta* **964**, 303-8.

Amara, J. F., Lederkremer, G. and Lodish, H. F. (1989). "Intracellular degradation of unassembled asialoglycoprotein receptor subunits: a pre-Golgi, nonlysosomal endoproteolytic cleavage." *J Cell Biol* **109**, 3315-24.

Amos, B. and Lotan, R. (1990). "Modulation of lysosomal-associated membrane glycoproteins during retinoic acid-induced embryonal carcinoma cell differentiation." *J Biol Chem* **265**, 19192-8.

Andres, D. A., Dickerson, I. M. and Dixon, J. E. (1990). "Variants of the carboxyl-terminal KDEL sequence direct intracellular retention." *J Biol Chem* **265**, 5952-5.

Aronson, N. J. and Kuranda, M. J. (1989). "Lysosomal degradation of Asn-linked glycoproteins." *Faseb J* **3**, 2615-22.

Bachmair, A., Finley, D. and Varshavsky, A. (1986). "In vivo half-life of a protein is a function of its amino-terminal residue." *Science* **234**, 179-86.

Bachmair, A. and Varshavsky, A. (1989). "The degradation signal in a short-lived protein." *Cell* **56**, 1019-32.

Ballard, F. J. (1987). Regulation of intracellular protein breakdown with special reference to cultured cells. In *Lysosomes: Their Role in Protein Breakdown.* H. Glaumann and F. J. Ballard eds. (London: Academic Press), pp 285-318.

Bapat, B., Ethier, M., Neote, K., Mahuran, D. and Gravel, R. A. (1988). "Cloning and sequence analysis of a cDNA encoding the β-subunit of mouse β-hexosaminidase." *Febs Lett* **237**, 191-5.

Baranski, T. J., Faust, P. L. and Kornfeld, S. (1990). "Generation of a lysosomal enzyme targeting signal in the secretory protein pepsinogen." *Cell* **63**, 281-91.

Barriocanal, J. G., Bonifacino, J. S., Yuan, L. and Sandoval, I. V. (1986). "Biosynthesis, glycosylation, movement through the Golgi system, and transport to lysosomes by an N-linked carbohydrate-independent mechanism of three lysosomal integral membrane proteins." *J Biol Chem* **261**, 16755-63.

Barton, N. W., Furbish, F. S., Murray, G. J., Garfield, M. and Brady, R. O. (1990). "Therapeutic response to intravenous infusions of glucocerebrosidase in a patient with Gaucher disease." *Proc Natl Acad Sci U S A* **87**, 1913-6.

Bishop, D. F., Calhoun, D. H., Bernstein, H. S., Hantzopoulos, P., Quinn, M. and Desnick, R. J. (1986). "Human α-galactosidase A: nucleotide sequence of a cDNA clone encoding the mature enzyme." *Proc Natl Acad Sci U S A* **83**, 4859-63.

Blobel, G. (1980). "Intracellular protein topogenesis." *Proc Natl Acad Sci U S A* **77**, 1496-500.

Block, M. R., Glick, B. S., Wilcox, C. A., Wieland, F. T. and Rothman, J. E. (1988). "Purification of an N-ethylmaleimide-sensitive protein catalyzing vesicular transport." *Proc Natl Acad Sci U S A* **85**, 7852-6.

Bonifacino, J. S., Cosson, P. and Klausner, R. D. (1990a). "Colocalized transmembrane determinants for ER degradation and subunit assembly explain the intracellular fate of TCR chains." *Cell* **63**, 503-13.

Bonifacino, J. S., Suzuki, C. K. and Klausner, R. D. (1990b). "A peptide sequence confers retention and rapid degradation in the endoplasmic reticulum." *Science* **247**, 79-82.

Booth, C. and Koch, G. L. (1989). "Perturbation of cellular calcium induces secretion of luminal ER proteins." *Cell* **59**, 729-37.

Bourne, H. R. (1988). "Do GTPases direct membrane traffic in secretion?" *Cell* **53**, 669-71.

Braun, M., Waheed, A. and von Figura, K. (1989). "Lysosomal acid phosphatase is transported to lysosomes via the cell surface." *Embo J* **8**, 3633-40.

Brodsky, F. M. (1988). "Living with clathrin: its role in intracellular membrane traffic." *Science* **242**, 1396-402.

Brodsky, F. M., Hill, B. L., Acton, S. L., Nathke, I., Wong, D. H., Ponnambalam, S. and Parham, P. (1991). "Clathrin Light Chains - Arrays of Protein Motifs That Regulate Coated-Vesicle Dynamics." *Trends in Biochemical Sciences* **16**, 208-213.

Canfield, W. M., Johnson, K. F., Ye, R. D., Gregory, W. and Kornfeld, S. (1991). "Localization of the signal for rapid internalization of the bovine cation-independent mannose 6-phosphate/insulin-like growth factor-II receptor to amino acids 24-29 of the cytoplasmic tail." *J Biol Chem* **266**, 5682-88.

Canfield, W. M. and Kornfeld, S. (1989). "The chicken liver cation-independent mannose 6-phosphate receptor lacks the high affinity binding site for insulin-like growth factor II." *J Biol Chem* **264**, 7100-3.

Chao, H. H., Waheed, A., Pohlmann, R., Hille, A. and von Figura, K. (1990). "Mannose 6-phosphate receptor dependent secretion of lysosomal enzymes." *Embo J* **9**, 3507-13.

Chau, V., Tobias, J. W., Bachmair, A., Marriott, D., Ecker, D. J., Gonda, D. K. and Varshavsky, A. (1989). "A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein." *Science* **243**, 1576-83.

Chen, W. J., Goldstein, J. L. and Brown, M. S. (1990). "NPXY, a sequence often found in cytoplasmic tails, is required for coated pit-mediated internalization of the low density lipoprotein receptor." *J Biol Chem* **265**, 3116-23.

Chiang, H. L. and Schekman, R. (1991). "Regulated import and degradation of a cytosolic protein in the yeast vacuole." *Nature* **350**, 313-8.

Ciechanover, A. and Schwartz, A. L. (1989). "How are substrates recognized by the ubiquitin-mediated proteolytic system?" *Trends Biochem Sci* **14**, 483-8.

Collard, M. W., Sylvester, S. R., Tsuruta, J. K. and Griswold, M. D. (1988). "Biosynthesis and molecular cloning of sulfated glycoprotein 1 secreted by rat Sertoli cells: sequence similarity with the 70-kilodalton precursor to sulfatide/GM1 activator." *Biochemistry* **27**, 4557-64.

Connolly, T. and Gilmore, R. (1989). "The signal recognition particle receptor mediates the GTP-dependent displacement of SRP from the signal sequence of the nascent polypeptide." *Cell* **57**, 599-610.

Conzelmann, E. and Sandhoff, K. (1983). "Partial enzyme deficiencies: residual activities and the development of neurological disorders." *Dev Neurosci* **6**, 58-71.

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**, 7131-40.

Diaz, R., Mayorga, L. S., Weidman, P. J., Rothman, J. E. and Stahl, P. D. (1989). "Vesicle fusion following receptor-mediated endocytosis requires a protein active in Golgi transport." *Nature* **339**, 398-400.

Dice, J. F. (1987). "Molecular determinants of protein half-lives in eukaryotic cells [published erratum appears in FASEB J 1988 Apr;2(7):2262]." *Faseb J* **1**, 349-57.

Dice, J. F. (1990). "Peptide sequences that target cytosolic proteins for lysosomal proteolysis." *Trends Biochem Sci* **15**, 305-9.

Dong, J. M., Prence, E. M. and Sahagian, G. G. (1989). "Mechanism for selective secretion of a lysosomal protease by transformed mouse fibroblasts." *J Biol Chem* **264**, 7377-83.

Driscoll, J. and Goldberg, A. L. (1990). "The proteasome (multicatalytic protease) is a component of the 1500-kDa proteolytic complex which degrades ubiquitin-conjugated proteins." *J Biol Chem* **265**, 4789-92.

Duden, R., Griffiths, G., Frank, R., Argos, P. and Kreis, T. E. (1991). "β-COP, a 110 kd protein associated with non-clathrin-coated vesicles and the Golgi complex, shows homology to β-adaptin." *Cell* **64**, 649-65.

Dunn, W. J. (1990a). "Studies on the mechanisms of autophagy: formation of the autophagic vacuole." *J Cell Biol* **110**, 1923-33.

Dunn, W. J. (1990b). "Studies on the mechanisms of autophagy: maturation of the autophagic vacuole." *J Cell Biol* **110**, 1935-45.

Eakle, K. A., Bernstein, M. and Emr, S. D. (1988). "Characterization of a component of the yeast secretion machinery: identification of the SEC18 gene product." *Mol Cell Biol* **8**, 4098-109.

Edman, J. C., Ellis, L., Blacher, R. W., Roth, R. A. and Rutter, W. J. (1985). "Sequence of protein disulphide isomerase and implications of its relationship to thioredoxin." *Nature* **317**, 267-70.

Elbein, A. D. (1987). "Inhibitors of the biosynthesis and processing of N-linked oligosaccharide chains." *Annu Rev Biochem* **56**, 497-534.

Ellis, R. J., van der Vies, S. M. and Hemmingsen, S. M. (1989). "The molecular chaperone concept." *Biochem Soc Symp* **55**, 145-53.

Ellis, R. J. and van der Vies, S. M. (1991). "Molecular Chaperones." *Annual Review of Biochemistry* **60**, 321-47.

Erickson, A. H. (1989). "Biosynthesis of lysosomal endopeptidases." *J Cell Biochem* **40**, 31-41.

Fagan, J. M., Waxman, L. and Goldberg, A. L. (1986). "Red blood cells contain a pathway for the degradation of oxidant-damaged hemoglobin that does not require ATP or ubiquitin." *J Biol Chem* **261**, 5705-13.

Fagan, J. M., Waxman, L. and Goldberg, A. L. (1987). "Skeletal muscle and liver contain a soluble ATP + ubiquitin-dependent proteolytic system." *Biochem J* **243**, 335-43.

Faust, P. L. and Kornfeld, S. (1989). "Expression of the yeast aspartyl protease, proteinase A. Phosphorylation and binding to the mannose 6-phosphate receptor are altered by addition of cathepsin D sequences." *J Biol Chem* **264**, 479-88.

Faust, P. L., Kornfeld, S. and Chirgwin, J. M. (1985). "Cloning and sequence analysis of cDNA for human cathepsin D." *Proc Natl Acad Sci U S A* **82**, 4910-4.

Febbraio, M. and Silverstein, R. L. (1990). "Identification and characterization of LAMP-1 as an activation-dependent platelet surface glycoprotein." *J Biol Chem* **265**, 18531-7.

Firon, N., Eyal, N., Kolodny, E. H. and Horowitz, M. (1990). "Genotype assignment in Gaucher disease by selective amplification of the active glucocerebrosidase gene." *Am J Hum Genet* **46**, 527-32.

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

Fisher, K. J. and Aronson, N. N. (1991). "Characterization of the mutation responsible for aspartylglucosaminuria in three finnish patients - amino acid substitution Cys163->Ser abolishes the activity of lysosomal glycosylasparaginase and its conversion into subunits." *Journal of Biological Chemistry* **266**, 12105-12113.

Freedman, R. B. (1989). "Protein disulfide isomerase: multiple roles in the modification of nascent secretory proteins." *Cell* **57**, 1069-72.

Fujiwara, T., Tanaka, K., Orino, E., Yoshimura, T., Kumatori, A., Tamura, T., Chung, C. H., Nakai, T., Yamaguchi, K., Shin, S. and et, a. I. (1990). "Proteasomes are essential for yeast proliferation. cDNA cloning and gene disruption of two major subunits." *J Biol Chem* **265**, 16604-13.

Furst, W., Machleidt, W. and Sandhoff, K. (1988). "The precursor of sulfatide activator protein is processed to three different proteins." *Biol Chem Hoppe Seyler* **369**, 317-28.

Gahl, W. A., Bashan, N., Tietze, F., Bernardini, I. and Schulman, J. D. (1982). "Cystine transport is defective in isolated leucocyte lysosomes from patients with cystinosis." *Science* **217**, 1263-65.

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

Ghoda, L., van Daalen-Wetters, T., Macrae, M., Ascherman, D. and Coffino, P. (1989). "Prevention of rapid intracellular degradation of ODC by a carboxyl-terminal truncation." *Science* **243**, 1493-5.

Gilmore, R. and Blobel, G. (1983). "Transient involvement of signal recognition particle and its receptor in the microsomal membrane prior to protein translocation." *Cell* **35**, 677-85.

Gilmore, R. and Blobel, G. (1985). "Translocation of secretory proteins across the microsomal membrane occurs through an environment accessible to aqueous perturbants." *Cell* **42**, 497-505.

Gilmore, R., Walter, P. and Blobel, G. (1982a). "Protein translocation accross the endoplasmic reticulum I. Detection in the microsomal membrane of a receptor for the signal recognition particle." *J Cell Biol* **95**, 463-69.

Gilmore, R., Walter, P. and Blobel, G. (1982b). "Protein translocation across the endoplasmic reticulum. II. Isolation and characterization of the signal recognition particle receptor." *J Cell Biol* **95**, 470-77.

Glickman, J. N., Conibear, E. and Pearse, B. M. (1989). "Specificity of binding of clathrin adaptors to signals on the mannose-6-phosphate/insulin-like growth factor II receptor." *Embo J* **8**, 1041-7.

Gorvel, J. P., Chavrier, P., Zerial, M. and Gruenberg, J. (1991). "rab5 Controls Early Endosome Fusion Invitro." *Cell* **64**, 915-925.

Gottschalk, S., Waheed, A., Schmidt, B., Laidler, P. and von Figura, K. (1989). "Sequential processing of lysosomal acid phosphatase by a cytoplasmic thiol proteinase and a lysosomal aspartyl proteinase." *Embo J* **8**, 3215-9.

Graham, T. R., Zassenhaus, H. P. and Kaplan, A. (1988). "Molecular cloning of the cDNA which encodes β-N-acetylhexosaminidase A from Dictyostelium discoideum. Complete amino acid sequence and homology with the human enzyme." *J Biol Chem* **263**, 16823-9.

Granger, B., Green, S. A., Gabel, C. A., Howe, C. L., Mellman, I. and Helenius, A. (1990). "Characterization and cloning of lgp110, a lysosomal membrane glycoprotein from mouse and rat cells." *J Biol Chem* **265**, 12036-43.

Griffiths, G., Hoflack, B., Simons, K., Mellman, I. and Kornfeld, S. (1988). "The mannose 6-phosphate receptor and the biogenesis of lysosomes." *Cell* **52**, 329-41.

Griffiths, G. and Simons, K. (1986). "The trans Golgi network: sorting at the exit site of the Golgi complex." *Science* **234**, 438-43.

Gronostajski, R. M., Pardee, A. B. and Goldberg, A. L. (1985). "The ATP dependence of the degradation of short- and long-lived proteins in growing fibroblasts." *J Biol Chem* **260**, 3344-9.

Gropper, R., Brandt, R. A., Elias, S., Bearer, C. F., Mayer, A., Schwartz, A. L. and Ciechanover, A. (1991). "The Ubiquitin-Activating Enzyme, E1, Is Required for Stress- Induced Lysosomal Degradation of Cellular Proteins." *Journal of Biological Chemistry* **266**, 3602-3610.

Gruenberg, J., Griffiths, G. and Howell, K. E. (1989a). "Characterization of the early endosome and putative endocytic carrier vesicles in vivo and with an assay of vesicle fusion in vitro." *J Cell Biol* **108**, 1301-16.

Gruenberg, J. and Howell, K. E. (1989b). "Membrane traffic in endocytosis: insights from cell-free assays." *Annu Rev Cell Biol* **5**, 453-81.

Haas, I. G. and Wabl, M. (1983). "Immunoglobulin heavy chain binding protein." *Nature* **306**, 387-9.

Haig, D. and Graham, C. (1991). "Genomic imprinting and the strange case of the insulin-like growth factor II receptor." *Cell* **64**, 1045-6.

Hasilik, A. and Neufeld, E. F. (1980a). "Biosynthesis of lysosomal enzymes in fibroblasts. Phosphorylation of mannose residues." *J Biol Chem* **255**, 4946-50.

Hasilik, A. and Neufeld, E. F. (1980b). "Biosynthesis of lysosomal enzymes in fibroblasts. Synthesis as precursors of higher molecular weight." *J Biol Chem* **255**, 4937-45.

Hasty, P., Ramirez, S. R., Krumlauf, R. and Bradley, A. (1991). "Introduction of a subtle mutation into the Hox-2.6 locus in embryonic stem cells." *Nature* **350**, 243-6.

Heinemeyer, W., Kleinschmidt, J. A., Saidowsky, J., Escher, C. and Wolf, D. H. (1991). "Proteinase YscE, the yeast proteasome/multicatalytic-multifunctional proteinase: mutants unravel its function in stress induced proteolysis and uncover its necessity for cell survival." *EMBO J* **10**, 555-562.

Hers, H. G. (1965). "Inborn lysosomal diseases." *Gastroenterology* **48**, 625-33.

Hicke, L. and Schekman, R. (1990). "Molecular machinery required for protein transport from the endoplasmic reticulum to the Golgi complex." *Bioessays* **12**, 253-8.

Ho, M. W. and O'Brien, J. S. (1971). "Gaucher's disease: deficiency of 'acid' β-glucosidase and reconstitution of enzyme activity in vitro." *Proc Natl Acad Sci USA* **68**, 2810-13.

Hoefsloot, L. H., Hoogeveen-Westerveld, M., Kroos, M. A., van Beeumen, J., Reuser, A. J. and Oostra, B. A. (1988). "Primary structure and processing of lysosomal α-glucosidase; homology with the intestinal sucrase-isomaltase complex." *Embo J* **7**, 1697-704.

Hoflack, B. and Kornfeld, S. (1985). "Purification and characterization of a cation-dependent mannose 6-phosphate receptor from murine P388D1 macrophages and bovine liver." *J Biol Chem* **260**, 12008-14.

Holtschmidt, H., Sandhoff, K., Kwon, H. Y., Harzer, K., Nakano, T. and Suzuki, K. (1991). "Sulfatide activator protein - alternative splicing that generates three messenger RNAs and a newly found mutation responsible for a clinical disease." *Journal of Biological Chemistry* **266**, 7556-7560.

Hopkins, C. R., Gibson, A., Shipman, M. and Miller, K. (1990). "Movement of internalized ligand-receptor complexes along a continuous endosomal reticulum." *Nature* **346**, 335-9.

Horst, M. and Hasilik, A. (1991). "Expression and maturation of human cathepsin-D in baby- hamster kidney cells." *Biochemical Journal* **273**, 355-361.

Hough, R., Pratt, G. and Rechsteiner, M. (1987). "Purification of two high molecular weight proteases from rabbit reticulocyte lysate." *J Biol Chem* **262**, 8303-13.

Hurtley, S. M. and Helenius, A. (1989). "Protein oligomerization in the endoplasmic reticulum." *Annu Rev Cell Biol* **5**, 277-307.

Idriss, J.-M. and Jonas, A. J. (1991). "Vitamin B12 transport by rat liver lysosomal membrane vesicles." *J Biol Chem* **266**, 9438-41.

Jackson, M. R., Nilsson, T. and Peterson, P. A. (1990). "Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum." *Embo J* **9**, 3153-62.

Jentsch, S., Seufert, W. and Hauser, H. P. (1991). "Genetic Analysis of the Ubiquitin System." *Biochimica et Biophysica Acta* **1089**, 127-139.

Jentsch, S., Seufert, W., Sommer, T. and Reins, H. A. (1990). "Ubiquitin-conjugating enzymes: novel regulators of eukaryotic cells." *Trends Biochem Sci* **15**, 195-8.

Johnson, K. F., Chan, W. and Kornfeld, S. (1990). "Cation-dependent mannose 6-phosphate receptor contains two internalization signals in its cytoplasmic domain [published erratum appears in Proc Natl Acad Sci U S A 1991 Feb 15;88(4):1591]." *Proc Natl Acad Sci U S A* **87**, 10010-4.

Kaplan, A., Acord, D. T. and Sly, W. S. (1977). "Phosphohexosyl components of a lysosomal enzyme are recognized by pinocytosis receptors on human fibroblasts." *Proc Natl Acad Sci USA* **74**, 2026-30.

Keen, J. H. (1990). "Clathrin and associated assembly and disassembly proteins." *Annu Rev Biochem* **59**, 415-38.

Kelly, R. B. (1990). "Microtubules, membrane traffic, and cell organization." *Cell* **61**, 5-7.

Kirschke, H. and Barrett, A. J. (1987). Chemistry of lysosomal proteases. In *Lysosomes: Their Role in Protein Breakdown.* H. Glaumann and F. J. Ballard eds. (London: Academic Press), pp 193-238.

Klausner, R. D. and Sitia, R. (1990). "Protein degradation in the endoplasmic reticulum." *Cell* **62**, 611-4.

Klionsky, D. J., Herman, P. K. and Emr, S. D. (1990). "The fungal vacuole: composition, function, and biogenesis." *Microbiol Rev* **54**, 266-92.

Koch, G. L. (1987). "Reticuloplasmins: a novel group of proteins in the endoplasmic reticulum [editorial]." *J Cell Sci* **87**, 491-2.

Koch, G. L. (1990). "The endoplasmic reticulum and calcium storage." *Bioessays* **12**, 527-31.

Korneluk, R. G., Mahuran, D. J., Neote, K., Klavins, M. H., O'Dowd, B. F., Tropak, M., Willard, H. F., Anderson, M. J., Lowden, J. A. and Gravel, R. A. (1986). "Isolation of cDNA clones coding for the α-subunit of human β-hexosaminidase. Extensive homology between the α- and β-subunits and studies on Tay-Sachs disease." *J Biol Chem* **261**, 8407-13.

Kornfeld, R. and Kornfeld, S. (1985). "Assembly of asparagine-linked oligosaccharides." *Annu Rev Biochem* **54**, 631-64.

Kornfeld, S. and Mellman, I. (1989). "The biogenesis of lysosomes." *Annu Rev Cell Biol* **5**, 483-525.

Kornreich, R., Desnick, R. J. and Bishop, D. F. (1989). "Nucleotide sequence of the human α-galactosidase A gene." *Nucleic Acids Res* **17**, 3301-2.

Kreysing, J., von Figura, K. and Gieselmann, V. (1990). "Structure of the arylsulfatase A gene." *Eur J Biochem* **191**, 627-31.

Krivit, W., Shapiro, E., Kennedy, W., Lipton, M., Lockman, L., Smith, S., Summers, C. G., Wenger, D. A., Tsai, M. Y., Ramsay, N. K. and et, a. l. (1990). "Treatment of late infantile metachromatic leukodystrophy by bone marrow transplantation." *N Engl J Med* **322**, 28-32.

Kuranda, M. J. and Aronson, J., N. N. (1987). Receptor-mediated endocytosis and lysosoaml degradation of asialoglycoproteins by the liver. In *Lysosomes: Their Role in Protein Breakdown*. H. Glaumann and F. J. Ballard eds. (London: Academic Press), pp 241-82.

Landry, S. J. and Gierasch, L. M. (1991). "Recognition of nascent polypeptides for targeting and folding." *Trends Biochem Sci* **16**, 159-63.

Lau, M. M. and Neufeld, E. F. (1989). "A frameshift mutation in a patient with Tay-Sachs disease causes premature termination and defective intracellular transport of the α-subunit of β-hexosaminidase." *J Biol Chem* **264**, 21376-80.

Lazzarino, D. A. and Gabel, C. A. (1988). "Biosynthesis of the mannose 6-phosphate recognition marker in transport-impaired mouse lymphoma cells. Demonstration of a two-step phosphorylation." *J Biol Chem* **263**, 10118-26.

Lazzarino, D. A. and Gabel, C. A. (1990). "Protein determinants impair recognition of procathepsin L phosphorylated oligosaccharides by the cation-independent mannose 6-phosphate receptor." *J Biol Chem* **265**, 11864-71.

Lederkremer, G. Z. and Lodish, H. F. (1991). "An alternatively spliced miniexon alters the subcellular fate of the human asialoglycoprotein receptor H2 subunit. Endoplasmic reticulum retention and degradation or cell surface expression." *J Biol Chem* **266**, 1237-44.

Lemons, R. M. and Thoene, J. G. (1991). "Mediated calcium transport by isolated human fibroblasts lysosomes." *J Biol Chem* **266**, 14378-82.

Lewis, M. J. and Pelham, H. R. (1990a). "A human homologue of the yeast HDEL receptor." *Nature* **348**, 162-3.

Lewis, M. J., Sweet, D. J. and Pelham, H. R. (1990b). "The ERD2 gene determines the specificity of the luminal ER protein retention system." *Cell* **61**, 1359-63.

Li, S. C. and Li, Y. T. (1976). "An activator stimulating the enzymic hydrolysis of sphingoglycolipids." *J Biol Chem* **251**, 1159-63.

Lippincott-Schwartz, J., Bonifacino, J. S., Yuan, L. C. and Klausner, R. D. (1988). "Degradation from the endoplasmic reticulum: disposing of newly synthesized proteins." *Cell* **54**, 209-20.

Lippincott-Schwartz, J., Donaldson, J. G., Schweizer, A., Berger, E. G., Hauri, H. P., Yuan, L. C. and Klausner, R. D. (1990). "Microtubule-dependent retrograde transport of proteins into the ER in the presence of brefeldin A suggests an ER recycling pathway." *Cell* **60**, 821-36.

Lippincott-Schwartz, J., Yuan, L. C., Bonifacino, J. S. and Klausner, R. D. (1989). "Rapid redistribution of Golgi proteins into the ER in cells treated with brefeldin A: evidence for membrane cycling from Golgi to ER." *Cell* **56**, 801-13.

Lobel, P., Fujimoto, K., Ye, R. D., Griffiths, G. and Kornfeld, S. (1989). "Mutations in the cytoplasmic domain of the 275 kd mannose 6-phosphate receptor differentially alter lysosomal enzyme sorting and endocytosis." *Cell* **57**, 787-96.

Ma, Z., Grubb, J. H. and Sly, W. S. (1991). "Cloning, sequencing and functional characterization of the murine 46-kDa mannose 6-phosphate receptor." *J Biol Chem* **266**, 10589-95.

Malhotra, V., Orci, L., Glick, B. S., Block, M. R. and Rothman, J. E. (1988). "Role of an N-ethylmaleimide-sensitive transport component in promoting fusion of transport vesicles with cisternae of the Golgi stack." *Cell* **54**, 221-7.

Mancini, G. M. S., Beerens, C. E. M. T. and Verheijen, F. W. (1990). "Glucose transport in lysosomal membrane vesicles. Kinetic demonstration of a carrier for neutral hexoses." *J Biol Chem* **265**, 12380-87.

Mancini, G. M. S., de Jonge, H. R., Galjaard, H. and Verheijen, F. W. (1989). "Characterization of a proton-driven carrier for sialic acid in the lysosomal membrane." *J Biol Chem* **264**, 15247-54.

Marzella, L. and Glaumann, H. (1987). Autophagy, microautophagy and crinophagy as mechanisms for protein degradation. In *Lysosomes: Their Role in Protein Breakdown*. H. Glaumann and F. J. Ballard eds. (London: Academic Press), pp 319-67.

Masuno, M., Tomatsu, S., Sukegawa, K. and Orii, T. (1990). "Non-existence of a tight association between a 444leucine to proline mutation and phenotypes of Gaucher disease: high frequency of a NciI polymorphism in the non-neuronopathic form." *Hum Genet* **84**, 203-6.

Meier, E. M., Schwarzmann, G., Furst, W. and Sandhoff, K. (1991). "The Human GM2 Activator Protein - A Substrate Specific Cofactor of β-Hexosaminidase-A." *Journal of Biological Chemistry* **266**, 1879-1887.

Meyer, D. I., Krause, E. and Dobberstein, B. (1982). "Secretory protein translocation across membranes-the role of the 'docking protein'." *Nature* **297**, 647-50.

Miller, R. D., Hoffmann, J. W., Powell, P. P., Kyle, J. W., Shipley, J. M., Bachinsky, D. R. and Sly, W. S. (1990). "Cloning and characterization of the human β-glucuronidase gene." *Genomics* **7**, 280-3.

Morgan, D. O., Edman, J. C., Standring, D. N., Fried, V. A., Smith, M. C., Roth, R. A. and Rutter, W. J. (1987). "Insulin-like growth factor II receptor as a multifunctional binding protein [published erratum appears in Nature 1988 Jul;20(7):442]." *Nature* **329**, 301-7.

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* **264**, 20655-63.

Mortimore, G. E. (1987). Mechanism and regulation of induced and basal protein degradation in liver. In *Lysosomes: Their Role in Protein Breakdown.* H. Glaumann and F. J. Ballard eds. (London: Academic Press), pp 415-43.

Mortimore, G. E. and Khurana, K. K. (1990). "Regulation of protein degradation in the liver." *Int J Biochem* **22**, 1075-80.

Munro, S. and Pelham, H. R. (1986). "An Hsp70-like protein in the ER: identity with the 78 kd glucose-regulated protein and immunoglobulin heavy chain binding protein." *Cell* **46**, 291-300.

Munro, S. and Pelham, H. R. (1987). "A C-terminal signal prevents secretion of luminal ER proteins." *Cell* **48**, 899-907.

Myerowitz, R., Piekarz, R., Neufeld, E. F., Shows, T. B. and Suzuki, K. (1985). "Human β-hexosaminidase α-chain: coding sequence and homology with the β-chain." *Proc Natl Acad Sci U S A* **82**, 7830-4.

Nakano, T., Sandhoff, K., Stumper, J., Christomanou, H. and Suzuki, K. (1989). "Structure of full-length cDNA coding for sulfatide activator, a Co-β-glucosidase and two other homologous proteins: two alternate forms of the sulfatide activator." *J Biochem (Tokyo)* **105**, 152-4.

Neote, K., Bapat, B., Dumbrille, R. A., Troxel, C., Schuster, S. M., Mahuran, D. J. and Gravel, R. A. (1988). "Characterization of the human HEXB gene encoding lysosomal β-hexosaminidase." *Genomics* **3**, 279-86.

Neufeld, E. F. (1991). "Lysosomal Storage Diseases." *Annual Review of Biochemistry* **60**, 257-80.

Nguyen, T. H., Law, D. T. and Williams, D. B. (1991). "Binding protein BiP is required for translocation of secretory proteins into the endoplasmic reticulum in Saccharomyces cerevisiae." *Proc Natl Acad Sci U S A* **88**, 1565-9.

Nicchitta, C. V., Migliaccio, G. and Blobel, G. (1991). "Biochemical fractionation and assembly of the membrane components that mediate nascent chain targeting and translocation." *Cell* **65**, 587-98.

Nilsson, T., Jackson, M. and Peterson, P. A. (1989). "Short cytoplasmic sequences serve as retention signals for transmembrane proteins in the endoplasmic reticulum." *Cell* **58**, 707-18.

Nishimura, Y., Rosenfeld, M. G., Kreibich, G., Gubler, U., Sabatini, D. D., Adesnik, M. and Andy, R. (1986). "Nucleotide sequence of rat preputial gland β-glucuronidase cDNA and in vitro insertion of its encoded polypeptide into microsomal membranes." *Proc Natl Acad Sci U S A* **83**, 7292-6.

O'Brien, J. S., Kretz, K. A., Dewji, N., Wenger, D. A., Esch, F. and Fluharty, A. L. (1988). "Coding of two sphingolipid activator proteins (SAP-1 and SAP-2) by same genetic locus." *Science* **241**, 1098-101.

Ohkuma, S. (1987). The lysosomal proton pump and its effect on protein breakdown. In *Lysosomes: Their Role in Protein Breakdown.* H. Glaumann and F. J. Ballard eds. (London: Academic Press), pp 115-48.

Orci, L., Ravazzola, M., Amherdt, M., Perrelet, A., Powell, S. K., Quinn, D. L. and Moore, H. P. (1987). "The trans-most cisternae of the Golgi complex: a compartment for sorting of secretory and plasma membrane proteins." *Cell* **51**, 1039-51.

Orci, L., Tagaya, M., Amherdt, M., Perrelet, A., Donaldson, J. G., Lippincott-Schwartz, J., Klausner, R. D. and Rothman, J. E. (1991). "Brefeldin A, a drug that blocks secretion, prevents the assembly of non-clathrin-coated buds on Golgi cisternae." *Cell* **64**, 1183-95.

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). "Cloning, sequencing, and expression of cDNA for human β-glucuronidase." *Proc Natl Acad Sci U S A* **84**, 685-9.

Paabo, S., Bhat, B. M., Wold, W. S. and Peterson, P. A. (1987). "A short sequence in the COOH-terminus makes an adenovirus membrane glycoprotein a resident of the endoplasmic reticulum." *Cell* **50**, 311-7.

Paulson, J. C. and Colley, K. J. (1989). "Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation." *J Biol Chem* **264**, 17615-8.

Pearse, B. M. (1987). "Clathrin and coated vesicles." *Embo J* **6**, 2507-12.

Pelham, H. R. (1989). "Control of protein exit from the endoplasmic reticulum." *Annu Rev Cell Biol* **5**, 1-24.

Pelham, H. R., Hardwick, K. G. and Lewis, M. J. (1988). "Sorting of soluble ER proteins in yeast." *Embo J* **7**, 1757-62.

Peters, C., Braun, M., Weber, B., Wendland, M., Schmidt, B., Pohlmann, R., Waheed, A. and von Figura, K. (1990). "Targeting of a lysosomal membrane protein: a tyrosine-containing endocytosis signal in the cytoplasmic tail of lysosomal acid phosphatase is necessary and sufficient for targeting to lysosomes." *Embo J* **9**, 3497-506.

Peters, C., Geier, C., Pohlmann, R., Waheed, A., von Figura, K., Roiko, K., Virkkunen, P., Henttu, P. and Vihko, P. (1989). "High degree of homology between primary structure of human lysosomal acid phosphatase and human prostatic acid phosphatase." *Biol Chem Hoppe Seyler* **370**, 177-81.

Peters, C., Schmidt, B., Rommerskirch, W., Rupp, K., Zuhlsdorf, M., Vingron, M., Meyer, H. E., Pohlmann, R. and von Figura, K. (1990). "Phylogenetic conservation of arylsulfatases. cDNA cloning and expression of human arylsulfatase B." *J Biol Chem* **265**, 3374-81.

Petersen, G. M., Rotter, J. I., Cantor, R. M., Field, L. L., Greenwald, S., Lim, J. S. T., Roy, C., Schoenfeld, V., Lowden, J. A. and Kaback, M. M. (1983). "The Tay-Sachs disease gene in North American Jewish populations: geographic variation and origin." *Am J Hum Genet* **35**, 1258-69.

Pisoni, R. L. (1991). "Characterization of a Phosphate Transport System in Human Fibroblast Lysosomes." *Journal of Biological Chemistry* **266**, 979-985.

Pohlmann, R., Krentler, C., Schmidt, B., Schroder, W., Lorkowski, G., Culley, J., Mersmann, G., Geier, C., Waheed, A., Gottschalk, S., Grzeschik, K.-H., Hasilik, A. and von Figura, K. (1988). "Human lysosomal acid phosphatase: cloning, expression and chromosomal assignment." *Embo J* **7**, 2343-50.

Powell, P. P., Kyle, J. W., Miller, R. D., Pantano, J., Grubb, J. H. and Sly, W. S. (1988). "Rat liver β-glucuronidase. cDNA cloning, sequence comparisons and expression of a chimeric protein in COS cells." *Biochem J* **250**, 547-55.

Proia, R. L. (1988). "Gene encoding the human β-hexosaminidase β–chain: extensive homology of intron placement in the α- and β-chain genes." *Proc Natl Acad Sci U S A* **85**, 1883-7.

Proia, R. L. and Soravia, E. (1987). "Organization of the gene encoding the human β-hexosaminidase α-chain [published erratum appears in J Biol Chem 1987 Nov 5;262(31):15322]." *J Biol Chem* **262**, 5677-81.

Quintern, L. E., Schuchman, E. H., Levran, O., Suchi, M., Ferlinz, K., Reinke, H., Sandhoff, K. and Desnick, R. J. (1989). "Isolation of cDNA clones encoding human acid sphingomyelinase: occurrence of alternatively processed transcripts." *Embo J* **8**, 2469-73.

Rechsteiner, M. (1987). "Ubiquitin-mediated pathways for intracellular proteolysis." *Annu Rev Cell Biol* **3**, 1-30

Reiner, O., Wigderson, M. and Horowitz, M. (1988). "Structural analysis of the human glucocerebrosidase genes." *DNA* **2**, 107-16.

Rijnboutt, S., Aerts, H. M., Geuze, H. J., Tager, J. M. and Strous, G. J. (1991). "Mannose 6-phosphate-independent membrane association of cathepsin D, glucocerebrosidase, and sphingolipid-activating protein in HepG2 cells." *J Biol Chem* **266**, 4862-8.

Ritonja, A., Colic, A., Dolenc, I., Ogrinc, T., Podobnik, M. and Turk, V. (1991). "The Complete Amino Acid Sequence of Bovine Cathepsin-S and a Partial Sequence of Bovine Cathepsin-L." *Febs Letters* **283**, 329-331.

Rivett, A. J. (1989a). "High molecular mass intracellular proteases." *Biochem J* **263**, 625-33.

Rivett, A. J. (1989b). "The multicatalytic proteinase of mammalian cells." *Arch Biochem Biophys* **268**, 1-8.

Rivett, A. J. (1990). "Intracellular protein degradation." *Essays Biochem* **25**, 39-81.

Robertson, D. A., Freeman, C., Nelson, P. V., Morris, C. P. and Hopwood, J. J. (1988). "Human glucosamine-6-sulfatase cDNA reveals homology with steroid sulfatase." *Biochem Biophys Res Commun* **157**, 218-24.

Rochefort, H., Capony, F., Garcia, M., Cavailles, V., Freiss, G., Chambon, M., Morisset, M. and Vignon, F. (1987). "Estrogen-induced lysosomal proteases secreted by breast cancer cells: a role in carcinogenesis?" *J Cell Biochem* **35**, 17-29.

Rogers, S., Wells, R. and Rechsteiner, M. (1986). "Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis." *Science* **234**, 364-8.

Rogers, S. W. and Rechsteiner, M. (1988a). "Degradation of structurally characterized proteins injected into HeLa cells. Basic measurements." *J Biol Chem* **263**, 19833-42.

Rogers, S. W. and Rechsteiner, M. (1988b). "Degradation of structurally characterized proteins injected into HeLa cells. Effects of intracellular location and the involvement of lysosomes." *J Biol Chem* **263**, 19843-9.

Rogers, S. W. and Rechsteiner, M. (1988c). "Degradation of structurally characterized proteins injected into HeLa cells. Tests of hypotheses." *J Biol Chem* **263**, 19850-62.

Rothman, J. E. (1989). "Polypeptide chain binding proteins: catalysts of protein folding and related processes in cells." *Cell* **59**, 591-601.

Rothman, J. E. (1991). "Enzymology of intracellular membrane fusion." *Klin Wochenschr* **69**, 98-104.

Sahagian, G. G., Distler, J. and Jourdian, G. W. (1981). "Characterization of a membrane-associated receptor from bovine liver that binds phosphomannosyl residues of bovine testicular β-galactosidase." *Proc Natl Acad Sci U S A* **78**, 4289-93.

San Segundo, B., Chan, S. J. and Steiner, D. F. (1986). "Differences in cathepsin B mRNA levels in rat tissues suggest specialized functions." *Febs Lett* **201**, 251-6.

Savitz, A. J. and Meyer, D. I. (1990). "Identification of a ribosome receptor in the rough endoplasmic reticulum." *Nature* **346**, 540-4.

Schauer, R. (1985). "Sialic acids and their role as biological masks." *Trends Biochem Sci* **10**, 357-60.

Schmid, S. L., Fuchs, R., Male, P. and Mellman, I. (1988). "Two distinct subpopulations of endosomes involved in membrane recycling and transport to lysosomes." *Cell* **52**, 73-83.

Schuchman, E. H., Jackson, C. E. and Desnick, R. J. (1990). "Human arylsulfatase B: MOPAC cloning, nucleotide sequence of a full-length cDNA, and regions of amino acid identity with arylsulfatases A and C." *Genomics* **6**, 149-58.

Schuchman, E. H., Suchi, M., Takahashi, T., Sandhoff, K. and Desnick, R. J. (1991). "Human acid sphingomyelinase. Isolation, nucleotide sequence and expression of the full-length and alternatively spliced cDNAs." *J Biol Chem* **266**, 8531-9.

Semenza, J. C., Hardwick, K. G., Dean, N. and Pelham, H. R. (1990). "ERD2, a yeast gene required for the receptor-mediated retrieval of luminal ER proteins from the secretory pathway." *Cell* **61**, 1349-57.

Serafini, T., Stenbeck, G., Brecht, A., Lottspeich, F., Orci, L., Rothman, J. E. and Wieland, F. T. (1991). "A coat subunit of Golgi-derived non-clathrin-coated vesicles with homology to the clathrin-coated vesicle coat protein β-adaptin." *Nature* **349**, 215-20.

Simon, S. M. and Blobel, G. (1991). "A protein-conducting channel in the endoplasmic reticulum." *Cell* **65**, 371-80.

Sitia, R., Neuberger, M., Alberini, C., Bet, P., Fra, A., Valetti, C., Williams, G. and Milstein, C. (1990). "Developmental regulation of IgM secretion: the role of the carboxy-terminal cysteine." *Cell* **60**, 781-90.

Sloane, B. F., Rozhin, J., Hatfield, J. S., Crissman, J. D. and Honn, K. V. (1987). "Plasma membrane-associated cysteine proteinases in human and animal tumors." *Exp Cell Biol* **55**, 209-24.

Stein, C., Gieselmann, V., Kreysing, J., Schmidt, B., Pohlmann, R., Waheed, A., Meyer, H. E., O'Brien, J. S. and von Figura, K. (1989a). "Cloning and expression of human arylsulfatase A." *J Biol Chem* **264**, 1252-9.

Stein, C., Hille, A., Seidel, J., Rijnbout, S., Waheed, A., Schmidt, B., Geuze, H. and von Figura, K. (1989b). "Cloning and expression of human steroid-sulfatase. Membrane topology, glycosylation, and subcellular distribution in BHK-21 cells." *J Biol Chem* **264**, 13865-72.

Stein, M., Zijderhand-Bleekemolen, J. E., Geuze, H., Hasilik, A. and von Figura, K. (1987). "Mr 46,000 mannose 6-phosphate specific receptor: its role in targeting of lysosomal enzymes." *Embo J* **6**, 2677-81.

Stoller, T. J. and Shields, D. (1989). "The propeptide of preprosomatostatin mediates intracellular transport and secretion of α-globin from mammalian cells." *J Cell Biol* **108**, 1647-55.

Stoorvogel, W., Strous, G. J., Geuze, H. J., Oorschot, V. and Schwartz, A. L. (1991). "Late Endosomes Derive from Early Endosomes by Maturation." *Cell* **65**, 417-427.

Thomas, K. R. and Capecchi, M. R. (1987). "Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells." *Cell* **51**, 503-12.

