INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road. Ann Arbor. MI 48106-1346 USA
313 761-4700 800 521-0600

Order Number 9304727

Characterization of the human prosaposin locus

Rorman, Efrat Gavrieli, Ph.D.

City University of New York, 1992

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

Characterization of the Human Prosaposin Locus

by

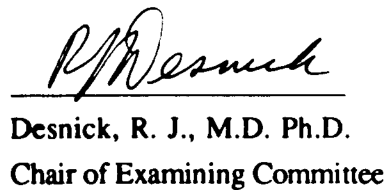
Efrat Gavrieli Rorman

A dissertation submitted to the Graduate Faculty in Biomedical Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1992

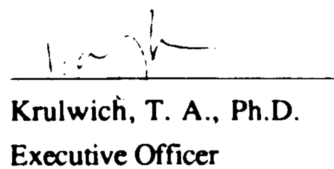
This manuscript has been read and accepted by the Graduate Faculty in Biomedical Sciences in satisfaction of the dissertation for the degree of Doctor of Philosophy.

6/16/92
Date



Desnick, R. J., M.D. Ph.D.
Chair of Examining Committee

6/29/92
Date



Krulwich, T. A., Ph.D.
Executive Officer

Grabowski, G. A., M.D.

Schuchman, E. H., Ph.D.

Bishop, D. F., Ph.D.

Gatt, S., Ph.D.

Supervisory Committee

The City University of New York

Abstract

Characterization of the Human Prosaposin Locus

by

Efrat Gavrieli Rorman

Adviser: Gregory A. Grabowski, M.D.

Many lysosomal hydrolases involved in the catabolism of complex sphingolipids require interactions with activator proteins for optimal hydrolytic activity. Such a small, heat stable, glycoprotein was described to enhance the hydrolytic rate of glucosyl ceramide by acid β-glucosidase *in vitro*. Abnormal enzymatic activity of acid β-glucosidase causes Gaucher disease. A similar disease is caused by the deficiency of the activator protein of acid β-glucosidase, saposin C. This work describes the cloning and characterization of a cDNA encoding four saposins, as well as the genomic organization of this locus. Based on the molecular data the following is presented: a model for the evolution of this gene, as well as the steps involved in maturation from prosaposin to four proteins and their functions.

Screening of two human cDNA libraries using mixed oligonucleotides encoding amino acid sequence of saposin C identified a 2767 bp long, full-length cDNA. This includes 38 bp of 5' untranslated region, a 1672 bp open reading frame and 1157 bp of 3' untranslated region. The coding region included a hydrophobic signal peptide and four, highly homologous, protein domains termed saposin A, B, C, and D which are proteolyticaly clipped from a precursor protein, prosaposin. An ~ 80% amino acid similarity was found between prosaposin and a rat sertoli cell encoding sulfated glycoprotein-1. These results suggested that the human and rat genes derived from a single ancestral precursor. To examine this hypothesis the genomic organization of the human gene is described. A region of 19,985 bp-long which includes the four protein domains and the carboxy-terminus of the signal peptide was sequenced. This sequence comprised 13 exons ranging in size from 59 to 119 bp. Each of the intron/exon boundaries conformed to the gt/ag rules. Only two *Alu* consensus sequences were found. The location of the introns in respect to

the exonic sequence was used to explain the evolution of this locus by duplication events and a double crossover.

The function of prosaposin is not known yet but, it might be involved in glycolipid binding and have a role in cell differentiation processes.

Dedication

This work is dedicated to the memory of my good friend Lin Shafarman. Her strength, sense of humor and positive attitude while struggling with Gaucher Disease and life will remain a source of inspiration. I am sorry she could not see this work completed.

Acknowledgments

I would like to thank Dr. Gregory Grabowski for his efforts to train me as a thinking as well as a bench scientist. I have learned much from his scientific knowledge and perspective.

It was a pleasure to work with Dr. Vladimir Scheinker on parts of my projects, and I thank him for his input.

I wish to thank Dr. Marie Grace for fruitful discussions which I believe are an important part of science.

Finally I want to thank my husband, Dan, who has always been with me during these years and has taken a major part in our family and genetic projects. Hadas, our daughter, has helped me to keep perspective outside science.

Table of Contents

Title Page i

Approval Page ii

Abstract iii

Dedication v

Acknowledgment vi

Table of Contents vii

List of Figures and Tables ix

Introduction

A. Gaucher Disease and Acid β-Glucosidase 1

B. Activator Proteins for Acid β-Glucosidase 8

Experimental Results

Molecular Cloning of a Human Co-β-Glucosidase cDNA: Evidence That Four Sphingolipid Hydrolase Activator Proteins Are Encoded by a Single Gene in Human and Rats 14

Structure and Evolution of the Human Prosaposin Chromosomal Gene. 33

Conclusions
A. Genomic Organization and Evolution of the Prosaposin Gene 56
B. Post-translational Processing 58
C. Function of the Four Saposins and the Precursor Protein 60
References 66

List of Figures and Tables

Figure 1 Schematic diagram of the human prosaposin cDNA and the sertoli rat Sulfated Glycoprotein-1 cDNA. 22

Figure 2 Amino acids similarity and conservation of the four protein domains in the human proactivator and rat Sulfated Glycoprotein-1 cDNA. 26

Figure 3 Schematic diagram of the human and murine prosaposin cDNA. 37

Figure 4 Genomic organization, sequencing strategy and predicted restriction map for the prosaposin chromosomal gene. 40

Figure 5 Nucleic acid sequence of the prosaposin chromosomal gene. 43

Table 1 Intron/Exon boundaries of prosaposin gene. 47

Figure 6 The alignment of the intron positions in the human saposin A, B, C and D amino acid sequences. 48

Figure 7 An evolutionary model to account for the maintenance of the intronic positions in the amino acid sequences of the saposins. 51

Table 2 Mutations causing saposin deficiency diseases. 62

Introduction

Abnormal enzymatic hydrolysis of glucosyl ceramide by acid β-glucosidase [N-acyl-sphingosyl-1-0-β-D-glucoside: glucohydrolase (E.C.3.2.1.45)], is the cause of Gaucher disease [1,2]. Acid β-glucosidase is the membrane-associated lysosomal hydrolase which requires activator protein(s), termed saposin(s), for cleavage of glucosyl ceramide to β-glucose and ceramide [3]. *In vitro* assays to determine acid β-glucosidase activity, are usually performed in the presence of detergents which circumvent the effect of the activator proteins. In such assays acid β-glucosidase levels in tissues of Gaucher disease patients has 2-20% activity as compared to cells from normal individuals [1,2,4-7]. In comparison, in the lysosomal environment small activators, saposins, are required for efficient cleavage of glucosyl ceramide. Although, one activator protein, saposin C, was described in 1971 as an activator "factor", its physiologic role was in question until recently. In three cases of Gaucher disease-like variants the patients had nearly normal acid β-glucosidase activity *in vitro* as measured in the presence of a detergents, but no cross-reactive-immunologic-material, CRIM, with rabbit antibodies raised to human saposin C [8,9,10].

Much of the enzymology and the molecular biology of Gaucher disease has been elucidated in the last two decades. However, the molecular biology of the saposins have not been extensively explored. The focus of these studies is on the molecular organization of the unique locus which was found, during this work, to give rise to the four mature saposins .

A. Gaucher Disease and Acid β-Glucosidase

Gaucher disease, the most prevalent lysosomal storage disease, is inherited as an autosomal recessive trait [11]. In this disease there is a progressive accumulation of acid β-glucosidase's major natural substrate, glucosyl ceramide, primarily within lysosomes of the monocyte-macrophage system. Clinically, there are three major types of Gaucher disease:

Gaucher disease Type 1 is a chronic, non-neuronopathic disease that may first manifest from birth to old age. It has an increased frequency among the Ashkenazi Jewish population, q ~ 0.04 [12,13], but also occurs in Blacks, Hispanics and other ethnic groups [14]. The major characteristics of Gaucher disease Type 1 are hepatosplenomegaly, bony disease and the absence of neuronopathic manifestations [1,15]. There is a remarkable variability in the phenotype of Type 1; i.e., in the age at clinical onset and degree of involvement of bone marrow, liver, spleen and bony complications .

Gaucher disease Type 2 is an infantile, acute neuronopathic disease. This variant is a rare, panethnic variant with an infantile onset, typically before 6 months, which is fatal by two years of age [15]. This disease is characterized by hepatosplenomegaly, failure to thrive and progressive psychomotor degeneration.

Gaucher disease Type 3 is a juvenile, subacute neuronopathic disease. This variant was first described in Swedish individuals of Norrbottnian descent and occurs in high frequency in this population [16-18], although it is panethnic. Visceral organs and the central nervous system are involved but it is distinguished from Type 2 by its later onset and slower rate of neuronopathic progression.