Tsuji, S., Choudary, P. V., Martin, B. M., Stubblefield, B. K., Mayor, J. A., Barranger, J. A. and Ginns, E. I. (1987). "A mutation in the human glucocerebrosidase gene in neuronopathic Gaucher's disease." *N Engl J Med* **316**, 570-5.

Tsuji, S., Martin, B. M., Barranger, J. A., Stubblefield, B. K., LaMarca, M. E. and Ginns, E. I. (1988). "Genetic heterogeneity in type 1 Gaucher disease: multiple genotypes in Ashkenazic and non-Ashkenazic individuals [published erratum appears in Proc Natl Acad Sci U S A 1988 Aug;85(15):5708]." *Proc Natl Acad Sci U S A* **85**, 2349-52.

Tsuji, S., Yamauchi, T., Hiraiwa, M., Isobe, T., Okuyama, T., Sakimura, K., Takahashi, Y., Nishizawa, M., Uda, Y. and Miyatake, T. (1989). "Molecular cloning of a full-length cDNA for human α-N-acetylgalactosaminidase (α-galactosidase B)." *Biochem Biophys Res Commun* **163**, 1498-504.

Vaux, D., Tooze, J. and Fuller, S. (1990). "Identification by anti-idiotype antibodies of an intracellular membrane protein that recognizes a mammalian endoplasmic reticulum retention signal." *Nature* **345**, 495-502.

Vigers, G. P., Crowther, R. A. and Pearse, B. M. (1986a). "Location of the 100 kd-50 kd accessory proteins in clathrin coats." *Embo J* **5**, 2079-85.

Vigers, G. P., Crowther, R. A. and Pearse, B. M. (1986b). "Three-dimensional structure of clathrin cages in ice." *Embo J* **5**, 529-34.

von Heijne, G. (1986). "A new method for predicting signal sequence cleavage sites." *Nucl Acids Res* **14**, 4683-90.

Waheed, A., Gottschalk, S., Hille, A., Krentler, C., Pohlmann, R., Braulke, T., Hauser, H., Geuze, H. and von Figura, K. (1988). "Human lysosomal acid phosphatase is transported as a transmembrane protein to lysosomes in transfected baby hamster kidney cells." *Embo J* **7**, 2351-8.

Waheed, A., Hille, A., Junghans, U. and von Figura, K. (1990a). "Quaternary structure of the Mr 46,000 mannose 6-phosphate specific receptor: effect of ligand, pH, and receptor concentration on the equilibrium between dimeric and tetrameric receptor forms." *Biochemistry* **29**, 2449-55.

Waheed, A. and von Figura, K. (1990b). "Rapid equilibrium between monomeric, dimeric and tetrameric forms of the 46-kDa mannose 6-phosphate receptor at 37 degrees C. Possible relation to the function of the receptor." *Eur J Biochem* **193**, 47-54.

Walter, P. and Blobel, G. (1980). "Purification of a membrane-associated protein complex required for protein translocation across the endoplasmic reticulum." *Proc Natl Acad Sci U S A* **77**, 7112-6.

Walter, P. and Blobel, G. (1982). "Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum." *Nature* **299**, 691-8.

Wang, A. M., Bishop, D. F. and Desnick, R. J. (1990). "Human α-N-acetylgalactosaminidase-molecular cloning, nucleotide sequence, and expression of a full-length cDNA. Homology with human α-galactosidase A suggests evolution from a common ancestral gene." *J Biol Chem* **265**, 21859-66.

Wang, A. M. and Desnick, R. J. (1991). "Structural Organization and Complete Sequence of the Human α-N-Acetylgalactosaminidase Gene - Homology with the α-Galactosidase-A Gene Provides Evidence for Evolution from a Common Ancestral Gene." *Genomics* **10**, 133-142.

Watanabe, H., Grubb, J. H. and Sly, W. S. (1990). "The overexpressed human 46-kDa mannose 6-phosphate receptor mediates endocytosis and sorting of β-glucuronidase." *Proc Natl Acad Sci U S A* **87**, 8036-40.

Waters, M. G., Serafini, T. and Rothman, J. E. (1991). " 'Coatomer': a cytosolic protein complex containing subunits of non-clathrin-coated Golgi transport vesicles." *Nature* **349**, 248-51.

Wigderson, M., Firon, N., Horowitz, Z., Wilder, S., Frishberg, Y., Reiner, O. and Horowitz, M. (1989). "Characterization of mutations in Gaucher patients by cDNA cloning." *Am J Hum Genet* **44**, 365-77.

Wileman, T., Kane, L. P., Carson, G. R. and Terhorst, C. (1991). "Depletion of cellular calcium accelerates protein degradation in the endoplasmic reticulum." *J Biol Chem* **266**, 4500-7.

Williams, M. A. and Fukuda, M. (1990). "Accumulation of membrane glycoproteins in lysosomes requires a tyrosine residue at a particular position in the cytoplasmic tail." *J Cell Biol* **111**, 955-66.

Wilson, D. W., Wilcox, C. A., Flynn, G. C., Chen, E., Kuang, W. J., Henzel, W. J., Block, M. R., Ullrich, A. and Rothman, J. E. (1989). "A fusion protein required for vesicle-mediated transport in both mammalian cells and yeast." *Nature* **339**, 355-9.

Wilson, P. J., Morris, C. P., Anson, D. S., Occhiodoro, T., Bielicki, J., Clements, P. R. and Hopwood, J. J. (1990). "Hunter syndrome: isolation of an iduronate-2-sulfatase cDNA clone and analysis of patient DNA." *Proc Natl Acad Sci U S A* **87**, 8531-5.

Woods, C. M. and Lazarides, E. (1985). "Degradation of unassembled α- and β-spectrin by distinct intracellular pathways: regulation of spectrin topogenesis by β-spectrin degradation." *Cell* **40**, 959-69.

Xie, B., Mcinnes, B., Neote, K., Lamhonwah, A. M. and Mahuran, D. (1991). "Isolation and Expression of a Full-Length cDNA Encoding the Human GM2 Activator Protein." *Biochem Biophys Res Commun* **177**, 1217-1223.

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 a human acid β-galactosidase cDNA." *DNA Cell Biol* **9**, 119-27.

Yamauchi, T., Hiraiwa, M., Kobayashi, H., Uda, Y., Miyatake, T. and Tsuji, S. (1990). "Molecular cloning of two species of cDNAs for human α-N-acetylgalactosaminidase and expression in mammalian cells." *Biochem Biophys Res Commun* **170**, 231-7.

Yen, P. H., Allen, E., Marsh, B., Mohandas, T., Wang, N., Taggart, R. T. and Shapiro, L. J. (1987). "Cloning and expression of steroid sulfatase cDNA and the frequent occurrence of deletions in STS deficiency: implications for X-Y interchange." *Cell* **49**, 443-54.

Zhang, X. L., Rafi, M. A., DeGala, G. and Wenger, D. A. (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., Rafi, M. A., Degala, G. and Wenger, D. A. (1991). "The Mechanism for a 33-Nucleotide Insertion in Messenger RNA Causing Sphingolipid Activator Protein (SAP-1)- Deficient Metachromatic Leukodystrophy." *Human Genetics* **87**, 211-215.

Zot, A. S. and Fambrough, D. M. (1990). "Structure of a gene for a lysosomal membrane glycoprotein (LEP100). Housekeeping gene with unexpected exon organization." *J Biol Chem* **265**, 20988-95.

2. EXPERIMENTAL WORK

2.1.1 Introduction

Lysosomal protective protein was discovered through its deficiency in the distinct metabolic storage disorder galactosialidosis, a disease characterized by the severely reduced activities of the enzymes β-galactosidase and neuraminidase. In order to understand the relationship among the three proteins some of the relevant properties of the latter two glycosidases are reviewed first.

Lysosomal β-D-galactosidase is involved in the turnover of a variety of natural substrates such as glycoproteins, gangliosides and glycosaminoglycans, from which it hydrolyzes the non-reducing terminal galactose residues (for review see O'Brien, 1989). As mentioned previously, the *in vitro* degradation of G_{M1}-ganglioside by β-galactosidase is accelerated in the presence of a particular activator protein, SAP-1. The enzyme will also hydrolyze several artificial substrates, among which is the fluorescent compound 4-methylumbelliferyl-β-D-galactoside, utilized in most studies.The gene encoding human lysosomal β-galactosidase has been localized on chromosome 3 (Shows *et al,* 1978) and mutations in this locus result in the lysosomal storage disorders designated G_{M1}-gangliosidosis and Morquio B syndrome. In the former the major accumulated substrate is the ganglioside G_{M1} (see Figure 2), whereas in Morquio B syndrome β-galactosidase affinity for the glycosaminoglycan keratan sulfate is abnormally low (Paschke and Kresse, 1982; van der Horst *et al,* 1983).

β-Galactosidase activity has been detected in a number of tissues and species and depending upon the extraction procedure and source of material enzymatic activity is recovered in mono-, di- , tetra- and multimeric forms (e.g. Hoeksema *et al,* 1979; Hoogeveen *et al,* 1983; Norden *et al,* 1974; Potier *et al,* 1990; Scheibe *et al,* 1990; Verheijen *et al,* 1982; 1985; Yamamoto *et al,* 1982; Yamamoto and Nishimura, 1987). The mature enzyme can be conveniently isolated using the affinity matrix p-aminophenylthiogalactoside-CH-Sepharose (van Diggelen *et al,* 1981). Using this purification method high and low molecular weight aggregates were also found to persist.

In human cultured fibroblasts β-galactosidase is synthesized as a high molecular weight precursor of 85 kDa, which is processed into a mature lysosomal protein of 64 kDa (d'Azzo *et al,* 1982), with a half life of about 10 days (van Diggelen *et al,* 1981). In mutant human fibroblasts with an isolated β-galactosidase deficiency heterogeneity in biosynthesis and processing is observed (Hoogeveen *et al,*

1984). It has been calculated that in the normal enzyme purified from human liver three carbohydrate chains per β-galactosidase molecule are present (Overdijk *et al,* 1986). A similar biosynthetic processing pattern compared to human fibroblasts is detectable in mouse macrophages (Skudlarek and Swank, 1979). The half life of β-galactosidase in the latter cell type was calculated to be 2.9-3.5 days. In this calculation turnover is defined as the sum of enzyme degradation as well as secretion, each of which contributes in roughly equal amount to the disappearance of newly synthesized molecules (Skudlarek and Swank, 1981; Tropea *et al,* 1988). Mature β-galactosidase, derived from mouse macrophages also carries three oligosaccharide chains (Tropea *et al,* 1988).

The cDNAs and genes encoding human and mouse β-galactosidase have recently been isolated (see publication 2; Morreau *et al,* 1991; Nanba and Suzuki, 1990; 1991; Oshima *et al,* 1988; Yamamoto *et al,* 1990). The structure of the human enzyme will be dealt with in publication 2. Murine β-galactosidase, for which the corresponding cDNA was cloned after its human counterpart, is about 80 % identical to the human enzyme. Similarity is only significantly lower towards the C-terminus of the two proteins (Nanba and Suzuki,1990). As mentioned previously, the human gene gives rise to alternatively spliced transcripts, that encode the classic, catalytically active β-galactosidase protein, but also a non-lysosomal form, which is inactive towards the artificial substrate used (publication 2). A similar β-galactosidase-related protein can not be encoded by the mouse gene, since the reading frame of the putative alternatively spliced message would contain several stop codons and give rise to a truncated polypeptide (Morreau *et al,* 1991). This indicates that the β-galactosidase-related protein, if physiologically functional, performs this role only in man.

Lysosomal N-acetyl-α-neuraminidase hydrolyzes terminal sialic acid residues linked to oligosaccharides, glycoproteins, glycolipids and the artificial substrate 4-methylumbelliferyl-N-acetylneuraminic acid. A single neuraminidase deficiency gives rise to the lysosomal storage disorder sialidosis (for review see Beaudet and Thomas, 1989). Urine and fibroblasts of patients with this disorder were found to contain mainly sialyloligosaccharides as secreted or stored products, respectively (reviewed by Cantz and Ulrich-Bott, 1990). The enzyme is encoded by a gene on human chromosome 10 (Mueller *et al,* 1986).

Mammalian lysosomal neuraminidase has been difficult to purify because of its lability upon extraction from most tissues and its apparent membrane-bound character. There appear to be several mammalian neuraminidases, that so far

could only be distinguished by specific biochemical parameters and because of the lysosomal enzyme deficiency in the disorder sialidosis (see review by Corfield *et al,* 1981; and Lieser *et al,* 1989; Mendla and Cantz, 1984; Samollow *et al,* 1990; Schneider-Jakob and Cantz, 1991; Usuki *et al,* 1988; Verheijen *et al,* 1983). However, it is possible to purify a soluble form of acid neuraminidase together with β-galactosidase, the protective protein and other polypeptides in a stable high molecular weight complex from bovine testis and other tissues and species (Hiraiwa *et al,* 1991; Potier *et al,* 1990; Scheibe *et al,* 1990; Verheijen *et al,* 1982; 1985; Yamamoto *et al,* 1982; Yamamoto and Nishimura, 1987). Photoaffinity labeling studies have demonstrated that within the bovine testis complex the neuraminidase polypeptide carrying the active site is a 55 kDa protein (van der Horst *et al,* 1990). In contrast, the complexed human placental enzyme was shown to be a heavily glycosylated polypeptide of 66 kDa (van der Horst *et al,* 1989), or 61 kDa (Warner *et al,* 1990). Nothing is known yet about the synthesis, routing and maturation of this highly intriguing enzyme.

There is a distinct lysosomal storage disorder that differs from the isolated β-galactosidase and neuraminidase deficiencies mentioned above in that both glycosidase activities are severely affected. After the first report of a patient with this combined lysosomal enzyme deficiency (Wenger *et al,* 1978) more of these cases were shown to exist among patients with different clinical phenotypes. The "new" disorder, termed galactosialidosis (Andria *et al,* 1981), is rare and is transmitted in an autosomal recessive mode (see reviews by Andria *et al,* 1981; Galjaard *et al,* 1984; O'Brien, 1989; Suzuki *et al,* 1984). Based on age of onset of the disease three phenotypes are distinguished: 1) a severe early infantile form, which leads to death at or soon after birth and is characterized by CNS involvement, macular cherry-red spots, visceromegaly, renal insufficiency, coarse facies and skeletal abnormalities; 2) a milder late infantile type, that usually manifests itself at 6-12 months of age but remarkably does not result in mental retardation, at least in those patients thus far examined and followed up (Chitayat *et al,* 1988; Strisciuglio *et al,* 1990); 3) a juvenile/adult form, which is mainly found in Japan and is characterized by slowly progressive CNS symptoms, including motor disturbance and mental retardation, skeletal abnormalities, dysmorphism, macular cherry-red spots and angiokeratoma. Investigations aimed at deciphering the nature of accumulated products in placenta and urine from galactosialidosis patients revealed them to be mainly sialylated oligosaccharides, as in the disorder sialidosis (Okada *et al,* 1978; van Pelt *et al,* 1988a;1988b; 1989).

The defective gene in galactosialidosis was demonstrated by complementation analysis to differ from those encoding β-galactosidase or neuraminidase (Galjaard *et al,* 1975; Hoogeveen *et al,* 1980) and to be localized either on chromosome 20 (Mueller *et al,* 1986) or 22 (Sips *et al,* 1985). Further evidence for the participation of another factor in the regulation of β-galactosidase and neuraminidase came from "uptake" studies. The activity of both glycosidases in galactosialidosis fibroblasts is restored to near normal levels when these cells are cultured in medium, suplemented with secreted proteins derived from medium of control, G_{M1}-gangliosidosis or sialidosis fibroblasts, but not of galactosialidosis cells. Uptake of a "corrective factor" is impaired if concomitantly with the secreted material 1 mM M6P is added to the medium of the recipient cells (Hoogeveen *et al,* 1981). As is now known, the latter compound competitively inhibits MPR-mediated endocytosis of lysosomal protein precursors. Indeed the "corrective factor" later turned out to be a phosphorylated protein (Hoogeveen *et al,* 1986).

In galactosialidosis cells β-galactosidase has normal hydrolytic properties but its half life is less than one day, compared to ten days in normal cells (van Diggelen *et al,* 1981; 1982). Biosynthetic labeling and immunoprecipitation analysis revealed that fibroblasts derived from an early infantile galactosialidosis patient synthesize a normal quantity of the 85 kDa β-galactosidase precursor. The latter undergoes delayed endoproteolytic processing and the mature form is rapidly turned over (d'Azzo *et al,* 1982). Only after addition of leupeptin to the medium of these cells could a 66 kDa intermediate be visualized, that was also detected in leupeptin-treated normal cells. These data indicated that in galactosialidosis cells β-galactosidase is normally synthesized but it is rapidly degraded upon maturation. Surprisingly, proteins of 54 and 32 kDa that were immunoprecipitated together with β-galactosidase from control fibroblast extracts, were completely absent in the early infantile galactosialidosis cell lysates. The 54 kDa form turned out to be the precursor of the 32 kDa polypeptide. Since its deficiency in galactosialidosis leads to the increased turnover of β-galactosidase and to a complete absence of neuraminidase activity (d'Azzo *et al,* 1982), the "corrective factor" was later named "protective protein" (Hoogeveen *et al,* 1983).

The multimeric form of β-galactosidase that is detected in extracts of normal tissues is absent in galactosialidosis fibroblast lysates (Hoeksema *et al,* 1979; Hoogeveen *et al,* 1983). However, if the latter contain normal protective protein, taken up by receptor-mediated endocytosis, β-galactosidase is again found in a multimeric aggregate. It was postulated that the 32 kDa protective protein forms high molecular weight complexes with β-galactosidase in lysosomes and that this

event prevents the rapid proteolytic degradation of β-galactosidase molecules (Hoogeveen *et al,* 1983). Immuno electron microscopy has provided indirect evidence for this hypothesis (Willemsen *et al,* 1986). The model was extended to include neuraminidase. This enzyme copurifies in complex with β-galactosidase, the protective protein and other polypeptides after p-aminophenylthiogalactoside affinity chromatography of bovine and porcine testis glycoprotein preparations (Verheijen *et al,* 1982; Yamamoto and Nishimura, 1987). Using human placenta similar results are obtained with the notable difference that neuraminidase activity has to be generated by concentration and 37 ^{o}C incubation of the glycoprotein preparation (Verheijen *et al,* 1985). In the stabilized sample monospecific antibodies against the 32 kDa form of the protective protein coprecipitate both β-galactosidase and neuraminidase activities. Thus, *in vitro* a complex exists of the three aforementioned glycoproteins. Given the combined hydrolase deficiency in galactosialidosis it was proposed that this complex is a normal constituent of lysosomes and the function of the protective protein is to stabilize and activate the other two enzymes by virtue of the association.

In normal fibroblasts proteolytic conversion of precursor to mature protective protein starts 30 min after synthesis (Palmeri *et al,* 1986). Aside from the early infantile fibroblasts mentioned above, several galactosialidosis cell strains synthesize a mutated 54 kDa precursor in varying quantities, but no mature form is immunoprecipitated after a chase of 2 hr (Palmeri *et al,* 1986). In a late infantile cell strain minute amounts of immunoprecipitable 32 kDa polypeptide were detected after treatment of the fibroblasts with the protease inhibitor leupeptin (Palmeri *et al,* 1986). Similar results were obtained using other strains derived from patients with the same phenotype. This has led to the suggestion that the presence of residual 32 kDa polypeptide is a distinct feature in late infantile galactosialidosis (Strisciuglio *et al,* 1988).

We have cloned the cDNAs encoding human, mouse and chicken protective proteins and human β-galactosidase in order to investigate their respective primary structures, the function(s) of the protective protein, its interaction with β-galactosidase, and to identify mutation(s) underlying the different lysosomal storage disorders. Our findings and main conclusions are reported in publications 1-6. In publication 7 the gene encoding human protective protein is unequivocally localized on chromosome 20, by using *in situ* hybridization.

2.1.2 References

Andria, G., Strisciuglio, P., Pontarelli, G., Sly, W. S. and Dodson, W. E. (1981). Infantile neuraminidase and β-galactosidase deficiencies (galactosialidosis) with mild clinical courses. In *Sialidases and Sialidoses*. G. Tettamanti, P. Durand and S. DiDonato eds. (Milan: Edi. Ermes), pp 379-95.

Beaudet, A. L. and Thomas, G. H. (1989). Disorders of glycoprotein degradation: mannosidosis, fucosidosis, sialidosis, and aspartylglycosaminuria. In *The Metabolic Basis of Inherited Disease*. C. R. Scriver, A. L. Beaudet, W. S. Sly and D. Valle eds. (New York: McGraw-Hill), ed. 6, pp 1603-21.

Cantz, M. and Ulrich-Bott, B. (1990). "Disorders of glycoprotein degradation." *J Inherited Metab Dis* **13**, 523-37.

Chitayat, D., Applegarth, D. A., Lewis, J., Dimmick, J. E., McCormick, A. Q. and Hall, J. G. (1988). "Juvenile galactosialidosis in a white male: a new variant." *Am J Med Genet* **31**, 887-901.

Corfield, A. P., Michalski, J. C. and Schauer, R. (1981). The substrate specificity of sialidases from microorganisms and mammals. In *Sialidases and Sialidoses*. G. Tettamanti, P. Durand and S. DiDonato eds. (Milan: Ermes), pp 1-70.

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

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

Galjaard, H., Hoogeveen, A. T., Keijzer, W., de Wit-Verbeek, H. A., Reuser, A. J. J., Ho, M. W. and Robinson, D. (1975). "Genetic heterogeneity in GM1-gangliosidosis." *Nature* **257**, 60-2.

Hiraiwa, M., Uda, Y., Tsuji, S., Miyatake, T., Martin, B. M., Tayama, M., O'Brien, J. S. and Kishimoto, Y. (1991). "Human Placental Sialidase Complex - Characterization of the 60 kDa Protein That Cross-Reacts with Anti-Saposin Antibodies." *Biochem Biophys Res Comm* **177**, 1211-1216.

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

Hoogeveen, A., d'Azzo, A., Brossmer, R. and Galjaard, H. (1981). "Correction of combined β-galactosidase/neuraminidase deficiency in human fibroblasts." *Biochem Biophys Res Commun* **103**, 292-300.

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

Hoogeveen, A. T., Reuser, A. J., Kroos, M. and Galjaard, H. (1986). "GM1-gangliosidosis. Defective recognition site on β-galactosidase precursor." *J Biol Chem* **261**, 5702-4.

Hoogeveen, A. T., Verheijen, F. W., d'Azzo, A. and Galjaard, H. (1980). "Genetic heterogeneity in human neuraminidase deficiency." *Nature* **285**, 500-2.

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**, 12143-6.

Lieser, M., Harms, E., Kern, H., Bach, G. and Cantz, M. (1989). "Ganglioside GM3 sialidase activity in fibroblasts of normal individuals and of patients with sialidosis and mucolipidosis IV. Subcellular distribution and and some properties." *Biochem J* **260**, 69-74.

Mendla, K. and Cantz, M. (1984). "Specificity studies on the oligosaccharide neuraminidase of human fibroblasts." *Biochem J* **218**, 625-8.

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

Mueller, O. T., Henry, W. M., Haley, L. L., Byers, M. G., Eddy, R. L. and Shows, T. B. (1986). "Sialidosis and galactosialidosis: chromosomal assignment of two genes associated with neuraminidase-deficiency disorders." *Proc Natl Acad Sci U S A* **83**, 1817-21.

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

Nanba, E. and Suzuki, K. (1991). "Organization of the mouse acid β-galactosidase gene." *Biochem Biophys Res Comm* **178**, 158-164.

Norden, A. G. W., Tennant, L. and O'Brien, J. S. (1974). "Ganglioside GM1 β-galactosidase A: purification and studies of the enzyme from human liver." *J Biol Chem* **249**, 7969-76.

O'Brien (1989). β-galactosidase deficiencies (GM1-gangliosidosis, galactosialidosis, and Morquio syndrome type B); ganglioside sialidase deficiency (mucolipidosis type IV). In *The Metabolic Basis of Inherited Disease*. C. R. Scriver, A. L. Beaudet, W. S. Sly and D. Valle eds. (New York: McGraw-Hill), ed. 6, pp 1797-806.

Okada, S., Kato, T., Miura, S., Yabuuchi, H., Nishigaki, M., Kobata, A., Chiyo, H. and Furuyama, J. (1978). "Hypersialyloligosaccharidurla in mucolipidosis: a method for diagnosis." *Clin Chim Acta* **86**, 159-67.

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**, 238-44.

Overdijk, B., Hiensch-Goormachtig, E. F. J., Beem, E. P., van Steijn, G. J., Trippelvitz, L. A. W., Lisman, J. J. W., van Halbeek, H., Mutsaers, J. H. G. M. and Vliegenthart, J. F. G. (1986). "The carbohydrate structures of β-galactosidase from human liver." *Glyconjugate J* **3**, 339-50.

Palmeri, S., Hoogeveen, A. T., Verheijen, F. W. and Galjaard, H. (1986). "Galactosialidosis: molecular heterogeneity among distinct clinical phenotypes." *Am J Hum Genet* **38**, 137-48.

Paschke, E. and Kresse, H. (1982). "Morquio disease, type B: activation of GM1-β-galactosidase by GM1-activator protein." *Biochem Biophys Res Commun* **109**, 568-75.

Potier, M., Michaud, L., Tranchemontagne, J. and Thauvette, L. (1990). "Structure of the lysosomal neuraminidase-β-galactosidase-carboxypeptidase multienzymic complex." *Biochem J* **267**, 197-202.

Samollow, P. B., Ford, A. L. and VandeBerg, J. L. (1990). "Biochemical characteristics and subcellular localizations of rat liver neuraminidase isozymes: a paradox resolved." *Biochem Genet* **28**, 283-98.

Scheibe, R., Hein, K. and Wenzel, K. W. (1990). "Lysosomal β-galactosidase from rat liver: purification, molecular forms and association with neuraminidase." *Biomed Biochim Acta* **49**, 547-56.

Schneider-Jakob, H. R. and Cantz, M. (1991). "Lysosomal and plasma membrane ganglioside GM3 sialidases of cultured human fibroblasts - differentiation by detergents and inhibitors." *Biol Chem Hoppe Seyler* **372**, 443-450.

Shows, T. B., Scrafford-Wolff, L., Brown, J. A. and Meisler, M. (1978). "Assignment of a β-galactosidase gene (β-Gal A) to chromosome 3 in man." *Cytogenet Cell Genet* **22**, 219-22.

Sips, H. J., de Wit-Verbeek, H. A., de Wit, J., Westerveld, A. and Galjaard, H. (1985). "The chromosomal localization of human β-galactosidase revisited: a locus for β-galactosidase on human chromosome 3 and for its protective protein on human chromosome 22." *Hum Genet* **69**, 340-4.

Skudlarek, M. D. and Swank, R. T. (1979). "Biosynthesis of two lysosomal enzymes in macrophages. Evidence for a precursor of β-galactosidase." *J Biol Chem* **254**, 9939-42.

Skudlarek, M. D. and Swank, R. T. (1981). "Turnover of two lysosomal enzymes in macrophages." *J Biol Chem* **256**, 10137-44.

Strisciuglio, P., Parenti, G., Giudice, C., Lijoi, S., Hoogeveen, A. T. and d'Azzo, A. (1988). "The presence of a reduced amount of 32-kd 'protective' protein is a distinct biochemical finding in late infantile galactosialidosis." *Hum Genet* **80**, 304-6.

Strisciuglio, P., Sly, W. S., Dodson, W. E., McAlister, W. H. and Martin, T. C. (1990). "Combined deficiency of β-galactosidase and neuraminidase: natural history of the disease in the first 18 years of an American patient with late infantile onset form." *Am J Med Genet* **37**, 573-7.

Suzuki, Y., Sakuraba, H., Yamanaka, T., Ko, Y. M., Limori, Y., Okamura, Y. and Hoogeveen, A. T. (1984). Galactosialidosis: a comparative study of clinical and biochemical data on 22 patients. In *The Developing Brain and its Disorders*. M. Arina, Y. Suzuki and H. Yabuuchi eds. (Tokyo: University of Tokyo Press), pp 161-75.

Tropea, J. E., Swank, R. T. and Segal, H. L. (1988). "Effect of swainsonine on the processing and turnover of lysosomal β-galactosidase and β-glucuronidase from mouse peritoneal macrophages." *J Biol Chem* **263**, 4309-17.

Usuki, S., Lyu, S. C. and Sweeley, C. C. (1988). "Sialidase activities of cultured human fibroblasts and the metabolism of GM3 ganglioside." *J Biol Chem* **263**, 6847-53.

van der Horst, G. T. J., Galjart, N. J., d'Azzo, A., Galjaard, H. and Verheijen, F. W. (1989). "Identification and in vitro reconstitution of lysosomal neuraminidase from human placenta." *J Biol Chem* **264**, 1317-22.

van der Horst, G. T. J., Kleijer, W. J., Hoogeveen, A. T., Huijmans, J. G., Blom, W. and van Diggelen, O. P. (1983). "Morquio B syndrome: a primary defect in β-galactosidase." *Am J Med Genet* **16**, 261-75.

van der Horst, G. T. J., Rose, U., Brossmer, R. and Verheijen, F. W. (1990). "Photoaffinity labeling of the lysosomal neuraminidase from bovine testis." *Febs Lett* **277**, 42-4.

van Diggelen, O. P., Hoogeveen, A. T., Smith, P. J., Reuser, A. J. and Galjaard, H. (1982). "Enhanced proteolytic degradation of normal β-galactosidase in the lysosomal storage disease with combined β-galactosidase and neuraminidase deficiency." *Biochim Biophys Acta* **703**, 69-76.

van Diggelen, O. P., Schram, A. W., Sinnott, M. L., Smith, P. J., Robinson, D. and Galjaard, H. (1981). "Turnover of β-galactosidase in fibroblasts from patients with genetically different types of β-galactosidase deficiency." *Biochem J* **200**, 143-51.

van Pelt, J., Hard, K., Kamerling, J. P., Vliegenthart, J. F., Reuser, A. J. and Galjaard, H. (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., Kamerling, J. P., Vliegenthart, J. F., Hoogeveen, A. T. and Galjaard, H. (1988a). "A comparative study of the accumulated sialic acid-containing oligosaccharides from cultured human galactosialidosis and sialidosis fibroblasts." *Clin Chim Acta* **174**, 325-35.

van Pelt, J., van Kuik, J. A., Kamerling, J. P., Vliegenthart, J. F., van Diggelen, O. P. and Galjaard, H. (1988b). "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., 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**, 868-75.

Verheijen, F. W., Janse, H. C., van Diggelen, O. P., Bakker, H. D., Loonen, M. C., Durand, P. and Galjaard, H. (1983). "Two genetically different MU-NANA neuraminidases in human leucocytes." *Biochem Biophys Res Commun* **117**, 470-8.

Verheijen, F. W., Palmeri, S., Hoogeveen, A. T. and Galjaard, H. (1985). "Human placental neuraminidase. Activation, stabilization and association with β-galactosidase and its protective protein." *Eur J Biochem* **149**, 315-21.

Warner, T. G., Louie, A. and Potier, M. (1990). "Photolabeling of the α-neuraminidase/β-galactosidase complex from human placenta with a photoreactive neuraminidase inhibitor." *Biochem Biophys Res Commun* **173**, 13-9.

Wenger, D. A., Tarby, T. J. and Wharton, C. (1978). "Macular cherry-red spots and myoclonus with dementia: coexistent neuraminidase and β-galactosidase deficiencies." *Biochem Biophys Res Commun* **82**, 589-95.

Willemsen, R., Hoogeveen, A. T., Sips, H. J., van Dongen J. and Galjaard, H. (1986). "Immunoelectron microscopical localization of lysosomal β-galactosidase and its precursor forms in normal and mutant human fibroblasts." *Eur J Cell Biol* **40**, 9-15.

Yamamoto, Y., Fujie, M. and Nishimura, K. (1982). "The interrelation between high- and low-molecular-weight forms of GM1-β-galactosidase purified from porcine spleen." *J Biochem (Tokyo)* **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 a human acid β-galactosidase cDNA." *DNA Cell Biol* **9**, 119-27.

Yamamoto, Y. and Nishimura, K. (1987). "Copurification and separation of β-galactosidase and sialidase from porcine testis." *Int J Biochem* **19**, 435-42.

2. Publications

Publication 1

Cell **54** (1988), 755-764

Expression of cDNA Encoding the Human "Protective Protein" Associated with Lysosomal β-Galactosidase and Neuraminidase: Homology to Yeast Proteases

Niels J. Galjart,* Nynke Gillemans,* Alan Harris,† Gijsbertus T. J. van der Horst,* Frans W. Verheijen,* Hans Galjaard,* and Alessandra d'Azzo*
* Department of Cell Biology and Genetics
Erasmus University
P. O. Box 1738
3000 DR Rotterdam, The Netherlands
† Laboratory of Protein Structure
National Institute for Medical Research
The Ridgeway, Mill Hill
London NW7 1AA, England

Summary

The "protective protein" is a glycoprotein that associates with lysosomal β-galactosidase and neuraminidase and is deficient in the autosomal recessive disorder galactosialidosis. We have isolated the cDNA encoding human "protective protein." The clone recognizes a 2 kb mRNA in normal cells that is not evident in fibroblasts of an early infantile galactosialidosis patient. The cDNA directs the synthesis of a 452 amino acid precursor molecule that is processed in vivo to yield mature "protective protein," a heterodimer of 32 kd and 20 kd polypeptides held together by disulfide bridges. This mature form is also biologically functional since it restores β-galactosidase and neuraminidase activities in galactosialidosis cells. The predicted amino acid sequence of the "protective protein" bears homology to yeast carboxypeptidase Y and the *KEX1* gene product. This suggests a protease activity for the "protective protein."

Introduction

The lysosomal enzymes β-D-galactosidase (E.C. 3.2.1.23) and N-acetyl-α-neuraminidase (sialidase, E.C. 3.2.1.18) are hydrolytic glycoproteins responsible for the catabolism of a variety of natural and synthetic substrates (for reviews see Corfield et al., 1981; O'Brien, 1983). Mutations at their structural genes cause deficient or severely altered enzyme activities with consequent accumulation of undegraded substrate(s) in the lysosomes. The resulting metabolic storage disorders associated with a single β-galactosidase deficiency are G_{M1}-gangliosidosis and Morquio B disease (Groebe et al., 1980; O'Brien, 1983), whereas sialidosis is due to an isolated neuraminidase deficiency (Cantz et al., 1984).

There is, however, another autosomal recessive disease, genetically distinct from G_{M1}-gangliosidosis and sialidosis, that is currently designated galactosialidosis (Andria et al., 1981) because of the coexistent deficiency of β-galactosidase and neuraminidase in patients with this disorder (Wenger et al., 1978). Different clinical phenotypes have been observed, ranging from very severe early infantile (E.I.) forms, fatal within childhood, to milder late infantile (L.I.) and juvenile/adult variants (Andria et al., 1981; Suzuki et al., 1984; Galjaard et al., 1984). In all types of patients, the responsible molecular defect has been attributed to the deficiency of a 32 kd glycoprotein, referred to as "protective protein," which appears to be essential for the full biological activity of both β-galactosidase and neuraminidase (d'Azzo et al., 1982).

In human liver and cultured fibroblasts, active β-galactosidase is present as a 64 kd monomeric form and as a high molecular weight aggregate (600–700 kd) (Norden et al., 1974; Hoeksema et al., 1979). In galactosialidosis fibroblasts, the residual β-galactosidase activity is about 10% of the normal level and is the result of enhanced intralysosomal degradation of the enzyme, which is present only in the monomeric form (van Diggelen et al., 1982; Hoogeveen et al., 1983).

Studies on the biosynthesis of lysosomal β-galactosidase in cultured fibroblasts have shown that the enzyme precipitates with anti-human β-galactosidase antiserum together with two other polypeptides, of 32 kd and 20 kd (d'Azzo et al., 1982). Furthermore, the majority of active β-galactosidase can be resolved with the 32 kd and 20 kd proteins in the aforementioned high molecular weight aggregate (Hoogeveen et al., 1983). The 32 kd component is synthesized as a 54 kd precursor that is also recovered extracellularly (d'Azzo et al., 1982). This secreted form, taken up by galactosialidosis fibroblasts via the mannose-6-phosphate receptor, is processed and promotes multimerization of β-galactosidase monomers, which prevents rapid proteolytic degradation of the enzyme (Hoogeveen et al., 1981, 1983; Willemsen et al., 1986). Although the involvement of the 32 kd protein in the aggregation of monomeric β-galactosidase became apparent, the biologial significance of the 20 kd polypeptide has remained unclear until now.

In concurrent studies on lysosomal neuraminidase, Verheijen et al. (1982, 1985) found that in bovine testis and human placenta this enzyme copurifies with the "protective protein" and β-galactosidase and that its activity depends on the presence of the "protective protein." Taken together, these findings strongly suggest that the three glycoproteins—β-galactosidase, neuraminidase, and the "protective protein"—exist in lysosomes as a functional complex, but the stoichiometry of this complex is not yet understood.

We have isolated cDNA encoding the "protective protein" to study the primary structure, function, and expression of this protein as well as to investigate the molecular nature of the mutations in galactosialidosis patients. The cDNA clone directs the synthesis of "protective protein" precursor, which becomes biologically active after being endocytosed and processed in galactosialidosis fibroblasts. The amino acid sequence deduced from the cDNA reveals that the active "protective protein" consists of a heterodimer of 32 kd and 20 kd polypeptides. These components are held together by disulfide bridges. The predicted amino acid sequence bears homology to car-

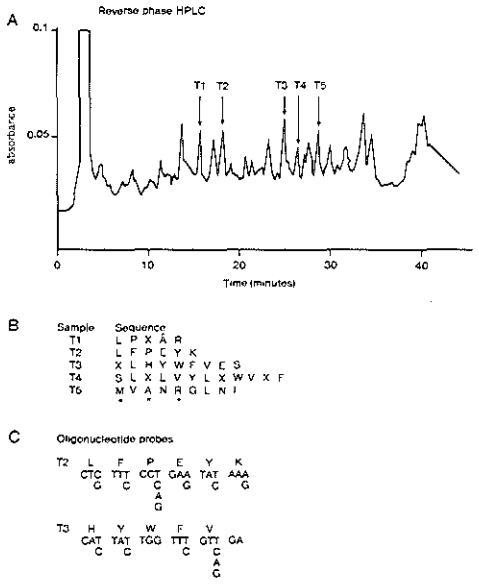


Figure 1. Separation of Tryptic Peptides from the 32 kd Protein: Amino Acid Sequences and Oligonucleotide Probes

The 32 kd protein, purified from human placenta, was digested to completion with trypsin. Resulting peptides were separated on a reverse-phase HPLC column using a gradient of 5%–60% acetonitrile (A). The numbered peaks, T1–T5, represent peptides used for automated Edman degradation. Corresponding amino acid sequences are shown in (B). Asterisks refer to discrepancies between deduced and predicted amino acid sequences; unassigned residues are indicated by the letter X. Two amino acid sequences were suitable for synthesizing oligonucleotide probes, whose sequences are depicted in (C).

boxypeptidase Y (CPY), a yeast vacuolar protease, and the *KEX1* gene product, a yeast processing enzyme with carboxypeptidase B–like activity.

Results

Partial Amino Acid Sequence and Isolation of Antibodies

The high molecular weight complex of β-galactosidase, neuraminidase, and "protective protein" can be isolated from human placenta using concanavalin A–Sepharose chromatography and affinity chromatography for β-galactosidase (Verheijen et al., 1985). One of these preparations was used in this study to obtain purified 32 kd protein. For this purpose the different components of the complex were separated by SDS–polyacrylamide gel electrophoresis (SDS-PAGE) under reducing and denaturing conditions. Electroeluted 32 kd protein was used to raise monospecific polyclonal antibodies in rabbit. This antibody preparation, tested in biosynthetic labeling experiments as well as on Western blots, recognized only the 32 kd protein and its 54 kd precursor. In a separate experiment a gel slice containing the 32 kd protein was treated in situ with trypsin, and the released peptides were fractionated by reverse-phase high-pressure liquid chromatography (HPLC) (Figure 1A). Five of the oligopeptides were subjected to automated Edman degradation (Figure 1B). The N-terminal sequence of the intact 32 kd protein was obtained using the microsequencing method of Aebersold et al. (1986). We also determined the N-terminal sequence of the 20 kd polypeptide since this protein is consistently present in purified preparations of the complex. Two stretches, of 26 and 28 amino acids, were obtained.

Isolation of cDNA Clones and Nucleotide Sequence Analysis

Two of the five peptide sequences were used to synthesize oligonucleotide probes (Figure 1C). They were made complementary to the mRNA in such a way that all codon usage possibilities were represented. Polyclonal anti-32 kd antibodies and two oligonucleotide probes were used independently to screen a human testis cDNA library in the expression vector λgt11 (Young and Davis, 1983). Both screening procedures yielded several recombinant clones, the longest of which, λHu54, contained an insert of 1.8 kb. This clone was recognized by the antibodies as well as by the two oligonucleotide probes. These findings supported the identity of this cDNA.

The 1.8 kb insert, subcloned into pTZ19 (pHu54), was subjected to restriction endonuclease mapping (Figure 2). The insert was sequenced using the dideoxy chain-termination method of Sanger et al. (1977). Figure 2 depicts the nucleotide sequencing strategy used. The complete sequence of the pHu54 clone is shown in Figure 3A. The cDNA contains 1825 nucleotides. A potential ATG translation initiation codon is encountered only 6 nucleotides from the 5′ end of the cDNA. This ATG is the begin-

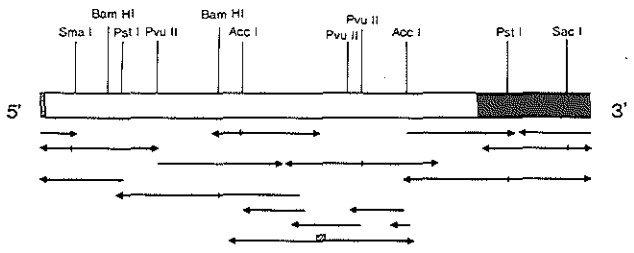


Figure 2. Partial Restriction Map and Sequencing Strategy of the Hu54 cDNA Clone

All the restriction sites used for subcloning and sequencing are shown, except for HaeIII, which has multiple restriction sites. Arrows indicate the extent and direction of sequencing reactions. The hatched box on the lowermost arrow indicates a location where synthetic oligonucleotides other than the universal primer were used. Hatched bar, 5′ noncoding region; open bar, coding region; solid bar, 3′ untranslated region.

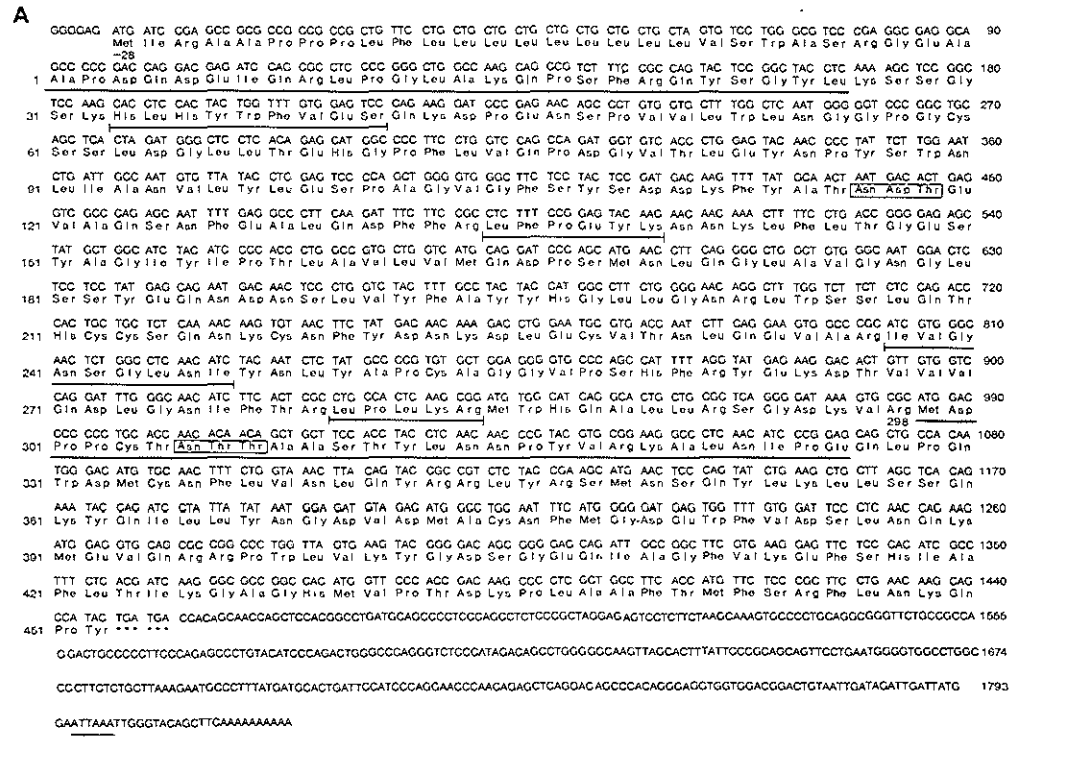


Figure 3. Nucleotide and Predicted Amino Acid Sequences of the Hu54 cDNA

(A) The predicted amino acid sequence is shown, with amino acids −28 to −1 representing the signal sequence and amino acids 1 to 452 representing the precursor polypeptide of the "protective protein." Tryptic peptides (thin underlines) and N-terminal sequences (thick underlines) of the 32 kd and 20 kd proteins are marked. Potential N-linked glycosylation sites are boxed. A putative polyadenylation signal in the 3' untranslated region is underlined.

(B) Alignment of the amino acid sequences of the "protective protein" precursor (HU54), yeast CPY (Valls et al., 1987), and the *KEX1* gene product (Dmochowska et al., 1987). Identical residues in the proteins are boxed. The asterisk designates the active-site serine residue of CPY. Numbers on the left refer to positions of the amino acids within the sequences.

ning of an open reading frame of 1440 nucleotides interrupted by two consecutive stop codons, and it is flanked at the 3′ end by a 363 nucleotide untranslated region. The nucleotide sequence terminates with a short poly(A) tail, and a potential polyadenylation signal (ATTAAA) is found at position 1796.

The putative ATG initiation codon is in a context that is not optimal for ribosome binding and initiation of translation (Kozak, 1986). To establish the usage of this ATG, pHu54 was transcribed in vitro. The resulting mRNA, translated in a rabbit reticulocyte lysate system, gave rise to a primary product of 54 kd immunoprecipitable with anti-32 kd antibodies. The size of this molecule correlated well with that of the protein immunoprecipitated after translation of total or polysomal RNA from normal fibroblasts (data not shown). These results suggest that the cDNA contains the entire coding region for the "protective protein" precursor.

Predicted Amino Acid Sequence of the "Protective Protein" and Homology to Yeast Carboxypeptidase Y and *KEX1* Gene Product

The open reading frame of 1440 nucleotides encodes a protein of 480 amino acids. Four of the five tryptic peptides are found in the amino acid sequence deduced from the Hu54 cDNA (Figure 3A). Few discrepancies exist between the primary structure predicted here and the amino acid sequencing data (Figure 1C). These include Ile-238, Gly-240, Ser-243, and Lys-283. One of the tryptic peptides (T4) could not be located within the predicted amino acid sequence. The localization of the N-terminus of the 32 kd protein at residues 1–26 (Figure 3A, indicated with a thick line) allows for the assignment of a signal sequence of 28 amino acids immediately preceding the 32 kd polypeptide. It contains the three domains—basic N-terminal, hydrophobic central, and polar C-terminal regions—typical of signal sequences (von Heijne, 1986). Moreover, the alanine and glycine residues at positions −1 and −3 conform to statistically determined rules for amino acids at those positions with respect to signal sequence cleavage sites (von Heijne, 1986). Surprisingly, the N-terminus of the 20 kd protein is also located within the predicted amino acid sequence at residues 299–326 (Figure 3A). Thus, the primary structure of the "protective protein" includes a signal sequence of 28 amino acids followed by 298 and 154 residues constituting the 32 kd and the 20 kd domains, respectively.

There are two potential N-linked glycosylation sites, at positions 117 and 305. The predicted molecular mass of the "protective protein" precursor is 54,496 daltons, which is reduced to 51,421 daltons after removal of the signal sequence. Assuming glycosylation at the aforementioned residues and therefore addition of approximately 4,000 daltons, the estimated molecular mass of the precursor molecule will be 55 kd. This corresponds to the size experimentally calculated after SDS-PAGE.

Searches of different protein data bases demonstrated homology between the "protective protein" and the yeast vacuolar protease CPY. The gene for CPY has recently been cloned and sequenced (Valls et al., 1987). We have used the sequence of the mature form of CPY, residues 112–532 (Valls et al., 1987), for alignment with the "protective protein" (Figure 3B). On the other hand, Dmochowska et al. (1987) have reported sequence homology between CPY and another yeast protein, the *KEX1* gene product. We therefore also compared the "protective protein" sequence with residues 34–489 of the *KEX1*-encoded protein (Figure 3B). Identity is predominantly confined to the N-terminal portion of all three proteins. This region is flanked by a sequence tract that does not show any significant similarity. Homology among the three proteins is again found over a stretch of about 100 amino acids, positioned at the C-termini of Hu54 and CPY and residues 390–489 of the *KEX1* gene product. A conserved region of six amino acids (Gly-Glu-Ser-Tyr-Ala-Gly), present in all three proteins, spans the serine residue shown to be in the active site of CPY (Hayashi et al., 1973) and possibly of the *KEX1* gene product (Dmochowska et al., 1987). Two other regions, also identical, one of ten amino acids (Trp-Leu-Asn-Gly-Gly-Pro-Gly-Cys-Ser-Ser) including a Cys residue, and one of four amino acids (His-Met-Val-Pro) containing a His residue, are positioned at the beginning and the end of the three aligned sequences (Figure 3B). These observations imply that the "protective protein" might function as a carboxypeptidase.

Nine cysteines within the predicted amino acid sequence of the "protective protein" are present, four of which cluster between residues 212 and 228. This suggests intramolecular covalent binding by disulfide bridges between the 32 kd and 20 kd polypeptides. To test this hypothesis, the different components of the purified complex (β-galactosidase, neuraminidase, and "protective protein") were fractionated by SDS-PAGE under reducing and nonreducing conditions (Figure 4). Proteins were either directly stained with Coomassie blue or Western blotted

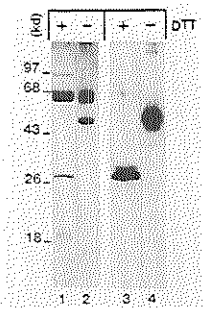


Figure 4. Functional State of the "Protective Protein" in the Purified Complex

Purified protein components of the complex were either reduced and denatured by boiling in SDS sample buffer in the presence of 20 mM dithiothreitol (+, lanes 1 and 3), or simply boiled without reduction (−, lanes 2 and 4). Samples were resolved by SDS-PAGE. One part of the gel was stained for total proteins with Coomassie brilliant blue (lanes 1 and 2); the other part was subjected to Western blotting. The nitrocellulose filter was incubated with anti-32 kd antibodies and ^{125}I-protein A. Radiolabeled bands were visualized by autoradiography (lanes 3 and 4).

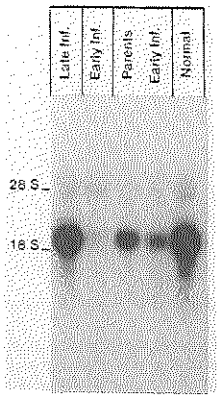


Figure 5. Northern Blot Analysis of Fibroblast RNA

Samples of total fibroblast RNA (15 μg) from a normal individual, the E.I. (Early Inf.) and L.I. (Late Inf.) galactosialidosis patients, and the parents of the E.I. patient were fractionated on a formaldehyde–agarose gel and probed with the Hu54 cDNA. Ribosomal RNA markers are indicated. Exposure time was 24 hr.

and probed with anti-32 kd antibodies. In the presence of the reducing agent, 32 kd and 20 kd polypeptides are visible in the Coomassie blue–stained gel, whereas a 54 kd form is detected under nonreducing conditions (Figure 4, lanes 1 and 2). Both the 32 kd and 54 kd proteins react with the antiserum (Figure 4, lanes 3 and 4). We conclude from these results that the 32 kd and 20 kd components of the "protective protein" are held together by disulfide bridges and function as a heterodimer.

RNA Hybridization Studies

To determine the size of the "protective protein" transcript as well as its abundance in galactosialidosis patients, total RNA was isolated from normal cultured fibroblasts and from fibroblasts of the L.I. galactosialidosis patient, the E.I. galactosialidosis patient, and the parents of the latter. A Northern blot containing these RNAs was hybridized with ^{32}P-labeled Hu54 cDNA (Figure 5). In normal fibroblasts an mRNA species of about 2 kb is present. This implies that the 5′ untranslated region measures about 200 bp. The 2 kb mRNA is not evident in the E.I. galactosialidosis patient, whereas it is present in normal amounts in fibroblasts of the L.I. galactosialidosis patient. When 10 μg of poly(A)$^+$ RNA was used, traces of the 2 kb message could be seen in the E.I. galactosialidosis patient, but only upon long exposures of the autoradiographs (data not shown). The parents of this patient have drastically reduced amounts of the 2 kb mRNA. To make certain that equal amounts of the different RNA samples were present on the Northern blot, the filter was rehybridized with a probe recognizing the glyceraldehyde-3-phosphate dehydrogenase mRNA, which is abundant in fibroblasts (Benham et al., 1984). In all lanes, approximately the same amount of this 1.2 kb message was detected (data not shown). These results provide additional proof of the identity of the cDNA clone and they also show different mutations in the two clinical forms of galactosialidosis tested.

Transient Expression of Hu54 cDNA in COS-1 Cells

The Hu54 cDNA was cloned in two orientations into a derivative of the mammalian expression vector pCD-X (Okayama and Berg, 1983). In one construct, pCDHu54-sense, the 5′ end of the cDNA was oriented toward the SV40 promoter, whereas in the pCDHu54-antisense construct the orientation of the cDNA was reversed. Transfection experiments were carried out on COS-1 cells (Gluzman, 1981), since transient expression of exogenous DNA in this cell type has proved to be successful in previous studies.

The pCDHu54-sense and -antisense constructs were transfected separately. Two days after transfection, newly synthesized proteins were labeled with [^{35}S]methionine in the presence of 10 mM NH_4Cl to induce maximal secretion of lysosomal protein precursors (Hasilik and Neufeld, 1980). Radiolabeled proteins from cell lysates and medium concentrates were immunoprecipitated with anti-32 kd antibodies. The results are shown in Figure 6. Untransfected COS-1 cells contain an endogenous 54 kd protein that is barely detectable by anti-32 kd antibodies and is partially secreted into the medium (Figure 6, lanes 1 and 2). Whether this protein represents the COS-1 cell equivalent of the "protective protein" precursor is at present not clear. After transfection of the cells with the pCDHu54-sense plasmid, a large amount of the 54 kd precursor is synthesized (Figure 6, lanes 3 and 4). This precursor protein is recognized by the antibodies and it is present in both cells and medium. Its estimated molecular weight correlates with that observed for the glycosylated "protective protein" precursor immunoprecipitated in human cells. The synthesis of this precursor molecule is not accomplished in COS-1 cells transfected with the antisense construct (Figure 6, lanes 5 and 6).

The cDNA-derived 54 kd precursor protein is apparently not processed to 32 kd and 20 kd polypeptides in COS-1 cells. Although the experiments illustrated in Figure 6 were performed in the presence of NH_4Cl, similar results were obtained in the absence of this reagent (data not shown). These findings demonstrate that the pCDHu54-sense construct, when transfected into COS-1 cells, directs the synthesis of the 54 kd "protective protein" precursor, which is also secreted into the medium.

Uptake and Processing of COS-1 Cell-Derived 54 kd Precursor by Human Fibroblasts

To show that the cDNA-encoded precursor yields a biologically active "protective protein," uptake studies were performed. Medium from [^{35}S]methionine-labeled COS-1 cells, transfected with either of the two pCDHu54 constructs, was collected and concentrated. Aliquots of each of the different medium concentrates were added to the medium of cultured fibroblasts from a normal individual, the E.I. galactosialidosis patient, and the L.I. galactosialidosis patient. After a further 3 days, immunoprecipitation was carried out on cell extracts and medium concentrates

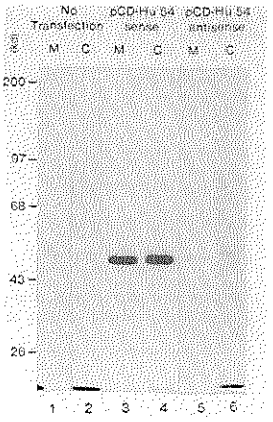


Figure 6. Transient Expression of pCDHu54 in COS-1 Cells

Hu54 cDNA was cloned in two orientations into the mammalian expression vector pCD-X. The resulting pCDHu54 constructs were transfected into COS-1 cells. Control cells were treated in the same manner as transfected cells but no DNA was added. After 48 hr, cells were incubated with [^{35}S]methionine for an additional 16 hr in the presence of 10 mM NH_4Cl. Labeled proteins from cells (C) and media (M) were immunoprecipitated with anti-32 kd antibodies, analyzed on a 10% polyacrylamide gel under reducing and denaturing conditions, and visualized by fluorography. Molecular size markers are indicated at left. Exposure time for lanes 2–4 was 24 hr; lanes 1, 5, and 6 were exposed three times as long.

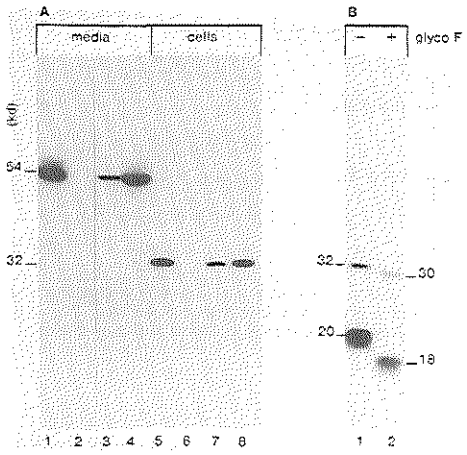


Figure 7. Uptake, Processing, and Glycopeptidase F Treatment of COS-1 Cell-Derived "Protective Protein" Precursor in Human Fibroblasts

(A) Uptake and processing. ^{35}S-labeled 54 kd precursor protein, produced by COS-1 cells transfected with pCDHu54 in the sense orientation, was added to culture medium of normal fibroblasts and fibroblasts from the E.I. and L.I. galactosialidosis patients. After 72 hr, labeled proteins from media and cells were immunoprecipitated with anti-32 kd antibodies, separated by 12.5% SDS-PAGE, and visualized by fluorography. Lanes 1 and 5, E.I. fibroblasts; lanes 3 and 7, normal fibroblasts; lanes 4 and 8, L.I. fibroblasts. Lanes 2 and 6 represent the control experiment in which proteins secreted by COS-1 cells transfected with pCDHu54-antisense were fed to E.I. galactosialidosis fibroblasts. Molecular sizes were calculated by comparison with protein markers.

(B) Glycopeptidase F treatment. COS-1 cell-derived, labeled 54 kd precursor was fed to E.I. galactosialidosis cells. Immunoprecipitation was carried out with anti-complex antiserum (lane 1). Part of the immunoprecipitated material was digested with glycopeptidase F (lane 2). Samples were analyzed as in (A).

using anti-32 kd antibodies. Media collected from the three different strains after 72 hr still contain 54 kd precursor protein (Figure 7A, lanes 1, 3, and 4). This precursor form is absent in the medium of galactosialidosis fibroblasts containing proteins secreted by COS-1 cells originally transfected with the antisense plasmid (Figure 7A, lane 2). The 54 kd precursor derived from COS-1 cells is endocytosed by all three cell strains and it is rapidly processed into an immunoprecipitable 32 kd protein (Figure 7A, lanes 5, 7, and 8). The 20 kd polypeptide is not recognized by the anti-32 kd antibodies used in this experiment.

The predicted amino acid sequence of the 54 kd precursor protein (Figure 3A) has potential glycosylation sites at two Asn residues, one within the 32 kd and one within the 20 kd polypeptide. These two components of the "protective protein" exist as a heterodimer (Figure 4). They can only be visualized separately after reduction and by using an antiserum originally raised against the total complex (d'Azzo et al., 1982). The use of the two potential glycosylation sites in vivo could be demonstrated by performing uptake experiments similar to those described above but using anti-complex antiserum. Galactosialidosis fibroblasts of the E.I. type were used as recipient cells. Three days after uptake of COS-1 cell-derived 54 kd precursor, immunoprecipitation was performed followed by digestion with glycopeptidase F, an enzyme that cleaves all types of Asn-bound N-glycans (Plummer et al., 1984; Tarentino et al., 1985). As shown in Figure 7B, the 32 kd and 20 kd proteins undergo deglycosylation, resulting in a shift in apparent mass of about 2000 daltons (lanes 1 and 2). Thus, glycosylation of the 54 kd precursor in COS-1 cells takes place at the two predicted Asn residues. From these data we conclude that the COS-1 cell-derived 54 kd precursor contains all signals necessary for uptake by human fibroblasts and intracellular processing.

We finally tested whether the processed, cDNA-encoded "protective protein" was able to restore lysosomal β-galactosidase and neuraminidase activities in both mutant cell types. As shown in Table 1, COS-1 cell-derived 54 kd precursor, taken up and processed by E.I. and L.I. galactosialidosis fibroblasts, restores both enzyme activities. In the L.I. cells the relative increase in neuraminidase activity is less evident because of the higher endogenous residual activity. To rule out a possible effect of other secreted COS-1 proteins rather than the 54 kd precursor, galactosialidosis cells were also cultured for the same period of time in conditioned medium from COS-1 cells originally transfected with the antisense plasmid. Table 1 shows that β-galactosidase and neuraminidase activities in these fi-

Table 1. Restoration of Enzyme Activities in Galactosialidosis Fibroblasts after Uptake of COS-1 Cell-Derived "Protective Protein" Precursor

Cell Strain	Addition of COS-1 Cell-Derived Proteins	Transfection in COS-1 Cells	Activity (nmol/hr per mg of protein)	
			β-Galactosidase	Neuraminidase
E.I. Galactosialidosis	+	pCDHu54-sense	193	13
	+	pCDHu54-antisense	44	0.9
	–	–	50	0.9
L.I. Galactosialidosis	+	pCDHu54-sense	117	10
	+	pCDH54-antisense	56	5
	–	–	60	6
Normal	+	pCDHu54-sense	233	60
	–	–	294	104

Proteins secreted by COS-1 cells, after transfection with pCDHu54-sense or -antisense, were added to the medium of normal fibroblasts and fibroblasts from E.I. and L.I. galactosialidosis patients. After 72 hr, cells were harvested by trypsinization, and enzyme activities were measured in cell homogenates using synthetic 4-methylumbelliferyl substrates.

broblasts do not increase. The presence or absence of proteins derived from COS-1 cells in the medium of normal cells does not cause any significant difference in enzyme activities. The levels of other lysosomal enzymes tested as internal controls in the different cell homogenates remain unchanged (data not shown).

Discussion

We have isolated and characterized the cDNA encoding the 54 kd precursor of the "protective protein" known to be deficient in patients with the autosomal recessive disease galactosialidosis. In normal human fibroblasts, the cDNA recognizes an mRNA of approximately 2 kb that is not evident in fibroblasts of the E.I. galactosialidosis patient. Parents of this patient, who are consanguineous (Kleijer et al., 1979) and therefore likely to carry the same allelic mutation, have markedly reduced amounts of the 2 kb mRNA. These data demonstrate heterozygosity for the "protective protein" deficiency, for which no direct enzymatic assay is available. The L.I. patient has a normal amount of the 2 kb mRNA species. This is consistent with previous data from immunoprecipitation studies that established the absence of cross-reacting material for the "protective protein" in fibroblasts of the E.I. patient, but the presence of a normal amount of precursor molecules in cells of the L.I. patient (d'Azzo et al., 1982; Palmeri et al., 1986).

The N-terminal sequences of the 32 kd and 20 kd polypeptides are found in the predicted amino acid sequence of the "protective protein" precursor. From their locations we could infer that the two proteins are positioned next to each other within the precursor molecule and we could demonstrate that, after the initial cleavage of the precursor, the two polypeptides remain held together by disulfide bridges. These findings clarify the presence of the 20 kd component in purified complex preparations and show that it is an integral part of the mature "protective protein." The latter is a heterodimer of 32 kd and 20 kd polypeptides.

The predicted molecular masses of the glycosylated precursor (55 kd) and the 20 kd component correlate well with their estimated sizes after SDS-PAGE. There is, however, a difference of 3–4 kd between the predicted size (35.5 kd) and the estimated size of the 32 kd polypeptide. One explanation may be that C-terminal processing of the 32 kd component takes place after initial cleavage of the precursor between residues 298 and 299 (Arg-Met). Conversely, N-terminal processing of the 20 kd component may occur after cleavage of the precursor molecule somewhat before residue 299. The proteolytic event, if it occurs, must be confined to residues 284–298, accounting for about 1.7 kd, since residue 284 is the C-terminus of tryptic peptide T1.

The Hu54 cDNA directs the synthesis in COS-1 cells of the "protective protein" precursor, and the protein is also secreted in substantial amounts. COS-1 cells are not able to process this form intracellularly into 32 kd and 20 kd polypeptides. Nevertheless, the presence of these components has been shown in other mammalian tissues, e.g., bovine testis (Verheijen et al., 1982) and porcine spleen and testis (Yamamoto et al., 1982; Yamamoto and Nishimura, 1987). At present we do not know whether COS-1 cells fail to recognize the processing site of human 54 kd precursor or whether the transfection procedure and subsequent production of the precursor molecule somehow interfere with this cleavage event. In contrast, correct intracellular processing of COS-1 cell-derived 54 kd precursor takes place in normal human fibroblasts as well as in both E.I. and L.I. galactosialidosis cells. This result implies that this precursor protein acquires all molecular characteristics for further posttranslational modifications.

The cDNA-encoded precursor also yields a biologically functional "protective protein" capable of correcting β-galactosidase and neuraminidase activities after being endocytosed and processed by mutant fibroblasts. This indirect approach of transient expression in COS-1 cells and uptake by deficient fibroblasts offers several opportunities for testing the different functions of the "protective protein." For instance, expression of in vitro mutagenized cDNAs may prove useful for determining the structural domains in the "protective protein" involved in the association with β-galactosidase and neuraminidase.

The most striking observation in this report is the homology of the "protective protein" to yeast CPY and the *KEX1* gene product. CPY is a vacuolar protease involved in the degradation of small peptides (for review see Jones, 1984). It is synthesized as a larger, inactive precursor (Hasilik and Tanner, 1978) that is converted to the mature, active form prior to or upon delivery to the vacuole (Hemmings et al., 1981). The *KEX1* gene product in yeast is involved in the proteolytic processing of killer toxin and α-pheromone precursors, probably at the Golgi stage in the secretory pathway (Dmochowska et al., 1987). It completes maturation of subunits of these proteins by the removal of dibasic amino acid residues from C-termini. This proteolytic step is essential for the activation of mature killer toxin and α-pheromone. The carboxypeptidase B–like activity of the *KEX1* product is exerted only after cleavage of the precursor molecules by an endopeptidase.

Homology between CPY, the *KEX1* gene product, and the "protective protein" is extensive, and the stretch of amino acids that spans the active site serine in CPY (Hayashi et al., 1973) is completely identical in the three proteins. Dmochowska et al. (1987) have demonstrated that mutation of this serine to alanine in the *KEX1* gene product impairs its activity, strongly suggesting that *KEX1* is, like CPY, a serine protease. It is of interest that the two proteases can apparently catalyze similar reactions using a conserved active site but are involved in different proteolytic pathways and function in different subcellular compartments.

The conserved stretch of six amino acids including the active-site serine is located within the 32 kd polypeptide of the "protective protein." The homology suggests a putative protease activity for the "protective protein." Moreover, not only is the "serine domain" completely identical in the three proteins, but there are also two other identical stretches, one containing a cysteine and one a histidine residue. A sulfhydryl group as well as a histidine residue has been suggested to contribute to the enzymatic activity of CPY (Hayashi et al., 1973; Kuhn et al., 1974). It is noteworthy that, in the case of the "protective protein," His-429 is located in the 20 kd polypeptide. This implies that this component might also be involved in the putative protease activity of the "protective protein," which supports the proposed model of the heterodimeric state of its mature form. We can speculate that this protease activity is responsible for specific proteolytic modifications of β-galactosidase and neuraminidase at their C-termini, resulting in the stabilization of the former and in the activation of the latter enzyme. The "protective protein" could therefore be the first example of a processing enzyme with carboxypeptidase B–like activity in man.

The characterization of the cDNA encoding the "protective protein" has enabled a better understanding of the protein's primary structure and possible mode of action. At the same time, the sequence homologies revealed here lead us to speculate on the paradox of a protective function named to a potential proteolytic activity. Studies of the tertiary structure of the "protective protein" and its interaction with β-galactosidase and neuraminidase will become feasible once the genes coding for the latter two components are cloned.

Note: We have now established that COS-1 cell-derived "protective protein" precursor, taken up by untreated COS-1 cells, is correctly converted into the heterodimeric form. This implies that the impaired processing of the precursor molecule in transfected cells is probably attributable to the transfection procedure.

Experimental Procedures

Cell Culture

Human skin fibroblasts from normal individuals and a patient with the E.I. form of galactosialidosis (Kleijer et al., 1979) were obtained from the Rotterdam Cell Repository (Dr. M. F. Niermeijer). Cells from the patient with the L.I. form of galactosialidosis (Andria et al., 1978) were provided by Dr. G. Andria, Dept. of Pediatrics, University of Naples, Italy. Fibroblasts were maintained in Dulbecco's modified Eagle's medium–Ham's F10 medium (1:1 vol/vol) supplemented with antibiotics and 10% fetal calf serum.

Protein Sequence Analysis

The complex of β-galactosidase, neuraminidase, and "protective protein" was purified from human placenta as previously described (Verheijen et al., 1985). The different components were separated by SDS-PAGE under reducing and denaturing conditions according to Laemmli (1970) with minor modifications (Hasilik and Neufeld, 1980). The 32 kd protein band was digested in situ with TPCK-trypsin (Worthington Diagnostic Systems Inc., U.K.). Tryptic peptides were fractionated by HPLC (Waters 6000 System) using a Du Pont Zorbax C8 4.6 mm × 25 cm reverse-phase column (Anachem Ltd., U.K.) and a 5%–60% gradient of acetonitrile in 0.1% trifluoroacetic acid; detection was at 216 nm.

Sequence analysis of peptide fractions was performed by automated Edman degradation on an Applied Biosystems 470A gas-phase peptide sequenator (Hunkapiller and Hood, 1983). PTH-amino acids were analyzed on-line using an Applied Biosystems 120A analyzer (Southan et al., 1987). 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.) as described by Towbin et al. (1979). Filter pieces containing either the 32 kd or the 20 kd protein were cut out and used as starting material for automated Edman degradation (Aebersold et al., 1986). Amino acid sequencing was performed at the Biotechnology Instrumentation Facility (University of California, Riverside).

cDNA Library Screening

A human testis cDNA library in λgt11 (Young and Davis, 1983; 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 PFU per 90 mm plate and screened with anti-32 kd antibodies as previously described (Huynh et al., 1985). The same library, plated out at a density of 2×10^5 PFU per 22 × 22 cm plate, was screened with oligonucleotide probes labeled at the 5′ end with ^{32}P using [γ-^{32}P]ATP and polynucleotide kinase (Maniatis et al., 1982). The probes were synthesized on an Applied Biosystems 381A oligonucleotide synthesizer. Hybridization and washing conditions were as described by Wood et al. (1985).

DNA Sequencing

The Hu54 cDNA insert and its restriction fragments were subcloned into the plasmid vectors pTZ18 and pTZ19. Nucleotide sequences on both strands were obtained by the dideoxy chain-termination method (Sanger et al., 1977). M13 universal reverse primer as well as synthetic oligonucleotides were used. Sequence data were analyzed using the program of Staden (1986). Homology searches of the EMBL (release 14.0; 1988), NBRF Protein (release 14.0; 1987), and SWISS-PROT (release 5.0; 1987) data bases were performed using the program of Lipman and Pearson (1985). Protein alignment was done with the sequence analysis sofware package of the University of Wisconsin Genetics Computer Group (Devereux et al., 1984). Sequences have

been deposited in the EMBL/GenBank data base (accession no. J03159).

Northern Blot Analysis

Total RNA was isolated from cultured fibroblasts according to the procedure of Auffray and Rougeon (1980). RNA samples were electrophoresed on a 0.8% agarose gel containing 2.2 M formaldehyde (Maniatis et al., 1982) and blotted onto a nitrocellulose filter. The filter was hybridized with the cDNA probe labeled according to the procedure of Feinberg and Vogelstein (1983).

Transient Expression of cDNA in COS-1 Cells

The 1.8 kb Hu54 cDNA EcoRI insert was subcloned in two orientations in a derivative of the mammalian expression vector pCD-X (Okayama and Berg, 1983) in which the PstI–BamHI fragment 3′ of the SV40 promoter was substituted by a polylinker containing an EcoRI site (a gift from N. Heisterkamp, University of Southern California). COS-1 cells (Gluzman, 1981) were maintained in the same culturing medium as human fibroblasts but supplemented with only 5% fetal calf serum. Two days before transfection, cells were seeded in 100 mm Petri dishes and grown to 30% confluency.

Transfections with pCDHu54 constructs and biosynthetic labeling with [^{35}S]methionine were performed according to the method of van Heuvel et al. (1986) except that cells were incubated in methionine-free medium for 1 hr before labeling. The latter was carried out for 16 hr in the presence or absence of 10 mM NH_4Cl (Hasilik and Neufeld, 1980). The preparation of cell extracts and medium concentrates and the immunoprecipitation of the "protective protein" and its precursor were performed as reported earlier (Proia et al., 1984). Staphylococcus aureus cells, used to precipitate antigen–antibody complexes, were treated before use as recommended by the supplier, thereby introducing reducing agent in the immunoprecipitation assay. Immunoprecipitated proteins were resolved by SDS-PAGE under reducing and denaturing conditions. Radioactive bands were visualized by fluorography of gels impregnated with En^{3}Hance (New England Nuclear). Apparent molecular weights were calculated by comparison with conventional marker proteins.

Uptake Studies in Human Cells

Medium (5 ml) from transfected and ^{35}S-labeled COS-1 cells cultured in 25 cm^2 flasks was concentrated 2000-fold by ammonium sulfate precipitation and was desalted afterward through a Sephadex G-50 column (Proia et al., 1984). This concentrated material (250 μl) was divided into two aliquots. Each aliquot was added to fresh medium of confluent human fibroblasts plated in 60 mm dishes. The cell strains used were from a normal individual and from the E.I. and L.I. galactosialidosis patients. After an additional 3 days of subculture, cell extracts and medium concentrates were prepared for immunoprecipitation analysis as described by Proia et al. (1984). Samples were incubated with anti-32 kd antibodies for 16 hr at room temperature and were subsequently treated with S. aureus suspension (10% wt/vol) for 30 min on ice. Immunoprecipitated radiolabeled proteins were separated by SDS-PAGE and visualized by fluorography as above. In parallel experiments, identically treated cells were harvested by trypsinization and homogenized by vortexing in double-distilled water (Galjaard, 1980). Enzyme activities were measured in cell homogenates with artificial 4-methylumbelliferyl substrates using standard assay conditions (Galjaard, 1980). Total protein concentrations were determined by the method of Lowry (1951). In some instances, immunoprecipitation of radiolabeled proteins was carried out using an antibody preparation recognizing the components of the complex, including the 20 kd polypeptide (d'Azzo et al., 1982). Immunoprecipitated material was divided into two aliquots: one was left untreated, and the other was subjected to glycopeptidase F digestion for 16 hr at 37°C using conditions recommended by the supplier. Treated and untreated proteins were resolved and visualized as described above.

Acknowledgments

We are indebted to Dr. Gerard Grosveld for continuous support and good suggestions. We would like to thank Dr. Mike Geisow and Dr. Gary Hathaway for their expert help with the protein sequencing, Dr. Frank Grosveld for providing a good start on this project, and Prof. Don Robinson for stimulating discussions and critical reading of the manuscript. We also thank Jaap Bosveld for advice on the transfection experiments, Sjozef van Baal for assistance with the computer data analysis, Cor van Dijk and Mirko Kuit for the photographic work, and Nellie van Sluysdam and Diana Helmsius for editing the manuscript.

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.

Received March 28, 1988; revised May 23, 1988.

References

Aebersold, R. H., Teplow, D. B., Hood, L. E., and Kent, S. B. H. (1986). Electroblotting onto activated glass. J. Biol. Chem. *261*, 4229–4238.

Andria, G., Del Giudice, E., and Reuser, A. J. J. (1978). Atypical expression of β-galactosidase deficiency in a child with Hurler-like features but without neurological abnormalities. Clin. Genet. *14*, 16–23.

Andria, G., Strisciuglio, P., Pontarelli, G., Sly, W. S., and Dodson, W. E. (1981). Infantile neuraminidase and β-galactosidase deficiencies (galactosialidosis) with mild clinical courses. In Sialidases and Sialidoses, G. Tettamanti, P. Durand, and S. DiDonato, eds. (Milan: Edi. Ermes), pp. 379–395.

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.

Benham, F. J., Hodgkinson, S., and Davies, K. E. (1984). A glyceraldehyde-3-phosphate dehydrogenase pseudogene on the short arm of the human X chromosome defines a multigene family. EMBO J. *3*, 2635–2640.

Cantz, M., Mendla, K., and Baumkötter, J. (1984). Sialidosis and related disorders: the nature of the sialidase defect. In Recent Progress in Neurolipidoses and Allied Disorders, M. T. Vanier, ed. (Lyon, France: Fondation M. Mérieux), pp. 137–145.

Corfield, A. P., Michalski, J. C., and Schauer, R. (1981). The substrate specificity of sialidases from microorganisms and mammals. In Sialidases and Sialidoses, G. Tettamanti, P. Durand, and S. DiDonato, eds. (Milan: Edi. Ermes), pp. 1–70.

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.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. *12*, 387–395.

Dmochowska, A., Dignard, D., Henning, D., Thomas, D. Y., and Bussey, H. (1987). Yeast *KEX1* gene encodes a putative protease with a carboxypeptidase B-like function involved in killer toxin and α-factor precursor processing. Cell *50*, 573–584.

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

Galjaard, H. (1980). Genetic Metabolic Disease: Early Diagnosis and Prenatal Analysis. (Amsterdam: Elsevier North Holland).

Galjaard, H., d'Azzo, A., Hoogeveen, A. T., and Verheijen, F. W. (1984). Combined β-galactosidase–sialidase deficiency in man: genetic defect of a "protective protein." In Molecular Basis of Lysosomal Storage Disorders, J. A. Barranger and R. O. Brady, eds. (New York: Academic Press), pp. 113–132.

Gluzman, Y. (1981). SV40-transformed simian cells support the replication of early SV40 mutants. Cell *23*, 175–182.

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.

Hasilik, A., and Neufeld, E. F. (1980). Biosynthesis of lysosomal enzymes in fibroblasts. J. Biol. Chem. *255*, 4937–4945.

Hasilik, A., and Tanner, W. (1978). Biosynthesis of the vacuolar yeast

glycoprotein carboxypeptidase Y. Conversion of precursor into the enzyme. Eur. J. Biochem. *85*, 599–608.

Hayashi, R., Moore, S., and Stein, W. (1973). Serine at the active center of yeast carboxypeptidase. J. Biol. Chem. *248*, 8366–8369.

Hemmings, B. A., Zubenko, G. S., Hasilik, A., and Jones, E. W. (1981). Mutant defective in processing of an enzyme located in the lysosome-like vacuole of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA *78*, 435–439.

Hoeksema, H. L., van Diggelen, O. P., and Galjaard, H. (1979). Intergenic complementation after fusion of fibroblasts from different patients with β-galactosidase deficiency. Biochim. Biophys. Acta *566*, 72–79.

Hoogeveen, A. T., d'Azzo, A., Brossmer, R., and Galjaard, H. (1981). Correction of combined β-galactosidase/neuraminidase deficiency in human fibroblasts. Biochem. Biophys. Res. Commun. *103*, 292–300.

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.

Hunkapiller, M. W., and Hood, L. E. (1983). Protein sequence analysis: automated microsequencing. Science *219*, 650–659.

Huynh, T. V., Young, R. A., and Davis, R. W. (1985). Constructing and screening cDNA libraries in λgt10 and λgt11. In DNA Cloning: A Practical Approach, Vol. 1, D. M. Glover, ed. (Oxford: IRL Press), pp. 49–78.

Jones, E. W. (1984). The synthesis and function of proteases in *Saccharomyces*: genetic approaches. Annu. Rev. Genet. *18*, 233–270.

Kleijer, W. J., Hoogeveen, A. T., Verheijen, F. W., Niermeijer, M. F., Galjaard, H., O'Brien, J. S., and Warner, T. G. (1979). Prenatal diagnosis of sialidosis with combined β-galactosidase and neuraminidase deficiency. Clin. Genet. *16*, 60–61.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell *44*, 283–292.

Kuhn, R. W., Walsh, K. A., and Neurath, H. (1974). Isolation and partial characterization of an acid carboxypeptidase from yeast. Biochemistry *13*, 3871–3877.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680–685.

Lipman, D. J., and Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. Science *227*, 1435–1441.

Lowry, O. H., Rosebrough, N. J., Farr, A. L., and Randall, R. J. (1951). Protein measurement with the Folin phenol reagent. J. Biol. Chem. *193*, 265–275.

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

Norden, A. G. W., Tennant, L., and O'Brien, J. S. (1974). Ganglioside G_{M1} β-galactosidase A: purification and studies of the enzyme from human liver. J. Biol. Chem. *249*, 7969–7976.

O'Brien, J. S. (1983). The gangliosidoses. In The Metabolic Basis of Inherited Disease, J. B. Stanbury, J. B. Wijngaarden, D. S. Fredrickson, J. L. Goldstein, and M. S. Brown, eds. (New York: McGraw-Hill), pp. 945–969.

Okayama, H., and Berg, P. (1983). A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. Mol. Cell. Biol. *3*, 280–289.

Palmeri, S., Hoogeveen, A. T., Verheijen, F. W., and Galjaard, H. (1986). Galactosialidosis: molecular heterogeneity among distinct clinical phenotypes. Am. J. Hum. Genet. *38*, 137–148.

Plummer, T. H., Jr., Elder, J. H., Alexander, S., Phelan, A. W., and Tarentino, A. L. (1984). Demonstration of peptide:N-glycosidase F activity in endo-β-N-acetylglucosaminidase F preparations. J. Biol. Chem. *259*, 10700–10704.

Prola, R. L., d'Azzo, A., and Neufeld, E. F. (1984). Association of α- and β-subunits during biosynthesis of β-hexosaminidase in cultured fibroblasts. J. Biol. Chem. *259*, 3350–3354.

Sanger, F. G., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Southan, C., Aitken, A., Childs, R. A., Abbott, W. M., and Feizi, T. (1987). Amino acid sequence of β-galactosidase-binding bovine heart lectin. Member of a novel class of vertebrate proteins. FEBS Lett. *214*, 301–304.

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

Suzuki, Y., Sakuraba, H., Yamanaka, T., Ko, Y. M., Iimori, Y., Okamura, Y., and Hoogeveen, A. T. (1984). Galactosialidosis: a comparative study of clinical and biochemical data on 22 patients. In The Developing Brain and Its Disorders, M. Arima, Y. Suzuki, and H. Yabuuchi, eds. (Tokyo: University of Tokyo Press), pp. 161–175.

Tarentino, A. L., Gomez, C. M., and Plummer, T. H., Jr. (1985). Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F. Biochemistry *24*, 4665–4671.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA *76*, 4350–4354.

Valls, L. A., Hunter, C. P., Rothman, J. H., and Stevens, T. H. (1987). Protein sorting in yeast: the localization determinant of yeast vacuolar carboxypeptidase Y resides in the propeptide. Cell *48*, 887–897.

van Diggelen, O. P., Hoogeveen, A. T., Smith, P. J., Reuser, A. J. J., and Galjaard, H. (1982). Enhanced proteolytic degradation of normal β-galactosidase in the lysosomal storage disease with combined β-galactosidase and neuraminidase deficiency. Biochim. Biophys. Acta *703*, 69–76.

van Heuvel, M., Bosveld, I. J., Mooren, A. T. A., Trapman, J., and Zwarthoff, E. C. (1986). Properties of natural and hybrid murine alpha interferons. J. Gen. Virol. *67*, 2215–2222.

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.

Verheijen, F. W., Palmeri, S., Hoogeveen, A. T., and Galjaard, H. (1985). Human placental neuraminidase. Activation, stabilization and association with β-galactosidase and its protective protein. Eur. J. Biochem. *149*, 315–321.

Von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. Nucl. Acids Res. *14*, 4683–4690.

Wenger, D. A., Tarby, T. J., and Wharton, C. (1978). Macular cherry-red spots and myoclonus with dementia: coexistent neuraminidase and β-galactosidase deficiencies. Biochem. Biophys. Res. Commun. *82*, 589–595.

Willemsen, R., Hoogeveen, A. T., Sips, H. J., van Dongen, J. M., and Galjaard, H. (1986). Immunoelectron microscopical localization of lysosomal β-galactosidase and its precursor forms in normal and mutant human fibroblasts. Eur. J. Cell Biol. *40*, 9–15.

Wood, W. I., Gitschier, J., Lasky, L. A., and Lawn, R. M. (1985). Base composition-independent hybridization in tetramethylammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. Proc. Natl. Acad. Sci. USA *82*, 1585–1588.

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 interrelation between high- and low-molecular-weight forms of G_{M1} β-galactosidase purified from porcine spleen. J. Biochem. *92*, 13–21.

Young, R. A., and Davis, R. W. (1983). Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. USA *80*, 1194–1198.

Publication 2

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 post-translationally 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. T_4 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. T_4 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}, II^{3}NeuAc-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 GTG GCC GTG GAC GGT ATC CCC	(1)
	Q G V L E R CAG GGC GTG CTG GAG CGC	
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*.