1. Properties of Normal β-Glucosidase

In 1965 Brady et al. [1] and Patrick [2] described the common cause of the three clinically distinguished Gaucher disease types as a deficiency in the enzymatic activity of acid β-glucosidase. Since then, much research has been focused at understanding the biochemical and molecular basis of each Gaucher disease variant, particularly because of the importance for defining patient prognosis and for pre-natal diagnosis. Additionally, these efforts have provided insight into this enzyme's mechanism of action and to correcting the defective enzyme found in Gaucher disease patients. In the following, the biochemical and molecular characteristics of the normal enzyme will be presented first and then, defective forms will be discussed.

The calculated molecular weight of acid β-glucosidase as predicted from its cDNA is 55.4 kDa [19]. This value is in agreement with the deglycosylated molecular weight of pure acid β-glucosidase obtained by SDS-PAGE analysis [20]. The post-translational processing of the protein has been evaluated by several laboratories. As is typical for a protein in this secretory path, acid β-glucosidase is synthesized and transported through the endoplasmic reticulum membrane, at the same time its signal peptide is clipped [21]. No other proteolytic processing occurs during its maturation and transport to the lysosome [22]. Pulse-chase processing experiments with normal fibroblasts showed that the enzyme is first detected after passing through the membrane of the endoplasmic reticulum as a glycosylated 63-64 kDa form. Passage through the Golgi apparatus results in remodeling of the oligosaccharides to a complex type and the formation of a 69 kDa sialyated form. The half time of this process is about 3 hours [22]. The final glycosylated form of the enzyme in fibroblasts has a molecular weight of 58-59 kDa [20]. Glycosylation studies of the mature porcine enzyme showed occupancy of four of the enzyme's five possible N-glycosylation sites [21]. Expression studies using site-directed-mutagenesis of the human cDNA also showed that only four out of the five potential glycosylation sites are normally occupied [23]. Unlike most soluble lysosomal enzymes targeting of the enzyme to the lysosome is independent of the mannose-6-phosphate receptor system [24,25]. This was concluded from experiments using fibroblasts from I-cell disease patients whose fibroblasts are genetically deficient in the enzyme N-acetylglucosamine-1-phosphotransferase [24,26] which is required for the synthesis of a mannose-6-phosphate group. Most lysosomal proteins are secreted into the media of I-cell fibroblasts due to the lack of a mannose-6-phosphate group, whereas acid β-glucosidase and acid phosphatase are targeted to the lysosome even in I-cells [25]. In addition, as determined by content analysis of acid β-glucosidase carbohydrate groups, there is no phosphate group on any of its sugar moieties [27]. The above findings suggest that an alternative targeting mechanism is used to direct acid β-glucosidase to the lysosome.

Acid β-glucosidase cleaves the β-glucoside linkage of glucosyl ceramide [1,2,3] as well as synthetic, water soluble substrates including p-nitrophenyl-β-D-glucoside and 4-methyllumbelliferyl β-D-glucopyranoside (4MU-Glc) [20,28,29]. The

enzyme has been isolated from a variety of tissues, including human placenta, spleen and liver, as well as bovine brain and rat liver [for review see 30]. Extraction and solubilization of the enzyme from the lysosomal membrane requires the use of detergents [18,31]. Organic solvents are required to delipidate the enzyme in order to make it partially water soluble [22,32,33].

For its activity the delipidated enzyme requires negatively charged lipids (e.g. phosphatydilserine or gangliosides) or detergents. In addition, when no detergents are present the enzyme requires one or two naturally occurring acidic activator proteins (see section B). Many *in vitro* assays to evaluate enzymatic activity have been devised for use with purified enzyme or crude cellular extracts. Acid β-glucosidase has a pH optimum of 4.2 to 6.0 depending on the precise assay conditions. However, in most systems pH 5.5-5.7 provides maximal hydrolytic rates for artificial or glucosyl ceramide substrates [for review see 30]. From pH optimum profiles, the catalytic rate constant is dependent on the ionization state of two active site residues with pKa(app) of about 4.5 and 6.5 [34]. Affinity labeling of purified enzyme with an active site directed inhibitor, conduritol B epoxide, and isolation of labeled peptides was used to localize the active site to the carboxy-terminal 1/3 of the enzyme. The binding site for conduritol B epoxide was identified as Asp^{443} [35]. Studies with various enzyme modifiers suggested that the acid β-glucosidase active site has at least three binding sites with different specificities [34,36]: 1.) A glycon binding site which is a hydrophilic pocket that recognizes β-glucosyl moieties, and conduritol B epoxide, 2.) An aglycon binding site that is hydrophobic and has an affinity for the fatty acid acyl chains of glucosyl ceramide and modulates V_{max}, and 3.) A site which interacts with sphingosyl moieties and modulates K_m [34,36]. Additional site(s) on the enzyme are required for its interaction with the activator protein(s) or negatively charged lipid activators [36 and references therein]. The organization of the active site is similar to binding sites for globotriaocylceramide or phospholipids to their respective binding proteins, saposin B [37] or phospholipid transfer protein [38] . These similarities suggest a common structure of binding sites for lipids which has been preserved during development of lipid degradation pathways.

Comparison of the amino acid sequence of acid β-glucosidase between the mouse and human proteins showed 86% identity [39]. Within the carboxyl-terminal of the enzyme there are stretches of 100% sequence identity (from amino acid 315 to 375, 377 to 406, and 411 to 434). However, the immediate region of the suggested binding site, from amino acid 440 to 451 has less than 50% identity between the two proteins although, Asp^{443} and Leu^{444} are conserved [39]. It is still possible that the overall similarity at the carboxy-terminal region compensate for the local differences.

Studies at the molecular genetic level mapped the human acid β-glucosidase gene to chromosome 1q21 [40,41]. Analysis of the structural human gene revealed that it comprises 11 exons and 10 introns, a total of 7.6 kb. In comparison, a highly homologous pseudogene, located 16 kb from the structural one, is only 5.8 kb, due to intronic deletions. There is a 96% nucleotide sequence homology between the exonic region of the structural gene and the corresponding sequence of the pseudogene [42]. In addition there are numerous single base differences in the exons of the structural gene and the corresponding sequence on the pseudogene [42]. This will be discussed later as a potential source of mutations in Gaucher disease patients (see Introduction section A2). Characterization of the promoter for acid β-glucosidase showed a TATA and a CAAT like boxes upstream of the translation initiation site [43]. Functional analysis using the bacterial CAT gene as a reporter revealed very low, if any, activity of the pseudogene promoter as compared to the structural gene [42]. Using the CAT reporter gene and lymphoid cells, which have low levels of acid β-glucosidase mRNA, poor promoter activity was measured, suggesting tissue specific transcriptional regulation[43]. Studies of acid β-glucosidase mRNA from HeLa cells and skin fibroblasts of normal and Gaucher disease individuals showed three RNA species, 5.6, 2.5 and 2.0 kb in length [43,44]. The longest transcript was thought to represent an unspliced genomic transcript, while the shortest one was shown by S1 nuclease analysis to have arisen from alternative transcription initiation sites as well as alternative polyadenylation [43]. The 2.5 kb seems to be the full length mRNA, as demonstrated by primer extension studies [45]. In the longest cloned cDNAs, there is a stretch of 280-300 bases in the 5' untranslated region followed by two in-frame ATG translation initiation codons located 57 bases apart. The most 5' ATG does not exhibit a

consensus Kozak sequence while the second one does. At the 3' untranslated region, a stretch of 550 bases followed by 100 bases of a poly (A) tail [45] is present. There is no consensus polyadenylation signal (AAUAAA), but rather a CUUAAA hexamer, positioned upstream to the 3' end of the mRNA [43,44].

2. Acid ß-Glucosidase Mutant Forms

Enzymatic studies of tissues from Gaucher disease patients had shown extensive heterogeneity between and among the different types of Gaucher disease [46]. It was thought that analyses at the DNA level would provide insight into the etiology of the different phenotypes of Gaucher disease. Today, about 30 different causative mutations have been defined by various researchers, most of them are rare or private point mutations. Only two mutations have high frequency in the Gaucher disease population, they account for about 60-70% of all the Gaucher disease patients alleles [47].

About 50% of the total alleles in Gaucher disease patients, contain a mutation which leads to a substitution of serine for asparagine at amino acid 370 (N370S). Interestingly, this point mutation occurs only in Gaucher disease type 1 even when present as a heteroallele in affected patients [47,48,51]. Homozygous, N370S, individuals may have a milder phenotype [48-50]. Expression of mutant cDNA containing this mutation resulted in a kinetically altered but stable protein, with a 2-5 fold decrease in CRIM-specific activity as compared to the normal allele [51]. It was also reported that under specific assay conditions the N370S mutated allele is poorly stimulated by saposin C but it can be stimulated by phosphatidylserine to a greater extent than the normal allele [52].

The second common mutation accounts for more than 10% of the total Gaucher disease patients alleles. This is a substitution of leucine to proline at amino acid 444 (L444P). This mutation has been found in all types of Gaucher disease patients but homozygous individuals were found to have Gaucher disease type 3 [53, 54]. Expression of this mutation resulted in an unstable protein with about 20 fold decreased catalytic rate constant, as compared to the normal [51,52].

In the Norrbottnian and the Ashkenazi-Jewish population almost all of the mutated alleles have been defined. All the Norrbottnian type 3 Gaucher disease patients that have been studied, are homozygous for the L444P substitution [53,54]. From extensive pedigree analyses and population studies, the commonalty of this mutation has been shown to be due to a founder effect [53-55].

In the type 1 Ashkenazi Jewish population 65-75% of the mutated Gaucher disease alleles are N370S, whereas 6% are L444P. A G insertion at DNA base number 84 accounts for about 12%, and a G to A substitution causing the aberrant splicing of exon 2 accounts for 4.5% [56,57]. Thus, about 88-98% of the Gaucher disease alleles in the Ashkenazi-Jewish population has been identified.

A unique and interesting phenomenon found among the mutated forms of acid β-glucosidase is the duplication of pseudogene base changes, either singly or in groups, in cDNA clones from patients with Gaucher disease. These relatively frequent occurrences suggest that the rearrangement between the structural gene and the pseudogene contributes to the pathogenesis of Gaucher disease. These rearrangements have probably been facilitated by the close proximity (16 kb) and high exonic homology between these loci. An example of this event is the L444P mutation where the normal sequence of the structural gene has a T at position 6433 while the mutated form as well as the pseudogene contain a C at this position. A more dramatic demonstration of this phenomenon is the occurence of complex alleles. In these cases several point mutations are found on the same mutated allele of the structural gene, all of them correspond to the sequence from the pseudogene. An example is an allele found in a Gaucher disease type 3 which contains three amino acid substitutions: an Asp to His at amino acid 409, a Leu to Pro at amino acid 444, an Ala to Pro at amino acid 456, and a silent point mutation 35 bases downstream from base number 6433. These changes account for all the differences between the structural gene and the pseudogene in this region [53]. One example of a fusion protein was reported where the 5' end is the structural gene and the 3' end is the pseudogene [47]. Such rearrangements can probably occur by a mechanism of gene conversion or homologous recombination.

Although, much has been learned about the molecular defects in the acid β-glucosidase alleles which cause Gaucher disease, there is a very limited correlation between the known genotypes and the phenotypic heterogeneity observed in Gaucher disease patients. As described above, mutation N370S correlates with Gaucher disease type 1, and homozygous patients have a milder phenotype [48-50]. This allele is considered as a protective allele, if found in a patient the diagnosis is non-neuronopathic disease. When the L444P allele is found in the homozygous state the diagnosis is type 3 Gaucher disease [53,54]. There is no other clear correlation between a known genotype and a specific phenotype. Moreover, individuals with the same genotype do not necessarily show similar manifestation of the disease. This observation may suggests that another locus is involved in determining the various phenotypes of Gaucher disease. The only other protein that is known to participate in the degradation of glucosyl ceramide is saposin C, which can be considered as a candidate gene for modulating the phenotype in Gaucher disease.

B. Activator Proteins for Acid β-Glucosidase

Similar to other sphingolipids, the enzymatic hydrolysis of glucosyl ceramide *in vivo* requires an activator protein. In 1971 Ho and O'Brien [58] first described a heat stable factor which enhanced acid β-glucosidase activity. Since then this activator has been referred to by numerous names - factor P, co-β-glucosidase, SAP-2, heat stable factor, and saposin C, with the latter term becoming more generally used. More recently another protein, saposin A, was shown to activate acid β-glucosidase *in vitro* [59]. Only the cloning of the cDNA revealed that these two saposins as well as two others were encoded in tandem by the same transcript [60-63].

I. Saposin C

Saposin C is a small, heat stable, acidic glycoprotein. The molecular weight of the mature protein is 8.5 kDa, and it contains one N-glycosylation site which is occupied by carbohydrates of either the complex or the hybrid type [64].

As measured *in vitro*, saposin C can increase the hydrolytic rate of both the natural and the water soluble substrates of acid β-glucosidase. It can also increase *in vitro* hydrolytic rate of galactosyl ceramide by β-galactosidase, and some researchers have reported enhancement of sphingomyelinase activity [65-67]. All these enzymes participate in degradation of sphingolipids to ceramide. In order to achieve activation of acid β-glucosidase, saposin C requires the presence of an acidic phospholipid [68], the best of which is phosphatidylserine [69,70]. However, when the enzyme and saposin C are both well delipidated and no lipids are added back, there is very little activation of acid β-glucosidase by saposin C alone [69]. Addition of phosphatidylserine, or lipids that were extracted from the enzyme, restores the sensitivity of the enzyme to the effects of saposin C [69]. The primary effects of the activator are a two to five fold increase of V_{max}, a reduction of the reaction's optimal pH from the 5.5-6.0 range to 4.5 and a small decrease of K_m [32, 69].

Sphingolipid activator proteins such as saposin B, have been shown to form a 1:1 mole ratio complex with their substrates and present them to the respective enzymes [71,72]. In contrast, saposin C binds directly to acid β-glucosidase [69], apparently in conjunction with an acidic phospholipid. Evidence for the formation of the ternary complex of enzyme, activator and phospholipid includes affinity chromatography where immobilized enzyme bound saposin C [69,74], and the demonstration that saposin C protected the enzyme against inactivation by thiol reagents and by heat [69]. Additionally, the ternary complex exhibits a lowered K_m for its substrate as well as a higher V_{max} compared with the enzyme in the presence of phospholipids only [75,76]. Berent and Radin [69] suggested that the substrate may approach the enzyme by lateral diffusion in the lysosomal membrane, so that the saposin C does not play a role as a carrier protein but, together with acidic lipids, they make the enzyme accessible to the substrate. Based on experimental

data, three different modes of interaction of the enzyme with the activator at various concentrations of acidic lipids were proposed. At low lipid concentration, which probably mimics *in situ* conditions, the lipids help to present the enzyme so that it can bind and be activated by saposin C. In higher lipid concentration there is a competition for the binding site of the activator protein. When the lipids bind to this site on the enzyme they activate acid β-glucosidase but to a lesser extent than saposin C. At very high lipid concentration the enzyme is bound to a micellar form of the lipids and thus is less active [69]. The use of monoclonal antibodies raised against acid β-glucosidase supports the theory of direct conformational effects of saposin C on the enzyme [77]. In these studies a specific monoclonal antibody mimicked the enzyme activation by saposin C. There was no requirement for phosphatidylserine in order to obtain binding of the monoclonal antibody to the enzyme but, higher activation rates were obtained only in the presence of phosphatidylserine. Addition of saposin A to the enzyme, phosphatidylserine, and saposin C or the monoclonal antibody mixture provided a small enhancement of hydrolysis, though the major effect was a decrease in the total concentration of activators required for maximal enzyme activation [77]. In order to characterize the properties of saposin C that are important for its binding and activation, it was shown that the same degree of activation by saposin C takes place after deglycosylation of saposin C [64] and extraction with an organic solvent [78]. It was also demonstrated that a peptide isolated after a CNBr cleavage of saposin C had the same degree of activation as saposin C [79]. A preliminary study showed that a short synthetic peptide composed of amino acids 41-80 of the mature saposin C keeps the effect of the activator protein [80]. These findings suggest that only part of the protein is required for activation of acid β-glucosidase, and that the carbohydrate component is not important for this effect.

The physiologic significance of saposin C has been demonstrated by its deficiency state. Lack of saposin C causes a Gaucher disease-like variant [8,9,10]. Two Gaucher disease-like variants diagnosed and characterized had normal *in vitro* acid β-glucosidase activity, as measured in the presence of detergent. But, the fibroblasts from these patients were CRIM negative to saposin C and accumulated only glucosyl ceramide. The patients died at 14 and 15 years [8,10]. In a third

case, a boy who died at 16 weeks of age showed no CRIM with antibodies raised against saposin C and accumulated glucosyl ceramide and ceramide [9].

To evaluate the post-translational processing of saposin C to its mature form of 8.5 kDa, pulse chase studies were done using skin fibroblasts and rabbit anti-human saposin C. After 15 minutes of ^{35}S-methionine metabolic labeling, the major and high molecular weight product was formed, 68 kDa glycoprotein. Within one hour the precursor was converted to a 73 kDa form, and the ratio between the two forms was about 1:1. A third form with a molecular weight of 50 kDa was present at all times. Over a period of 2 to 3 hours chase, the amount of the 73, 68 and 50 kDa forms decreased and an increase in the 9 kDa and 12 kDa forms was obtained. By the end of 5 hours the 12 kDa form disappeared, while the 9 kDa form remained the major product [81]. A very similar pattern of processing was described for another activator protein named saposin B [82]. This similarity in processing would be expected since it has now been demonstrated that saposin A,B,C, and D are all encoded by the same gene on chromosome 10 [60-63,83,84].