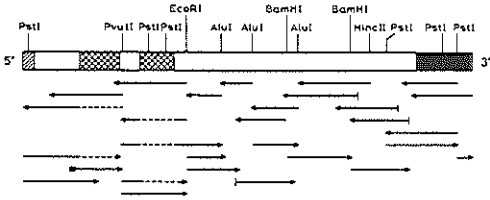


FIG. 2. **Composite restriction maps and sequencing strategy of HβGa cDNA clones.** All restriction enzyme sites in the cDNAs used for subcloning are shown. *Arrows* indicate the direction and extent of sequencing reactions. *Cross-hatched boxes* represent the two protein-coding regions not present in 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. 1*A*), 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

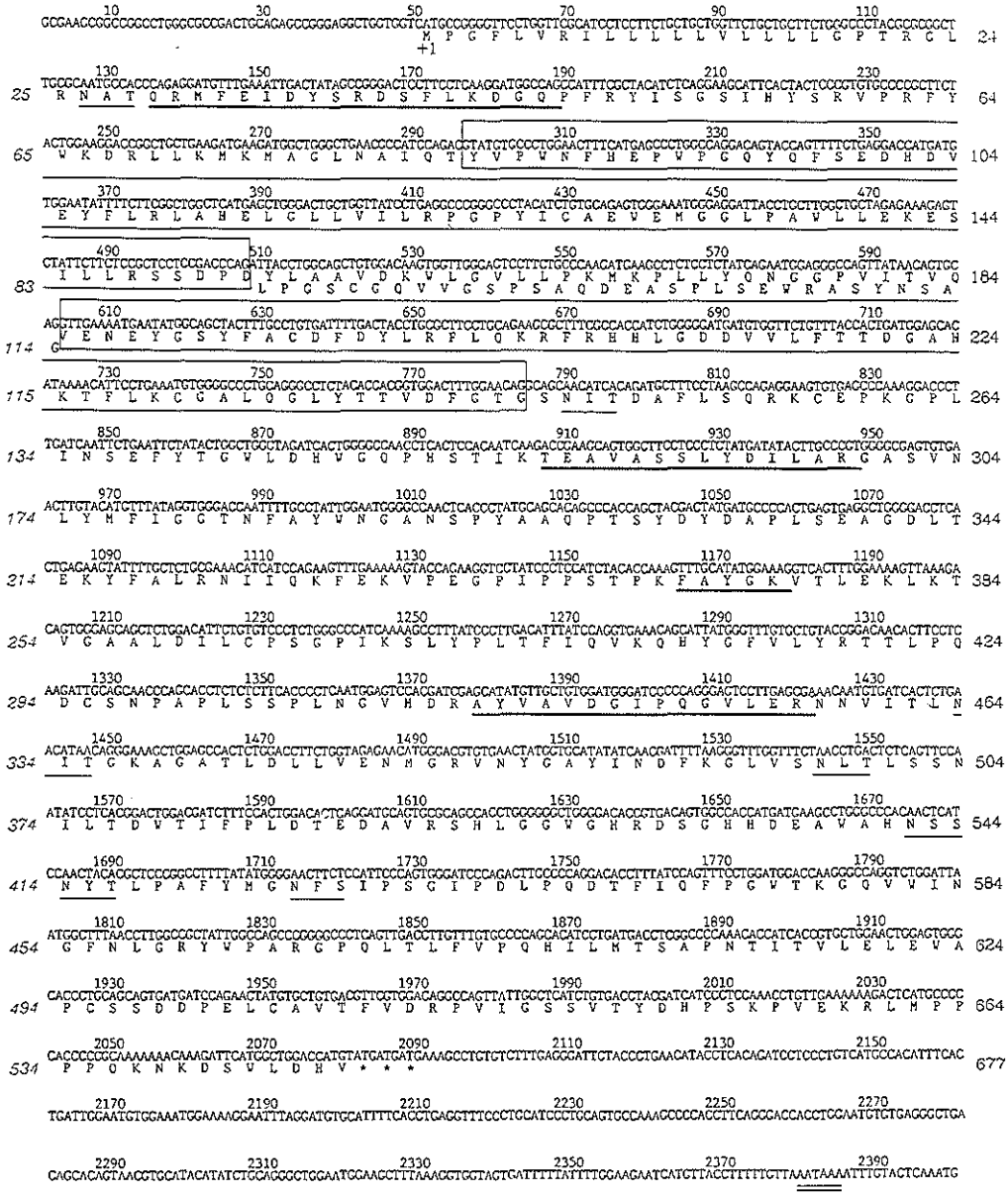


FIG. 3. **Combined nucleotide and predicted amino acid sequences of β-galactosidase cDNAs.** The numbers *above* the nucleotide sequence refer to HβGa(L) cDNA. Numbers on the *right* and on the *left* specify the amino acids of the long and short β-galactosidase-predicted sequences, respectively. Amino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

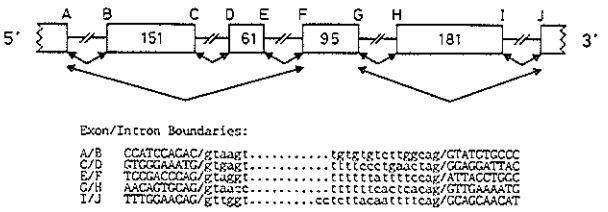


FIG. 4. **Exons involved in alternative splicing, generating human β-galactosidase long and short mRNAs.** *Boxes* represent separate exons; *numbers* specify their length in base pairs. *Interrupted lines* depict intronic sequences of unknown size. The sequences of all exons have been determined except 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 mole-

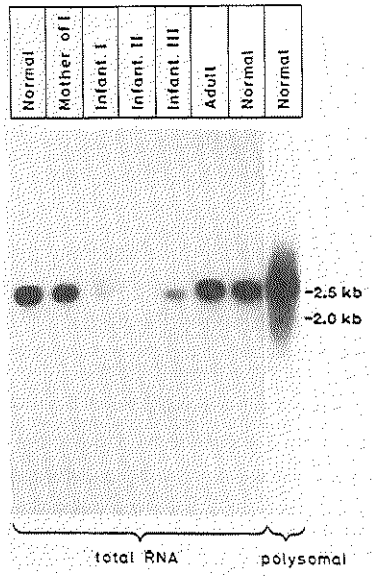


FIG. 5. **Northern blot analysis of fibroblast RNA.** Total and polysomal fibroblast RNA from three normal individuals and total fibroblast RNA from an adult and three infantile G_{M1}-gangliosidosis patients as well as one heterozygote were fractionated on a formaldehyde-agarose gel and probed with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

cule 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

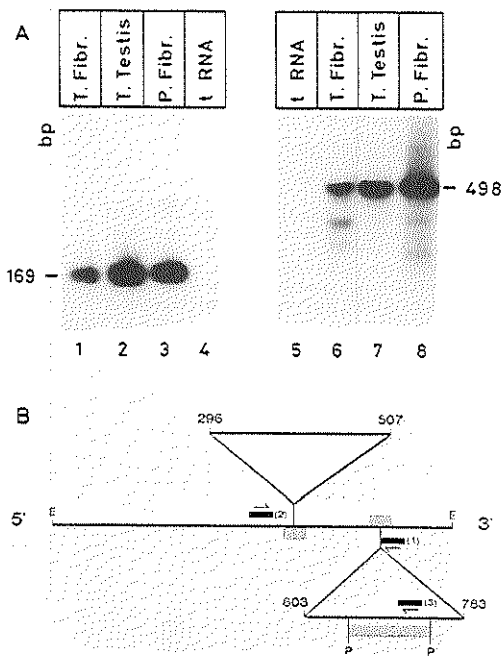


FIG. 6. **Detection of two mRNAs for β-galactosidase by PCR amplification.** *A*, total (*T. Fibr.*) and polysomal fibroblast (*P. Fibr.*) RNA and total testis (*T. Testis*) RNA were used to synthesize single-stranded cDNAs that were subsequently subjected to 24 rounds of amplification. Amplified products 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 *8*). 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-

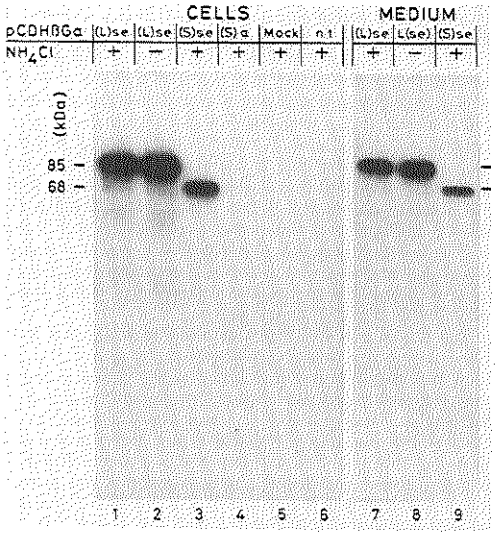


FIG. 7. **Transient expression of pCDHβGa constructs in COS-1 cells.** The pCDHβGa(L) and pCDHβGa(S) cDNA constructs in the sense (*se, lanes 1–3* and *7–9*) and antisense (*a, lane 4*) orientations were transfected to COS-1 cells. A mock transfection was carried out without the addition of DNA (*lane 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 5-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[a]/mg protein
pCDHβGa(L)-sense	8.9
pCDHβGa(S)-sense	1.4
pCDHβGa(S)-anti	1.8
Mock-transfected	1.7
Not transfected	1.5

[a] 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

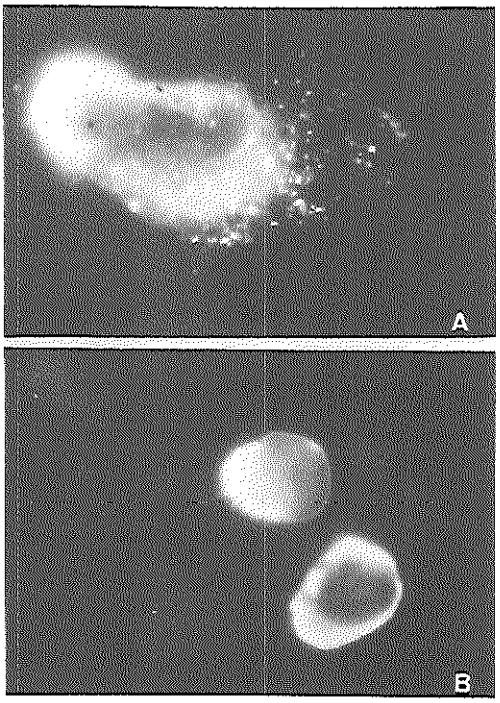


FIG. 8. **Immunocytochemical localization of β-galactosidase proteins in transiently transfected COS-1 cells.** *A*, transfection with pCDHβGa(L) cDNA construct; *B*, transfection with pCDH-βGa(S) cDNA construct.

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., Wiessman, 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. Breathnach, 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 3

J. Biol. Chem. **265** (1990), 4678-4684

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1990 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 265, No. 8, Issue of March 15, pp. 4678–4684, 1990
Printed in U.S.A.

Mouse "Protective Protein"

cDNA CLONING, SEQUENCE COMPARISON, AND EXPRESSION*

(Received for publication, November 3, 1989)

Niels J. Galjart‡, Nynke Gillemans‡, Dies Meijer, and Alessandra d'Azzo§

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

The "protective protein" is the glycoprotein that forms a complex with the lysosomal enzymes β-galactosidase and neuraminidase. Its deficiency in man leads to the metabolic storage disorder galactosialidosis. The primary structure of human protective protein, deduced from its cloned cDNA, shows homology to yeast serine carboxypeptidases.

We have isolated a full-length cDNA encoding murine protective protein. The nucleotide sequences as well as the predicted amino acid sequences are highly conserved between man and mouse. Domains important for the protease function are completely identical in the two proteins. Both human and mouse mature protective proteins covalently bind radiolabeled diisopropyl fluorophosphate. Transient expression of the murine cDNA in COS-1 cells yields a protective protein precursor of 54 kDa, a size characteristic of the glycosylated form. This cDNA-encoded precursor, endocytosed by human galactosialidosis fibroblasts, is processed into a 32- and a 20-kDa heterodimer and corrects β-galactosidase and neuraminidase activities. A tissue-specific expression of protective protein mRNA is observed when total RNA from different mouse organs is analyzed on Northern blots.

Galactosialidosis is a human autosomal recessive disorder characterized by a combined deficiency of lysosomal β-D-galactosidase (EC 3.2.1.23) and *N*-acetyl-α-neuraminidase (EC 3.2.1.18) because of a primary defect in the protective protein (1–3). Different clinical phenotypes exist in patients with this disease, ranging from severe infantile forms to milder late infantile and juvenile/adult types (1, 4).

In cultured human fibroblasts the first immunoprecipitable form of the protective protein is a glycosylated precursor of 54 kDa, which undergoes post-translational processing (3). The primary structure of human protective protein has been determined recently through cDNA cloning (5). From the deduced amino acid sequence we could infer that the precursor molecule is proteolytically cleaved into a mature heterodimer of 32- and 20-kDa polypeptides, held together by disulfide bridges (5).

It was shown earlier that in human cells the protective protein associates with lysosomal β-galactosidase since it is resolved together with the latter enzyme in a high molecular mass aggregate (600–700 kDa). This association promotes multimerization of β-galactosidase monomers, which, in turn, stabilizes the enzyme (6). In addition, neuraminidase can be purified in an active and stable form together with β-galactosidase and the protective protein (7, 8). Although it is likely that the catalytic site of human placental neuraminidase resides on a 66-kDa polypeptide, the presence of the protective protein is essential for expression of neuraminidase activity (9). Taken together, these studies indicate that a complex of β-galactosidase, neuraminidase, and protective protein may be functionally present in lysosomes, but so far no data on the stoichiometry of this complex have accumulated.

The primary structure of human protective protein (5) shares homology with yeast carboxypeptidase Y (10), the KEX1 gene product (11), and certain plant serine carboxypeptidases (12–14). All these proteins have in common three conserved amino acid stretches that include the serine and the histidine residues essential for their catalytic activity (15; for review see Ref. 16 and references therein). This implies that the protective protein might function as a serine protease.

In order to define domains involved in the proteolytic activity of the protective protein and/or its binding to β-galactosidase and neuraminidase, we have isolated the cDNA encoding murine protective protein. Our results show that mouse and human cDNA sequences are largely homologous, giving rise to almost identical proteins.

EXPERIMENTAL PROCEDURES

Materials—Restriction endonucleases were purchased from the following companies: Boehringer Mannheim, Bethesda Research Laboratories, New England Biolabs, Pharmacia LKB Biotechnology Inc., and Promega Biotec. DNA polymerase, Klenow fragment, was from Promega Biotec. M13 universal and reverse sequencing primers and pTZ18 and 19 plasmid vectors were obtained from Pharmacia. The synthetic oligonucleotide, used for sequence determination, was synthesized on an Applied Biosystems 381A oligonucleotide synthesizer. Sequenase and sequencing kit were purchased from United States Biochemical Corp. Immunoprecipitin and prestained molecular weight markers were from Bethesda Research Laboratories. BCA protein assay reagents were purchased from Pierce Chemical Co. [α-^{32}P]dATP, [α-^{32}P]dCTP (3000 Ci/mmol), [α-^{35}S]dATPαS[1] (>1000 Ci/mmol), [^{35}S]methionine (>1000 Ci/mmol), and [^{3}H]DFP (3.0 Ci/mmol) were obtained from Amersham Corp. and Du Pont-New England Nuclear. All other reagents were from standard commercial suppliers, if not specified otherwise.

Cell Culture—Fibroblasts from an early infantile galactosialidosis patient (5) were obtained from the European Human Cell Bank, Rotterdam (Dr. W. J. Kleijer). They were maintained in Dulbecco's modified Eagle's medium-Ham's F-10 medium (1:1, v/v), supple-

* 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) J05261.

‡ These authors contributed equally to the results described in this report.

§ 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: dATPαS, deoxyadenosine 5′-*O*-(3-thiotriphosphate); DFP, diisopropyl fluorophosphate; SDS, sodium dodecyl sulfate; kb, kilobase(s).

mented with 10% fetal bovine serum and antibiotics. COS-1 cells (17) were maintained in the same culture medium but supplemented with only 5% fetal bovine serum.

Isolation of cDNA Clones—Construction of the mouse testis cDNA library in λgt10 has been described previously (18). The mouse brain cDNA library was prepared by D. M. according to the same procedure. The testis cDNA library was plated out at a density of 2–3 × 10^5 plaque-forming units/20 × 20-cm plate. Screening was performed using the heterologous Hu54 cDNA as a probe (5, 19). Standard hybridization and washing conditions were used except that the washing temperature was lowered to 60 °C. Positive phages were isolated after two further steps of plaque purification. The 1.2-kb insert of clone λT2 was used to screen the brain cDNA library, plated out as described above. This library was subsequently screened with a 5′ end fragment of the Hu54 cDNA (nucleotides 1–660 (5)). One clone (λB1) positive in both hybridizations was selected for further investigation.

DNA Sequence Analysis—Suitable restriction endonuclease sites in the inserts of λB1 and λT2 were used for subcloning of fragments into plasmid vectors pTZ18 and pTZ19. Nucleotide sequences were determined using the dideoxy chain termination method on single-stranded DNA (20) and double-stranded DNA (21). The reaction conditions were modified for use with [α-^{35}S]dATPαS and Sequenase, as recommended by the supplier. Sequence data were analyzed with the programs of Staden (22) and the University of Wisconsin Genetics Computer Group (23). DNA and protein alignments were done with the latter software package.

Transfections in COS-1 Cells—A *Bam*HI restriction site, shared by B1 and T2 cDNAs, was used to construct full-length mouse protective protein cDNA (Mo54). The resulting composite *Eco*RI insert was subcloned in two orientations into a derivative of the mammalian expression vector pCD-X as described previously (5). Transfections in COS-1 cells, metabolic labeling of transfected cells with [^{35}S] methionine, and preparation of cell extracts were performed as reported earlier (5, 24). Secreted proteins were concentrated 20-fold by $(NH_4)_2SO_4$ precipitation and desalted afterward on a Sephadex G-50 column (24). An aliquot (about 1%) was taken for direct analysis on SDS-polyacrylamide gel electrophoresis. The rest of the concentrated medium from different transfection experiments was either used for immunoprecipitation (about 25%) or in uptake studies (about 75%; see below). Cell extracts and medium concentrates were immunoprecipitated with anti-human 32-kDa antibodies and resolved on 12.5% polyacrylamide gels under reducing conditions (25). Radioactive proteins were visualized by fluorography of dried gels that had been treated with Amplify (Amersham Corp.).

Uptake of Protective Protein Precursors in Human Cells—Aliquots of medium concentrates from transfected COS-1 cells were added to the medium of fibroblasts from an early infantile galactosialidosis patient. After 3 days of uptake, cells were processed for either immunoprecipitation analysis or enzyme activity assays as reported previously (5). The anti-human complex antibodies used in these experiments have been described earlier (3).

DFP-binding Assay—COS-1 cells, seeded in six-well plates, were transfected with various pCD constructs as described above and maintained afterward in normal culture medium for 72 h. The cells were then harvested by trypsinization, rinsed 3 × with phosphate-buffered saline, and homogenized in ice-cold water. Confluent human fibroblasts from a normal individual and the early infantile galactosialidosis patient were cultured in six-well plates for 6 days and harvested as above. Cell lysates were adjusted to 0.01 M sodium phosphate, pH 6.8, and freeze-thawed once. To these homogenates 3 μCi of [^{3}H]DFP was added, and binding of the inhibitor was allowed for 1 h at room temperature. Radiolabeled proteins were immunoprecipitated using the anti-human complex antibodies (3) and subsequently visualized by SDS-polyacrylamide gel electrophoresis and fluorography.

Northern Blot Analysis—Adult BCBA mice and fetuses at days 13, 16, or 19 of gestation were used as sources of tissues for RNA extraction. Adult mice were killed by cervical dislocation; tissues were removed and immediately frozen in liquid nitrogen. Total RNA was extracted using the method of Auffray and Rougeon (26). 10–15 μg of RNA from tissues and embryos was dissolved in loading buffer containing ethidium bromide and applied on a 0.8% agarose gel containing 0.66 M formaldehyde (27). After electrophoresis, the RNA was visualized by UV illumination at 300 nm and subsequently blotted onto a Zeta-Probe membrane (Bio-Rad) (27). Standard hybridization and washing conditions were applied (28).

RESULTS

Isolation of cDNA Clones and Nucleotide Sequence Analysis—Human protective protein cDNA Hu54 (5) was used as a probe to screen a mouse testis λgt10 cDNA library (18). Out of 2–3 × 10^6 plaques, six positive clones were isolated, all with an insert size of about 900 base pairs which hybridized with the 3′ end of the Hu54 cDNA (nucleotides 1200–1800) but not with a 5′ end probe. One clone, λT2 (Fig. 1), was selected for further analysis, and its insert was used to screen a mouse brain λgt10 cDNA library. Out of 1.6 × 10^6 plaques, 33 recombinant clones were detected, one of which (λB1) strongly hybridized with a fragment of the human cDNA spanning base pairs 1–660 (5). Its insert was mapped relative to Hu54 and T2 cDNAs. As shown in Fig. 1, B1 and T2 overlap partially and share a common *Bam*HI restriction site that was used to create a composite cDNA. The total length of the two inserts (2.0 kb) was sufficient to contain all coding information for a mouse protective protein having a size of 54 kDa.

B1 and T2 cDNAs were subcloned separately and sequenced on both strands using the strategy depicted in Fig. 1. The complete nucleotide sequence and predicted amino acid sequence are shown in Fig. 2. The mouse protective protein cDNA Mo54 is 1987 nucleotides long. It can be divided into a 5′-untranslated region, an open reading frame, and a 3′-noncoding region of 213, 1422, and 352 nucleotides, respectively. In the 5′-untranslated region, a putative translation initiation codon is found at position 12. This ATG, however, is immediately followed by a stop codon. A second ATG, which starts at position 214, is in the correct context for translation initiation (29) and could therefore represent the true initiator, also by comparison with the human cDNA. In the 3′-untranslated region, a putative polyadenylation signal (ATTAAA) is present at position 1966–1971.

When Mo54 and Hu54 cDNAs are aligned at the nucleotide level, the most striking feature is an identity of 72% throughout their entire 3′-noncoding regions (Fig. 3). This is not due to the presence of an open reading frame shared by human and mouse cDNAs. Identity in the coding regions of the two cDNAs is 85%.

Predicted Amino Acid Sequence of Mouse Protective Protein—The open reading frame of 1,422 nucleotides in Mo54 cDNA encodes a protein of 474 amino acids (Fig. 2) with a predicted M_r of 53,844. The amino acid sequence of mouse protective protein is highly homologous to its human counterpart, the identity being 87%. The alignment shown in Fig.

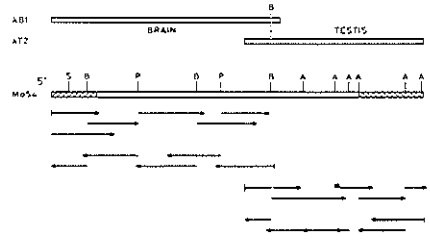


FIG. 1. **Partial restriction map and sequencing strategy of full-length Mo54 cDNA.** Overlapping clones, λB1 and λT2, were isolated from mouse brain and testis cDNA libraries, respectively. Their position with regard to the full-length Mo54 cDNA of 2.0 kb (*lower bar*) is indicated by the *top bars*. Restriction sites used for sequencing are shown. *Arrows* indicate extent and direction of sequencing reactions. The *black box* on one *arrow* represents a synthetic oligonucleotide used for sequencing. *Hatched bars*, 5′- and 3′-noncoding regions.

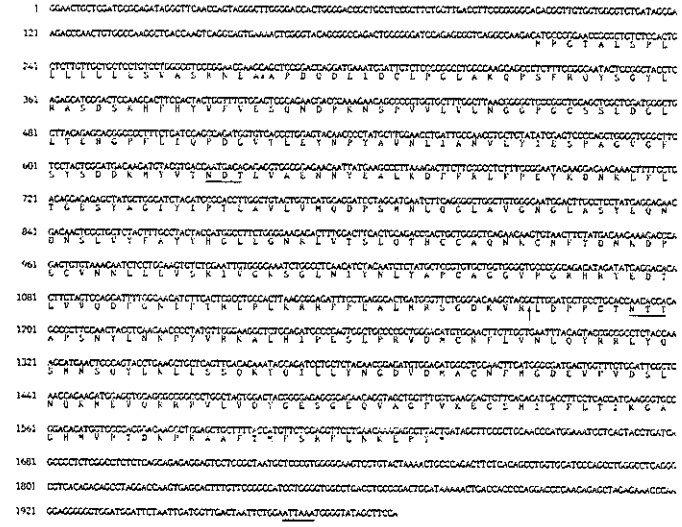


FIG. 2. **Nucleotide and predicted amino acid sequences of Mo54 cDNA.** The composite nucleotide sequence is derived from λB1 and λT2. Nucleotides are *numbered* on the *left*; the predicted amino acid sequence is *numbered* on the *right*. Residue 1 is the first amino acid (alanine) of precursor mouse protective protein. The proposed proteolytic cleavage sites are indicated by *vertical arrows*. *N*-Linked glycosylation sites and the putative polyadenylation signal are *underlined*. Only one of the two consecutive stop codons is marked by an *asterisk*.

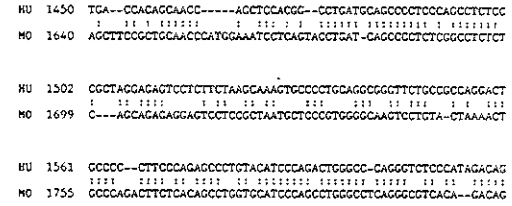


FIG. 3. **Comparison of mouse and human protective protein cDNAs at their 3′-untranslated regions.** Identical nucleotides are indicated by *double dots*. *Numbers* to the *left* refer to the position of the first nucleotide in that line, within mouse (*MO*) and human (*HU*) cDNA sequences.

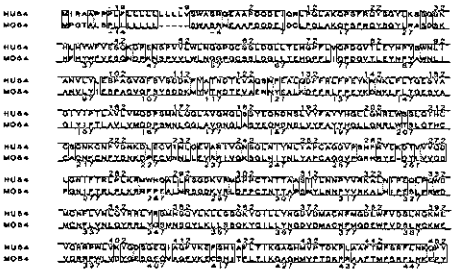


FIG. 4. **Alignment of predicted amino acid sequences of mouse and human preproforms.** Identical residues are *boxed*; *numbers* above and below the sequences refer to amino acid positions.

4 reveals that only two gaps need to be introduced in the Mo54 sequence in order to obtain maximal similarity. Therefore, it is likely that proteolytic processing in the mouse preproform occurs in a similar fashion as in its human homologue. Based on this assumption, the most probable site for signal peptidase cleavage is between alanines −1 and +1 (Fig. 2). Moreover, the amino acids surrounding these residues conform best to the consensus sequence for the peptidase cleavage site (30). Thus, mouse protective protein can be divided into three major domains: a signal sequence of 23 amino acids (−23 to −1 in Fig. 2), and a 32-kDa and a 20-kDa component of 297 and 154 residues, respectively. Their identity with the corresponding sequences in the human protein is 70% (signal sequence), 88% (32 kDa), and 91% (20 kDa).

Domains that are implicated in the protease activity of known serine carboxypeptidases isolated from yeast and plants are maintained in mouse and human protective proteins. These domains, spanning amino acids 53–62, 148–153, and 427–431 in the mouse protein (Figs. 2 and 4), include the essential residues Cys^{60}, Ser^{150}, and His^{428}. The corresponding serine residues in different yeast carboxypeptidases have been proven to be present in the active site (11, 15, 31).

Two potential glycosylation sites are present in the mouse protective protein (*underlined* in Fig. 2) which are located at positions identical to the human protein.

A notable difference between Mo54 and Hu54 predicted amino acid sequences is the presence of 2 extra cysteines in the mouse protective protein, which are found at the amino

and carboxyl termini of 32- and 20-kDa polypeptides, respectively.

Expression of Mo54 cDNA in COS-1 Cells and Uptake of cDNA-encoded Precursor by Human Cells—In order to demonstrate that the Mo54 sequence encodes a functional protective protein, the cDNA was cloned in two orientations (pCDMo54 sense and antisense) into a derivative of the mammalian expression vector pCD-X and transfected into COS-1 cells (5). For comparison, parallel transfections were performed using pCDHu54 constructs (5). Transient expression was detected 2 days after transfection by [^{35}S]methionine labeling of newly synthesized proteins in the presence of 10 mM NH_4Cl to induce maximal secretion of lysosomal protein precursors (25). Radiolabeled proteins from cell lysates and medium concentrates were immunoprecipitated using anti-human 32-kDa antibodies. As shown in Fig. 5, the Mo54 cDNA directs the synthesis in COS-1 cells of a protein of about 54 kDa which is partly secreted into the medium (Fig. 5, *lanes 1* and *5*). This polypeptide appears somewhat bigger than its human counterpart (Fig. 5, *lanes 3* and *7*), and it seems to be synthesized in less amount. However, when a small aliquot of total secreted proteins from transfected COS-1 cells was directly analyzed by SDS-polyacrylamide gel electrophoresis, it became clear that the quantity of mouse protective protein precursor present in the medium is comparable to that of its human counterpart (Fig. 5, *lanes 9–12*). Apparently, the anti-human 32-kDa antibodies precipitate the mouse protective protein to a much lesser extent. COS-1 cells transfected with antisense constructs do not give rise to major immunoprecipitable products (Fig. 5, *lanes 2, 4, 6,* and *8*).

Previous experiments have demonstrated that the Hu54 cDNA-encoded precursor, secreted by COS-1 cells, is endocytosed by human galactosialidosis fibroblasts, resulting in correction of β-galactosidase and neuraminidase activities in these cells (5). To compare the biological activity of mouse and human protective proteins, similar uptake studies were performed using early infantile galactosialidosis fibroblasts as recipient cells. Secreted proteins from COS-1-transfected cells were concentrated and added to the medium of confluent human fibroblasts. After 3 days of uptake, cells were harvested for either immunoprecipitation analysis or enzyme activity assays. As shown in Fig. 6, COS-1 cell-derived mouse protective protein precursor is endocytosed and processed intracellularly to a 32- and 20-kDa heterodimer (Fig. 6, *lane 1*). The mouse 32-kDa polypeptide is bigger than its human homologue, whereas the two 20-kDa forms are similar in size (Fig. 6, *lanes 1* and *3*). The anti-human complex antibodies used in this experiment also show a reduced affinity for the mouse protective protein components. Taken together, the data on sequence homology, transient expression, and uptake suggest that mouse protective protein is synthesized, proteolytically processed, and glycosylated in a way comparable to the human protein.

In fact, cDNA-encoded mouse protective protein precursor

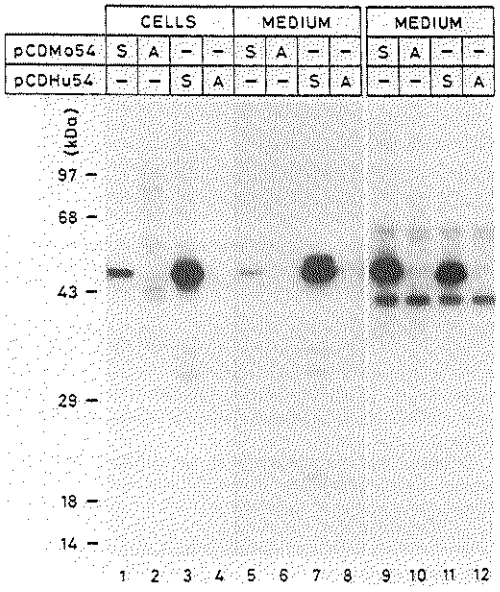


FIG. 5. **Transient expression of mouse and human cDNAs.** COS-1 cells were transfected with pCDMo54 and pCDHu54 sense (*S*) and antisense (*A*) constructs. After transfection, newly synthesized proteins were labeled with [^{35}S]methionine in the presence of 10 mM NH_4Cl. Labeled proteins from cells and media were immunoprecipitated using anti-human 32-kDa antibodies (*lanes 1–8*). Small aliquots of secreted proteins were kept for direct analysis (*lanes 9–12*). Molecular size markers are indicated at *left*. Exposure time for *lanes 1–4* was 16 h; for *lanes 5–8*, 48 h; and for *lanes 9–12*, 72 h.

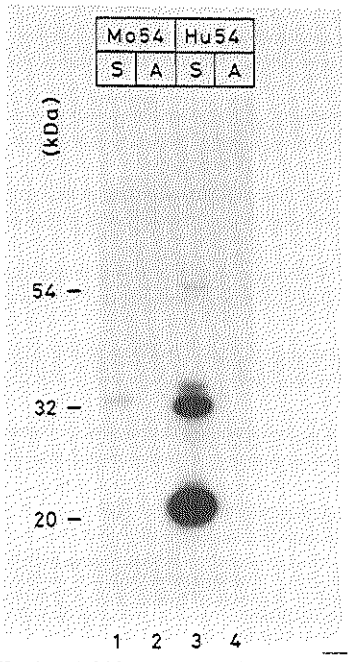


FIG. 6. **Uptake of COS-1 cell-derived mouse and human protective protein precursors by human cells.** Labeled proteins, secreted by transfected COS-1 cells, were added to the medium of fibroblasts from an early infantile galactosialidosis patient. After 3 days of uptake, cells were lysed and proteins immunoprecipitated using anti-human complex antibodies. Four separate uptake experiments were performed by adding concentrated medium from COS-1 cells transfected with pCDMo54 sense (*lane 1*), pCDMo54 antisense (*lane 2*), pCDHu54 sense (*lane 3*), and pCDHu54 antisense (*lane 4*). Molecular sizes were calculated by comparison with protein markers. Exposure time was 10 days.

can substitute for its human counterpart since it is able to correct β-galactosidase and neuraminidase activities after endocytosis by early infantile galactosialidosis cells (Table I). The increase in β-galactosidase activity is comparable irrespective of the fact that mouse or human protective proteins are used. On the contrary, correction of neuraminidase activity seems to depend upon which protein is added to the culture medium.

Protective Proteins Bind to the Serine Protease Inhibitor DFP—From our data it is clear that mouse protective protein preserves one of the biological functions of the human protein, namely the interaction with β-galactosidase and neuraminidase. Given the high degree of homology with certain known serine carboxypeptidases, we investigated whether both mouse and human protective proteins are a member of this family of enzymes. In general, serine proteases can be distinguished from other proteases by their sensitivity to the inhibitor DFP.

We made use of [^{3}H]DFP to detect directly the binding of this inhibitor to the two protective proteins. Experiments were performed using COS-1 cells transfected with pCDMo54 or pCDHu54 plasmids. As already seen in Fig. 5, protective protein precursors are synthesized intracellularly in large amounts, but they are very poorly or not at all proteolytically processed. Using slightly different experimental conditions, partial proteolytic cleavage of the precursor molecules to their respective mature forms can be observed (data not shown). Unlabeled cell extracts of COS-1 cells transfected with the aforementioned sense or antisense constructs were incubated with [^{3}H]DFP. Human fibroblasts from a normal individual and the early infantile galactosialidosis patient were treated in the same way. Radiolabeled proteins were immunoprecipitated afterward with anti-human complex antibodies, resolved by SDS-polyacrylamide gel electrophoresis, and visualized by fluorography.

As shown in Fig. 7, only the 32-kDa component of heterodimeric mouse and human protective proteins binds [^{3}H]DFP. After transfection with pCDHu54, a large amount of radiolabeled 32-kDa polypeptide is immunoprecipitated with anti-human complex antibodies (Fig. 7, *lane 3*). As observed before (Fig. 6), the mouse 32-kDa form is bigger in size and is precipitated less efficiently with these antibodies (Fig. 7, *lane 1*). The faint band visible throughout *lanes 1–4* represents endogenous COS-1 mature protective protein which is of the same size as the human protein. The [^{3}H]DFP labeling method is sensitive enough to detect human 32-kDa protective protein in normal fibroblasts (Fig. 7, *lane 5*), and it is highly specific since no labeled protein can be seen in the early

TABLE I

Correction of β-galactosidase and neuraminidase activities in galactosialidosis fibroblasts after uptake of COS-1 cell-derived human and mouse protective protein precursors

Transfection in COS-1 cells	Addition of COS-1 cell-derived proteins	Activity in galactosialidosis fibroblasts	
		β-Galactosidase	Neuraminidase
		milliunits[a]*/mg protein*	*microunits*[b]*/mg protein*
pCDMo54 (sense)	+	4.9	147
pCDHu54 (sense)	+	5.0	250
pCDMo54 (antisense)	+	1.1	0
pCDHu54 (antisense)	+	1.0	13
Mock	+	1.0	10
	−	0.9	8

[a] One milliunit of enzyme activity is defined as the activity that releases 1 nmol of 4-methylumbelliferone/min.

[b] One microunit of enzyme activity is defined as the activity that releases 1 pmol of 4-methylumbelliferone/min.

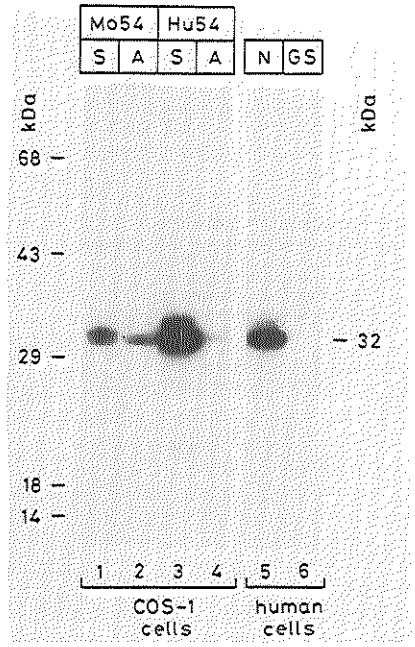


FIG. 7. **Labeling of protective proteins with [^{3}H]DFP.** Cell extracts of COS-1 cells transfected with pCDMo54 sense (*lane 1*) or antisense (*lane 2*), and with pCDHu54 sense (*lane 3*) or antisense (*lane 4*) as well as extracts of normal (*N, lane 5*) and galactosialidosis (*GS, lane 6*) fibroblasts were incubated with [^{3}H]DFP. Labeled proteins were immunoprecipitated with anti-human complex antibodies, resolved on a 12.5% SDS-polyacrylamide gel, and visualized by fluorography. Molecular size markers are indicated at *left*. Position of the 32-kDa components is indicated at *right*. Exposure times for *lanes 1–4* was 48 h and for *lanes 5 and 6*, 6 days.

infantile galactosialidosis fibroblasts (Fig. 7, *lane 6*). It is noteworthy that the respective precursor forms cannot bind the inhibitor. From these results we can infer that mouse and human protective proteins function also as serine proteases, and they are active after conversion to their heterodimeric state.

Tissue-specific Expression of Mouse Protective Protein mRNA—To determine the size and abundance of mouse protective protein mRNA, we analyzed total RNA isolated from several mouse tissues on a Northern blot. Samples were applied in equal amounts, and the blot was hybridized using the complete Mo54 cDNA as a probe. As shown in Fig. 8, all tissues investigated contain a protective protein transcript of about 2.0 kb. Thus, the composite Mo54 cDNA is probably full-length. In spleen, brain, heart, and ovarium, a transcript of slightly longer size is also detected (Fig. 8, *lanes 2, 4, 7,* and *11*). In two of these tissues, brain and heart, this mRNA is present in nearly equal amounts as the major 2.0-kb species. Faint bands of much bigger size, visible only in few of the samples, may represent precursor mRNA. A tissue-specific expression of the 2.0-kb protective protein mRNA is observed, with high amounts present in kidney and placenta and low amounts in testis (Fig. 8, *lanes 6, 10,* and *12*). The latter result explains the low yield of recombinant phages isolated from the testis cDNA library.

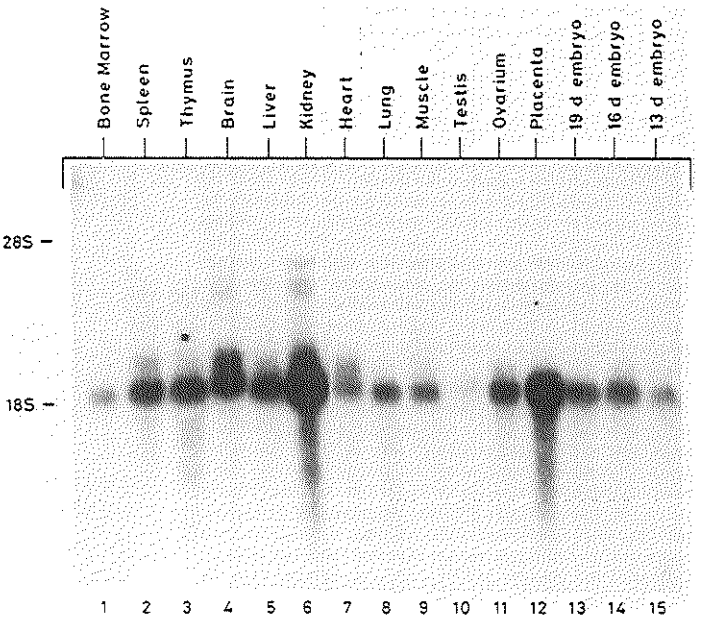


FIG. 8. **Expression of protective protein mRNA in different mouse tissues.** Total RNA was isolated from different mouse tissues and embryos at indicated days of gestation. Samples of 10–15 μg were fractionated on a formaldehyde-agarose gel, transferred to Zeta-Probe membranes, and probed with the composite Mo54 cDNA. Ribosomal RNA markers are indicated. Exposure time was 3 days.

DISCUSSION

The purpose of the work presented here was to characterize the mouse counterpart of human protective protein in order to gain insight in the primary structure of the protein and to select out conserved domains responsible for its function(s). We have isolated a full-length mouse cDNA encoding a biologically functional protective protein.

At the amino acid sequence level, mouse and human proteins are highly homologous. This stringent conservation agrees with a multifunctional role for the protective protein, namely stabilization of β-galactosidase, activation of neuraminidase, and a putative protease activity that might relate to the other two functions. The mouse protective protein contains 2 extra cysteines located at the amino terminus and carboxyl terminus of 32- and 20-kDa components, respectively. These residues might be responsible for the formation of another disulfide bridge that, in turn, could influence the folding of the protein. The remaining cysteines are in identical positions in mouse and human protective proteins. Since their total number is uneven, at least 1 free sulfhydryl group should be present in both protective proteins. The mouse 54-kDa precursor as well as the mature 32-kDa component are bigger in size than the corresponding human forms. The simplest explanation is that differences in amino acid composition cause retarded electrophoretic mobility. Nonetheless, we cannot exclude that signal sequence cleavage occurs amino-terminal to the predicted site, since amino-terminal sequencing of mouse precursor protein has not been performed.

We have shown that both mouse and human protective proteins can bind the inhibitor DFP. These results strongly suggest that these proteins are either serine proteases or esterases, although direct inhibition studies with DFP could not be performed since an enzymatic assay for the protective protein is not yet available. Radiolabeled DFP can bind only to the 32-kDa and not to the 20-kDa component of the mature heterodimer. This conforms with the location of the serine active site within the 32-kDa part of the polypeptide. Furthermore, the results indicate that the precursor molecule is not proteolytically active, hence behaving like a zymogen.

Based on their homology with well characterized proteases from lower organisms (10, 11, 15) and plants (12–14), the two protective proteins can be included in a newly delineated family of serine carboxypeptidases (15, 16, 32). For carboxypeptidase Y and the KEX1 gene product, the active site serine has been determined (11, 31). This residue resides in a stretch of 6 amino acids (Gly-Glu-Ser-Tyr-Ala-Gly) that are identical in all proteins of this family (7 in total) but differ from the amino acids surrounding the active site serine in serine endopeptidases (Gly-Asp-Ser-Gly-Gly-(Pro)) (33). It seems, therefore, that the Gly-Glu-Ser-Tyr-Ala-Gly domain is a prerequisite for serine carboxypeptidase activity. Moreover, it has been suggested that, like in endopeptidases, the essential serine residue in carboxypeptidases is activated by the concerted action of 2 other residues: a histidine and an aspartic acid (16). These 3 amino acids form the so-called "catalytic triad." Comparison of the 7 serine carboxypeptidases at the amino acid level shows that His^{428}, Asp^{369}, and Asp^{371} in the mouse protective protein are the only histidine and aspartic acid residues conserved throughout the family. The involvement of any of these amino acids in the formation of the catalytic triad is currently investigated by site-directed mutagenesis. Finally, it is remarkable that the protective proteins show structural similarity with plant serine carboxypeptidases, which are also heterodimeric proteins, with subunits of size similar to the 32- and 20-kDa components (16).

The uptake studies described here have shown that mouse protective protein can replace its human homologue in early infantile galactosialidosis cells as far as the correction of β-galactosidase and neuraminidase activities is concerned. Our results support the data of Mueller *et al.* (34), who constructed

mouse-human somatic cell hybrids to localize the genes encoding human neuraminidase and protective protein to chromosomes 10 and 20, respectively. These authors observed in hybrids retaining chromosome 10 but lacking chromosome 20 an increase in human neuraminidase activity over the mouse background. This is very likely due to the presence of endogenous mouse protective protein. It is important to notice that the increase of β-galactosidase activity is apparently not influenced by the type of protective protein added, indicating that mouse and human precursors are as efficiently endocytosed by the human cells. The correction of neuraminidase activity instead seems to be species dependent. This might reflect a differential interaction between the latter enzyme and either of the two protective proteins, due to changes in their tertiary structures. From these data we have a further indication that the determinants responsible for the interaction of protective protein with β-galactosidase and neuraminidase are clearly distinct (9).

A surprising finding in the nucleotide sequence comparison is the high degree of similarity throughout the 350-base pair 3′-untranslated regions of protective protein cDNAs. There are examples of mRNAs that contain sequences in their 3′-noncoding regions which regulate mRNA turnover (35–37). No obvious homology to the consensus sequence is found in the 3′-noncoding regions of the protective protein transcripts, yet their similarity suggests a possible involvement of these sequences in the folding and stability of the mRNA. Whether this feature relates to the observed tissue-specific expression of mouse protective protein mRNA remains to be investigated. In addition, we need to clarify the nature of a transcript slightly bigger than the major 2.0-kb mRNA present in considerable amounts in spleen, brain, heart, and ovarium. This species might arise from alternative splicing of the precursor mRNA. A number of genes encoding lysosomal proteins, including β-galactosidase, have recently been shown to generate diversity through such a mechanism (38–40). It will also be relevant to examine whether the expression of the protective protein transcript in different tissues correlates with the expression of β-galactosidase and neuraminidase mRNAs.

Acknowledgments—We are grateful to Professor Hans Galjaard for constant support and Dr. Gerard Grosveld for useful discussions and suggestions. We very much appreciate the help of Dr. Klaus Breddam (Dept. of Chemistry, Carlsberg Laboratory, Denmark), who provided us with detailed information on plant carboxypeptidases. We would like to thank Christine Troelstra and Marieke von Lindern for extracting RNA from mouse tissues and providing the Northern blot, Pim Visser and Mirko Kuit for excellent photographic work, Sjozef van Baal for his help with the computer data analysis, and Jeannette Lokker for typing the manuscript.

REFERENCES

1. 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
2. Wenger, D. A., Tarby, T. J., and Wharton, C. (1978) *Biochem. Biophys. Res. Commun.* **82**, 589–595
3. 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
4. Suzuki, Y., Sakuraba, H., Yamanaka, T., Ko, Y. M., Limori, Y., Okumaru, Y., and Hoogeveen, A. T. (1984) in *The Developing Brain and Its Disorders* (Arina, M., Suzuki, Y., and Yabuuchi, H., eds) pp. 161–175, University of Tokyo Press, Tokyo
5. 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
6. Hoogeveen, A. T., Verheijen, F. W., and Galjaard, H. (1983) *J. Biol. Chem.* **258**, 12143–12146
7. Verheijen, F. W., Brossmer, R., and Galjaard, H. (1982) *Biochem. Biophys. Res. Commun.* **108**, 868–875
8. Verheijen, F. W., Palmeri, S., Hoogeveen, A. T., and Galjaard, H. (1985) *Eur. J. Biochem.* **149**, 315–321
9. 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
10. Valls, L. A., Hunter, C. P., Rothman, J. H., and Stevens, T. H. (1987) *Cell* **48**, 887–897
11. Dmochowska, A., Dignard, D., Henning, D., Thomas, D. Y., and Bussey, H. (1987) *Cell* **50**, 573–584
12. Sørensen, S. B., Breddam, K., and Svendsen, I. (1986) *Carlsberg Res. Commun.* **51**, 475–485
13. Sørensen, S. B., Svendsen, I., and Breddam, K. (1987) *Carlsberg Res. Commun.* **52**, 285–295
14. Breddam, K., Sørensen, S. B., and Svendsen, I. (1987) *Carlsberg Res. Commun.* **52**, 297–311
15. Cooper, A., and Bussey, H. (1989) *Mol. Cell. Biol.* **9**, 2706–2714
16. Breddam, K. (1986) *Carlsberg Res. Commun.* **51**, 83–128
17. Gluzman, Y. (1981) *Cell* **23**, 175–182
18. Meijer, D., Hermans, A., von Lindern, M., van Agthoven, T., de Klein, A., Mackenbach, P., Grootegoed, A., Talarico, D., Della Valle, G., and Grosveld, G. (1987) *EMBO J.* **6**, 4041–4048
19. Feinberg, A. P., and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13
20. Sanger, F. G., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
21. Murphy, G., and Kavanagh, T. (1988) *Nucleic Acids Res.* **16**, 5198
22. Staden, R. (1986) *Nucleic Acids Res.* **14**, 217–223
23. Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395
24. Proia, R. L., d'Azzo, A., and Neufeld, E. F. (1984) *J. Biol. Chem.* **259**, 3350–3354
25. Hasilik, A., and Neufeld, E. F. (1980) *J. Biol. Chem.* **255**, 4937–4945
26. Auffray, C., and Rougeon, F. (1980) *Eur. J. Biochem.* **107**, 303–314
27. Fourney, R. M., Miyakoshi, J., Day, R. S., III, and Paterson, M. C. (1988) *Focus* **10**, 5–7
28. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
29. Kozak, M. (1987) *Nucleic Acids Res.* **20**, 8125–8148
30. Von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690
31. Hayashi, R., Moore, S., and Stein, W. H. (1973) *J. Biol. Chem.* **248**, 8366–8369
32. Hayashi, R., Moore, S., and Stein, W. H. (1973) *J. Biol. Chem.* **248**, 2296–2302
33. Brenner, S. (1988) *Nature* **334**, 528–530
34. Mueller, O. T., Henry, W. M., Haley, L. L., Byers, M. G., Eddy, R. L., and Shows, T. B. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 1817–1821
35. Shaw, G., and Kamen, R. (1986) *Cell* **46**, 659–667
36. Casey, J. L., Hentze,-M. W., Koeller, D. M., Caughman, S. W., Rouault, T. A., Klausner, R. D., and Harford, J. B. (1988) *Science* **240**, 924–928
37. Bernstein, P., and Ross, J. (1989) *Trends Biochem. Sci.* **14**, 373–377
38. 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
39. Quintern, L. E., Schuchman, E. H., Levran, O., Suchi, M., Ferlinz, K., Reinke, H., Sandhoff, K., and Desnick, R. J. (1989) *EMBO J.* **8**, 2469–2473
40. 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

Publication 4

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

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 266, No. 22, Issue of August 5, pp. 14754-14762, 1991
Printed in U.S.A.
© 1991 by The American Society for Biochemistry and Molecular Biology, Inc.

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^{60} 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^{3}NeuAc-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.[2] 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^{299} 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-

[2] 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[a]/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

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

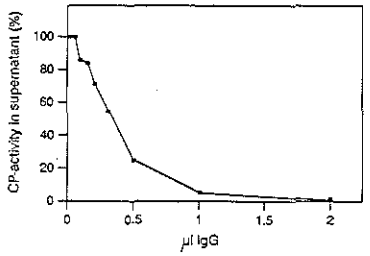


FIG. 1. **Immunotitration of cathepsin A-like activity.** Increasing amounts of an IgG antibody fraction, raised against human native protective protein precursor, were added to a cell extract of normal human fibroblasts. Antibody-antigen complexes were removed by addition of *S. aureus* cells and the 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

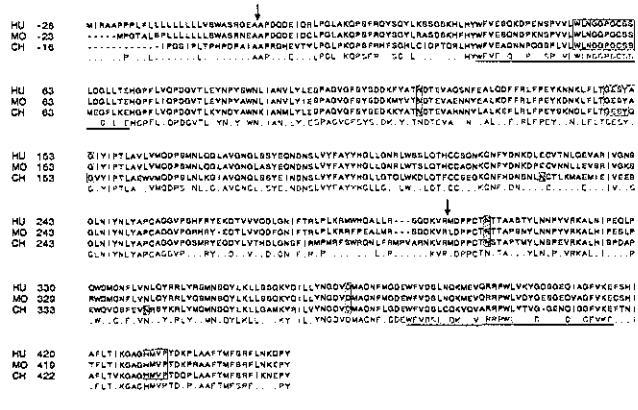


FIG. 2. **Alignment of predicted amino acid sequences of human, mouse, and chicken protective proteins.** Amino acid sequences of previously characterized human (*HU*) and mouse (*MO*) protective proteins are shown aligned with the chicken (*CH*) homologue. Residues on the *fourth line* (consensus sequence) 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.

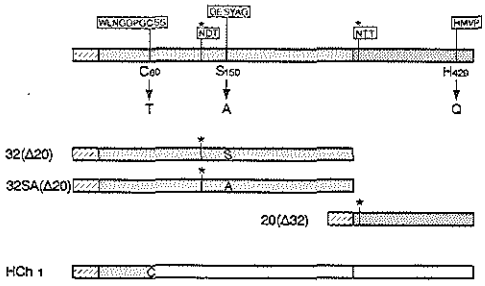


FIG. 3. ***In vitro* mutagenized and hybrid protective proteins.** The preprofrom of human protective protein is represented by the *upper bar*. *Hatched part*, signal sequence; *stippled part*, 32-kDa subunit; *cross-hatched part*, 20-kDa subunit. Single amino acid substitutions are indicated. The *three 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.

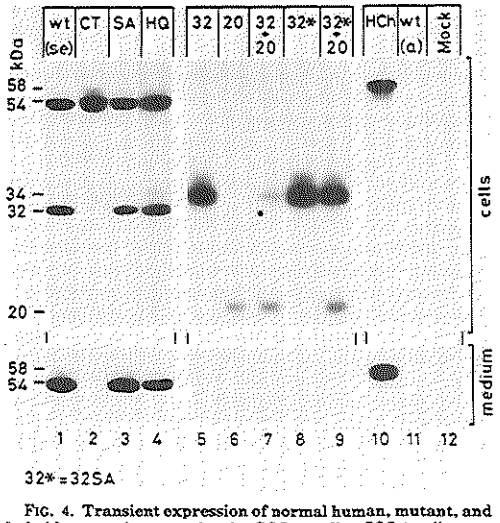


FIG. 4. Transient expression of normal human, mutant, and hybrid protective proteins in COS-1 cells. COS-1 cells were transfected with various pCD constructs: *wt(sc)* and *(a)* represent human protective protein cDNA in the sense and antisense orientation, respectively; *CT*, *SA*, and *HQ* are the 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^{150} to Ala^{150} 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^{429} mutant protein behaves similarly (not shown). In contrast, the CT^{60} 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

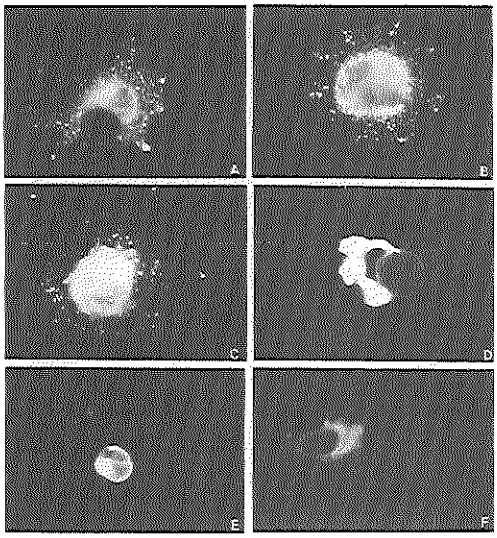


FIG. 5. Immunocytochemical localization of normal, mutant, and hybrid protective proteins in transfected COS-1 cells. COS-1 cells were treated with trypsin 48 h after transfection and reseeded on coverslips. 16 h later, cells were fixed and incubated with anti-32 antibodies. The intracellular distribution of normal human protective protein (*A*), HCh1 hybrid (*B*), and mutants SA^{150} (*C*), CT^{60} (*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^{160} 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^{160}, 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^{160}, 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

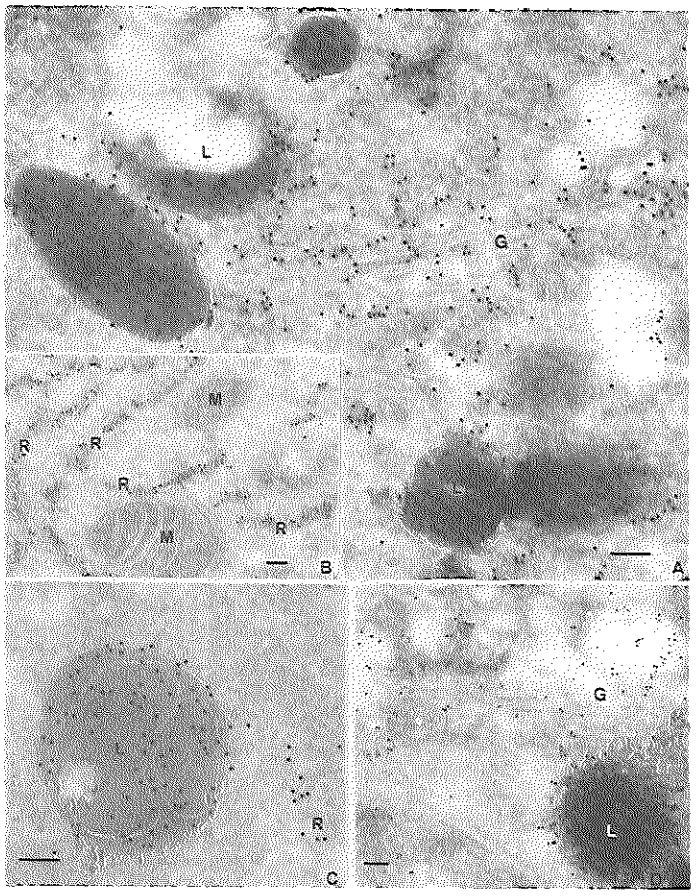


FIG. 6. **Cryosections of COS-1 cells, transfected with normal human (*A* and *B*), HCh1 hybrid (*C*) or SA^{150} mutant (*D*) protective proteins and labeled with anti-32 antibodies and goat anti-(rabbit IgG)-gold.** *A*, shows extensive labeling of the Golgi complex (*G*) and lysosomes (*L*), in cells 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.

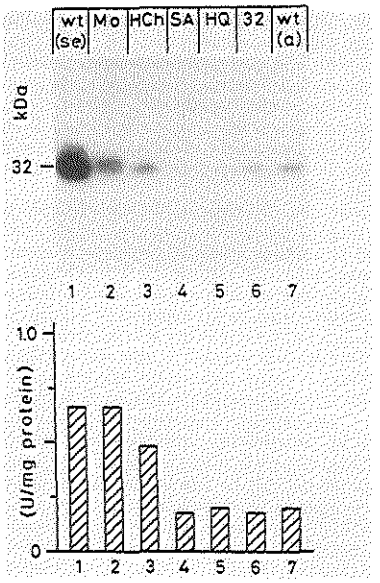


FIG. 7. **[^{3}H]DFP labeling and cathepsin A activity of normal, mutant, and hybrid protective proteins.** Cell extracts of COS-1 cells, transfected with selected constructs, were incubated with [^{3}H]DFP, followed by immunoprecipitation using anti-54 antibodies (*upper panel*). Molecular size is indicated 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[a]*/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

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

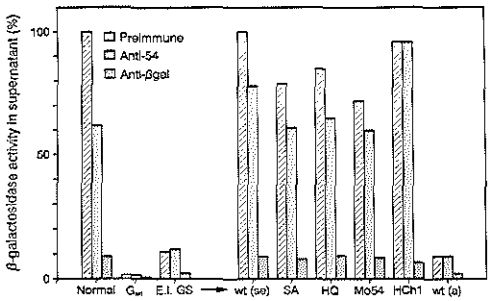


FIG. 8. **Precipitation of β-galactosidase activity in galactosialidosis fibroblasts after uptake of various COS-1 cell-derived protective protein precursors.** Early infantile galactosialidosis (*E.I.GS*) cell homogenates from the uptake experiment described in Table II were each divided in three aliquots 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, Sjozëf 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 Erdös, 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**, 265–295
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

(submitted for publication)

ANALYSIS OF THE GLYCOSYLATION, INTRACELLULAR TRANSPORT AND STRUCTURE OF HUMAN LYSOSOMAL PROTECTIVE PROTEIN.

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

From the MGC Department of Cell Biology and Genetics and the [¶]Department of Clinical Genetics, Erasmus University Rotterdam, The Netherlands

Summary

In lysosomes the acid hydrolases b-galactosidase and neuraminidase acquire a stable and active conformation through their association with a third glycoprotein, the protective protein. The latter is synthesized as a 54 kDa precursor that is processed in endosomes or lysosomes into a 32/20 kDa two-chain product. The protein has amino acid sequence homology to serine carboxypeptidases and is identical in the N-termini of the two chains to a deamidase/carboxypeptidase isolated from human platelets. It actually resembles closely a previously characterized lysosomal enzyme, cathepsin A. Site-directed mutagenesis experiments have demonstrated that its catalytic activity and protective function are distinct. In order to delineate domains and signals on the human protective protein that govern its intracellular transport, processing and quaternary structure we overexpressed normal as well as *in vitro* mutagenized protective proteins in COS-1 cells. The results indicate that all mutated precursors are withheld, either partially or completely, in the ER. Complete retention leads to the degradation of the mutant proteins. Of the two oligosaccharide chains present on human protective protein the one on the 32 kDa subunit acquires the mannose-6-phosphate (M6P) recognition marker, whereas the one on the 20 kDa subunit appears to be essential for the stability of the mature two-chain protein. Wildtype human protective protein precursors predominantly form homodimers of 85 kDa at neutral pH, indicating that dimerization could occur at the level of the ER.

Lysosomal protective protein has two thus far identified modes of action: a catalytic activity overlapping with the one of a previously isolated lysosomal enzyme

cathepsin A and a protective function towards lysosomal β-D-galactosidase (EC 3.2.1.23) and N-acetyl-α-neuraminidase (EC 3.2.1.18). It copurifies from different tissues of various species in a high molecular weight complex with the two glycosidases (Verheijen *et al,* 1982; 1985; Yamamoto *et al,* 1982; Yamamoto and Nishimura, 1987) and its presence in lysosomes is essential for their stabilisaton and activation (d'Azzo *et al,* 1982; Hoogeveen *et al,* 1983; van der Horst *et al,* 1989). Mutations that interfere with the protective function of the protein result in a severe combined β-galactosidase and neuraminidase deficiency (d'Azzo *et al,* 1982; Wenger *et al,* 1978), the hallmark of the rare metabolic storage disorder galactosiali dosis (Andria *et al,* 1981).

The primary structures of human, mouse and chicken protective proteins, as determined from their cloned cDNAs, are homologous to the yeast and plant serine carboxypeptidase family of enzymes (Breddam, 1986; Galjart *et al,* 1988; 1990; 1991). In addition, it has also been shown that human platelets, upon thrombin stimulation, release a deamidase/carboxypeptidase that is likely identical to the protective protein (Jackman *et al,* 1990) and may function in the local (in)activation of bioactive peptides. The carboxypeptidase activity, optimal at acid pH, has been compared with that of cathepsin A (Jackman *et al,* 1990; McDonald and Barrett, 1986). In accordance with these results we have given direct evidence that the protective protein maintains cathepsin A-like activity (Galjart *et al,* 1991). Moreover, it was shown that *in vitro* mutagenized protective proteins, deficient in their cathepsin A-like activity, retain completely their protective function, indicating that these two roles are distinct (Galjart *et al,* 1991).

In biosynthetic labeling studies the first immunoprecipitable form of human protective protein is a glycosylated 54 kDa precursor (d'Azzo *et al,* 1982), that is proteolytically converted within one hour after synthesis to a mature two-chain product of disulfide linked 32 and 20 kDa subunits (d'Azzo *et al,* 1982; Galjart *et al,* 1988). The two-chain form binds the serine protease inhibitor diisopropylfluorophosphate[1] demonstrating that maturation serves primarily to release the carboxypeptidase activity (Galjart *et al,* 1990). Whether the protective function is also inhibited in the 54 kDa zymogen remains to be investigated. After the initial endoproteolytic cleavage of the precursor, the 32 kDa subunit appears to undergo additional carboxyterminal processing of about 1-2 kDa (Galjart *et al,* 1988; 1991).

Complexes of the protective protein/β-galactosidase/neuraminidase 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;

Yamamoto and Nishimura, 1987). Immunochemical detection of the protective protein in one of these studies revealed that the majority of the two-chain form is not resolved in the complex (Hoogeveen *et al,* 1983). The quaternary structure of cathepsin A has also been probed by detection of its enzymatic activity in different aggregation states (Kawamura *et al,* 1974; 1975). The purified two-chain deamidase/carboxypeptidase is a homodimer of 94 kDa (Jackman *et al,* 1990), whereas, using different experimental conditions, its toad skin homologue can be resolved both as a homodimer and as a multimeric aggregate (Simmons and Walter, 1980). Together, structural and functional analyses indicate that pools of free and complexed protective protein/cathepsin A exist, that may have different functions in intra-or extracellular compartments (Galjart *et al,* 1991). Interestingly, the two-chain plant serine carboxypeptidases which have subunits of a size similar to the ones of the protective protein/cathepsin A, are also found as homodimers, but never in multimeric forms (Breddam,1986).

In this report we have analyzed the intracellular transport and processing in transfected COS-1 cells of normal and *in vitro* mutagenized protective proteins, independently of β-galactosidase or neuraminidase. The purpose of this work was to get insight in the domains and signals present in the protective protein that govern these aforementioned events.

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 wildtype cDNA segments using suitable restriction enzyme sites. In the final construct the DNA derived from PCR amplification was verified by double 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).

Transfections in COS-1 cells

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 out 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 described previously (Galjart *et al,* 1988; Proia *et al,* 1984). 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 . Immunoprecipitation methods using fixed *Staphylococcus aureus* cells (Immunoprecipitin, BRL) have been described (Proia *et al,* 1984), as have the antibodies that recognize all forms (54, 32, 20 kDa) of the protective protein and are designated anti-54 antibodies (Galjart *et al,* 1991). Normally we pretreat Immunoprecipitin as suggested by the supplier, a step that introduces reducing and denaturing agents into the cell extracts at an early stage. In the 16 hr pulse experiment, however, the Immunoprecipitin used was extensively washed with immunoprecipitation buffer after pretreatment, such that reducing agent was removed. 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.

Limited proteolysis with trypsin

In order to obtain larger quantities of secreted proteins COS-1 cells were transfected in 100 mm dishes. Cells were labeled with [^{3}H]leucine 48 hr after transfection in medium without fetal bovine serum. 16 hr later media were collected, centrifuged for 5 min at roomtemperature and 1000 rpm, to remove detached cells, after which bovine serum albumin (BSA, Boehringer Mannheim) was added to the supernatant 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 ^{0}C. Trypsin was inactivated by the addition of 3 µg bovine pancreas trypsin inhibitor (Sigma). In the 0 min time point the inhibitor was actually 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 using the N-blocked dipeptide benzyloxycarbonyl-phenylalanyl-alanine (Z-Phe-Ala) and a modified procedure (Galjart *et al,* 1991) that is based on 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.

Gelfiltration

COS-1 cell-derived protein precursors were obtained from the medium of [^{3}H]leucine labeled COS-1 cells, transfected in 100 mm dishes and labeled in the presence of 2.5% dialyzed fetal bovine serum. Media were concentrated and desalted (Proia *et al,* 1984) and 20% of this material was diluted fivefold in 50 mM 2-[N-morpholino]ethanesulfonic acid (MES) pH 6.95, containing 100 mM NaCl . This sample was applied to a column (85 × 1.53 cm, Pharmacia), containing Sephacryl S200 HR (Pharmacia), preequilibrated in the same buffer. Elution was carried out at a flow rate of 5.4 ml/hr and fractions of 0.9-0.95 ml were collected. All steps were performed at 4 ^{0}C. After gelfiltration fractions were prepared for immunoprecipitation by preclearing each fraction once with 100 µl pretreated Immunoprecipitin. Radiolabeled proteins were immunoprecipitated using anti-54 antibodies only, or a mixture of these and a monospecific antiserum against denatured human 64 kDa β-galactosidase (Morreau *et al,* 1989). Radiolabeled proteins were separated and visualized as described above. The column was calibrated with the following set of globular protein markers (Pharmacia): ferritin (440 kDa), aldolase (158 kDa), albumin (67 kDa), ovalbumin (43 kDa) and chymotrypsinogen (25 kDa).

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). The antibodies against the 32 kDa denatured chain of human protective protein (anti-32 antibodies) have been described (Galjart *et al,* 1988). This monospecific antiserum recognizes under reducing and denaturing conditions the 54 kDa precursor as well as the 32 kDa mature subunit of the protective protein.

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 intro-

duced *in vitro*. As summarized in Fig. 1 (lower 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.

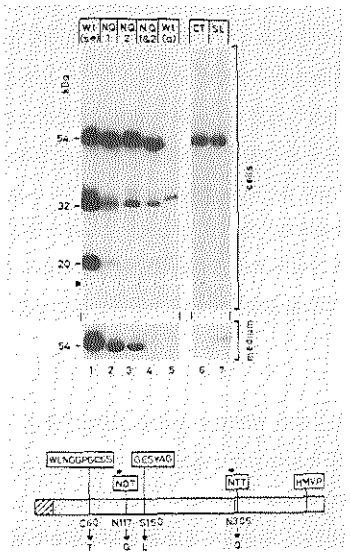


Fig. 1 *Transient expression of normal and in vitro mutagenized human protective proteins in COS-1 cells. The preproform of human protective protein is represented by the bar in the lower half of the figure, the hatched part being the signal sequence. The start of the 20 kDa subunit is indicated by a vertical line in the bar. The position and sequence of three domains, highly conserved in serine carboxypeptidases, is drawn above the bar, as are the two potential glycosylation sites found in the protective protein. Single amino acid substitutions are depicted, numbers refer to the position of the residues within the protective protein. The CT_{60} mutant has been described previously (Galjart et al, 1991). COS-1 cells were transfected with pCD-constructs containing the various cDNAs: wt (se) and (a) represent human protective protein cDNA cloned in the sense and antisense orientation respectively; NQ1 is NQ_{117}, NQ2 is NQ_{305}, NQ1&2 is the double glycosylation mutant $NQ_{(117\ \&\ 305)}$; CT and SL are the CT_{60} and SL_{150} mutants respectively. Newly synthesized proteins were labeled 2 days after transfection for an additional 16 hr with [^{3}H]leucine. Immunoprecipitations from cell extracts and media were carried out using anti-54 antibodies. Radiolabeled proteins were resolved by SDS-PAGE under reducing conditions and visualized by fluorography. Exposure times for cell samples, 4 days (lanes 1-5) or 2 days (lanes 6 and 7), for all media samples, 2 days. The molecular sizes of wildtype human precursor and mature polypeptides are shown at left. The arrowhead indicates the position of the 18 kDa subunit generated after proteolytic cleavage of the NQ_{305} mutant protein.*

Transient expression of normal and mutant protective proteins in COS-1 cells.
The various newly synthesized mutant cDNAs were cloned into the eukaryotic expression vector pCD-X, transfected into COS-1 cells and transiently expressed. Intracellular transport and processing of the mutant proteins were followed 48 hr after transfection and compared to wildtype human protective protein by labeling COS-1 cells with [^{3}H]leucine for an additional 16 hr. Radiolabeled proteins from cell lysates and media were immunoprecipitated with anti-54 antibodies. The immunoprecipitated material was resolved on a 12.5 % SDS-polyacrylamide gel, under reducing conditions (Fig. 1, upper part). Wildtype 54 kDa protective protein precursor is detected both intra- and extracellularly, whereas the mature subunits of 32 and 20 kDa are visualized in substantial amounts within the cells only (Fig. 1, lane 1). The two single glycosylation mutants, NQ_{117} and NQ_{305}, are synthesized as precursors of slightly reduced sizes, that are secreted but to a lesser extent than the wildtype protein (Fig. 1, lanes 1-3). Deletion of both oligosaccharide chains in the $NQ_{(117\ \&\ 305)}$ mutant precursor results in the synthesis of a polypeptide of even smaller size (Fig. 1, lanes 2-4), that is not secreted at all (Fig. 4, lane 4). Tunicamycin treatment of transfected COS-1 cells prior to and during labeling revealed that the protein moieties of wildtype, NQ_{117}, NQ_{305} and $NQ_{(117\ \&\ 305)}$ have an identical electrophoretic mobility (not shown). Of the three glycosylation mutants only NQ_{305} can be detected in a processed form, consisting of 32 and 18 kDa subunits (Fig. 1, lane 3). This indicates that endoproteolytic processing of this mutant precursor takes place at the correct site but, as expected, lack of the oligosaccharide chain gives rise to a small subunit of 18 kDa.

Both CT_{60}- and SL_{150}- precursors are not detected in a processed form and only the SL_{150} mutant can be immunoprecipitated from the culture medium, albeit in minute amounts (Fig. 1, lanes 6 and 7). Endogenous COS-1 protective protein is visible throughout the fluorograph (Fig. 1). Together the data demonstrate that each targeted amino acid substitution has a profound effect on the processing and secretion of the protective protein.

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 using anti-54 antibodies. 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),

demonstrating that it is the sugar moiety on the 32 kDa subunit that carries the M6P recognition marker. 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.

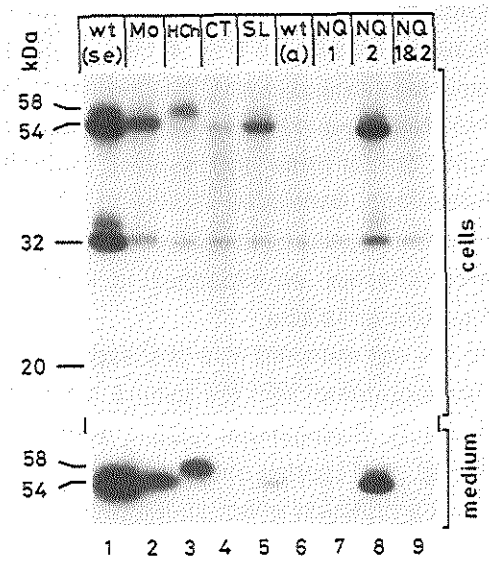


Fig. 2 *Phosphorylation of transiently expressed protective proteins.*
COS-1 cells were transfected with the pCD-constructs described in Fig.1 and with two other expression plasmids containing the cDNAs encoding mouse protective protein (Mo) and a human-chicken hybrid protective protein (HCh). Two days 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.

Turnover of normal and mutant protective proteins in transfected COS-1 cells.
To further investigate the fate of the normal and mutant protective proteins pulse-chase experiments were carried out. Transfected COS-1 cells were labeled with [^{3}H]leucine for 30 min, and chased for a maximal period of 6 hr. As shown in Fig. 3, processing of the 54 kDa wildtype precursor to its 32/20 kDa two-chain form is detected after 30-60 min chase (Fig. 3, wt(se), lanes 1-5). The majority of the

overexpressed precursor, however, is secreted into the culture medium. Secreted molecules are visible after 30 min chase. Accumulation of radioactive material in the medium reaches a maximum level at 3 hr and declines after 6 hr, indicating that internalisation of some of the labeled precursor has taken place.

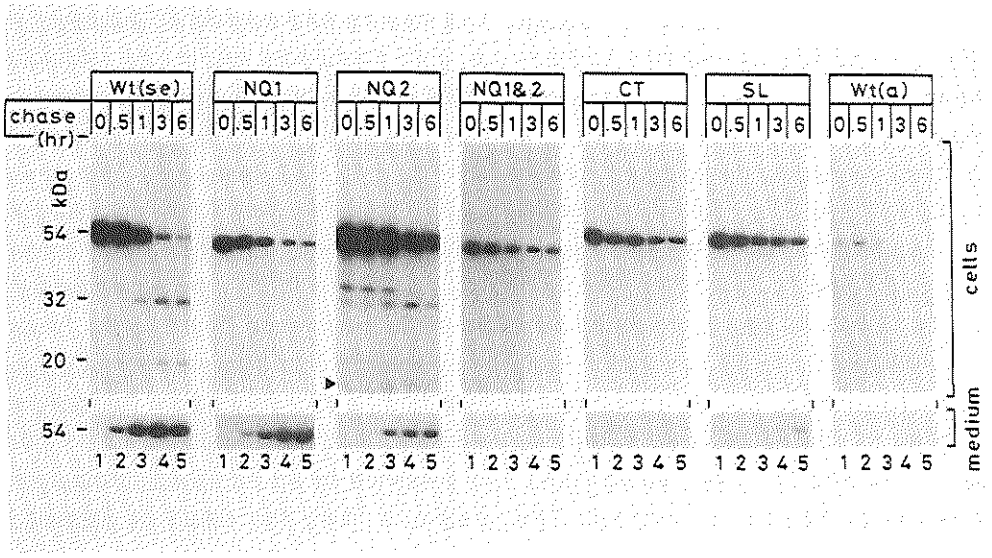


Fig. 3. *Pulse-chase analysis of wildtype and mutant protective proteins.*
COS-1 cells, transfected with the pCD-constructs explained in Fig. 1, were labeled for 30 min with [^{3}H]leucine and chased for the periods of time indicated above the lanes. Labeled proteins were immunoprecipitated from cell 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 kDa subunit of the NQ_{305} mutant protective protein. Exposure times of the fluorographs, cell samples: wt(se), 8 days; N_{Q1}, $N_{Q1\&2}$, CT, SL, 4 days; N_{Q2}, wt(a), 20 days; all media samples: 2 days.

Deletion of the oligosaccharide chain on the 32 kDa subunit results in the synthesis of a mutant precursor, which is not proteolytically processed, but is secreted (Fig. 3, NQ1, lanes 1-5). However, release of this precursor from transfected cells is more gradual compared to wildtype protein and no decline is visible after 6 hr chase, suggesting that once the NQ_{117} precursor arrives in the medium it is stably stored and not endocytosed. The NQ_{305} mutant protein instead behaves differently. Intracellularly, the processing to a 32/18 kDa two-chain form is

followed by degradation of this aberrant mature protein (Fig. 3, NQ2, lanes 1-5). Thus, either the deletion 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 the NQ_{305} precursor appears to be delayed in comparison to wildtype and NQ_{117} precursors and the amount of immunoprecipitable material is reduced. The kinetics and degree of secretion of NQ_{117}- and NQ_{305} mutants suggest that both precursors are in part withheld intracellularly. This effect is apparently additive since the double glycosylation mutant is not detected in the culture medium. Instead it is slowly degraded after synthesis without undergoing any processing (Fig. 3, NQ1&2, lanes 1-5). The same happens to 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 hr chase (Fig. 3, CT and SL, lanes 1-5). Processing of endogenous COS-1 protective protein can be visualized after a prolonged exposure of the fluorograph (Fig. 3, wt(a), lanes 1-5). Intracellularly, its maturation pattern is similar to that observed for wildtype protective protein, but no immunoprecipitable form is recovered from the culture medium. In the wt(se) and NQ_{305} fluorographs, that were exposed for longer periods (8 and 20 days, respectively), the presence of an additional proteolytic fragment of 35 kDa is detected. This protein is likely the result of aspecific proteolysis.

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

Amino terminal sequencing of the 20 kDa chain derived from purified placental protective protein revealed that its first amino acid is preceded in the precursor molecule by an arginine at position 298, a residue conserved in mouse and chicken protective proteins (Galjart *et al*, 1988; 1990; 1991). Therefore, we assessed whether the normal endoproteolytic processing of precursor protective protein could be mimicked by trypsin and yield a mature and active two-chain enzyme. Initial experiments demonstrated that limited proteolysis of COS-1 cell-derived 54 kDa precusor with trypsin gives rise to a two-chain molecule whose big subunit binds radiolabeled DFP (not shown). On SDS-PAGE this subunit is slightly larger than the normal 32 kDa chain, whereas the trypsin-generated 20 kDa subunit is identical in size to the wildtype (Zhou *et al*, 1991). We 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 precursors were aliquotted and

each sample incubated with a fixed amount of trypsin for increasing periods of time. Reactions were stopped by the addition of trypsin inhibitor. Afterwards cathepsin A activity was measured in each aliquot followed by immunoprecipitation with anti-54 antibodies. As shown in Fig. 4, 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. 4, lanes 1-6). Concomitant with the appearance of the two-chain product(s) 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=o small amounts of precursor protein are converted (Fig. 4, lane 2). The residual cathepsin A-like activity in the aliquot that was not treated with trypsin (Fig. 4, 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 trypsin digestion as wildtype precursor (Fig. 4, lanes 7-12). The two-chain product has a large subunit of about 30 kDa, but a normally sized 20 kDa subunit. The lack of an oligosaccharide moiety does not impair cathepsin A-like activity. However, deletion of the second oligosaccharide chain has a drastic effect on protease resistance (Fig.. 4, 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 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.

From their behaviour on SDS-PAGE it appears as if the three secreted precursors are all cleaved after the same residue(s). However, we have not determined the exact site of hydrolysis by trypsin, nor whether multiple cleavages can occur. Thus, even a small and undetected difference in proteolytic cleavage of the three precursor proteins could have a profound effect on catalytic activity of the corresponding two-chain forms. This would explain why a trypsin-cleaved two-chain NQ_{305} mutant protein is catalytically less active. In addition, differences in amount of secreted protein and/or resistance to trypsin will also account for the observed differences in enzymatic activities.

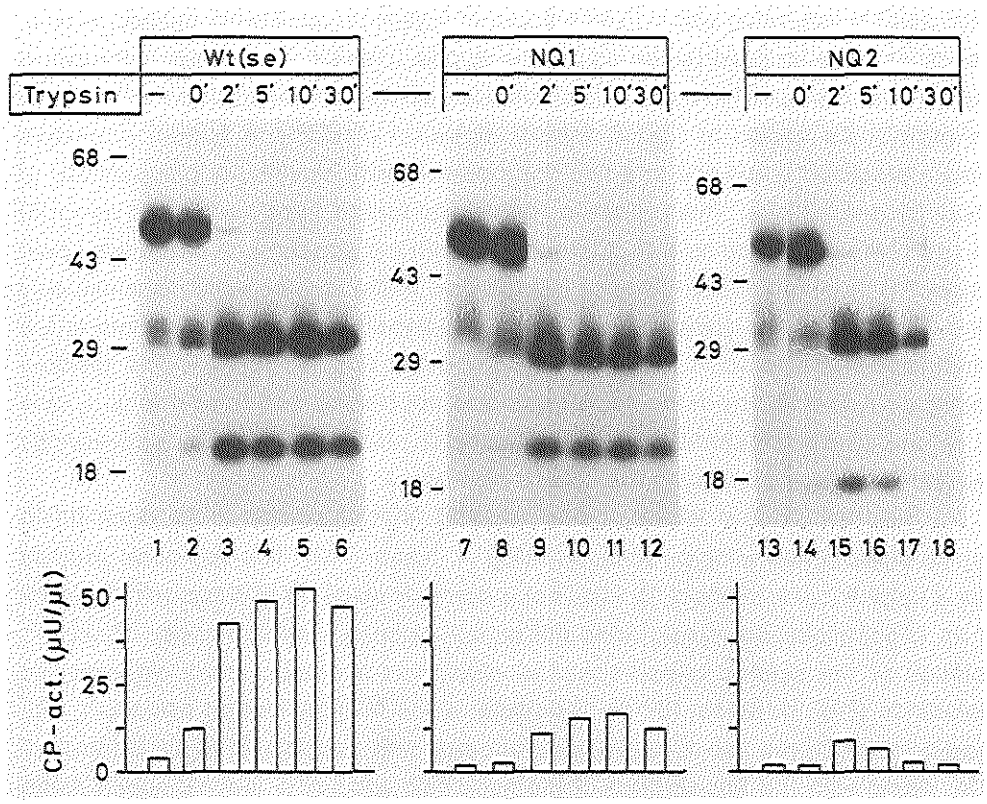


Fig. 4. *Limited proteolysis with trypsin of wildtype and mutant protective protein precursors.*

COS-1 cells were transfected with pCD-constructs encoding wildtype human, NQ1- or NQ2- mutant protective proteins. Two days later cells were labeled with [^{3}H]leucine, after which secreted proteins were collected by ammoniumsulphate precipitation,using 1 mg/ml BSA as carrier, and desalted. Aliquots of the concentrated preparations were incubated at 37 ^{0}C with 1.5 μg trypsin for the indicated periods of time. Reactions were stopped with 3 μg 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 was used for immunoprecipitation using anti-54 antibodies. Labeled proteins were further treated and visualized as described in Fig. 1. Molecular sizes of prestained protein standards are indicated to the left of each panel.