Targeting of saposin C to the lysosome is via the mannose-6-phosphate receptor system, as was shown in I-cell disease cells. No activator was found in cell extracts from I-cell disease fibroblasts [85], in another study the precursor form of saposin C was secreted into the media of I-cell disease fibroblasts [81]. Also, the precursor form was found in the media of normal fibroblasts treated with NH_4Cl [81]. These finding suggest that the precursor protein is being targeted to the lysosome using the mannose-6-phosphate receptor and that the proteolytic clipping into the mature saposin C form takes place in the lysosome. This process might be tissue specific since in some tissues the unclipped protein is secreted into body fluids including: cerebrospinal, semen, milk, pancreatic juice, bile, and plasma [86-88] (see Discussion section B).

II. Saposin A

Saposin A which is also referred to as Protein A, was shown *in vitro* to activate acid β-glucosidase with very similar kinetic properties to saposin C [59]. The only difference between the two is that saposin A maintained its activating effect in the

presence of Triton X-100 while saposin C did not. Similar to saposin C, saposin A was found to be active when deglycosylated [88]. It also increased the enzymatic hydrolysis of galactosyl-ceramide by β-galactosidase, but unlike saposin C, saposin A inhibited sphingomyelinase activity by 30% [59]. There is no report · ·t of saposin A deficiency.

III. Accumulation of Activator Proteins

Interestingly, even before the physiologic role of saposin C was clear, many laboratories reported elevated levels of saposin C in Gaucher disease patients [58, 75, 89, 90]. Recently, researchers have reported different levels of accumulation of all four saposins in various tissues from patients with lysosomal storage diseases [88,91]. For example, saposin C accumulates in spleens from Gaucher disease patients 16 fold over normal, while saposin A and D accumulate 60 and 17 fold, respectively, over normal [91]. An 80-fold accumulation of saposin A was found in brain of Tay-Sachs patients [91].

Thus far, there is no animal model for Gaucher disease but, injection of mice with conduritol B epoxide, a covalent inhibitor of acid β-glucosidase, rapidly produced elevated levels of saposin C in both the liver and brain of those mice [92]. Injection of glucosyl ceramide emulsified with Myrj 52 produced the same effects in the liver only [92].

These observations may suggests a relationship between the level of the active lysosomal enzymes and/or their substrates and the level of the saposins. Since it is now known that all four saposins are transcribed and translated together, different levels of saposins A, B, C, and D are likely to be regulate during or after post-translational processing, for example by specific susceptibility to proteolysis (see Discussion section B).

Experimental Results

This portion of the thesis has been presented as the following publications:

1. Rorman, E.G. and Grabowski, G.A., 1989: Molecular Cloning of a Human Co-β-glucosidase cDNA: Evidence That Four Sphingolipid Hydrolase Activator Proteins Are Encoded By single Genes in Humans and Rats. *Genomics* **5**:486-492.

2. Rorman, E.G., Scheinker, V. and Grabowski, G.A., 1992: Structure and Evolution of the Human Prosaposin Chromosomal Gene. *Genomics* **13**:312-318.

*Note: The references to the Experimental Results are as in the original publications.

Copyright permission was received from Academic Press Inc.

Molecular Cloning of a Human Co-β-Glucosidase cDNA: Evidence That Four Sphingolipid Hydrolase Activator Proteins Are Encoded by Single Gene in Humans and Rats

Efrat Gavrieli Rorman and Gregory A. Grabowski

Division of Medical and Molecular Genetics, Department of Pediatrics, Mount Sinai School of Medicine, New York

ABSTRACT

Authentic cDNAs encoding the activator protein for acid β-glucosidase (EC3.2.1.45), co-β-glucosidase , were cloned from the pCD and λgt11 human cDNA libraries. Initial screening with oligonucleotide mixtures encoding amino acid sequences of co–β–glucosidase identified partial cDNAs which were used to obtain a potentially full-length cDNA from the λgt11 library. This clone (2767bp), EGTISI, contained 5' (38bp) and 3' (1157bp) noncoding sequences, a translation initiation site, and an open reading frame encoding 524 amino acids which included a typical hydrophobic signal sequence (16 amino acids). Computer analyses identified three regions of high similarity to co–β–glucosidase encoded by tandem sequences in EGTISI. Searches revealed that two of these regions encoded peptides of known function; SAP1 (sphingolipid activator protein 1) and protein C (a new sphingolipid activator protein) were encoded by EGTISI sequences 5' and 3', respectively, to those of co–β–glucosidase. The third region of similarity, encoding a theoretical peptide (undefined function), was located most 5' in the cDNA. EGTISI and its encoded polypeptide had high similarity (77% nucleotide identity and about 80% amino acid similarity) to a rat Sertoli cell cDNA and its encoded sulfated glycoprotein-1. These results indicate that a single highly conserved gene encodes the precursor for four potential sphigolipid activator proteins in rat and man.

INTRODUCTION

Many lysosomal hydrolases involved in the catabolism of complex sphingolipids require interactions with activator proteins for optimal hydrolytic activity. A protein termed SAP1 (sphingolipid activator protein 1) (Fujibayashi and Wenger, 1986b) participates in the hydrolysis of sulfatide by arylsulfatase A, sphingomyelin by sphingomyelinase, and globotriaosylceramide by α-galactosidase A (Fischer and Jatzkewitz 1975; Wenger et al., 1982; Li et al., 1985). A distinct activator protein, co–β–glucosidase, optimizes the hydrolysis of glucosylceramide by acid β-glucosidase (Ho and O'Brien, 1971; Berent and Radin, 1981). A third activator protein is required for the degradation of GM2-ganglioside by β-hexosaminidase A (Conzelman and Sandhoff, 1978). The physiologic significance of these activator proteins has been demonstrated by their respective deficiency states, since disease analogs of the true enzymatic deficiency states result: i.e., a metachromatic leukodystrophy variant in SAP1 deficiency (Inui et al., 1983), a Tay-Sachs variant in β-hexosaminidase A activator deficiency (Conzelman and Sandhoff, 1979), and a Gaucher disease variant in co–β–glucosidase deficiency (Christomanou et al., 1986).

Two of these activator proteins, SAP1 and co–β–glucosidase, have similar biochemical and processing properties and their loci map to chromosome 10 (Inui et al., 1985; Kao et al., 1987). The remarkable resistance of these two activators to thermal denaturation facilitated their purification to homogeneity from several human tissue sources (e.g., Li et al., 1985; Sheh et al., 1985; Iver et al., 1983). By analytical polyacrylamide gel electrophoresis (PAGE) the purified proteins were small (Mr = 6K-22K), acidic glycoproteins. Partial SAP1 (Dewji et al., 1987) and complete co–β–glucosidase amino acid (Kleinschmidt et al., 1987) sequences demonstrated that the mature proteins from human spleen were distinct. Using antibodies specific for either activator, pulse-chase processing studies indicated that both proteins were synthesized initially as 68 to 73 kDa glycosylated precursors (Fujibayashi and Wenger, 1986a,b). The deglycosylated precursor were about 50 kDa for each activator. After extensive proteolytic processing, the mature glycopeptides were estimated to be about 9-12 kDa on SDS-PAGE. Immunohistologic studies localized these activators to the lysosomes (Tamaru et al., 1986), and studies of I-cell disease fibroblasts indicated that the co–β–glucosidase

requires the mannose 6- phosphate receptor for targeting to this organelle (Varon et al., 1982). Although the physical state of co–β–glucosidase in the lysosome has not been well characterized, SAP1 is thought to be present there as an active homodimer (Fujibayashi and Wenger, 1986b).

During the course of the present studies, the possibility that SAP1 and co–β–glucosidase could be encoded by the same gene was indicated by Furst et al. (1988). Using the chemically determined sequences of a few ganglioside-binding protein and co–β–glucosidase as well as the cDNA sequence encoding SAP1 (Dewji et al., 1987), these investigators indicated that with modifications to the nucleotide sequence three similar 80-amino-acid peptides could be encoded by that cDNA: SAP1, co–β–glucosidase, and protein C. Subsequently, O'Brien et al. (1988) confirmed this finding by cloning a partial cDNA which correctly predicted the three peptides and another similar sequence 5' to SAP1. However, this cDNA did not contain a 5' untranslated region, an initiating AUG, nor a signal sequence for transport across the endoplasmic reticulum.

In this communication we report the molecular cloning and characterization of a cDNA encoding four potential polypeptides with extensive self-similarity arranged tandemly. Three of these peptides, SAP1, co–β–glucosidase, and protein C, function in sphingolipid catalysis. In addtion, extensive amino acid sequence identity was predicted between the human precursor protein and a major sulfated glycoprotein (SGP-1) from rat Sertoli cells.