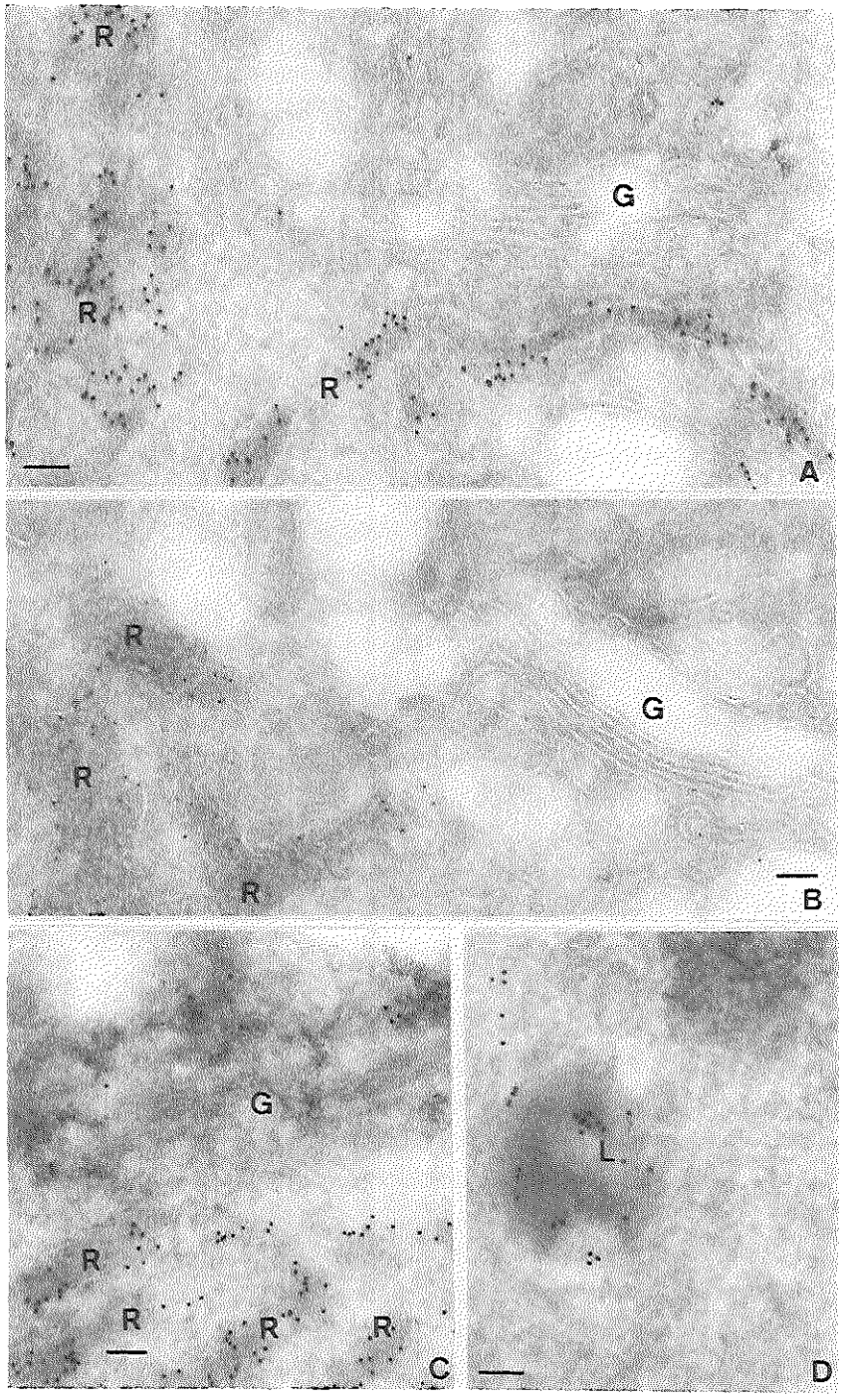

R
R
R
G
A
R
R
R
G
B
G
R
R
R
R
C
L
D

Fig. 5. *Subcellular 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 fo 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.

Localisation 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 compartimentalizes in structures corresponding to rough ER (RER), Golgi complex and lysosomes, whereas the CT_{60} mutant precursor was retained in the RER (Galjart *et al,* 1991). Fig. 5 shows that 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. 5 A and B, respectively). The same results were obtained with cells overexpressing the double glycosylation mutant (not shown). Thus, degradation of CT_{60}-, 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. 5C, D). This indicates that the NQ_{305} mutant protein reaches the lysosome, where it is subsequently processed and degraded.

Quaternary structure of human protective protein precursor

The native conformation of wildtype protective protein precursor was investigated by gelfiltration. A concentrated COS-1 cell-derived precusor preparation was diluted in buffer at neutral pH and applied to a Sephacryl S-200 HR column. To test the influence of β-galactosidase molecules on the conformation of human protective protein precursor, medium of COS-1 cells, cotransfected with human protective protein and β-galactosidase cDNAs, was also applied on the column. After elution fractions were immunoprecipitated with anti-54 antibodies and anti-human denatured β-galactosidase antibodies (Morreau *et al,* 1989). As shown in Fig. 6, under the conditions used, human protective protein precursor is

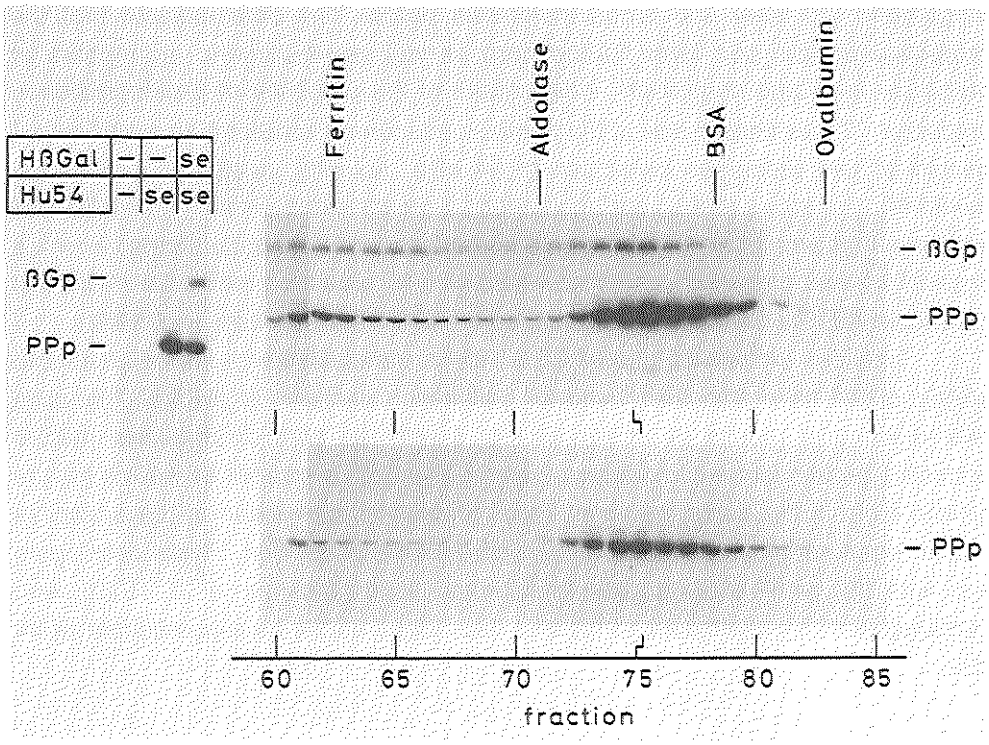


Fig. 6. *Gelfiltration of human protective protein precursor.*
COS-1 cells were either singly transfected with the pCD-construct encoding wildtype protective protein (Hu54) in the sense (se) orientation, or double transfected with this plasmid and a construct containing cDNA encoding human b-galactosidase (HbGal), also in the sense orientation. Labeling was as described in Fig. 1. Secreted proteins were concentrated and part of the preparation was applied on a Sephacryl S-200 HR column. After elution radiolabeled proteins in each fraction were immunoprecipated using anti-54 and anti-human b-galactosidase antibodies. About 10 % of the original concentrated preparation was taken along in the immunoprecipitation and is shown in the left part of the figure. Proteins were further treated as described in the previous figures. The elution position of the separately applied globular marker proteins is indicated above the fluorograph. PPp, protective protein precursor; bGp, b-galactosidase precursor. Exposure times, 3 days for left panel, 14 days for upper panel and 7 days for lower panel.

predominantly recovered from the column in a fraction corresponding to a molecular mass of 85 kDa. A small percentage of precursor molecules is detected in multimeric form (> 440 kDa). This distribution pattern is not influenced by the presence of β-galactosidase precursor. The latter is more equally divided over

monomeric (90 kDa) and broad multimeric peaks. The data indicate that the protective protein can form homodimers as well as multimeric aggregates at precursor level and neutral pH, independent of the presence of β-galactosidase precursor. The estimated size of the homodimeric form is smaller than would be expected of two precursor molecules of 54 kDa. Actually, peak fractions of homodimeric protective protein and monomeric β-galactosidase almost coincide. This indicates that protective protein dimers do not assume a globular shape, whereas β-galactosidase precursor monomers elute with a calculated molecular mass that correlates very well with the size estimated by SDS-PAGE (d'Azzo *et al,* 1982).

Total cell extracts of labeled COS-1 cells, transfected with normal protective protein cDNA have also been used for gelfiltration analysis, at pH 5.5. This gave rise to identical results as shown in Fig. 6 with the addition that also mature protective protein forms homodimers (not shown). Both precursor as well as mature protein elute in the same fractions. Thus, proteolytic conversion of the protective protein does not alter dramatically its hydrodynamic properties. These data agree well with the gelfiltration studies of Jackman et al (Jackman *et al,* 1990), who demonstrated that their purified platelet deamidase elutes as a homodimer of 94 kDa.

Discussion

We have analyzed the intracellular transport, processing and structure of protective proteins carrying targeted amino acid substitutions in order to get insight in the conformational characteristics of normal precursor and mature proteins, independently of β-galactosidase or neuraminidase. Experiments were carried out using the COS-1 cell system to transiently express normal and mutated proteins.

We have found that all aminoacid substitutions interfere with or impair completely the exit of the different mutant proteins from the ER. These results 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 processsing 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 β-hexosaminidase (Lau and Neufeld, 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 and Gabel, 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). It is interesting to note that an alanine substitution of the active site serine in the protective protein is well supported but replacement of the same amino acid with a more bulky leucine residue has drastic effects on the conformation of the precursor molecule.

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 β-hexosaminidase in transfected COS-1 cells (Lau and Neufeld, 1989), for human cathepsin D in baby-hamster kidney cells (Horst and Hasilik, 1991; Isidoro *et al,* 1991) and for mouse cathepsin L, which is the major excreted protein (MEP) in transformed mouse fibroblasts (Dong *et al,* 1989). Secretion of the α-subunit of β-hexosaminidase (Lau and Neufeld, 1989) and human cathepsin D (Horst and Hasilik, 1991; Isidoro *et al,* 1991) by the different cell types is not influenced by the addition of the lysosomotropic agent NH_4Cl, something we noticed as well in the case of overexpressed human protective protein (Galjart *et al,* 1988). It has been suggested that different protein sorting systems are present, that target lysosomal enzymes to their final destination with variable efficiency (Horst and Hasilik, 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. However, we were unable to demonstrate with immunoelectron microscopy the presence of NQ_{117} precursor protein in the Golgi complex in contrast to the NQ_{305} mutant. Perhaps transport of NQ_{117} molecules through this compartment is very rapid, because of the lack of a productive interaction with the phosphotransferase system and in turn with M6P-receptors. If the possible influence of an Asn to Gln amino acid substitution on the three dimensional structure of the protective protein is excluded, 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 acquisition of the M6P-recognition marker, it is however neither essential for the stability nor for the

catalytic activity of mature two-chain protective protein; 3) the chain on the 20 kDa subunit is crucial for the stability of the mature two-chain 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 could 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 (Winkler and Segal, 1984a; 1984b). 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 immediately 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.

The gelfiltration experiments suggest that dimerisation of wildtype human protective protein precursors takes place at neutral pH. Such an event could occur within the ER, although this should be demonstrated experimentally.The finding of a natural mutant protective protein that cannot form dimers and is partially retained in the ER (Zhou *et al*, 1991), is consistent with this hypothesis and would indicate that dimerization is a prerequisite for quick exit out of this compartment.

Multimerisation of protective protein precursors is a minor event under the conditions used here. Surprisingly, however, Zhou *et al* (1991) have found over 50 % of secreted wildtype human precursor molecules in multimeric form. The latter result was obtained by applying a more concentrated protein preparation on the column. These results together indicate that at neutral pH a protein concentration dependent equilibrium exists between dimeric and multimeric protective protein precursors. Since multimerisation takes place in the absence of labeled β-galactosidase precursors, it would seem that complex formation is a property of the protective protein precursor solely. Using purified recombinant human protective protein we are currently investigating in more detail the factors that influence its

multimerisation, whether it be in precursor state or in a mature two-chain conformation.

Footnotes

1) Abbreviations used are: DFP, diisopropylfluorophosphate; PCR, polymerase chain reaction; SDS, sodium dodecyl sulfate; M6P, mannose-6-phosphate; ER, endoplasmic reticulum, PAGE, polyacrylamide gel electrophoresis; Z, benzyloxycarbonyl.

Acknowledgements

We would like to thank Professor Hans Galjaard and the Rotterdam Foundation of Clinical Genetics for their continuous support and Dr. Gerard Grosveld for stimulating discussions. We are grateful to Pim Visser for the graphic work and Mirko Kuit for excellent photography.

References

Andria, G., Strisciuglio, P., Pontarelli, G., Sly, W. S. and Dodson, W. E. (1981). In *Sialidases and Sialidoses*. G. Tettamanti, P. Durand and S. DiDonato eds. (Milan: Edi. Ermes), pp 379-95.

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**, 4535-9.

Dong, J. M., Prence, E. M. and Sahagian, G. G. (1989). *J Biol Chem* **264**, 7377-83.

Frisch, A. and Neufeld, E. F. (1981). *J Biol Chem* **256**, 8242-6.

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**, 755-64.

Galjart, N. J., Gillemans, N., Meijer, D. and d'Azzo, A. (1990). *J Biol Chem* **265**, 4678-84.

Galjart, N. J., Morreau, H., Willemsen, R., Gillemans, N., Bonten, E. J. and d'Azzo, A. (1991). *J Biol Chem* **266**, 14754-62.

Gluzman, Y. (1981). *Cell* **23**, 175-82.

Hasilik, A. and Neufeld, E. F. (1980). *J Biol Chem* **255**, 4937-45.

Higuchi, R., Krummel, B. and Saiki, R. K. (1988). *Nucleic Acids Res* **16**, 7351-67.

Hoogeveen, A. T., Verheijen, F. W. and Galjaard, H. (1983). *J Biol Chem* **258**, 12143-6.

Horst, M. and Hasilik, A. (1991). *Biochemical Journal* **273**, 355-361.

Isidoro, C., Horst, M., Baccino, F. M. and Hasilik, A. (1991). *Biochemical Journal* **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**, 11265-72.

Kawamura, Y., Matoba, T., Hata, T. and Doi, E. (1974). *J Biochem* **76**, 915-24.

Kawamura, Y., Matoba, T., Hata, T. and Doi, E. (1975). *J Biochem* **77**, 729-37.

Lau, M. M. and Neufeld, E. F. (1989). *J Biol Chem* **264**, 21376-80.

Lazzarino, D. A. and Gabel, C. A. (1988). *J Biol Chem* **263**, 10118-26.

Lippincott-Schwartz, J., Bonifacino, J. S., Yuan, L. C. and Klausner, R. D. (1988). *Cell* **54**, 209-20.

McDonald, J. K. and Barrett, A. J. (1986). In *Mammalian Proteases: a Glossary and Bibliography.* (New York: Academic Press), pp 186-91.

Morreau, H., Galjart, N. J., Gillemans, N., Willemsen, R., van der Horst, G. T. J. and d'Azzo, A. (1989). *J Biol Chem* **264**, 20655-63.

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

Okayama, H. and Berg, P. (1982). *Mol Cell Biol* **2**, 161-70.

Palmeri, S., Hoogeveen, A. T., Verheijen, F. W. and Galjaard, H. (1986). *Am J Hum Genet* **38**, 137-48.

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**, 3350-4.

Roth, M. (1971). *Anal Chem* **43**, 880-2.

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**, 547-56.

Simmons, W. H. and Walter, R. (1980). *Biochemistry* **19**, 39-48.

Taylor, S. and Tappel, A. L. (1973). *Anal Biochem* **56**, 140-48.

van der Horst, G. T. J., Galjart, N. J., d'Azzo, A., Galjaard, H. and Verheijen, F. W. (1989). *J Biol Chem* **264**, 1317-22.

Verheijen, F., Brossmer, R. and Galjaard, H. (1982). *Biochem Biophys Res Commun* **108**, 868-75.

Verheijen, F. W., Palmeri, S., Hoogeveen, A. T. and Galjaard, H. (1985). *Eur J Biochem* **149**, 315-21.

Wenger, D. A., Tarby, T. J. and Wharton, C. (1978). *Biochem Biophys Res Commun* **82**, 589-95.

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. (1984a). *J Biol Chem* **259**, 1958-62.

Winkler, J. R. and Segal, H. L. (1984b). *J Biol Chem* **259**, 15369-72.

Yamamoto, Y., Fujie, M. and Nishimura, K. (1982). *J Biochem (Tokyo)* **92**, 13-21.

Yamamoto, Y. and Nishimura, K. (1987). *Int J Biochem* **19**, 435-42.

Zhou, X. Y., Galjart, N. J., Willemsen, R., Gillemans, N., Galjaard, H. and d'Azzo, A. (1991). *EMBO J* **(In press)**.

Publication 6

EMBO J. (1991), in press

A MUTATION IN A MILD FORM OF GALACTOSIALIDOSIS IMPAIRS DIMERIZATION OF THE PROTECTIVE PROTEIN AND RENDERS IT UNSTABLE

Xiao Yan Zhou, Niels J. Galjart, Rob Willemsen, Nynke Gillemans, Hans Galjaard and Alessandra d'Azzo

Dept. of Cell Biology and Genetics, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Summary

The lysosomal disorder galactosialidosis is caused by deficiency of the protective protein in the absence of which the activities of the enzymes β-galactosidase and neuraminidase are reduced. Aside from its protective function towards the two glycosidases this protein has cathepsin A-like activity. A point mutation in the protective protein gene, resulting in the substitution of Phe^{412} with Val in the gene product, was identified in two unrelated patients with the late infantile form of the disease. Expression in COS-1 cells of a protective protein cDNA with the base substitution resulted in the synthesis of a mutant protein that lacks cathepsin A-like activity. The newly made mutant precursor was shown to be partially retained in the endoplasmic reticulum. Only a fraction is transported to the lysosomes where it is degraded soon after proteolytic processing into the mature two-chain form. Since the mutant precursor, contrary to the wild type protein, does not form homodimers, the dimerization process might be a condition for the proper targeting and stable conformation of the protective protein. These results clarify the mechanism underlying the combined deficiency in these patients, and give new insight into the structure/function relationship of the wild type protein.

Galactosialidosis is an inherited metabolic storage disorder transmitted as an autosomal recessive trait (O'Brien, 1989). It is caused by mutation of the gene encoding the lysosomal protective protein (d'Azzo *et al.*, 1982) which normally associates with the enzymes β-D-galactosidase (E.C. 3.2.1.23) and N-acetyl-α-neuraminidase (sialidase, E.C. 3.2.1.18) and regulates their intralysosomal activity and stability (Verheijen *et al.*, 1982; Hoogeveen *et al.*, 1983; van der Horst *et al.*, 1989). Deficient or non functional protective protein causes impaired activities of the two glycosidases and consequent storage of predominantly sialylated oligosaccharides in tissues and urine (Wenger *et al.*, 1978; Okada *et al.*, 1977; van Pelt *et al.*, 1988a; 1988b;1989).

Although the disorder is rare, several dozens of patients of different ethnic origin have been described (Pinsky *et al.* 1974; Andria *et al.*, 1981; Suzuki *et al.* 1984; Loonen *et al.*, 1984; Sewell *et al.*, 1987; Chitayat *et al.*, 1988). They are clinically heterogeneous having either a very severe early onset form of the disease, mostly fatal at birth, or mild and slowly progressive late onset types. Patients with an early infantile phenotype have severe CNS involvement, macular cherry red spots, visceromegaly, skeletal abnormalities, renal dysfunction and coarse facies. Juvenile/adult forms, mainly of Japanese origin, are characterized by features like skeletal dysplasia, dysmorphism, macular cherry red spots, slowly progressive mental and motor deterioration and angiokeratoma. So far a defined, small group of patients represents the late infantile phenotype. These patients are all alive and have developed symptoms at 12-24 months of age with the main features being visceromegaly, dysostosis multiplex, heart involvement, but no mental retardation (Pinksy *et al.*, 1974; Andria *et al.*, 1978; 1981; Chitayat *et al.*, 1988; Strisciuglio *et al.*, 1990). This broad spectrum of clinical manifestations in distinct galactosialidosis phenotypes can only in part be correlated with differences in the level of expression of protective protein mRNA (Galjart *et al.*, 1988) and the amount and/or property of immunoprecipitated polypeptide (d'Azzo *et al.*, 1982; Palmeri *et al.*, 1986).

The human protective protein is synthesized as a 54 kDa precursor that is cleaved and modified in lysosomes into a mature 32/20 kDa two-chain product (d'Azzo *et al.*, 1982; Galjart *et al.*, 1988). Traces of 32 kDa polypeptide could be immunoprecipitated from fibroblasts of two late infantile galactosialidosis patients, but only when these cells were treated with the protease inhibitor leupeptin (Palmeri *et al.*, 1986; Strisciuglio *et al.*, 1988). These findings led to the supposition that the presence of a minimal amount of mature protective protein in these patients could account for their milder clinical course (Strisciuglio *et al.*, 1988).

The primary structures of human, mouse and chicken protective proteins have been found to be homologous to yeast and plant serine carboxypeptidases (Galjart *et al.*, 1988, 1990, 1991). This observation led to the discovery that the protective protein has cathepsin A-like activity (Galjart *et al.*, 1991). Clinically distinct galactosialidosis patients, reported to have reduced activity of a lysosomal carboxypeptidase called carb L (Tranchemontagne *et al.*, 1990; Kase *et al.*, 1990), are thus deficient in the cathepsin A-like enzyme (Galjart *et al.*, 1991). Although the physiological implications of these findings are as yet unclear, this lysosomal disorder should be viewed as the first one that is also associated with a protease deficiency. Jackman *et al.* (1990) have reported on a deamidase/carboxypeptidase enzyme purified from human platelets that enzymatically resembles cathepsin A and has sequence identity to the N-termini of the protective protein chains. Using protective protein mutants with targeted amino acid substitutions we have established that the protective and catalytic activities of the protein are distinct (Galjart *et al.*, 1991), and the latter can only be exerted after conversion of the inactive precursor (zymogen) into the mature and stable two-chain form (Galjart *et al.*, 1990).

We have identified different mutant alleles associated with the three clinical galactosialidosis phenotypes of mainly European patients. In the present study we focus on the mutation underlying two unrelated patients with the late infantile form of the disease. They both carry the same point mutation in the protective protein gene resulting in the substitution of Phe^{412} with Val in the gene product. The mutation was shown to impair the formation of protective protein precursor homodimers which in turn might be responsible for the observed partial retention of the mutant precursor polypeptide in the endoplasmic reticulum (ER). The fraction that does reach the lysosomes is catalytically inactive and undergoes rapid intralysosomal degradation.

Results

Point mutation identified in the protective protein/cathepsin A cDNA of two galactosialidosis patients.

The first patient analyzed is the child of healthy unrelated Italian parents. His disorder was diagnosed at the age of two and his clinical and biochemical characteristics have been described earlier (Andria *et al.*, 1978; 1981; Palmeri *et al.*, 1986). In total RNA preparations from the patient's fibroblasts a normal amount of 2 kb protective protein transcript was detected (Galjart *et al.*, 1988). In cultured fibroblasts

this mRNA was translated into a 54 kDa mutated precursor that was only poorly, if at all, processed to the mature form (Palmeri *et al.*, 1986). Consequently, no cathepsin A-like activity was measured in cell lysates from the patient (Galjart *et al.*, 1991).

To identify the genetic lesion underlying this phenotype, we first analyzed the cDNA derived from the patient's mRNA. Four overlapping fragments, encompassing the entire coding region, were reverse transcribed, amplified by the polymerase chain reaction (PCR), cloned and sequenced on both strands. The only difference between normal and mutant nucleotide sequences was a single base substitution, T to G, at position 1324 of the normal cDNA (Galjart *et al.*, 1988), changing Phe^{412} (TTC) to Val (GTC). Ten subclones of the cDNA portion including the base substitution showed the same mutation. The results were somewhat surprising since the parents of this patient were reported to be non consanguineous. However, direct sequencing of a 239 bp genomic region surrounding the mutation, amplified by PCR from total DNA of the patient and his parents, confirmed the cDNA data (Figure 1, upper panel). The patient is clearly homozygous for the T to G transversion, whereas both carriers are heterozygous for this mutation.

Since the T to G substitution generates a new *Acy*I restriction site in the mutant DNA, this point mutation can be easily detected by digesting the amplified 239 bp genomic fragment with *Acy*I. As shown in Figure 1 (lower panel), the patient's DNA releases upon digestion two fragments of 159 bp and 80 bp (lane 4), instead of the uncut fragment present in normal DNA (lane 1). Amplified samples from both parents display heterozygous patterns (lanes 2 and 3). We used this assay to screen for the T to G transversion DNA samples of seven clinically different galactosialidosis patients. In only one other unrelated Canadian late infantile patient (Pinsky *et al.*, 1974) the same mutation was found (not shown). It was reported earlier that in cultured fibroblasts from this patient a small amount of protective protein precursor was synthesized that again failed to mature (Palmeri *et al.*, 1986). As deduced from the *Acy*I digestion pattern, this patient carries the T to G substitution on one allele, whereas the other has a normal sequence in that region. Direct sequencing of the corresponding genomic DNA fragment validated the compound heterozygosity. Only the product of the T to G mutant allele was seen in direct sequences of the patient's cDNA (not shown), suggesting that a second mutation in the other allele either impairs its transcription or gives rise to an unstable mRNA. Our findings indicating compound heterozygosity explain why a reduced amount of protective protein transcript is detected in total RNA preparations from the patient's fibroblasts.

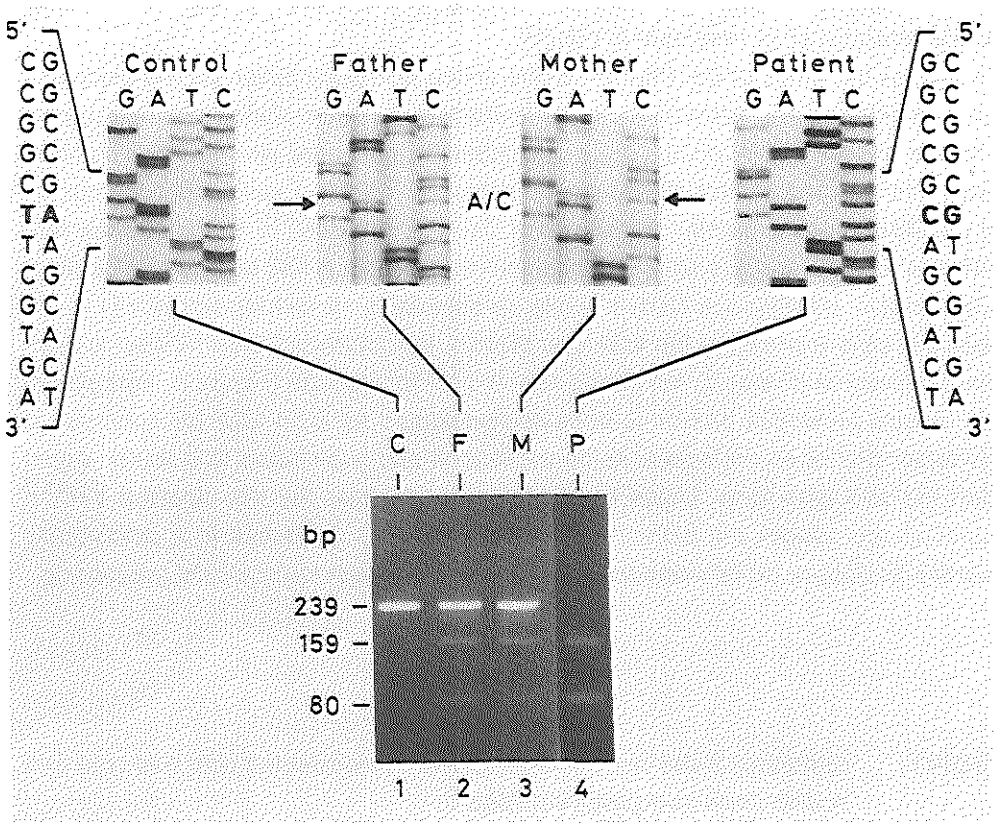


Figure 1. Upper.

Partial nucleotide sequence of the protective protein/cathepsin A gene from the Italian galactosialidosis patient and his parents. *Genomic DNA was isolated from fibroblasts of a normal individual, the patient and his parents and subjected to asymmetric PCR in the region containing the mutation. The products were directly sequenced. A portion of the autoradiograph from the sequencing gels is shown. The T to G transversion is indicated in bold. Arrows point to the double band at one position that is present in the nucleotide sequences of both parents. 5' and 3' refer to the orientation of the protective protein cDNA (Galjart et al, 1988).*

Lower.

Acyl restriction enzyme assay for the T to G mutation. *Genomic fragments, amplified from total DNA of the patient and his parents, were incubated with AcyI at 37°C for 2 hr. Restriction fragments were analyzed by electrophoresis on a 4% NuSieve agarose gel and stained with ethidium bromide. Molecular sizes are indicated at left.*

Val^{412} mutation impairs the catheptic activity of the mutant protein and causes partial retention of its precursor in the ER.

To assess the effect of the Phe^{412} to Val amino acid substitution on the biochemical behaviour of the protective protein/cathepsin A, the mutation was introduced into the normal cDNA (Hu54) by oligonucleotide mediated site-directed mutagenesis. Normal and mutant cDNAs were expressed in COS-1 cells under the control of the SV40 early promoter. As shown in Table 1, the Phe^{412} to Val change renders the enzyme inactive since no increase in cathepsin A-like activity can be measured above endogenous COS-1 levels; in contrast a 10-fold higher activity is detected in cells expressing the wild type protein. In transfected cells labeled with either [^{3}H]leucine (Figure 2, left panel) or with [^{32}P]phosphate (Figure 2, right panel) the mutant protective protein is immunoprecipitated as a 54 kDa phosphorylated precursor that does not appear to be processed and is secreted to a lesser extent than the wild type form. These results demonstrate that the site directed Val^{412} mutation reproduces the patient's phenotype in that it gives rise to a catalytically inactive protective protein/cathepsin A with impaired post-translational processing.

Table1. *Cathepsin A-like activity in COS-1 cells transfected with wild type and mutant protective protein cDNAs.*

Transfected DNA	Activity (Units/mg protein)
Hu54 (sense)	367
Hu54 Phe^{412} to Val	31
Hu54 (anti)	34

COS-1 cells were transfected with pCDHu54 sense (se), pCDHu54 Phe^{412} to Val or pCDHu54 antisense (a). After 72 hr cells were harvested by trypsinization and cathepsin A-like activity was measured in cell extracts using N-carbobenzoxy-L-phenylalanyl-L-alanine (Z-Phe-Ala) as a substrate. One unit of activity is defined as the activity that releases 1 nmol of alanine per minute.

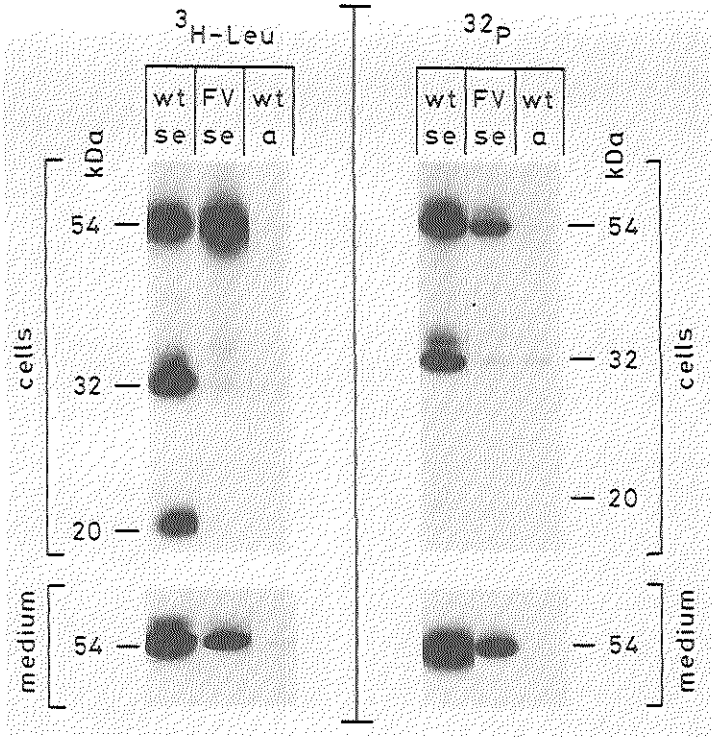
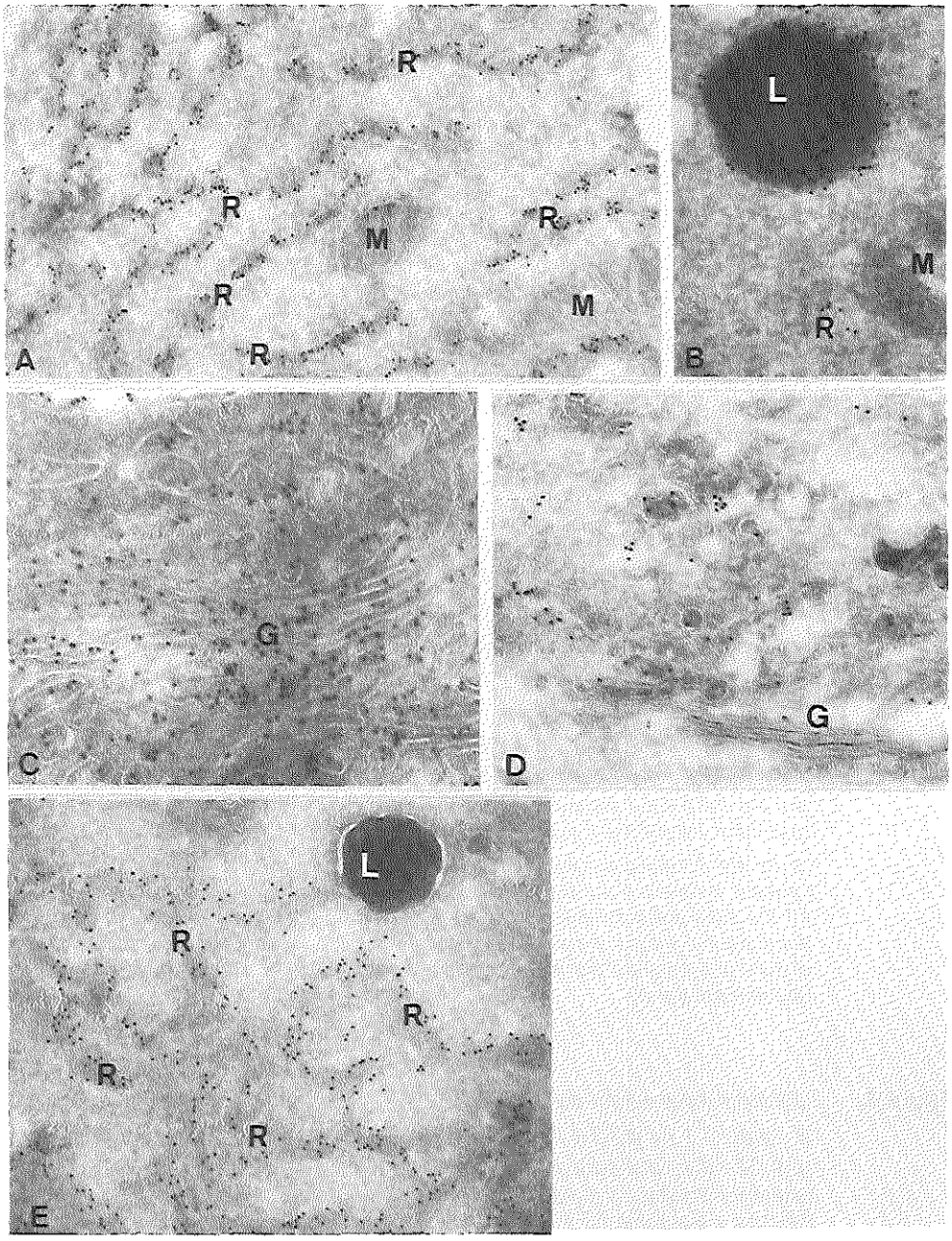


Figure 2. *Immunoprecipitation of normal and mutant protective proteins in transfected COS-1 cells.* *COS-1 cells were transfected with pCDHu54 sense (wt, se), pCDHu54 antisense (wt, a) and pCDHu54 Phe^{412} to Val (FV, se). Two days after transfection newly synthesized proteins were labeled for an additional 12 hr with [^{3}H]leucine (left panel) or for 7 hr with [^{32}P]phosphate (right panel). Labeled proteins from cells and media were immunoprecipitated, resolved on 12.5% SDS-polyacrylamide gels under reducing conditions, and visualized by fluorography (left) or by autoradiography (right). Molecular sizes of precusor and mature polypeptides are indicated. Exposure times were 5 days for left panel, 1 day for right panel.*

Figure 3. *Ultrathin cryo-sections of COS-1 cells, transfected with pCDHu54 (A, B and C) or pCDHu54 Phe^{412} to Val (D and E) and labeled with anti-32 kDa antibodies and goat anti-rabbit IgG-gold.* *In A, B, and C extensive labeling of rough endoplasmic reticulum structures (R), Golgi complex (G) and a lysosome (L) is shown, but not of a mitochondrion (M). In contrast, a weakly labeled Golgi complex (G) is seen in D. E shows strong labeling of rough endoplasmic reticulum (R) and a few gold particles in a lysosome (L). Magnifications were 46.000 x for A, 58.000 x for B, 93.000 x for C, 64.000 x for D and 65.000 x for E.*


R
R
M
R
R
M
A
L
M
R
B
G
C
G
D
L
R
R
R
R
E

Using immunoelectronmicroscopy, we have studied the intracellular distribution of Val^{412} protein overexpressed in COS-1 cells. As shown in Figure 3, the amount of gold particles present in the ER of cells expressing wild type and mutant protective proteins is similar (Figure 3, A and E). However, a drastically reduced labeling of the Golgi complex is found in cells expressing the Val^{412} mutant compared with those expressing the normal protein (Figure 3, C and D). Labeling of lysosomes is detected in cells transfected with either of the two cDNA constructs, but again the overal number of grains is distinctly diminished (less than 10%) in lysosomes of cells transfected with the mutant cDNA (Figure 3, B and E and Table 2). It therefore appears that the Phe^{412} to Val substitution causes retention of the precursor polypeptide in an early biosynthetic compartment. However, this retention can only be partial since some molecules of the total mutant precursor pool are routed to the Golgi complex, undergo phosphorylation as well as secretion and are correctly compartmentalized in lysosomes.

Table 2. *Quantitative data on the immunolabeling of lysosomes in different COS-1 transfected cells.*

Transfected DNA	Lysosomes counted	Gold particles*
Hu54 (sense)	30	125 (13)
Hu54 Phe^{412} to Val	30	7 (1)
Hu54 (anti)	55	1 (.3)

* The values represent the average number of gold particles per lysosome. The standard error of the mean (SEM) is indicated between brackets.

Mutant precursor does not dimerize at neutral pH.

It was previously reported that mature lysosomal cathepsin A, purified from different species, elutes on gel filtration as a dimer of about 95-100 kDa (Simmons and Walter, 1980; Jackman *et al.*, 1990). We have recently shown that the protective protein can also form homodimers at the precursor level and neutral pH (N.J. Galjart and A. d'Azzo, submitted), suggesting this to be an endoplasmic reticulum/early Golgi event. We have now tested the effect of the Phe^{412} to Val change on the elu-

tion pattern of the mutant precursor after gel filtration. COS-1 cells were transfected with either wild type or mutant protective protein cDNA constructs, and labeled with ^{3}H-leucine. Medium samples containing the secreted precursor forms, were concentrated, desalted, and applied on a Sephacryl S-200 column. Protective protein precursor was immunoprecipitated from each fraction using anti-54 kDa antibodies. As shown in Figure 4, about 50% of wild type precursor is resolved as a dimer of ~85 kDa, whereas the remaining ~50% forms aggregates of high molecular weight (upper panel). Multimerization seems to be concentration dependent (N.J. Galjart and A. d'Azzo, submitted). In contrast the majority of the mutant polypeptide is immunoprecipitated as a monomer of ~37 kDa and only a small fraction elutes as a multimer (Figure 4, lower panel). Both normal and mutant native precursors appear to be non globular proteins since they elute later than expected for polypeptides of 54 kDa. These findings suggest a correlation between the absence of dimerization of the mutant proform and its retention in an early biosynthetic compartment.

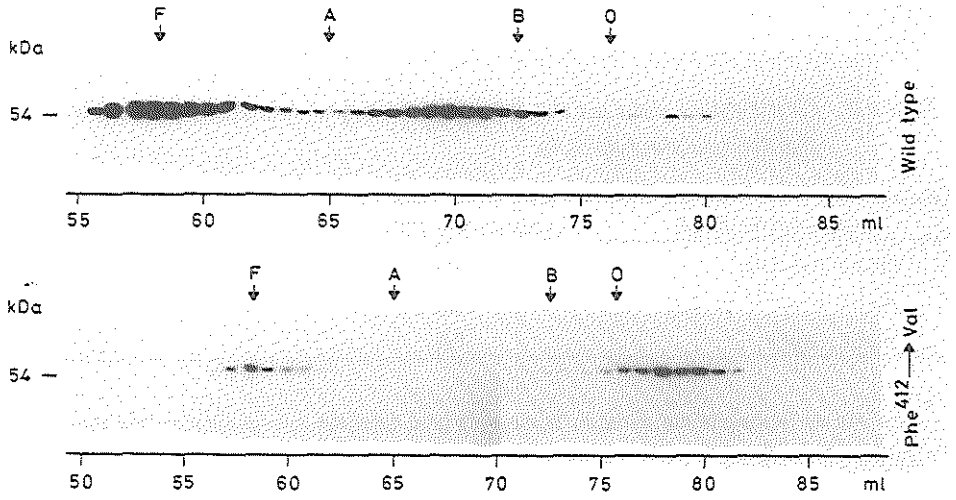


Figure 4. *Gel filtration analysis of normal and mutant protective protein precursors secreted by transfected COS-1 cells. Secreted radiolabeled proteins from COS-1 cells transfected with wild type or mutant (Phe^{412}_Val) cDNAs were concentrated, desalted and applied on a Sephacryl S-200 column. Precursors were immunoprecipitated, resolved and visualized as in Fig.2. The elution position of the separately applied globular protein markers is indicated above the fluorographs: ferritin (F), aldolase(A), BSA(B), and ovalbumin(O).*

Mutant 32/20 kDa two-chain product is rapidly degraded.

To test whether the amino acid substitution would affect or mask the recognition site of the endoprotease that normally converts the precursor to the two-chain form, the following experiment was performed. COS-1 cells, overexpressing either the normal or the mutant cDNA, secrete sufficient amounts of the corresponding 54 kDa precursors. These secreted forms were subjected to partial *in vitro* proteolysis using a fixed concentration of trypsin at increasing time points of incubation (Figure 5). After 2 minutes the normal precursor is already fully trimmed into 32 and 20 kDa subunits that remain rather stable during longer incubation times, though some degradation seems to occur after 30 minutes (Figure 5, left panel). The mutant precursor instead does undergo proteolytic cleavage but this is followed by rapid degradation of the 20 kDa subunit which is only barely visible after 2 minutes incubation. It is likely that this event renders the 32 kDa chain also unstable (Figure 5, right panel).