MATERIALS AND METHODS

Materials

The following were from commercial sources: Restriction endonucleases, polynucleotide kinase, M13 phage, and plasmid pUC19 (New England Biolabs, Beverly, MA); radiolabeled nucleotides (Amersham, Arlington Heights, IL); Sequenase (United Stated Biochemicals, Cleveland OH); and Bluescript (Stratagene, San Diego, CA). The human placental λgt11 cDNA library was from

Clontech (Palo Alto, CA). The human hepatic pCD cDNA library (Okayama and Berg, 1983) was a gift from Dr. H. Okayama at the National Institute of Health.

Cloning and Charachteriztion of cDNAs Encoding Co-ß-glucosidase

From the first 55 amino acids of the complete chemically determined amino acid sequence of co–β–glucosidase (Kleinschmidt et al., 1987), four sets of mixed oligonucleotides (20 to 26-mers, see below) were synthesized (Applied Biosystems Model 380B DNA synthesizer), end-labeled, and used as probes for the pCD cDNA library.

20-mer 128 mix
TA(C/T)TG(C/T)GA(G/A)GT(N*)TG(C/T)GA(A/G)TT

26-mer 512 mix
GA(C/T)AA(C/T)AA(C/T)AA(A/G)AC(N)GA(G/A)AA(A/G)GA(A/G)AT

20-mer 64 mix
GA(C/T)GA(N)TT(C/T)GA(C/T)AA(A/G)ATGTG

23-mer 512 mix
GA(A/G)GA(A/G)TG(C/T)CA(A/G)GA(A/G)GT(N)GT(N)GA.

*N refers to any nucleotide.

Using calculated Tm values, high-stringency hybridization and wash conditions were developed for screening of the pCD library (Maniatis et al., 1982). Sequence analyses were conducted by the dideoxy methods (Sanger et al., 1977) adapted for single or double-stranded templates (Chen and Seeburg, 1985).
Initial screening was accomplished with the 26-mer and 23-mer oligonucleotides. Positive colonies were rescreened with the two remaining probe mixtures. One colony was positive with all four oligonucleotide mixtures. This colony was picked, purified, and recloned into *EcoRI* sites of M13 and plasmid pUC19.

Nucleotide sequence analysis of these subclones, termed EG12a, indicated a length of about 1900 bp whose 5' end began 186 bases upstream from the co–β–glucosidase coding region. The 3' end contained a poly(A) tail of about 50 nucleotides in length. Computer searches of GenBank against the EG12a nucleotide sequence revealed extensive similarity to the 3' half of the cDNA encoding SAP1 sequence (accesion No. J03015) reported by Dewji et al. (1987).

Initial screening of the λgt11 human placental library was conducted with randomly primed (multiprime DNA labeling kit, Amersham) 5' (EG12aS, 669 bp) and 3' (EG12aL, ~1300 bp) *BamHI* fragments of EG12a which had been purified by electrophoresis in agarose gels (Maniatis et al., 1982) and subcloned into pUC19 and M13. Plaques that hybridized with both EG12aS and EG12aL were rescreened with unique 40-mer oligonucleotides complementary to bases 9-48 or 472-511 of the Dewji sequence. Plaques that hybridized with all four probes were purified. The two longest clones were subcloned into Bluescript and characterized.

Northern Blot Analyses

Total human RNA was prepared from a Gaucher disease patient's lymphoblastoid line or cultured skin fibroblasts obtained from two normal individuals and three patients with Gaucher disease. Briefly, the cells were harvested into guinidine thiocyanate, homogenized, and fractionated in a CsCl gradient (Maniatis et al., 1982). After isolation of the RNA by chloroform-phenol extraction and ethanol precipitation, RNA was electrophoresed on a 1% agarose/formaldehyde gel in MOPS buffer at 100 V for 4-5 hr. Blotting onto Zetabind, prehybridiztion and hybridiztion conditions washes were according to the manufacturer's instructions (Cuno Inc, Meriden, CT).

Computer -Assisted DNA and Protein Analyses

Searches of national and local DNA and protein databases as well as other analyses were conducted using the Wordsearch and other indicated programs of the Sequence Analysis Software of the Genetics Computer Group (University of Wisconsin, Madison, WI). Additonal analyses were conducted with the IDEAS (Integrated Database and Extended Analyses System for Nucleic Acids and

Proteins) from the Laboratory of Mathematical Biology and Advanced Sceintific Computing Laboratory of the National Cancer Instutute and the DNA/Protein Sequence Analysis System (IBI, New Haven, CT).

RESULTS

Characterization of the cDNAs Encoding Co–β–glucosidase

Partial sequence analyses of clone EG12a obtained from the pCD human liver library indicated a length of about 1900 bp. Authenticity of this clone was demonstrated by the identity of its deduced amino acid sequence with that chemically determined for co–β–glucosidase (Klienschmidt et al., 1987). Using this clone as a probe, Northern analyses of total fibroblast or lymphoblastoid RNA (data not shown) indicated a single co–β–glucosidase-specific mRNA of about 3.2-3.4 kb. To obtain longer clones the λgt11 placental cDNA library was screened with EG12aS and EG12aL and the unique oligonucleotide from the Dewji et al. (1987) sequence of SAP1 (see Materials and Methods). Two clones, EGTISI and EGTIS, of about 3.0 kb in size were isolated and sequenced. The longest clone, EGTISI, was 2767 bp in length and contained a long open reading frame of 1572 nucleotides, bases 39-1610, encoding 524 amino acids. The first 100 nucleotides of EGTISI indicating the initiating ATG are shown below and a reasonable consensus sequence for translation initiation (Kozak, 1986) included a G in the -3 (nucleotide 36) position and a T in position +4 (nucleotide 42). The other clone, EGTIS, was identical except that the 77 most 5' nucleotides in EGTISI were not present. Although neither EGTISI nor EGTIS contained a poly(A) tail , the most common polyadenylation signal, AAUAAA, was present at bases 2744-2750. The finding of a poly(A) tail of at least 50 bases in clone EG12a indicated that the longest predicted cDNA would be about 2850-2900 bp in the absence of additional non-coding 5' nucleotides.

1 GGGCGGGCGC ATTGCAGACT GCGGAGTCAG ACGGTGCTAT
41 GTACGCCCTC TTCCTCCTGG CCAGCCTCCT GGGCGCGGCT
81 CTAGCCGGCC CGGTCCTTGG

The authenticity of EGTISI and EGTIS was demonstrated by the finding that the complete chemically determined amino acid sequence of co–β–glucosidase was predicted exactly by bases 969 to 1208 of EGTISI. Furthermore, the complete protein C and partial SAP1 amino acid sequences were predicted correctly by nucleotides 1257-1490 and 621-674, respectively (Fig 1.). Bases 77 to 1614 of clone EGTISI and the corresponding sequences in EGTIS had essential identity to the partial cDNA reported by O'Brien (Accession No. J03086), except that the latter cDNA contained a G deletion at the base corresponding to 88 in the EGTISI sequence. This resulted in an apparent frameshift in the O'Brein et al. (1988) sequence, leading to incorrect prediction of the first four amino acids of their sequence. Searches of GenBank revealed no other significant similarity to included DNA sequences (January 1989). From a local database, the 1662-bp coding region, 554 amino acids, of the rat Sertoli cell sulfated glycoprotein 1 (SGP-1) had 77% nucleotide identity with the coding region of EGTISI. The major differences included an in-frame codon deletion corresponding to bases 481-483 of EGTISI and a 90-base in-frame insertion at base 1279 of EGTISI (Fig. 1). A single G deletion corresponging to position 30 in EGTISI was present in the 5' noncoding sequence of SGP-1. Following the respoctive stop codons (Fig.1), the rat and human cDNA sequences required numerous large deletions and insertions for optimal alignment (data not shown).

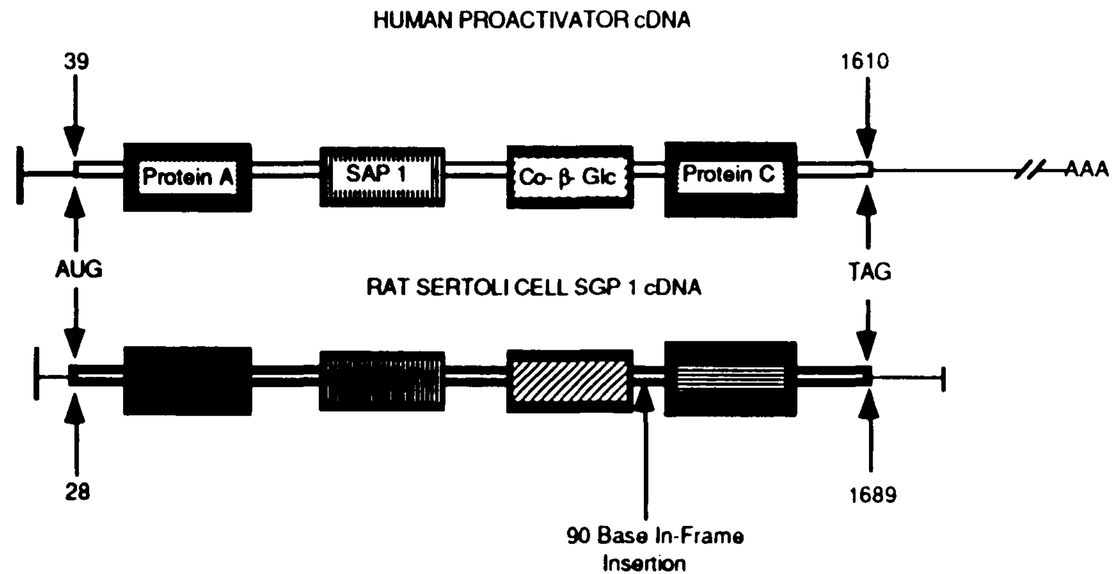


Figure 1.
Schematic diagram of the human proactivator cDNA (EGTISI) and the rat sertoli cell SGP-1 cDNA (4). The coding sequences are indicated by rectangles (open or shaded). Noncoding sequences are indicated by thin solid lines. The numbers refer to nucleotides for coding region. The similarly shaded rectangles indicate region of similarity encoding functionally assigned peptides: Protein A, SAP-1, co β-glucosidase, and protein C, arising from proteolytic processing of the human proactivator molecule.

Analysis of the Protein Encoded by the Co-β- glucosidase cDNA: Proactivator

A hydrophobic signal sequence was predicted to include 16 amino acids following the methionine encoded by bases 39-81 of EGTISI (Fig 2). As is typical for a signal sequence, the hydrophobic amino acids were predicted to constitute a transmembrane region by the ALOM (Allocation of Membrane Proteins) program (Klein et al., 1985) of the IDEAS package. The N-terminus of the mature protein precursor of the co–β–glucosidase was predicted by the weight-matrix method (von Heijne, 1986) to begin at Gly^{17}. This was consistent with the results obtained for the rat SGP-1, whose chemically determined N-terminal amino acid was Ser^{17} (Collars et al., 1988) (Fig. 2).

The predicted protein encoded by EGTISI, designated proactivator, contained four similar regions for an unknown protein ("Protein A"), SAP1, co–β–glucosidase, and Protein C. As shown in Figure 2, SGP-1 also had four internal regions of similarity (also see Collard et al., 1988). Computer analyses indicated that the first regions, Protein A, of the rat and human proteins had 85% amino acid identity (90% similarity). The corresponding SAP1 and Protein C regions had 81% (94% similarity) and 85% (94% similarity) identity, respectively. The co–β–glucosidase regions were the least conserved, since only 40% identity (71% similarity) was found.

Several structural features were conserved between each of the four regions in the deduced human and rat proteins (Fig 2). The relationships of the cysteine residues in each region were strictly preserved. A single N-glycosylation consensus sequence for each peptide was preserved precisely in alignment. Using a variety of programs available in the Genetics Computer Group and the IBI DNA/Protein Analysis packages for the prediction of secondary protein structure, hydrophobicity, and chemical properties of proteins, no major differences were evident among or between any of the eight potential peptides. For example, human (theoretical) Protein A, SAP1, co–β–glucosidase, and Protein C had calculated molecular weights of about 8.9 kDa and pI values between 5.1 and 5.4. Using different methods, O'Brien et al. (1988) reached complementary conclusions for the human peptides.

The major difference between the precursor protein, proactivator, and the SGP-1 precursor protein was that the latter is not proteolytically processed to small (~80 amino acid) peptides in Sertoli cells (Collard et al., 1988). To gain insight into the potential for amino acid sequence differences that could account for this processing difference (possibly tissue specific), the predicted amino acid sequences between the four regions were analyzed. In general, these sequences did not have greater divergence than those for the four major peptide regions, except for a predicted 30-amino-acid insertion in the SGP-1 sequence following the co–β–glucosidase region. Using the first cysteine in each region (Fig 2) as a reference, a lysine residue (i.e., a potential protease cleavage site) was located 4 to 6 amino acids upstream from each of these cysteines, except for a glutamate substitution in the co-β-glucosidase region of the SGP-1 sequence. For human SAP-1, co–β–glucosidase, and Protein C, the chemically determined N-termini were 1 to 2 amino acids downstream from these lysines. Similarly, at the carboxy-termini a basic residue (lysine or arginine) was 4 to 5 amino acids downstream from the last cysteine in each region of the human proteins (Fig.2). The charge of this amino acid was preserved in the rat SGP-1 regions except following the Protein A sequence where Glu (rat) to Lys (human) were present. Also, at the carboxyl-terminal end of each peptide region for the human sequence, the neighboring downstream amino acid from the lysine or arginine differed in charge or hydrophobicity from the corresponding residue in the rat sequence. For example, following the SAP-1 region, Arg^{275} in SGP-1 was Glu^{275} in the human sequence. From these data the N- and carboxy-termini of the theoretical peptide protein A encoded by the human cDNA were predicted to be Ser^{60} or Leu^{61} and Ser^{139}.

FIGURE 2. Similarity and conservation of functional and theoretical peptide coding regions of rat Sertoli cell SGP-1 (top sequences) and human proactivator (bottom sequences) corresponding to "protein A", SAP1, co-β-glucosidase, and protein C. The signal sequence for SGP-1 was predicted from the chemically determined N-terminus(4). That for the human proactivator was predicted by computer analyses. The boxed area indicates the four regions of the mature peptides in the human proactivator sequence. The solid diamonds indicate aligned cystein residues and the solid rectangle indicates conserved N-glycosylation sites. Shades areas indicated conserved similarities of amino acids with like properties. Alignment were conducted by repeated use of the Gap program (see Material and Methods) followed by minor manual adjustments.

SIGNAL SEQUENCE

MYALALLASLL.TALTSPVQDPKICSGGSAV

MYALFLLASLLTAALAGPVLGLKECTRGSAV

SAP1

CD-β-GLUCOSIDASE

TAARCNAVEHCKRHVWN

TAAQCNAVEHCKRHVWN

DISCUSSION

The present studies indicate that a single, highly conserved gene in man and rat encodes three, and potentially four, small "activator" proteins involved in the sphingolipid hydrolysis. These results with this human proactivator cDNA confirm and extend those of Furst et al. (1988), based on amino acid sequence, and those from partial cDNA sequence (O'Brien et al., 1988; Dewji et al., 1987). By the extensive similarity to the rat SGP-1 cDNA, the presence of a translation start site, and the prediction of the appropriate sized polypeptide precursor of co–β–glucosidase, Mr = 56,470, (Fujibayashi and Wenger, 1986a, b), the entire coding region was encoded by EGTISI. However, Northern analyses detected a 3.2- to 3.4-kb proactivator-specific mRNA in human fibroblast or lymphoblastoid lines. Consequently, a full-length transcript could contain an additional 300-400 bp of noncoding 5' sequence which were not present in EGTISI.

The striking similarity (80% overall) and identity (up to 90% locally) of the predected amino acid sequences of human proactivator and rat SGP-1 imply conserved function. However, the ultimate structures of the polypeptides from the respective cDNA differ dramatically. The resultant Mr ~56K or 59K proactivator or SGP-1 precursors, respectively, agree well with the initial protein observed in pulse-chase processing studies (Inui et al., 1985; Collard et al., 1988). By N-terminal amino acid sequencing as well as metabolic labeling and chase studies in rat sertoli cells, SGP-1 has been shown not to be processed proteolytically after the removal of the signal sequence (Collard et al., 1988). The fact that three peptides encoded by the human proactivator cDNA, SAP1, co–β–glucosidase, and protein C, have been purified to homogeneity from human tissues was consistent with the pulse-chase processing studies which indicated extensive proteolytic processing of a large precursor. The findings that the guinea pig (Sano and Radin, 1988) and bovine hepatic (Berent and Radin, 1988) co–β–glucosidase have molecular masses, ~8-12 kDa, very similar to those of the human fibroblast (Fujibayashi and Wenger, 1986a) or splenic proteins (Kleinschmidt et al., 1987) suggest tissue-specific processing at least in the rat. A comparison of the amino acid sequences between the four potentially encoded peptides in either SGP-1 or proactivator indicates that five substitutions have occured at or directly following basic amino acid (Fig. 2) at the end of the potential peptides. These alternations apparently

occur at strategic locations 3-5 residues upstream or downstream from the N- or carboxy-terminal cysteine residues of each mature human peptide, i.e., 1-2 residues from the N- or carboxy-termini of these each peptides. These findings, which could influence protease cleavage sites, suggest a species difference in processing due to either a general or a tissue-specific lack of a specific protease activity. Clearly, further studies will be required to resolve this problem and to determine whether the intact SGP-1 has any functional properties similar to those of SAP1, co–β–glucosidase, or protein C in sphingolipid catabolism.