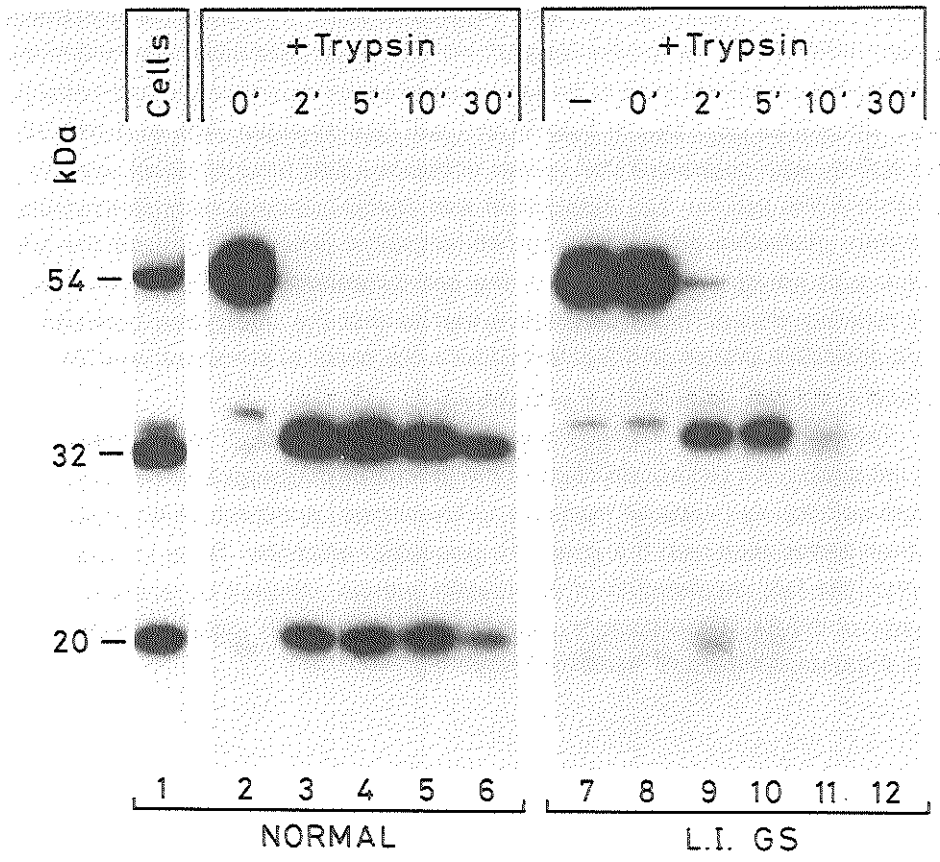


Figure 5. *Trypsin digest of normal and mutant protective protein precursors secreted by transfected COS-1 cells. Transfections were carried out with the pCD-constructs described in Table 1. Cells were labeled with [^{3}H]leucine 48 hr after transfection. Aliquots of medium concentrates containing the secreted precursor polypeptides were incubated with 1 mg of trypsin at 37^{0}C for 0,2,5,10,30 minutes, respectively. Trypsin was inactivated by addition of 3 mg trypsin inhibitor. Radiolabeled proteins were immunoprecipitated, resolved and visualized as in Fig. 2. Lane 1, immunoprecipitation of intracellular protective protein from pCDHu54 transfected COS-1 cells. Lanes 2-6, trypsin digest of normal protective protein precursor. Lane 7-12, trypsin digest of mutant precursor. Molecular sizes of precusor and mature polypeptides are indicated at left. Exposure time for lanes 1-6 was 48 hr, and for lanes 7-12 was 5 days.*

We next investigated whether the mutant protective protein retains the capacity to associate with β-galactosidase. For this purpose COS-1 cells were double transfected with constructs encoding either wild type protective protein and β-galactosidase, or Val^{412} protein and β-galactosidase. The physical association between either of the two protective proteins with β-galactosidase was determined by measuring co-precipitation of the latter enzyme with monospecific anti protective protein antibodies (anti-54). In both COS-1 cell lysates about 70% of β-galactosidase activity was co-precipitated, under conditions that brought down virtually all cathepsin A-like activity. The association was specific since in cells co-expressing the mutant protective protein and human lysosomal β-hexosaminidase the latter enzyme was not co-precipitated with the anti-54 antibodies. In addition, biosynthetic labeling studies have provided evidence that the interaction between the mutant protective protein and β-galactosidase can occur at precursor level (H. Morreau and A. d'Azzo, submitted for publication).

Taken together these results suggest that the Phe^{412} to Val substitution does not disturb the tertiary structure of the precursor polypeptide to the extent that it would completely inhibit its transport to lysosomes, endoproteolytic processing, and even association with the other two lysosomal enzymes. Once in lysosomes, however, the fast degradation of the mutant protein, initiated by endoproteolytic cleavage of the 54 kDa precursor, would explain the loss of protective function towards β-galactosidase and neuraminidase.

Discussion

Galactosialidosis patients with the late infantile form of the disease make up a defined subtype for the following reasons: they exhibit symptoms within the first two years of life, they have a slow and relatively mild progression of the disease, and, most important, the patients for whom a clinical update has been made (Chitayat *et al.*, 1988; Strisciuglio *et al.*, 1990, G. Andria and P. Strisciuglio, personal communication) do not show signs of mental retardation. The latter characteristic seems to be unique for this group and is in contrast with the severe neurological involvement observed in early infantile and even juvenile/adult cases. We were interested in identifying the genetic lesion(s) underlying this phenotype starting with the understanding of how a natural mutation in the gene impairs the gene product and at what subcellular level.

In this report we have described a point mutation in the protective protein gene present in two unrelated late infantile galactosialidosis patients of Italian and Canadian origin. The mutation consists of a T to G transversion at position 1324 of the human cDNA which results in the replacement of phenylalanine with valine at residue 412. The Italian patient is homozygous for this mutation whereas the Canadian patient is a compound heterozygote carrying the Val^{412} substitution together with an, as yet, unidentified lesion in the other allele. The unknown mutation apparently causes a significant decrease in the amount of protective protein mRNA suggesting a defect in transcription or mRNA stability. The T to G substitution instead does not interfere with the synthesis of an mRNA of correct size and quantity. In both patients' fibroblasts this transcript is translated into an abnormal precursor that is not or hardly processed to the mature two-chain product and consequently lacks cathepsin A-like activity (Palmeri *et al.*, 1986; Galjart *et al.*, 1991).

Using site directed mutants with the Phe^{412} to Val change we have demonstrated that this mutation is the one responsible for the *in vivo* phenotype affecting posttranslational modifications, intracellular transport and catalytic activity of the protective protein/cathepsin A. Overexpression of the mutant protein in COS-1 cells made it possible to follow its intracellular distribution for the first time. It became apparent that only a fraction of the newly synthesized precursor pool leaves the endoplasmic reticulum and gets compartmentalized. These molecules fold properly inasmuch as they acquire the mannose-6-phosphate marker, are routed to lysosomes and are also secreted. In addition, there is evidence that the amino acid change does not prevent the recognition of the mutant precursor by one of the other components of the complex, β-galactosidase. It is conceivable that *in vivo* a

fraction of Val^{412} precursor that exits the ER is associated with β-galactosidase and reaches the lysosomes, where it gets rapidly degraded in its processed two-chain form. This event would in turn provoke the loss of protective function towards β-galactosidase and neuraminidase, although the mode and site of interaction of the latter with the other two components is until now unknown. Our results raise the possibility that the protective function of the protein, contrary to its catalytic activity, could be exerted already at precursor level. They are also conformable with the presence of residual mature protective protein in lysosomes of patients' fibroblasts (Palmeri *et al.*, 1986; Strisciuglio *et al.*, 1988).

The fate of the intralysosomal Val^{412} precursor and its retention in the ER could be attributed to a common faulty event, the lack of oligomerization. During the last few years it has become increasingly clear from a variety of systems that not only a protein's tertiary structure but also oligomeric state determines its intracellular transport, final targeting and stability. These early steps in protein biosynthesis often take place and are regulated in the ER (for review see Hurtley and Helenius, 1989). The protective protein precursor by itself appears to exist as an oligomer, probably a dimer. In this form proteins thought to be identical or very homologous to the protective protein are also found in their mature and active state (Simmons and Walter, 1980; Jackman *et al.*, 1990). The dimerization process could play a very important role in the correct targeting and stable conformation of the protective protein, since we find that the monomeric Val^{412} mutant is retarded in the ER. It is unclear at this time whether the rapid degradation of the mutant protein upon proteolytic cleavage is solely due to its monomeric state or also to its inability to stably assemble into a two-chain polypeptide. The effects of the impaired dimerization of mutant protective protein resemble to some extend those caused by a number of natural mutations identified in the α-subunit of lysosomal β-hexosaminidase (Neufeld, 1989; Navon and Proia, 1991; Paw *et al.*, 1991).

It is noteworthy that phenylalanine at position 412 is highly conserved among protective proteins of different species (Galjart *et al.*, 1991). The mutation falls in a region of the 20 kDa subunit that is part of an internal repeat occurring once in each of the two chains and is characterized by recurring tryptophane residues that could be engaged in inter/intramolecular bonds. It has been postulated that initial folding of at least one domain of a subunit is required for correct oligomeric assembly (Hurtley and Helenius, 1989). The "repeated" motif in the protective protein chains might expose surface features essential for the recognition and binding of monomeric molecules with each other. In this context the Phe^{412} to Val change could have dramatic structural consequences.

Finally it is important to keep in mind that the protective protein exerts, at least *in vitro,* a cathepsin A-like activity towards a variety of bioactive peptides (Jackman *et al.*, 1990). This activity is distinct from its protective function toward β-galactosidase and neuraminidase and is deficient in different galactosialidosis patients. The consequences of the deficiency of the enzymatic activity on other metabolic pathways, could be clarified by the identification of the natural substrate(s) that are the target of this multifunctional protein *in vivo*.

Materials and methods

Cell culture

Human skin fibroblasts from normal individuals and the Canadian late infantile galactosialidosis patient (Pinsky *et al.*, 1974) were obtained from the European Cell Bank, Rotterdam (Dr. W.J. Kleijer). Cells from the Italian late infantile galactosialidosis patient (Andria *et al.*, 1978) 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 vol/vol) supplemented with antibiotics and 10% fetal bovine serum. COS-1 cells (Gluzman, 1981) were grown in the same medium, supplemented with 5% fetal bovine serum.

cDNA synthesis and cloning

For the synthesis and amplification of mutant cDNAs four sets of oligonucleotide primers were synthesized on an Applied Biosystems 381A oligonucleotide synthesizer according to the sequence of the human cDNA, Hu54 (Galjart *et al.*, 1988). The four sense primers correspond to nucleotide positions 5-24, 367-383, 701-720, and 1051-1070, respectively; the four antisense to positions 388-407, 750-731, 1080-1099, and 1450-1469, respectively. Eight bases were added to the 5' end of each primer to generate *Eco*RI sites and facilitate ligation of amplified DNA into a plasmid. Total RNA was isolated by the method of Auffray and Rougeon (1980). Four overlapping cDNA fragments, encompassing the entire coding region, were synthesized by reverse PCR (Hermans *et al.*, 1988) and amplified on a Perkin-Elmer Cetus thermocycler programmed for 25 cycles. The fragments were digested with *Eco*RI and subcloned into pTZ18 or pTZ19. Nucleotide sequence analysis on both strands was performed using the dideoxy chain termination method adapted for double-stranded DNA (Murphy and Kavanagh, 1988).

Analysis of amplified genomic DNA

The 239 bp region of genomic DNA surrounding the mutation was PCR amplified according to Saiki *et al.* (1988). The two oligonucleotide primers used in the reaction are derived from intronic.

sequences surrounding exon XIV (N.J. Galjart and A. d'Azzo, in preparation) in the protective protein gene that contains the mutation. The sense oligonucleotide primer sequence is 5'TCTTTCCTGGTGGGGCAGAT 3' (primer 1) and includes 2 bp of the exon; the antisense primer sequence is: 5'CCATACAGGGGCCAGATGGT 3' (primer 2) and is about 100 bp downstream of the exon. For direct DNA sequence analysis an aliquot of the amplified DNA sample was subjected to asymmetric PCR (Kadowaki *et al.*, 1990), using only primer 1 for another 35 cycles. The amplified DNAs were sequenced with primer 2. *Acy*I digestion of PCR amplified genomic fragments was carried out using standard conditions (Sambrook *et al.*, 1989). The resulting fragments were analyzed on a 4% NuSieve agarose gel.

Site-Directed Mutagenesis

In vitro mutagenesis of human protective protein cDNA was done as described by Higuchi *et al.* (1988), using the whole 1.8 kb protective protein cDNA as a template. After synthesis by PCR of a full length cDNA containing the T to G transversion, the entire fragment was verified by sequencing on both strands. The two oligonucleotides used for site directed mutagenesis are 5'ATTGCCGGCTTCGTGAAGGAG3' (sense) and 5'CTCCTTCACGAAGCCGGCAAT3' (antisense).

Transfection of cDNA into COS-1 cells

Normal and mutant protective protein cDNAs were subcloned into a derivative of the mammalian expression vector pCD-X (Okayama and Berg, 1983) as described earlier (Galjart *et al.*, 1988). Transfection of COS-1 cells, metabolic labeling and preparation of cell extracts and media were carried out as reported earlier (Proia *et al.*, 1984; Galjart *et al.*, 1988). Cells were labeled with [^{3}H]leucine (143 Ci/mmol; Amersham Corp.) or with [^{32}P]phosphate (carrier free, Amersham Corp.), for 12 hr or 7 hr, respectively. Labeled proteins were immunoprecipitated with anti-54 kDa antibodies that recognize precursor and mature protective protein (Galjart *et al.*, 1991). They were resolved on 12.5% SDS-polyacrylamide gels under reducing conditions and visualized by autoradiography ([^{32}P]labeled samples) or fluorography ([^{3}H]labeled samples). Transfected COS-1 cells, lysed in double distilled water were assayed for cathepsin A activity using the N-blocked dipeptide Z-Phe-Ala (Bachem Feinchemikalien AG, Bubendorf, Switzerland), using protocols adapted from the methods of Taylor and Tappel (1973) and Roth (1971). Total protein concentration was measured using the method of Smith *et al.* (1985).

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 methods for immunoelectron microscopy were as reported earlier (Willemsen *et al.*, 1986). The

antibodies against the 32 kDa denatured chain of human protective protein (anti-32 antibodies) have been described (Galjart *et al.*, 1988).

Gel filtration

Medium samples from transfected cells were collected, concentrated and desalted as described (Proia et al., 1984). 50% of this material was diluted in 50 mM 2-[N-morpholino] ethane sulfonic acid (MES) pH 6.95, containing 100 mM NaCl, and applied on a column (85x1.53 cm) of Sephacryl S-200 HR (Pharmacia) equilibrated in the same buffer. The whole procedure was performed at 4°C. After gel filtration radiolabeled proteins were immunoprecipitated using anti-54 antibodies and further processed as above. The column was separately run in the same buffer with the following set of globular protein markers (Pharmacia): ferritin (440 kDa), aldolase (158 kDa), BSA (67 kDa) and ovalbumin (43 kDa).

Limited proteolysis with trypsin

Transfections and metabolic labeling of transfected cells with [^{3}H]leucine was performed as above, with the exception that fetal bovine serum was omitted from the labeling medium. Secreted proteins were concentrated 20 fold with $(NH_4)_2SO_4$, in the presence of 1 mg/ml BSA, and desalted. Aliquots of 90 µl medium concentrate were diluted to 200 µl volume with 20 mM sodium phosphate pH 6.8. A modified protocol of Frisch and Neufeld (1981) was used for trypsin digestion. One sample was left on ice, the others were incubated with 1 µg of trypsin (Sigma) at 37°C for 0, 2, 5, 10, 30 minutes, respectively. Reactions were stopped by addition of 3 µg bovine pancreas trypsin inhibitor (Sigma). At t = 0 the inhibitor was added before the trypsin. Radiolabed proteins were immunoprecipitated and further processed as described previously.

Acknowledgements

We wish to thank Dr. G. Grosveld for useful discussions and suggestions; Prof.Dr. G. Andria and Dr. P. Strisciuglio (University of Napoli, Italy) for providing the cell strain from the Italian patient; Dr. W.J. Kleijer (European Cell Bank, Rotterdam) for making available the other galactosialidosis cell strains. We are grateful to Pim Visser for the graphic work, Mirko Kuit for excellent photography and Jeannette Lokker for typing and editing this manuscript.

References

Andria, G., Del Giudice, E. and Reuser, A.J.J. (1978) *Clin. Genet.,* 14, 16-23.

Andria, G., Strisciuglio, P., Pontarelli, G., Sly, W.S. and Dodson, W.E. (1981) In Tettamanti, G., Durand, P. and Di Donato, S. (eds.), *Sialidases and Sialidosis.* Milan: Edi. Ermes. pp. 379-395.

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

Chitayat, D., Applegarth, D.A., Lewis, J., Dimmick, J.E., McCormick, A.Q. and Hall, J.G. (1988) *Am. J. Hum. Genet.,* 31, 887-901.

d'Azzo, A., Hoogeveen, A.T., Reuser, A.J.J., Robinson, D. and Galjaard, H. (1982) *Proc. Natl. Acad. Sci. USA,* 79, 4535-4539.

Frisch, A. and Neufeld, E.F. (1981) *J. Biol. Chem.,* 256, 8242-8246.

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. and d'Azzo, A. (1991) *J. Biol. Chem.,* 266, 14754-14762.

Gluzman, Y. (1981) *Cell* 23, 175-182.

Hermans, A., Gow, J., Selleri, L., von Lindern, M., Hagemeijer, A., Wiedemann, L.M. and Grosveld, G.C. (1988) *Leukemia* 2, 628-633.

Higuchi, R., Kremmel, B. and Saiki, R.K. (1988) *Nucl. Acids Res.,* 16, 7351-7367.

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

Hurtley, S.M. and Helenius, A. (1989) *Annu. Rev. Cell Biol.,* 5, 277-307.

Jackman, H.L., Tan, F., Tamei, H., Beurling-Harbury, C., Li, X.-Y., Skidgel, R.A. and Erdös, E.G. (1990) *J. Biol. Chem.,* 265, 11265-11272.

Kadowaki, T., Kadowaki, H. and Taylor, S.I. (1990) *Proc. Natl. Acad. Sci. USA,* 87, 658-662.

Kase, R., Itoh, K., Takiyama, N., Oshima, A., Sakuraba, H. and Suzuki, Y. (1990) *Biocem. Biophys. Res. Commun.,* 172, 1175-1179.

Loonen, M.C.B., Reuser, A.J.J., Visser, P. and Arts, W.F.M. (1984) *Clin. Genet.,* 26, 139-149.

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

Navon, R. and Proia, R.L. (1991) *Am. J. Hum. Genet.,* 48, 412-419.

Neufeld, E.F. (1989) *J. Biol. Chem.*, 264, 10927-10930.

O'Brien, J.S. (1989) In Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D. (eds.), *The Metabolic Basis of Inherited Disease.* McGraw-Hill, New York, ed. 6, pp. 1797-1806.

Okada, S., Kato, T., Miura, S., Yabuuchi, H., Nishigaki, M., Kobata, A., Chiyo, H. and Furuyama, J. (1978) *Clin. Chim. Acta,* 86,159-167.

Okayama, H. and Berg, P. (1983) *Mol. Cell. Biol.,* 3, 280-289.

Palmeri, S., Hoogeveen, A.T., Verheijen, F.W. and Galjaard, H. (1986) *Am. J. Hum. Genet.,* 38, 137-148.

Paw, B.H., Kaback, M.M. and Neufeld, E.F. (1989) *Proc. Natl. Acad. Sci. USA*, 86, 2413-2417.

Paw, B.H., Wood, L.C. and Neufeld, E.F. (1991) *Am. J. Hum. Genet.*, 48, 1139-1146.

Pinsky, L., Miller, J., Shanfield, B., Watters, G. and Wolfe, L.S. (1974) *Am. J. Hum. Genet.,* 26, 563-577.

Proia, R.L., d'Azzo, A. and Neufeld, E.F. (1984) *J. Biol. Chem.,* 259, 3350-3354.

Roth, M. (1971) *Anal. Biochem.,*43, 880-882.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985) *Science,* 230, 1350-1354.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, New York, 2nd edition.

Sewell, A.C., Pontz, B.F., Weitzel, D. and Humburg, C. (1987) *Eur. J. Pediatr.,* 146, 528-531.

Simmons, W.H. and Walter, R. (1980) *Biochemistry*, 19, 39-48.

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 Klerk, D.C. (1985) *Anal. Biochem.*, 150, 76-85.

Strisciuglio, P., Parenti, G., Giudice, C., Lijoi, S., Hoogeveen, A.T. and d'Azzo, A. (1988) *Hum. Genet.*, 80, 329-332.

Strisciuglio, P., Sly, W.S., Dodson, W.E., McAlister, W.H. and Martin, T.C. (1990) *Am. J. Hum. Genet.,* 37, 573-577.

Suzuki, Y., Sakuraba, H., Yamanaka, T., Ko, Y.M., Iimori, Y., Okamura, Y. and Hoogeveen, A.T. (1984) In Arina, M., Suzuki, Y. and Yabuuchi, H. (eds.), *The Developing Brain and Its Disorders.* University of Tokyo Press, Tokyo. pp. 161-175.

Taylor, S. and Tappel, A.L. (1973) *Anal. Biochem.,*56, 140-148.

Tranchemontagne, J., Michaud, L. and Poitier, M. (1990) *Biochem. Biophys. Res. Commun.,* 168, 22-29.

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.

van Pelt, J., van Kuik, J.A., Kamerling, J.P., Vliegenthart, J.F., van Diggelen, O.P. and Galjaard, H. (1988a) *Eur. J. Biochem.*, 177, 327-338.

van Pelt, J., Kamerling, J.P., Vliegenthart, J.F., Hoogeveen, A.T. and Galjaard, H. (1988b) *Clin. Chim. Acta*, 174, 325-335.

van Pelt, J., Hard, K., Kamerling, J.P., Vliegenthart, J.F., Reuser, A.J.J. and Galjaard, H. (1989) *Biol. Chem. Hoppe. Seyler.*, 370, 191-203.

Verheijen, F.W., Brossmer, R. and Galjaard, H. (1982) *Biochem. Biophys. Res. Commun.*, 108, 868-875.

Wenger, D.A., Tarby, T.J. and Wharton, C. (1978) *Biochem. Biophys. Res. Commun.* 82,589-595.

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

Publication 7

Genomics **10** (1991), 345-349

GENOMICS 10, 345–349 (1991)

The Gene Encoding Human Protective Protein (PPGB) Is on Chromosome 20

JOOP WIEGANT,* NIELS J. GALJART,† ANTON K. RAAP,* AND ALESSANDRA D'AZZO†,[1]

*Department of Cytochemistry and Cytometry, University of Leiden, 2333 AL Leiden, The Netherlands; and †Department of Cell Biology and Genetics, Erasmus University, 3000 DR Rotterdam, The Netherlands

Received May 7, 1990; revised December 17, 1990

Normal lymphocyte prometaphase chromosome spreads were hybridized *in situ* using single- and double-color fluorescence techniques. The results obtained with either the 1.8-kb protective protein cDNA or a 12-kb genomic fragment of the human protective protein gene as probe demonstrate that the PPGB gene is localized on the long arm of chromosome 20. This assignment was confirmed by hybridization with whole chromosome DNA libraries. © 1991 Academic Press, Inc.

INTRODUCTION

Human protective protein is the glycoprotein required for stability and activity of the lysosomal enzymes β-galactosidase (EC 3.2.1.23) and neuraminidase (EC 3.2.1.18) (d'Azzo *et al.*, 1982; Hoogeveen *et al.*, 1983; Verheijen *et al.*, 1982, 1985). A primary deficiency of the protective protein (d'Azzo *et al.*, 1982) causes severely reduced β-galactosidase and neuraminidase activities and results in the metabolic storage disorder galactosialidosis (Wenger *et al.*, 1978; Okada *et al.*, 1978; Andria *et al.*, 1981; Suzuki *et al.*, 1984). In human cultured fibroblasts the protective protein is synthesized as a 54-kDa precursor, which is post-translationally processed into a mature heterodimer of 32- and 20-kDa polypeptides, held together by disulfide bridges (d'Azzo *et al.*, 1982; Galjart *et al.*, 1988). Cloning of human protective protein cDNA and analysis of its predicted amino acid sequence showed that the protective protein is homologous to yeast serine carboxypeptidases (Galjart *et al.*, 1988). Domains essential for the proteolytic activity of these enzymes are completely conserved in the human protein, suggesting that the protective protein also functions as a serine carboxypeptidase.

Two contradictory reports described the localization of the gene encoding human protective protein. Sips *et al.* (1985) localized it on chromosome 22, using an antibody preparation that recognized both human protective protein and β-galactosidase to screen human/mouse and human/Chinese hamster somatic cell hybrids. On the other hand, Mueller *et al.* (1986), by scoring for the increase in neuraminidase activity over the mouse background in human/mouse somatic cell hybrids, localized on chromosome 20 a gene that is defective in a galactosialidosis patient. They named it GSL (galactosialidosis). Here we show by double-fluorescence *in situ* hybridization that the gene encoding the protective protein is located on the long arm of chromosome 20. Our results are substantiated by the fact that both the human cDNA and a genomic DNA fragment containing almost the entire gene encoding the protective protein hybridize to the same locus.

EXPERIMENTAL PROCEDURES

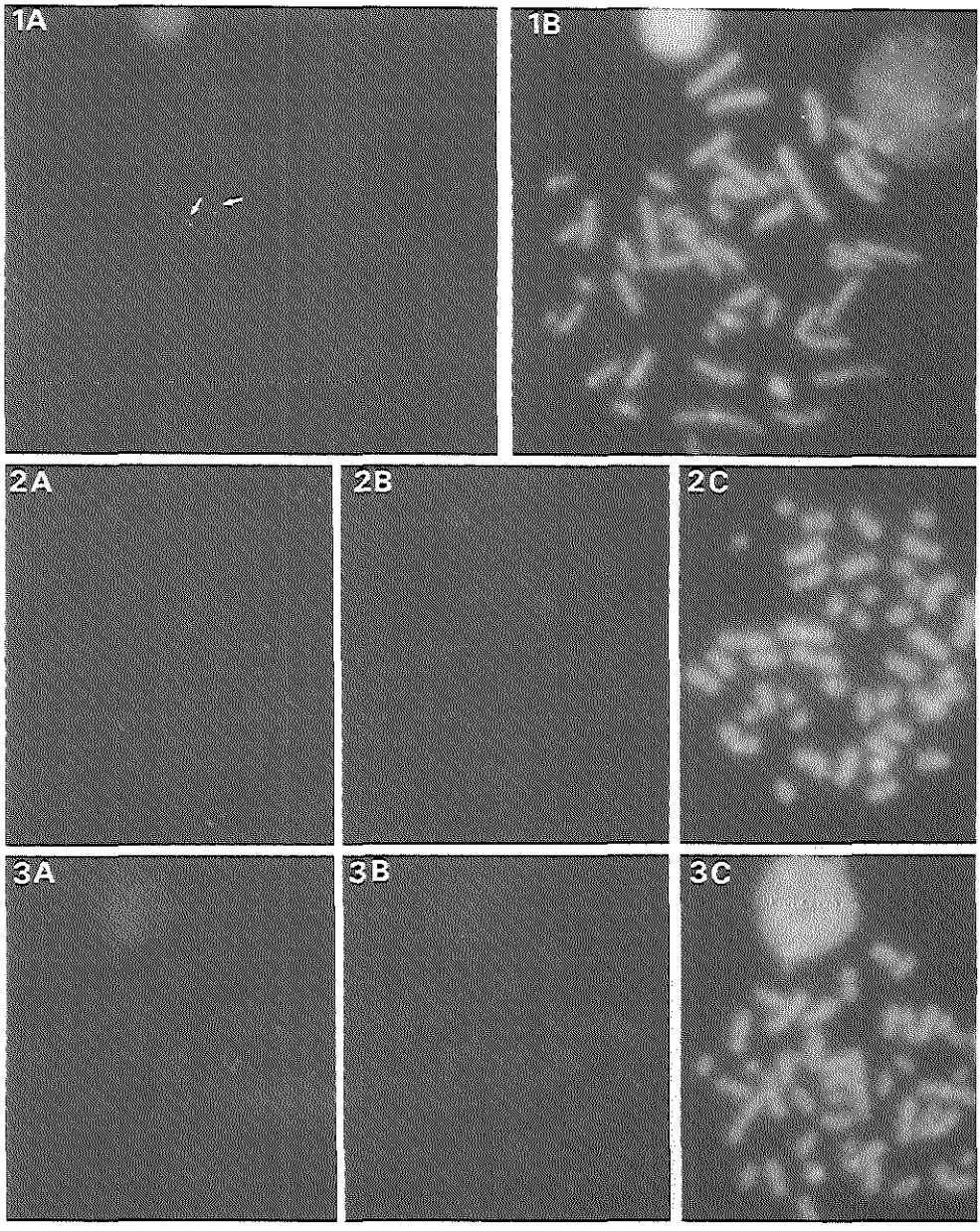
Materials

Reagents and enzymes were obtained from the following companies: RNase A, digoxigenin-11-dUTP, mouse anti-digoxigenin, and blocking reagent from Boehringer (Mannheim, Germany); biotin-11-dUTP, rabbit anti-mouse TRITC, goat anti-rabbit TRITC, propidium iodide, the antifading reagent 1,4-diazabicyclo-(2,2,2)-octane, and actinomycin D from Sigma (U.S.A.); avidin-D-FITC and biotinated goat anti-avidin from Vector Laboratories (U.S.A.); pepsin and 4',6-diamidino-2-phenylindole (DAPI) from Serva (Germany).

In Situ Hybridization

Routine prometaphase spreads obtained from a primary peripheral blood lymphocyte culture of a

[1] To whom correspondence should be addressed at Department of Cell Biology and Genetics, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. Fax: (0)10-4087212


1A
1B
2A
2B
2C
3A
3B
3C

healthy male were treated with RNase A (100 μg/ml 2× SSC) for 1 h at 37°C, using 100 μl solution under a 25 × 50-mm² coverslip. The slides were then rinsed three times for 5 min with 2× SSC and further incubated with pepsin (50 μg/ml 10 m*M* HCl) for 10 min at 37°C. After a rinse in PBS and a 5-min wash in PBS containing 50 m*M* $MgCl_2$, postfixation was performed with formaldehyde (1% (v/v)) in PBS containing 50 m*M* $MgCl_2$. This step was followed by a PBS rinse and gradual dehydration of the slides in ethanol. The cDNA (1.8 kb) and a fragment of genomic DNA (12 kb) were labeled with biotin-11-dUTP essentially according to Langer *et al.* (1981). The Bluescript DNA libraries specific for chromosomes 19, 20, 21, and 22 (generous gifts of Dr. J. W. Gray, Lawrence Livermore Laboratories, U.S.A.) were labeled with digoxigenin-11-dUTP using a similar nick-translation format.

For single hybridizations with the biotinated cDNA, the hybridization mixture contained 2 ng of the probe, 100 ng of sonicated salmon sperm DNA, and 100 ng of yeast tRNA in 1 μl of 50% formamide, 50 m*M* phosphate, 10% dextran sulfate, 2× SSC, pH 7.0. The probe mixture (10 μl) was applied to the slides, covered with a coverslip of 18 × 18 mm², and sealed with rubber cement. Probe and chromosomal DNA were denatured simultaneously by placing the slides on a 80°C metal plate in an incubator for 2.5 min. Hybridization was allowed to take place overnight in a moist chamber.

For double-fluorescence *in situ* hybridization (e.g., the biotinated genomic probe and a digoxiginated-chromosome library) the labeled probes were mixed in the hybridization buffer, which contained in addition a 500-fold excess of sonicated unlabeled total human placental DNA. This probe–competitor mixture was denatured prior to hybridization for 5 min at 75°C, quickly chilled, and subsequently incubated for 2–4 h at 37°C to compete out repetitive sequences in any of the probes. Chromosomal DNA was denatured separately on the slide by immersing it in 70% formamide, 10 m*M* phosphate, 2× SSC at 80°C for 2.5 min. The slides were then washed twice for 5 min with cold 70% ethanol (−20°C), dehydrated, and dried at 37°C. The 37°C hybridization mixture was subsequently applied under a coverslip, and hybridization was allowed to continue overnight.

After hybridizations, coverslips were removed and slides were washed three times for 20 min in 50% formamide, 2× SSC, pH 7.0, at 45°C, followed by three 5-min washes in 0.1× SSC at 60°C, and finally rinsed in 4× SSC, 0.05% Tween 20 (v/v) at room temperature (RT).

Immunocytochemical Detection and Banding

The detection of the biotinated cDNA hybridization was performed essentially according to the amplification method of Pinkel *et al.* (1986), using the high salt washes recommended by Lawrence *et al.* (1988). Briefly, slides were preincubated in 4× SSC, 5% nonfat dry milk (NFD) for 20 min at RT. After incubation with avidin-D-FITC in 4× SSC/NFD for 20 min at RT, the signals were amplified with biotinated goat anti-avidin-D in 4× SSC, 20 min at RT, followed by another incubation with avidin-D-FITC. Three 3-min washes with 4× SSC, 0.05% Tween-20 were performed at RT. Finally, the slides were rinsed once for 5 min with PBS, dehydrated through an ethanol series, and air-dried. The chromosomes were banded by an incubation with 100 ng DAPI/ml 0.2 *M* sodium phosphate/0.1 *M* citric acid, pH 7.0, for 25 min at RT. After two 2-min washes in H_2O, the slides were incubated with 0.3 mg actinomycin D/ml 10 m*M* sodium phosphate, pH 7.0, 1 m*M* EDTA for 18 min at RT. After two 2-min washes in H_2O, the slides were mounted in an antifade medium containing 9 parts glycerol, 1 part 1 *M* Tris · HCl, pH 7.5, 2% 1,4-diazabicyclo(2,2,2)-octane, and 0.02% thiomersal.

Detection of the signals from the simultaneous hybridization with digoxigenated library DNA and biotinated genomic DNA was performed as follows. Slides were preincubated in 4× SSC/NFD, washed with 4× SSC, 0.05% Tween-20, and incubated with avidin-D-FITC in 4× SSC/NFD for 20 min at RT. They were subsequently washed twice for 5 min with 4× SSC, 0.05% Tween-20 and once for 5 min with 0.1 *M* Tris · HCl, 0.15 *M* NaCl, 0.05% Tween 20, pH 7.5 (TNT). Afterward they were incubated with a mixture of biotinated goat anti-avidin-D and mouse monoclonal anti-digoxigenin, diluted in 0.1 *M* Tris · HCl, pH 7.5, 0.15 *M* NaCl, 0.5% blocking reagent (TNB), for 30 min at 37°C. After three 5-min washes with TNT, the slides were treated with a mix-

FIG. 1. Fluorescence *in situ* hybridization of the 1.8-kb biotinated protective protein cDNA. (A) Fluorescein image showing the specific 2 × 2 hybridization signals (arrows). (B) DAPI/actinomycin image of the microscopic field.

FIG. 2. Double-fluorescence *in situ* hybridization with biotinated 12-kb genomic protective protein DNA and digoxiginated chromosome 19 library DNA. (A) Fluorescein signals resulting from the protective protein–DNA *in situ* hybridization. (B) Rhodamine signals obtained by *in situ* hybridization with chromosome 19 DNA. (C) DAPI/actinomycin counterstaining.

FIG. 3. Double-fluorescence *in situ* hybridization with biotinated 12-kb genomic protective protein DNA and digoxigenated chromosome 20 library DNA. (A) Fluorescein signals resulting from the protective protein–DNA *in situ* hybridization. (B) Rhodamine signals obtained by *in situ* hybridization with chromosome 20 DNA. (C) DAPI/actinomycin counterstaining.

ture of avidin-D-FITC and rabbit anti-mouse TRITC diluted in TNB, for 30 min at 37°C. Finally, after three 5-min washes with TNT, the slides were incubated for 30 min at 37°C with goat anti-rabbit TRITC diluted in TNB. They were then washed three times for 5 min with TNT, dehydrated through an ethanol series, air-dried, and mounted in antifade medium containing 150 ng DAPI/ml, as a DNA counterstain.

Photomicrographs were taken on a Dialux (Leitz) microscope equipped for epillumination on a 640 asa color slide film (3M).

RESULTS AND DISCUSSION

The cDNA encoding human protective protein has been cloned and characterized (Galjart *et al.*, 1988). Using the cDNA as a probe we have analyzed on Southern blots restricted DNA from a number of human/mouse and human/Chinese hamster somatic cell hybrids. The locus for the protective protein, however, could not be unambiguously assigned, although in one hybrid cell line, pgMe25Nu (Geurts van Kessel *et al.*, 1981), that retained only human chromosome 22, no human-specific hybridization fragments could be detected (data not shown). Thus the gene encoding the protective protein could not be on chromosome 22. For a conclusive localization and to confirm either of the previously published reports, we decided to perform *in situ* hybridization. The entire 1.8-kb protective protein cDNA and a 12-kb genomic DNA fragment (containing all coding exons) were independently used as probes.

Figure 1A shows a single hybridization with the biotinated cDNA. Although chromosome banding with the counterstaining used is poorly resolved on the color micrograph (Fig. 1B), under the microscope the region carrying the *in situ* hybridization signal was unambiguously identified as 20q13.1. The frequency of metaphase plates carrying the expected 2 × 2 spots on the chromatids was about 0.1. The same chromosomal assignment was made with the single hybridization using the biotinated genomic DNA. The number of plates with four spots was higher (0.4) and, as a consequence of the larger target, the hybridization intensities were stronger (results not shown). To confirm the cytogenetic assignment, double-fluorescence *in situ* hybridization (Hopman *et al.*, 1986; Nederlof *et al.*, 1989, 1990) was performed with chromosome-specific libraries for 19, 20, 21, and 22. Figures 2 and 3 show the results of such hybridizations with the protective protein genomic DNA (2A and 3A) and the libraries of chromosome 19 (2B) and 20 (3B). The protective protein specific signals are clearly on chromosome 20.

Taken together, the results obtained with two different DNA probes and two independent ways of classifying chromosomes demonstrate that the PPGB gene encoding human protective protein is on chromosome 20q13.1. Furthermore, they confirm and extend the observation of Mueller *et al.* (1986), indicating that the PPGB gene is almost certainly the same as the GSL gene.

ACKNOWLEDGMENTS

We thank Dr. J. W. Wessels, Department of Human Genetics, University of Leiden, for his help with the cytogenetic analysis; Dr. J. W. Gray, Lawrence Livermore Laboratories, California, for the Bluescript chromosome libraries; and Miss Jeannette Lokker for typing and editing this manuscript.

REFERENCES

1. ANDRIA, G., STRISCIUGLIO, P., PONTARELLI, G., SLY, W. S., AND DODSON, W. E. (1981). Infantile neuraminidase and β-galactosidase deficiencies (galactosialidosis) with mild clinical courses. *In* "Sialidases and Sialidoses" (G. Tettamanti, P. Durand, and S. DiDonato, Eds.), pp. 379–395. Edi. Ermes, Milano.
2. 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.
3. 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.
4. GEURTS VAN KESSEL, A. H. M., DEN BOER, W. C., VAN ACTHOVEN, A. J., AND HAGEMEIJER, A. (1981). Decreased tumorigenicity of rodent cells after fusion with leukocytes from normal and leukemic cells. *Somatic Cell Genet.* 7: 645–656.
5. 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.
6. HOPMAN, A. H. N., WIEGANT, J., RAAP, A. K., LANDEGENT, J. E., VAN DER PLOEG, M., AND VAN DUIJN, P. (1986). Bi-color detection of two target DNAs by non-radioactive in situ hybridization. *Histochemistry* 85: 1–4.
7. JOHNSON, G. D., DAVIDSON, R. S., MCNAMEE, K. C., RUSSELL, G., GOODWIN, D., AND HOLBORROW, E. L. (1982). Fading of immunofluorescence during microscopy: A study of its phenomenon and its remedy. *J. Immunol. Methods* 55: 231–242.
8. LANGER, P. R., WALDROP, A. A., AND WARD, D. A. (1981). Enzymatic synthesis of biotin labeled polynucleotides: Novel nucleic acid affinity probes. *Proc. Natl. Acad. Sci. USA* 78: 6633–6637.
9. LAWRENCE, J. B., VILLNAVE, C. A., AND SINGER, R. H. (1988). Sensitive, high resolution chromatin and chromosome mapping in situ: Presence and orientation of two closely integrated copies of EBV in lymphoid line. *Cell* 52: 51–61.
10. MUELLER, O. T., HENRY, W. M., HALEY, L. L., BYERS, M. G., EDDY, R. L., AND SHOWS, T. B. (1986). Sialidosis and galactosialidosis: Chromosomal assignment of two genes associated

with neuraminidase-deficiency disorders. *Proc. Natl. Acad. Sci. USA* 83: 1817–1821.

11. NEDERLOF, P. M., ROBINSON, D., ABUKNESHA, R., WIEGANT, J., HOPMAN, A. H. N., TANKE, H. J., AND RAAP, A. K. (1989). Three color fluorescence in situ hybridization for the simultaneous detection of multiple nucleic acid sequences. *Cytometry* 10: 20–27.
12. NEDERLOF, P. M., VAN DER FLIER, S., WIEGANT, J., RAAP, A. K., TANKE, H. J., PLOEM, J. S., AND VAN DER PLOEG, M. (1990). Multiple fluorescence in situ hybridization. *Cytometry* 11: 126–131.
13. OKADA, S., KATO, T., MIURA, S., YABUUCHI, H., NISHIGAKI, M., KOBATA, A., CHIYO, H., AND FURUYAMA, J. (1978). Hypersialyloligosaccharidurla in mucolipidoses: A method for diagnosis. *Clin. Chim. Acta* 86: 159–167.
14. PINKEL, D., STRAUME, D., AND GRAY, J. W. (1986). Cytogenetic analysis using quantitive, high sensitivity fluorescence hybridization. *Proc. Natl. Acad. Sci. USA* 83: 2934–2938.
15. SIPS, H. J., DE WIT-VERBEEK, H. A., DE WIT, J., WESTERVELD, A., AND GALJAARD, H. (1985). The chromosomal localization of human β-galactosidase revisited: A locus for β-galactosidase on human chromosome 3 and for its protective protein on human chromosome 22. *Hum. Genet.* 69: 340–344.
16. SUZUKI, Y., SAKURABA, H., YAMANAKA, T., KO, Y. M., LIMORI, Y., OKAMURA, Y., AND HOOGEVEEN, A. T. (1984). Galactosialidosis: A comparative study of clinical and biochemical data on 22 patients. *In* "The Developing Brain and its Disorders" (M. Arina, Y. Suzuki, and H. Yabuuchi, Eds.), pp. 161–175. Univ. of Tokyo Press, Tokyo.
17. 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.
18. VERHEIJEN, F. W., PALMERI, S., HOOGEVEEN, A. T., AND GALJAARD, H. (1985). Human placental neuraminidase: Activation, stabilization and association with β-galactosidase and its protective protein. *Eur. J. Biochem.* 149: 315–321.
19. WENGER, D. A., TARBY, T. J., AND WHARTON, C. (1978). Macular cherry-red spots and myoclonus with dementia: Coexistent neuraminidase and β-galactosidase deficiencies. *Biochem. Biophys. Res. Commun.* 82: 589–595.

2.3 1 Discussion

The results presented in publications 1-7 demonstrate that human protective protein, encoded by a gene on chromosome 20, is synthesized as a precursor of 54 kDa, that is phosphorylated on a single Asn-linked oligosaccharide chain which allows its transport to lysosomes via the MPR targeting system. Probably in an endosomal/lysosomal compartment the precursor is proteolytically modified into a two-chain form consisting of disulfide-linked 32 and 20 kDa polypeptides. The 32 kDa component is subsequently further trimmed at its C-terminus. A schematic representation of these events is given in Figure 3.

Characterisation of the primary structure of the protective protein revealed that it maintains another function, namely as a serine carboxypeptidase, distinct from its protective role towards β-galactosidase and neuraminidase. The catalytic activity is unleashed only after conversion of the 54 kDa zymogen to the two-chain form. It still needs to be determined whether the precursor already has protective function or not.

Serine carboxypeptidases are a family of single and two-chain enzymes that are present in many different species, ranging from yeast to fungi, plants and man (for review see Breddam, 1986). Most of them are aspecific peptidases, involved in general protein turnover, but the KEX1 gene product of yeast has carboxypeptidase-B-like activity and participates in hormone processing (Dmochowska *et al,* 1987). Comparison of the structure of the protective protein with that of other serine carboxypeptidases allows several speculations to be made. First, some of the plant peptidases have been purified in an active two-chain form from which the amino acid sequences of the subunits were chemically determined (for a compendium of these sequences see Sorensen *et al,* 1989). In the case of barley carboxypeptidase I, these sequences were compared with the one of the precursor polypeptide that was deduced from the isolated cDNA. It was shown that the two subunits originate from a single chain precursor polypeptide that in addition contains a stretch of amino acids separating the two chains (Doan and Fincher, 1988). This highly resembles the biosynthesis and processing of the protective protein and provides a verified example in a homologous carboxypeptidase of C- (or N-) terminal trimming after the initial endoproteolytic processing step. The analogy in maturation events might indicate a similar function for the intermediate stretches of amino acids in the plant carboxypeptidase and protective protein precursors, namely to keep these forms in an inactive state. In Zn^{2+} carboxypeptidase B, such an "inactivation" function resides within its propeptide segment, which simply obstructs the access of a

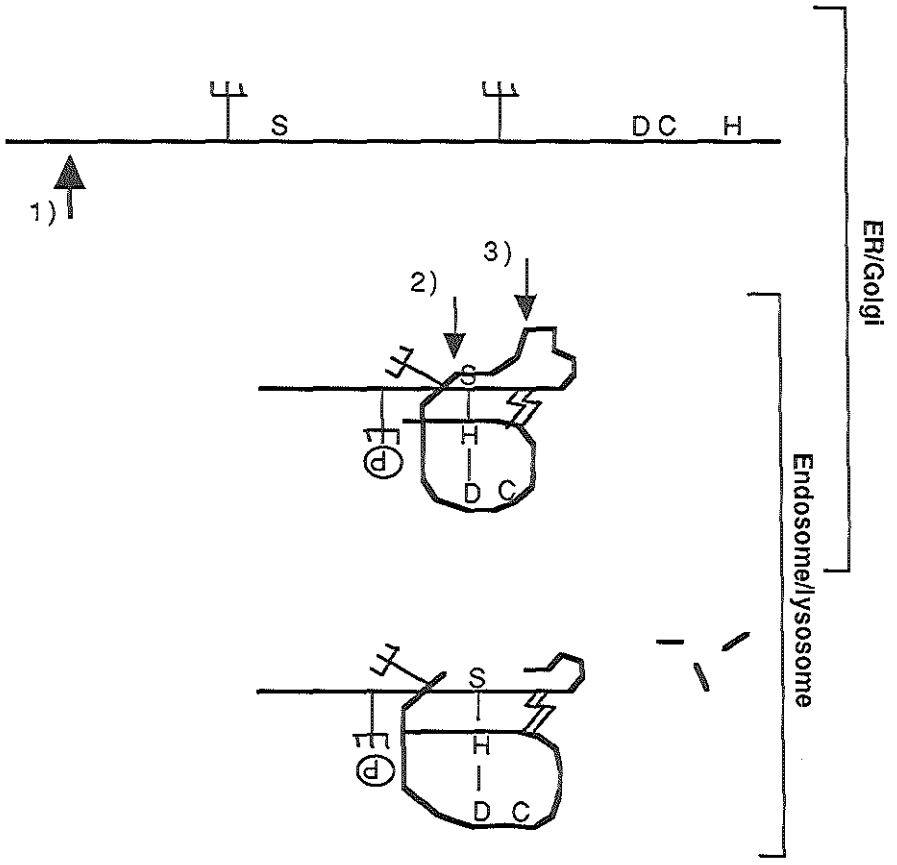


Figure 3. *Biosynthesis and processing of the protective protein.*
Three proteolytic processing steps are thought to occur during the life cycle of the protective protein, 1) signal sequence cleavage, 2) the first endosomal/lysosomal endoprotease cleavage and 3) C-terminal trimming of the 32 kDa polypeptide. Oligosaccharide side chains on the two subunits can be distinguished because the one on the 32 kDa polypeptide carries the M6P-recognition marker. In keeping with the predictions made in the text, the intermediate stretch of amino acids between 32 and 20 kDa chains is placed in front of the serine active site residue in the precursor molecule. Disulfide bridges connect the two chains. Of the nine cysteines in human protective protein cys375 is shown, since it might be in the catalytic site (see text).

substrate to the preformed active site (Coll *et al,* 1991). Limited proteolysis of pro-carboxypeptidase B with trypsin removes this inactivating segment and exposes the active site. It will be of interest to determine whether a similar activation mechanism exists for the two-chain serine carboxypeptidases.

Second, it is noteworthy that the two-chain plant serine carboxypeptidases, that bear such striking structural resemblance to the protective protein, also form dimers but not multimers (Breddam, 1986). Remarkably, and perhaps related to this is the fact that no neuraminidases exist in the plant kingdom. Maybe this implies that during evolution the protective protein acquired its catalytic activity first and its protective function later.

A third speculation that can be made is that Cys at amino acid position 375 in the active human protective protein is near the substrate binding site within the catalytic center, since this was shown to be the case for the homologous residue of yeast carboxypeptidase Y (Breddam and Svendsen, 1984; Winther and Breddam, 1987). Catalytic activity should then be modulated by reagents that interact with sulfhydryl groups, a prediction that is indeed corroborated in several reports (Jackman *et al,* 1990; Kawamura *et al,* 1974; 1975; Tranchemontagne *et al,* 1990). What remains are an even number of cysteines (8) in the human protective protein, that, as is the case for Cys^{375}, are conserved in the mouse and chicken homologues and that potentially could form intramolecular disulfide bridges. In addition, mouse and chicken protective proteins contain 2 extra cysteines, one on each subunit of their respective two-chain forms, but at different positions.

The protective protein is most likely identical to a carboxypeptidase/deamidase, secreted by human platelets upon thrombin stimulation (Jackman *et al,* 1990). *In vitro* this enzyme cleaves a number of important bioactive peptides among which are substance P, oxytocin and angiotensin I. Its deamidase activity is optimal at neutral pH, whereas its carboxypeptidase activity has an optimal pH of 5.0, with the artificial substrate furylacryloyl-Phe-Phe. The resemblance between the protective protein and the carboxypeptidase/deamidase is strengthened by the fact that both proteins maintain cathepsin A-like activity. Cathepsin A has been purified from many tissues and species, it is a carboxypeptidase which cleaves N-blocked dipeptides optimally at pH 5.5 and has a preference for hydrophobic residues in the (pen)ultimate position of the substrate protein. Other residues, but not lysine or arginine, may also be released (reviewed by McDonald and Barrett, 1986). Cathepsin A was shown to have deamidase activity as well (Matsuda, 1976). Because of its broad substrate specificity cathepsin A is thought to be involved in lysosomal protein turnover. If the protective protein and cathepsin A turn

out to be identical this hypothesis can be directly tested, by comparing cellular protein degradation in galactosialidosis and normal fibroblasts. Some authors have postulated a physiological role for cathepsin A in the conversion of angiotensin I to angiotensin II (Miller *et al,* 1988; 1991). Together, the data suggest that the role of the protective protein/cathepsin A might include protein turnover as well as the (in)activation of bioactive peptides and might be extended to an extralysosomal compartment. This could imply that some tissues have a differential need for the protective protein. Consistent with this is the observation that murine protective protein mRNA is differentially expressed in various mouse tissues (publication 2) and that its expression pattern differs from that of β-galactosidase mRNA (Morreau *et al,* 1991). A function of the protective protein outside the complex explains the previously unaddressed observation that in extracts of normal human fibroblasts the majority of the protective protein is not present in the high molecular weight complex together with β-galactosidase, but is instead immunoprecipitated from a "monomeric" peak (Hoogeveen *et al,* 1983).

Our results in publications 5 and 6 demonstrate that dimerization of the protective protein/cathepsin A occurs at precursor level and neutral pH, suggesting that it takes place in an early biosynthetic compartment. Dimerization 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 (publication 6). We have evidence that protective protein and β-galactosidase precursors associate at an early stage of biosynthesis, indicating that their assembly is not restricted to the lysosomal environment (H. Morreau and A. d'Azzo, manuscript in preparation). The advantage of this event in normal cells could be to 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 extent and in which stoichiometry association occurs in normal cells and tissues and whether this influences the dimerization of the protective protein precursor. It should be mentioned 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).

The finding that some *in vitro* mutagenized protective proteins, retained in the ER, prevent lysosomal targeting of ß-galactosidase (H. Morreau and A. d'Azzo, manuscript in preparation) has important consequences for the interpretation of the molecular events in galactosialidosis and perhaps even G_{M1}-gangliosidosis.

Natural mutants of the protective protein may 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. One could consider this kind of mutant protective proteins to have a "dominant" effect. The mutated form of the protective protein, present in the late infantile galactosialidosis patient described in publication 6, would have a "recessive" character instead. In the latter case the mutant protective protein could actually be rescued from premature ER degradation because its association to ß-galactosidase precursors would ensure its exit out of this compartment. An inverse situation, that could also be envisaged, is that there are "dominant" mutations in the ß-galactosidase protein that influence the protective protein. Remarkably, the proteolytic processing of the latter indeed seems altered in two cell lines derived from pateints with the infantile and adult forms of G_{M1}-gangliosidosis, respectively (Hoogeveen *et al,* 1984). These fibroblasts need to be reexamined with the monospecific antibodies against the protective protein that are now available.

Lysosomes have a pivotal function in the intracellular degradation of a multitude of macromolecules. They are the "dead end" route in a largely interconnected vesicular system that has its beginning in the ER. In spite of their well established digestive role the supramolecular structure of proteins within lysosomes is still unknown. Therefore, predictions of how intralysosomal degradation could be carried out in the most effective way are speculative. Several reasons given in the introduction explain why a stepwise macromolecular degradation is speeded up if the enzymes that act sequentially on a given substrate were to form complexes. In addition, in lysosomes complex formation could prevent the rapid degradation of the true lysosomal constituents by other hydrolases. On the other hand, there are many different types of substrates. If each would demand complexed enzymes for its quick degradation this would require as many complexes as there are substrates. Therefore, it may be that most lysosomal enzymes are only temporarily and loosely associated within complexes and that these vary constantly in enzymic composition, depending on incoming substrate.

A hypothetical model incorporating these considerations as well as data on the structure of the protective protein/cathepsin A is presented in Figure 4. In lysosomes (and perhaps other organelles) an equilibrium could exist between multimeric (matrix) and dimeric (free) protective proteins. In the absence of ß-galactosidase and neuraminidase, this balance could be dependent on the protective protein concentration, its conformation (precursor or mature) and other physiological parameters, such as pH, ion concentration and incoming substrate. In COS-1 cells, transiently transfected with the protective protein cDNA, overexpression of the pro-

tein could lead to premature multimerisation in an early biosynthetic compartment. Preexisting intralysosomal matrices of the protective protein specifically attract ß-galactosidase and/or neuraminidase molecules and when both enzymes are associated the complex is most stable. By maintaining, in most cells and tissues, a surplus amount of lysosomal protective protein with respect to ß-galactosidase and neuraminidase the existence of "empty" matrices that can easily be converted into dimeric cathepsin A is ensured. The function of the free pool of protective protein/cathepsin A in lysosomes would be the degradation of proteins and/or bioactive peptides (the extensively purified toad skin enzyme and the human deamidase are found solely as homodimers of 94-100 kDa (Jackman *et al,* 1990; Simmons *et al,* 1980)). It remains to be determined whether the complexed pool participates in the digestion of macromolecules. If so, the target substrates for the peptidase and the two glycosidases should be rather small because of the steric hindrance within the complex.

The model implies that the signals for matrix formation are embedded within the structure of the protective protein only. Because of this, in G_{M1}-gangliosidosis fibroblasts neither lysosomal neuraminidase nor cathepsin A activities are severely affected. Instead, in galactosialidosis the lack of a preexisting matrix would cause the impaired combined glycosidase deficiency. Since mouse and chicken protective proteins have maintained the capacity to form matrices in lysosomes, it is logical that they are able to complement a deficiency of human protective protein in galactosialidosis cells. Nevertheless, multimers formed by these heterologous proteins might have a reduced affinity for human neuraminidase.

The model is in keeping with the idea that there may be more than one type of complex in lysosomes. This would explain why in glycoprotein preparations from human placenta ß-galactosidase activity is detected in two multimeric peaks, one with and one without neuraminidase (Verheijen *et al,* 1985). It is not excluded that other lysosomal enzymes temporarily associate to the matrix, perhaps in varying amount, composition and in a loose conformation. Evidence in support of a multicomponent matrix is the recent finding that complexes purified through ß-galactosidase affinity chromatography also contain small amounts of lysosomal α-N-acetylgalactosaminidase (Tsuji *et al,* 1989). It should be mentioned, however, that this enzyme is not deficient in galactosialidosis (van Diggelen *et al,* 1988). Another indication for a multicomponent complex comes from the fact that cathepsin A has been copurified with prolylcarboxypeptidase, although rigorous further isolation methods separated the two enzyme activities (Odya *et al,* 1978).

All the steps depicted in Figure 4, some of which speculative, can and will be tested in the near future. However, a thorough understanding of the determinants on the protective protein that are responsible for its different conformations awaits the resolution of its three-dimensional structure by means of crystallographic analyses. Concomitantly with these studies, homologous recombination in embryonic stem cells (Thomas and Capecchi, 1987) in combination with the "hit and run" procedure (Hasty *et al,* 1991) will create animals with targeted protective protein/cathepsin A mutations, that will allow a better understanding of the molecular mechanisms underlying galactosialidosis and might also be of help in the identification of the natural substrates of the protective protein. Until these investigations have been concluded the name given to the protective protein/cathepsin A/deamidase can be chosen rather freely and will presumably be based on one's historic ties with this fascinating multifunctional protein.

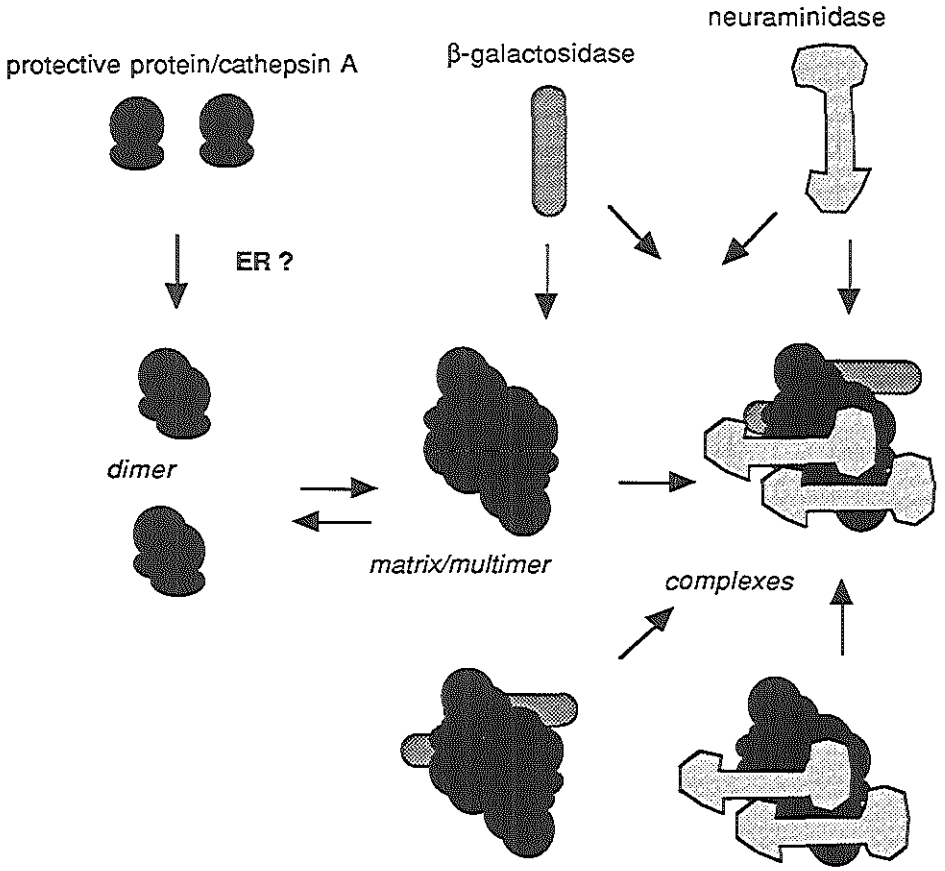


Figure 4. *Putative quaternary structures of the protective protein.*
It has been determined that the protective protein can form homodimers at precusor level, presumably within the ER. The model indicates that these homodimers may form matrices of lysosomal protective protein, to which subsequently other lysosomal enzymes could bind. No proteolytic processing steps are indicated in this figure. The quaternary structures of lysosomal β-galactosidase and neuraminidase and of the multimeric/complex forms depicted here are hypothetical.

2.3.2 References

Breddam, K. (1986). "Serine carboxypeptidases. A review." *Carlsberg Res Commun* **51**, 83-128.

Breddam, K. and Svendsen, I. (1984). "Identification of methionyl and cysteinyl residues in the substrate binding site of carboxypeptidase Y." *Carlsberg Res Commun* **49**, 639-45.

Coll, M., Guasch, A., Aviles, F. X. and Huber, R. (1991). "Three-dimensional structure of porcine procarboxypeptidase B: a structural basis of its inactivity." *Embo J* **10**, 1-9.

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

Dmochowska, A., Dignard, D., Henning, D., Thomas, D. Y. and Bussey, H. (1987). "Yeast KEX1 gene encodes a putative protease with a carboxypeptidase B-like function involved in killer toxin and α-factor precursor processing." *Cell* **50**, 573-84.

Doan, N. P. and Fincher, G. B. (1988). "The A- and B-chains of carboxypeptidase I from germinated barley originate from a single precursor polypeptide." *J Biol Chem* **263**, 11106-10.

Hasty, P., Ramirez, S. R., Krumlauf, R. and Bradley, A. (1991). "Introduction of a subtle mutation into the Hox-2.6 locus in embryonic stem cells." *Nature* **350**, 243-6.

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

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**, 12143-6.

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

Kawamura, Y., Matoba, T., Hata, T. and Doi, E. (1974). "Purification and some properties of cathepsin A of large molecular size from pig kidney." *J Biochem* **76**, 915-24.

Kawamura, Y., Matoba, T., Hata, T. and Doi, E. (1975). "Purification and some properties of cathepsin A of small molecular size from pig kidney." *J Biochem* **77**, 729-37.

Matsuda, K. (1976). "Studies on cathepsins of rat liver lysosomes. III Hydrolysis of peptides, an inactivation of anghiotensin and bradykinin by cathepsin A." *J Biochem* **80**, 659-69.

McDonald, J. K. and Barrett, A. J. (1986). Lysosomal carboxypeptidase A. In *Mammalian Proteases: a Glossary and Bibliography.* (New York: Academic Press), pp 186-91.

Miller, J. J., Changaris, D. G. and Levy, R. S. (1988). "Conversion of angiotensin I to angiotensin II by cathepsin A isoenzymes of porcine kidney." *Biochem Biophys Res Commun* **154**, 1122-9.

Miller, J. J., Changaris, D. G. and Levy, R. S. (1991). "Angiotensin carboxypeptidase activity in urine from normal subjects and patients with kidney damage." *Life Sciences* **48**, 1529-1535.

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

Odya, C. E., Marinkovic, D. V., Hammon, K. J., Stewart, T. A. and Erdös, E. G. (1978). "Purifiaction and properties of prolylcarboxypeptidase (angiotensinase C) from human kidney." *J Biol Chem* **253**, 5927-31.

Simmons, W. H. and Walter, R. (1980). "Carboxamidopeptidase: purification and characterization of a neurohypophyseal hormone inactivating peptidase from toad skin." *Biochemistry* **19**, 39-48.

Sorensen, S. B., Svendsen, I. and Breddam, K. (1989). "Primary structure of carboxypeptidase III from malted barley." *Carlsberg Res Commun* **54**, 193-202.

Thomas, K. R. and Capecchi, M. R. (1987). "Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells." *Cell* **51**, 503-12.

Tranchemontagne, J., Michaud, L. and Potier, M. (1990). "Deficient lysosomal carboxypeptidase activity in galactosialidosis." *Biochem Biophys Res Commun* **168**, 22-9.

Tsuji, S., Yamauchi, T., Hiraiwa, M., Isobe, T., Okuyama, T., Sakimura, K., Takahashi, Y., Nishizawa, M., Uda, Y. and Miyatake, T. (1989). "Molecular cloning of a full-length cDNA for human α-N-acetylgalactosaminidase (α-galactosidase B)." *Biochem Biophys Res Commun* **163**, 1498-504.

van Diggelen, O. P., Schindler, D., Willemsen, R., Boer, M., Kleijer, W. J., Huijmans, J. G. M., Blom, W. and Galjaard, H. (1988). "α-N-acetylgalactosaminidase deficiency , a new lysosomal storage disorder." *J Inher Metab Dis* **11**, 349-57.

Verheijen, F. W., Palmeri, S., Hoogeveen, A. T. and Galjaard, H. (1985). "Human placental neuraminidase. Activation, stabilization and association with β-galactosidase and its protective protein." *Eur J Biochem* **149**, 315-21.

Winther, J. R. and Breddam, K. (1987). "The free sulfhydryl group (Cys341) of carboxypeptidase Y: functional effects of mutational substitutions." *Carlsberg Res Commun* **52**, 263-73.

SUMMARY

Lysosomes are a class of heterogeneous acidic vesicles, surrounded by a single membrane, that play a key role in the degradation of a variety of macromolecules. Intralysosomal breakdown of substrates is carried out by a large number of hydrolases that work best at acid pH. Some of these enzymes require additional cofactors for full catalytic activity. Many of the hydrolases act in concert on a substrate, indicating the need for their organization into a higher order structure in order to speed up degradation. If, due to a mutation, a lysosomal protein is deficient a halt in the chain of catalytic events on a substrate occurs, resulting in its accumulation in the lysosome with consequent dysfunction of this organelle. In turn, this leads to the appearance of a lysosomal storage disorder.

Due to the vesicular structure of the lysosome both the macromolecular substrates as well as the lysosomal enzymes have to be transported to this compartment before degradation can occur. Several pathways that deliver lysosomal proteins to their destination have been described, the best characterized of which is the M6P targeting system that makes use of two related receptors recognizing soluble lysosomal proteins. In turn, the latter must contain determinants within their tertiary structure that allows their segregation from other non-lysosomal proteins. Lysosomes, their biogenesis and role in intracellular degradation, their constituents and the targeting of these proteins, are the subject of chapter 1.

In chapter 2 the experimental work is discussed, starting with an introduction (section 2.1) on the subject of this thesis: the protective protein. It is a lysosomal glycoprotein that was discovered through its deficiency in the metabolic storage disorder galactosialidosis. This rare autosomal disease is characterized by the severely reduced activities of the lysosomal enzymes β-galactosidase and neuraminidase. Galactosialidosis patients are heterogeneous with respect to their clinical manifestations, ranging in phenotype from a very severe early infantile form, that is fatal within childhood, to milder late infantile and juvenile/adult types. A high molecular weight complex in which β-galactosidase, neuraminidase and the protective protein are present, can be isolated from various tissues and species, but is absent in fibroblasts of galactosialidosis patients. Since in the latter cell type β-galactosidase is normally synthesized but rapidly degraded upon arrival in lysosomes it was postulated that in normal cells a high molecular weight complex of these three glycoproteins exists, which renders the two glycosidases active and stable within the acidic environment of the lysosome.

Biosynthetic labeling and immunoprecipitation studies in normal fibroblasts revealed that the protective protein is synthesized as a precursor of 54 kDa, that is proteolytically modified to a 32 kDa mature form. The cloning of the cDNA encoding human protective protein, described in publication 1 (section 2.2), tought us that the 54 kDa precursor polypeptide is actually processed into a two-chain form of 32 and 20 kDa subunits that are held together by disulfide bridges. Most strikingly, however, was the observation that the primary structure of the protective protein is homologous to those of yeast and plant serine carboxypeptidases. The catalytic triad Ser-His-Asp, responsible for the activation of the serine residue in these exopeptidases, is conserved in the protective protein and each amino acid is embedded in domains of high similarity to the other carboxypeptidases. Human and mouse protective proteins bind the serine protease inhibitor DFP, but only in their mature two-chain state, demonstrating that the 54 kDa precursor is a true zymogen (publication 3). Further studies, described in publication 4, provide direct evidence that the protective protein has carboxypeptidase activity and indicate that it might be identical to a previously partially characterized protease, cathepsin A. Using *in vitro* manipulated active site mutants of human protective protein we could demonstrate that in spite of their loss of catalytic activity these mutants retain their protective function towards β-galactosidase and neuraminidase. Thus, protective function and catalytic activity are distinct.

Analysis of the structure of the protective protein and of residues and domains that govern its correct folding as well as intracellular transport and processing is included in publication 5. It is demonstrated that human protective protein spontaneously forms homodimers already at precursor level and neutral pH. Furthermore, of the two oligosaccharide chains present on the human protein the one on the 32 kDa subunit acquires the M6P recognition marker, whereas the one on the 20 kDa subunit appears to be essential for the stability of the mature two-chain protein.

The gene encoding human protective protein is localized on chromosome 20 (publication 7). A natural mutation in this gene is found in two unrelated patients with the late infantile form of galactosialidosis (publication 6). Gelfiltration studies demonstrate that the mutant protective protein precursor, containing a Phe to Val amino acid substitution at position 412, cannot form homodimers. Moreover, the latter precursor is partially retained in the endoplasmic reticulum and those molecules that do reach the lysosomes are quickly degraded once they are cleaved into the 32/20 kDa mature form. We speculate therefore that dimerization of the protective protein might be an important event for its proper targeting and

stable conformation. Also, the presence of a residual lysosomal amount of protective protein in tissues of the two patients might account for their relatively mild clinical manifestations.

Our current model, which is described in section 2.3, is that the protective protein/cathepsin A, a pleiotropic member of the serine carboxypeptidase family, can either be assembled in complexes with β-galactosidase and/or neuraminidase, yet it might also function on its own. This equilibrium is dependent on as yet undefined factors but complex formation would be initiated by the protective protein. Once in complex its carboxypeptidase activity might be modulated by the other two enzymes. A recent report by the group of Dr. Erdös at the University of Chicago, extends the role of the protective protein even further. In an effort to purify a deamidase/carboxypeptidase released by human platelets this group came to the unexpected finding that their enzyme is identical at its N-termini to the protective protein 32/20 kDa chains. The deamidase/carboxypeptidase activity could be responsible for the local (in)activation of bioactive peptides.

The experiments described in this thesis have delineated a novel function for the protective protein, namely as a serine carboxypeptidase. They have also enlarged our knowledge on the biosynthesis, transport and proteolytic processing and on the structure of this protein, both in normal as well as in mutated state. In order to assess properly the different roles of this multifunctional lysosomal enzyme it is of importance to gain better insight into the mechanisms that control complex formation and the site(s) and mode(s)of association of the various components of the complex(es). In this respect the cloning of the cDNA encoding human β-galactosidase (publication 2) has allowed a more detailed study on its association with the protective protein, which apparently can already occur in an early biosynthetic compartment. Also, transgenic animals having targeted mutations in the protective protein gene will provide very useful model systems in which to study the contribution of the different activities to the aforementioned metabolic pathways.

SAMENVATTING

Lysosomen, een heterogene verzameling van zure vesikels omgeven door een enkel membraan, spelen een sleutelrol in de afbraak van een verscheidenheid aan macromoleculen. Intralysosomale degradatie wordt uitgevoerd door groot aantal hydrolasen met een zuur pH optimum. Sommige hebben niet enzymatische cofactoren nodig voor hun katalytische activiteit. Vaak breken lysosomale enzymen een substraat in een strikte volgorde af. Een versnelde afbraak van substraten wordt mogelijk gemaakt door een verzameling, bijvoorbeeld in een complex, van enzymen die na elkaar werken . Als door een mutatie een lysosomaal eiwit defect is stokt de afbraak van bepaalde substraten, hetgeen stapeling van deels gedegradeerde macromoleculen in lysosomen tot gevolg heeft. Uiteindelijk kunnen deze hierdoor niet meer naar behoren functioneren en ontwikkelt zich een lysosomale stapelingsziekte.

Vanwege de lysosomale structuur moeten zowel substraten als lysosomale enzymen via specifieke transport routes naar het lysosoom geleid worden. Een aantal van deze routes zijn recentelijk beschreven, de bekendste maakt gebruik van twee onderling verwante receptor eiwitten, die oplosbare lysosomale eiwitten herkennen en vervoeren. Op hun beurt bevatten lysosomale eiwitten specifieke signalen waardoor ze onderscheiden worden van de talloze andere eiwitten die niet bestemd zijn voor het lysosoom. Hoofdstuk 1 beschrijft lysosomen, hun ontstaan en hun rol in intracellulaire afbraak. Bovendien geeft het een overzicht van lysosomale eiwitten alsmede de manier waarop deze naar lysosomen vervoerd worden.

In hoofdstuk 2 wordt het experimentele werk besproken, beginnend met een inleiding over het 'protective protein', het onderwerp van dit proefschrift. Dit is een lysosomaal glycoproteïne dat is ontdekt doordat het defect is in de lysosomale stapelingsziekte galactosialidosis. Deze zeldzame aangeboren afwijking wordt gekenmerkt door een gereduceerde activiteit van de lysosomale enzymen β-galactosidase en neuraminidase. Galactosialidosis patienten worden klinisch ingedeeld in drie typen, al naar gelang de leeftijd waarop de ziekte zich openbaart en de hevigheid van de symptomen. De ernstigste of 'vroeg-infantiele' vorm van de ziekte leidt tot de dood kort na de geboorte, de andere twee vormen (de 'laat-infantiele' en 'juveniele/adulte' typen) zijn wat milder van aard.

Uit normale cellen en weefsels van verschillende organismen kan een complex geïsoleerd worden waarin zich β-galactosidase, neuraminidase en het 'protective protein' bevinden. Dit complex is afwezig in gekweekte fibroblasten van

galactosialidosis patienten. Bovendien wordt β-galactosidase in deze cellen wel normaal aangemaakt, maar eenmaal in het lysosoom wordt dit enzym versneld afgebroken. Deze waarnemingen leidden tot de hypothese dat normaliter in lysosomen een hoog moleculair complex van de drie voornoemde glycoproteïnen aanwezig is dat ervoor zorgdraagt dat β-galactosidase en neuraminidase actief en stabiel zijn in hun zure omgeving temidden van de andere hydrolasen.

Proeven waarbij radioactief aangemaakt 'protective protein' in normale fibroblasten werd geïmmunoprecipiteerd toonden aan dat het eiwit gesynthetiseerd wordt als een 54 kDa precursor vorm, die proteolytisch geklieft wordt tot een 32 kDa matuur eiwit. Klonering van het cDNA dat codeert voor het humane 'protective protein' (zie publicatie 1 in sectie 2.2) leerde ons dat de 54 kDa precursor feitelijk in twee ketens gesplitst wordt, van 32 en 20 kDa, die verbonden zijn door zwavel bruggen. Het meest verrassend was echter de ontdekking dat de aminozuur volgorde van het 'protective protein', voorspeld op basis van het gekloneerde cDNA, homoloog is aan die van gist en plante serine carboxypeptidasen. Drie specifieke aminozuren, Ser-His-Asp, die in al deze exopeptidasen verantwoordelijk zijn voor de activering van het serine residue in het katalytische centrum, zijn bewaard gebleven in het 'protective protein'. Bovendien liggen ze verankerd in van elkaar gescheiden domeinen in het eiwit, die qua aminozuur volgorde sterk lijken op overeenkomstige domeinen in de andere carboxypeptidasen. Het humane 'protective protein' is nauw verwant aan het overeenkomstige muize eiwit (publicatie3). Beide glycoproteïnen binden DFP, een specifieke remmer van serine proteasen, maar doen dit slechts in hun mature 32/20 kDa vorm. Dit wijst erop dat het 54 kDa polypeptide de zymogeen vorm van het 'protective protein' is (publicatie 3). Verdere experimenten, beschreven in publicatie 4, bewijzen direct dat het 'protective protein' serine carboxypeptidase activiteit heeft en duiden erop dat dit eiwit weleens identiek zou kunnen zijn aan lysosomaal cathepsine A, een enzym dat vroeger gedeeltelijk gekarakteriseerd is. *In vitro* gemaakte mutante 'protective proteins', die geen katalytische activiteit hebben, zijn nog steeds in staat tot bescherming van β-galactosidase en neuraminidase. Dit toont aan dat katalytische activiteit en de beschermende rol twee aparte functies zijn, verenigd in één eiwit (publicatie 4).

In publicatie 5 wordt de structuur van het 'protective protein' verder ontleed en worden signalen/domeinen op het eiwit gelokaliseerd die mede verantwoordelijk zijn voor goede eiwit vouwing, transport naar het lysosoom en proteolytische modificatie. We tonen aan dat het 'protective protein' reeds als precursor en bij neutrale pH spontaan homodimeren vormt. Van de twee suikerketens aan-

wezig op het precursor eiwit wordt de groep op het 32 kDa polypeptide gefosforyleerd, een modificatie die dient voor transport van het eiwit naar het lysosoom. De aanwezigheid van de suikerketen op de 20 kDa subeenheid is hoogstwaarschijnlijk belangrijk voor de stabiliteit van het mature 'protective protein'.

Het gen coderend voor het humane 'protective protein' is gelokaliseerd op chromosoom 20 (publicatie 7). Een mutatie in dit gen is gevonden in twee niet verwante galactosialidosis patienten, die beide lijden aan de milde, 'laat-infantiele' vorm van de ziekte (publicatie 6). De experimenten laten zien dat het gemuteerde 'protective protein' een aminozuur substitutie heeft ondergaan, waarbij op positie 412 in het eiwit de normale phenylalanine vervangen is door valine. Hierdoor kan het gemuteerde eiwit, in tegenstelling tot de normale precursor, geen homodimeren meer vormen. Bovendien wordt de gemuteerde precursor gedeeltelijk vastgehouden in het endoplasmatisch reticulum. Dié precursor moleculen die dit compartiment verlaten kunnen het lysosoom bereiken, maar eenmaal daar aangekomen en gesplitst in de 32/20 kDa mature vorm wordt het gemuteerde eiwit versneld afgebroken. Dit doet ons veronderstellen dat de vorming van dimeren door het 'protective protein' belangrijk is voor een stabiele conformatie, alsmede voor een correct en efficient transport van dit eiwit naar het lysosoom. De aanwezigheid van een sterk verlaagde hoeveelheid 'protective protein' in lysosomen van de twee hier beschreven patienten zou een verklaring kunnen geven voor de relatief milde vorm van galactosialidosis die bij hen gevonden wordt.

In sectie 2.3 wordt een nieuw model gepresenteerd, gebaseerd op de recentelijk verworven kennis van het 'protective protein', of cathepsine A. Het stelt voor dat dit eiwit alleen kan functioneren, als dimeer, maar dat het ook in eiwitcomplexen met de enzymen β-galactosidase en/of neuraminidase kan voorkomen. Dit evenwicht is afhankelijk van tot nu toe onbekende factoren, maar de vorming van complexen wordt door het 'protective protein' geïnitieerd. De carboxypeptidase activiteit van het eiwit zou, eenmaal in een complex, gereguleerd kunnen worden door de andere twee enzymen. Kort geleden heeft de groep van Dr. Erdös van de universiteit van Chicago een artikel gepubliceerd, dat een extra rol van het 'protective protein' suggereert. Deze auteurs vonden dat een deamidase/carboxypeptidase, dat zij gezuiverd hadden uit bloedplaatjes, dezelfde aminoterminale aminozuur volgorde had als de 32 en 20 kDa ketens van het 'protective protein'. De deamidase/carboxypeptidase activiteit zou verantwoordelijk kunnen zijn voor de plaatselijke (in)activering van bioactieve peptiden.

De experimenten beschreven in dit proefschrift hebben een nieuwe functie toegevoegd aan het 'protective protein', namelijk die van serine carboxypeptidase.

Ook hebben zij bijgedragen tot een beter inzicht in de biosynthese, transport en proteolytische modificatie en structuur van dit eiwit, zowel in normale als in gemuteerde vorm. Om de verschillende rollen van het 'protective protein' goed te kunnen beoordelen is het nodig dat precies uitgezocht wordt welke mechanismen en factoren van invloed zijn op complex vorming en hoe en op welke intracellulaire plaats(en) de verschillende componenten van de complexen met elkaar kunnen associëren. De klonering van het cDNA coderend voor humaan β-galactosidase (publicatie 2) heeft er toe geleid dat we konden vaststellen dat dit enzym al in een vroeg stadium en op precursor niveau kan associëren met het 'protective protein'. Transgene dieren met gerichte mutaties in het gen coderend voor het 'protective protein' zullen ook zeer waardevol blijken te zijn in het definiëren van de verschillende activiteiten van dit multifunctionele lysosomale enzym.

NAWOORD

Door de jaren heen ben ik, hier in Rotterdam, door een aantal mensen in meer of mindere mate geholpen met mijn onderzoek. Mijn inmiddels in kleine kring befaamde zeef-geheugen zal ertoe bijdragen dat ik sommigen van hen vergeet te noemen, dus bij deze een dankwoord aan een ieder die heeft bijgedragen aan het tot stand komen van dit proefschrift.

Mijn promotor, professor Galjaard, ben ik zeer erkentelijk voor zijn in mij gestelde vertrouwen en de door hem geboden mogelijkheden dit onderzoek te kunnen voltooien. Ook de andere leden van de promotiecommissie, de professoren Bootsma, Borst en Hasilik dank ik voor het kritisch lezen van dit manuscript en hun goede suggesties cq. kanttekeningen.

Mijn co-promotor, Dr. Alessandra d'Azzo, kan ik nauwelijks genoeg bedanken. Sandra, ik weet dat je niet zoveel om voetbal geeft, maar toch vind ik je de Johan Cruijff van de wetenschappelijke begeleiders ! Bedankt voor je nooit aflatende inzet, je inzicht en je vriendschap. Ik heb ondanks mijn koppigheid op alle vlakken van je geleerd en daardoor extra van deze jaren genoten. Nynke Gillemans, met jou en Sandra heb ik het langst samengewerkt. Volgens mij ben je enig in je soort, goed in wat je doet, fanatiek en onzelfzuchtig, wat een contrast met mij zou je bijna zeggen. Bedankt dat je mijn paranimf wilt zijn. Hans Morreau, die laatste opmerking geldt ook voor jou. Nu weet ik bovendien waar hoeven zich bevinden, wat knorrecte sokken zijn en hoe in units te meten. Dat is toch de essentie van het leven. Erik Bonten, bedankt voor de les, behalve in propjes gooien en Xiao Yan Zhou, you too, sje sje very much. Aarnoud van der Spoel, Robbert Rottier, Amelia Morrone en Arno van 't Hoog, helaas was er slechts een korte tijd om met jullie samen te werken, maar het was er niet minder prettig om.

En de anderen ? Gerard Grosveld, mochten de goden huur betalen om in het linkerbeen van Brian Roy te kunnen wonen, dan hebben ze het optrekje in jouw hoofd (en handen) nog niet bezichtigd ! Mijnheer "de Generaal", bedankt voor je advies, steun en goede gesprekken. Van Gerard's groep ben ik Dies Meijer en Marieke von Lindern erkentelijk voor hun hulp, met name 'in the early days', en natuurlijk Sjozèf van Baal waar het de EHBO (eerste hulp bij onbegrip) met de computer betreft. Harold Günther, jou bedank ik voor het gouden advies om een Macintosh aan te schaffen om daarmee het proefschrift te schrijven en voor je hulp bij de afwerking van dit manuscript. Rob Willemsen, jij hebt al het goede LM/EM werk gedaan, mijn dank voor deze bijzonder waardevolle bijdrage; Pim Visser, jij hebt een groot deel van de figuren getekend en Mirko Kuit, jij hebt bij mijn weten

vrijwel alle fotografie (behalve LM/EM) die hier gepubliceerd wordt verzorgd. Dank jullie wel. Jeanette Lokker, hartelijk bedankt voor al het type werk aan de artikel-units en, samen met Rita Boucke, voor de 'reisverslagen' en andere gesprekken. Van de 24ste verdieping ben ik Henk Janse erkentelijk voor advies bij en, soms, oplossingen voor het meten van enzymactiviteiten. De rest van de medewerkers van de vakgroep Celbiologie en Genetica, wetenschappelijk en niet-wetenschappelijk, van lab naar lab, van werkplaats naar keuken, van kweekhok naar bestellingen: bij deze een ieder bedankt voor zijn/haar bijdrage.

Tot slot is er Martine, maar jij staat voorop...

CURRICULUM VITÆ

27 juni 1961	geboren te Blaricum
juni 1979	eindexamen VWO, 'Het Wageningsch Lyceum'
september 1979	aanvang studie Moleculaire Wetenschappen
mei 1984	kandidaatsexamen Moleculaire Wetenschappen
november 1986	doctoraal examen Landouwwetenschappen, studierichting Moleculaire Wetenschappen, Landbouw Universiteit Wageningen hoofdvakken: Moleculaire Biologie, LUW, Biochemie, Universiteit van Amsterdam bijvak: Virologie, LUW
januari 1987	aanvang promotie-onderzoek als AIO op de afdeling Celbiologie en Genetica, bij Prof. Dr. H. Galjaard en Dr. A. d'Azzo
oktober 1991	aanvang EMBO 'long term fellowship', in het 'Laboratory of Gene Structure and Expression', NIMR, Mill Hill, London, England, bij Dr. F. Grosveld