The roles in sphingolipid catabolism of three of the four related peptides encoded by proactivator have been established by *in vitro* studies, and the physiologic importance of SAP1 and co–β–glucosidase deficiency has been proven by their deficiencies states which result in lysosomal storage disease (Inui et al., 1983; Conzelman and Sandhoff, 1979; Christomanou et al., 1986). However, the similarities in their primary or predicted secondary structures were reflected in the specificities or mechanisms of activation by SAP1, co–β–glucosidase, or protein C [Saposin D (O'Brien et al., 1988)]. *In vitro* studies have indicated that SAP1 increases the hydrolytic rates only of the natural sphingolipid substrates of arylsulfatase A (Fischer and Jatzkewitz, 1975), α-galactosidase A (Li et al., 1985), or GM1-ganglioside β-galactosidase (Li and Li, 1976). SAP1 apparently binds to the substrate(s) and forms a SAP1/substrate complex which is the true substrate. Potentially, this mechanism is similar to those of lipid transfer proteins. In comparison, co–β–glucosidase increases hydrolytic rates of water-soluble β-glucosides as well as a natural substrate, glucosylceramide, by acid β-glucosidase in the presence of phosphatidylserine or other nondetergent negatively charged lipids (e.g., Prence et al., 1985). Co-β-glucosedase binds directly to acid β-glucosidase (Ho and O'Brien, 1971) and has not been demonstrated to bind the lipid substrates. *In vitro*, co–β–glucosidase also increases the hydrolytic rates of sphingomelinase and galactosylceramide: β-galactosidase, respectively (Wenger et al., 1982). Indeed, the purification of co–β–glucosidase has been monitored with sphingomelinase activity (Kleinschmidt et al., 1987). However, only glucosylceramide was reported to accumulate in the single patient who had no immunologic cross-reacting co–β–glucosidase material (Christomanou et al., 1986). The recently described protein C was first purified by elution from a ganglioside affinity column (Furst et al., 1988). Subsequently, O'Brien and co-

workers (1988) demonstrated specific activation of sphingoomyelin hydrolysis by this protein, which they termed Saposin D. The mechanism of action and the *in vivo* specificity of this protein are unknown. Clearly, the diverse specificities and mechanisms of action of these activators of sphingolipid hydrolysis are hidden in subtle differences in their structures.

ACKNOWLEDGMENTS

This research was supported by grants from the NIH (R01 DK 36729), the NIH Division of Research Resources (RR-71), the March of Dimes Birth Defects Foundations (1-857), the National Gaucher Foundation (NGF 19), and Florence and Theodore Baumritter to the Mount Sinai Center for Jewish Genetic Disease. G.A.G is the recipient of an NIH Research Career Development Award (K04 DK 01351).

Note added in proof.

A similar cDNA to EGTISI recently has been reported by Nakano et al. (J. Biochem. 165: 152-154, 1989). "Protein A" has been purified to homogeneity from Gaucher disease splenic tissue (Morimoto et al., Proc. Natl. Acad. Sci. USA 86: 3389-3393, 1989). *In vitro* this peptide increases the hydrolytic rates of 4-methylumbelliferyl–β–D-glucopyranoside and glucosyl ceramide by acid β-glucosidase as well as galactosyl ceramide by β-galactosidase. The amino- and carboxy-terminal amino acids corresponded to Ser ^{60} and Lys, respectively.

REFERENCES

1.BERENT, S. L., AND RADIN, N. S. (1981). Mechanism of activation of glucocerebrosidase by co-β-glucosidase (glucosidase activator protein). *Biochim. Biophys. Acta* **664**:572-582.

2.CHEN, E.Y., AND SEEBURG, P.H. (1985). Supercoil sequencing: A fast and simple method of sequencing plasmid DNA. *DNA* **4**:165-170.

3.CHRISTOMANOU, H., AIGNESBERGER, A., AND LINKE, R. (1986). Immunological characterization of two activator proteins stimulating enzymic sphingomyelin degradation in vitro: Absence of one of them in a human Gaucher disease variant. *Biol. Chem. Hoppe-Seyler* **367**:879-890.

4.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/GM_1 activator. *Biochemistry* **27**:4557-4564.

5.CONZELMAN, E., AND SANDHOFF, K. (1979). Purification and characterization of an activator protein for the degradation of glycolipids G_{M2} and G_{A2} by hexosaminidase A. *Hoppe-Seyler's Z. Physiol. Chem.* **360**:1837-1849.

6.DEWJI, N.N., WENGER, D.A., AND O'BRIEN, J.S. (1987). Nucleotide sequence of cloned cDNA from human sphingolipid activator 1 precursor. Proc. Natl. Acad. Sci. USA **86**:8652-8656.

7.FISCHER, G., AND JATZKEWITZ, H. (1975). The activator of cerebroside sulfatase. Purification from human liver and identification as a protein. *Hoppe-Seyler's Z. Physiol. Chem.* **356**:605-613.

8.FUJIBAYASHI, S., AND WENGER, D.A. (1986a). Synthesis and processing of sphingolipid activator protein-2 in cultured human fibroblasts. *J. Biol. Chem.* **261**:15336-15343.

9.FUJIBAYASHI, S., AND WENGER, D.A. (1986b). Biosynthesis of the sulfatide/G_{M1} activator protein in control and mutant cultured skin fibroblasts. *Biochim. Biophys. Acta* **875**:554-562.

10.FUJIBAYASHI, S., AND KAO, F.-T., JONES, C., MORSE, H., LAW, M., AND WENGER, D.A. (1985). Assignment of the gene for human sphingolipid activator protein-2 (SAP-2) to chromosome 10. *Amer. J. Hum. Genet.* **37**:741-748.

11. 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**: 317-328.

12.HO, M.W., AND O'BRIEN, J.S. (1971). Gaucher disease: deficiency of acid-β-glucosidase and reconstitution of enzyme activity in vitro. *Proc. Natl. Acad. Sci. USA* **68**:2810-2813.

13.INUI, K., EMMETT, M., AND WENGER, D.A. (1983).Immunological evidence for deficiency in an activator protein for sulfatide sulfatase in a variant form of metachromatic leukodystrophy. *Proc. Natl. Acad. Sci. USA* **68**:2810-2813.

14.INUI, K., KAO, F.-T., FUJIBAYASHI, S., JONES, C., MORSE, H. G., LAW, M.L., AND WENGER, D.A. (1985). The gene coding for a sphingolipid activator protein, SAP-2, is on chromosome 10. *Hum. Genet.* **69**:197-200.

15.IVER, S. S., BERENT, S. L., AND RADIN, N. S. (1983). The cohydrolases in human spleen that stimulate glucosyl ceramide β-glucosidase. *Biochim. Biophys. Acta* **748**:1-7.

16.KAO, F.-T., LAW, M. L., HARTZ, J., JONES., ZHANG, X.-L., DEWJI, N.N., O'BRIEN, J.S., AND WENGER, D.A. (1987). Regional localization of the

gene coding for sphingolipid activator protein SAP-1 on human chromosome 10. *Somatic Cell Mol. Genet.* **13**: 685-693.

17.KLEIN, P., KANEHISHA, M., AND DELISI, C. (1985). The detection and classificaton of membrame spanning proteins. *Biochim. Biophys. Acta* **815**:468-476.

18.KLEINSCHMIDT, T., CHRISTOMANOU, H., AND BRAUNITZER, G. (1987). Complete amino-acid sequence and carbohydrate content of the naturally occurring glucosyl ceramide activator protein (A_1 activator) absent from a new human Gaucher disease variant. *Biol. Chem. Hoppe-Seyler* **368**:1571-1578.

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

20.LI, S.-C., KIHARA, H., SERIZAWA, S., LI,Y.-T., FLUHARTY, A.L., MAYER, J.S., AND SHAPIRO, L.J.(1985). Activator protein required for the enzymatic hydrolysis of cerebroside sulfate. *J. Biol. Chem.* **260**:1867-1871.

21.LI, S.-C., AND LI, Y.-T. (1986). An activator stimulating the enzymatic hydrolysis of sphingolipids. *J. Biol. Chem.* **25**: 1159-1165.

22.MANIATIS, T., FRITSCH, E. F., AND SAMBROOK, J. (1982). "Molecular Cloning : A Laboratory Manual" Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

23.MORIMOTO, S., MARTIN, B.M., KISHIMOTO, Y., AND O'BRIEN, J.S. (1988). Saposin D:A sphingomyelinase activator. *Biochem. Biophys. Res. Commun.* **156**:1098-1101.

24.O'BRIEN, J. S., KRETZ, K.A., DEWJI, N.N., WENGER, D.A., ESCH, F., AND FLUHARTY, A. L. (1988). Coding of two sphingolipid activator proteins (SAP-1 and SAP-2) by the same genetic locus. *Science* **214**:1098-1101.

25.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.

26.PRENCE, E., CHAKRAVORTI, S., BASU, A., CLARK, L. S., GLEW, R. H., AND CHAMBERS, J. A.(1985). Further studies on the activation of glucocerebrosidase by a heat-stable factor from Gaucher spleen. *Arch. Biochem. Biophys.* **236**:98-109.

27.SANGER, F., NICKLEN, S., AND COULSON, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.

28.SANO, A., AND RADIN, N.S. (1988). The carbohydrate moiety of the activator protein for glucosyl ceramide β-glucosidase. *Biochem. Biophys. Res. Commun* **154**:1197-1203.

29.SHEH, L., GLEW, R. H., BOTHNER-BY, A. A., AND MIRSHA, P.K.(1985). High resolution proton muclear magneric resonance studies of the glucocerebroside activator protien from Gaucher spleen. *Biochemistry* **24**:6645-6651.

30.TAMARU, T., FUJIBAYASHI, S., BROWN, W. R., AND WENGER, D.A. (1986).Immunocytochemical localization of sphingolipid activator protein-1, the sulfatide/GM1 ganglioside activator, to lysosomes in human liver and colon. *Histochemistry* **86**:195-200.

31.VARON, R., KLEINJER, W.J., THOMPSON, E.J., AND D'AZZO, A.(1982). Evidence for the deficiency of β-glucosidase-activating factor in fibroblasts of patients with I-cell disease. *Hum. Genet.* **62**:66-69.

32.VON HEIJNE, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res* **14**:4683-4690.

33.WENGER, D.A., SATTLER, M., AND ROTH, S. (1982). A protein activator of galactosylceramide β-galactosidase. *Biochim. Biophys. Acta* **712**:639-649.

Structure and Evolution of the Human Prosaposin Chromosomal Gene

Efrat Gavrieli Rorman, Vladimir Scheinker, and Gregory A. Grabowski

Division of Medical and Molecular Genetics, Department of Pediatrics, Mount Sinai School of Medicine, New York

ABSTRACT

The gene for prosaposin was characterized by sequence analysis of chromosomal DNA to gain insight into the evolution of this locus that encodes four highly conserved sphingolipid activator proteins or saposins. The 13 exons ranged in size from 59 to 119 bp, while the introns were from 91 to 3812 bp in length. The regions encoding saposins A, B and D each had three exons, while that for saposin C had only two. This sequence included the regions that encode the carboxy terminus of the signal peptide, the four mature prosaposin proteins and the 3' untranslated region. Primer extension studies indicated that over 99% of the coding sequence was contained in these 19,985 bp. Use of PCR and reverse PCR techniques indicated that the most extreme 5' coding ~140 bp contained large introns and at least two small exons. Analyses of the intronic positions in the saposin regions indicated that this gene evolved from an ancestral gene by two duplication events and at least one gene rearrangement involving a double crossover after introns had been inserted into the gene.

INTRODUCTION

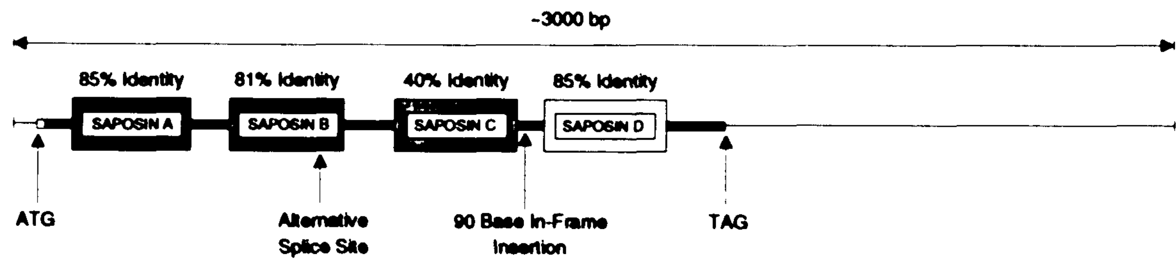
During the past decade, numerous laboratories have defined the importance of small molecular weight proteins as cofactors in the metabolism of lipids [Bernier and Jolles, (1987) for review]. In the degradation of complex glycosphingolipids, several enzymes have an essential requirement for specific small proteins, termed sphingolipid activator proteins (SAPs) or, more recently, saposins (Morimoto *et al.*, 1988). The isolated deficiencies of SAPs or saposins leads to lysosomal storage diseases that have phenotypes analogous to the specific enzyme defects. Thus, cerebroside sulfatide, glucosylceramide and G_{M2}-ganglioside storage diseases have resulted from deficiencies of arylsulfatase A (Jatzewitz and Mehl, 1969) or saposin B (Inui *et al.*, 1983), acid β-glucosidase (Brady *et al.*, 1965) or saposin C (Christomanou *et al.*, 1986) and β-hexosaminidase A (O'Brien *et al.*, 1970) or G_{M2}-activator protein (Conzelman *et al.*, 1979), respectively. Two additional saposins, A and D, have been shown to enhance the *in vitro* hydrolytic activity of acid β-glucosidase or sphingomyelinase (Furst *et al.*, 1988; Morimoto *et al.*, 1988), respectively. Biochemical studies have defined the interactions of these "activator" proteins and their respective enzymes and substrates (Ho and O'Brien, .971; Fischer and Jatzewitz, 1975; Fujibayashi and Wenger, 1986b; Morimoto *et al.*, 1989; Fabbro and Grabowski, 1991).

Intriguingly, several laboratories demonstrated that saposins A, B, C and D were encoded in tandem on a single ~3 kb mRNA species (O'Brien *et al.*, 1988; Nakano *et al.*, 1989; Rorman and Grabowski, 1989; Reiner *et al.*, 1989) and were derived from a single high-molecular-weight protein precursor by lysosomal proteolytic processing (Fujibayashi *et al.*, 1986a, 1986b). The mature saposins were small (MW~8.5 kDa), acidic glycoproteins [Grabowski *et al.*, (1990) for review]. In addition, the protein sequences of these saposins were highly similar with about 50% amino acid identity. Comparison of the analogous murine sequence for Sertoli cell sulfated glycoprotein 1 (SGP-1) (Collard *et al.*, 1988) showed ~80% amino acid identity (~90% similarity) between the saposin A, B and D regions of the two species. In the saposin C region, the analogous amino acid sequences had 40% identity and 71% similarity (Rorman and Grabowski, 1989) (Fig. 3). These results suggested that the human and murine genes encoding these saposins derived by duplication events from a single ancestral precursor. To examine this hypothesis at

a molecular level, we report the organization and sequence of the human chromosome 10 locus (Inui *et al.*, 1985; Kao *et al.*, 1987) encoding prosaposin.

FIGURE 3.

Schematic diagram of the cDNA sequences of human and murine prosaposin (adapted from Rorman and Grabowski, 1989). The regions designated saposin A, B, C and D indicate the mature proteolytically processed saposins as determined from the amino acid sequences of the purified proteins from human sources (Morimoto *et al.* 1988, 1989; Kleinschmidt *et al.* 1987; Dewji *et al.* 1987). The solid black regions indicate protein sequences that are removed during post-translational proteolytic processing. The sequence indicated by a solid open area near the ATG is the signal sequence of the prosaposin propeptide, which is removed during transport into the lumen of the endoplasmic reticulum. The thin lines indicate 5' and 3' untranslated sequences. The 90-bp additional exonic sequence in the murine cDNA is shown as "90 Base In-Frame Insertion" (Collard *et al.* 1988). The location of the intron in the saposin B region that contains an alternative splice site is labeled as "alternative splice site." The percentage amino acid identity of the homologous human and murine saposin regions is shown above each designated area.


~3000 bp
85% Identity
81% Identity
40% Identity
85% Identity
SAPOSIN A
SAPOSIN B
SAPOSIN C
SAPOSIN D
ATG
Alternative
Splice Site
90 Base In-Frame
Insertion
TAG

MATERIALS AND METHODS

The following were from commercial sources: Restriction endonucleases, Taq DNA polymerase and buffers, random priming kit, and pGEM-4Z (Promega, Madison, WI); radiolabeled nucleotides (Amersham, Arlington Heights, IL); polynucleotide kinase (New England Biolabs, Beverly, MA); Sequenase (U.S. Biochemicals, Cleveland, OH); Bluescript (Stratagene, San Diego, CA); Reverse Transcriptase (Gibco BRL, Gaithersburg, MD); chromosome 10 specific libraries (ATCC, Rockville, MD). The following total genomic libraries were screened: Two fibroblast DNA EMBL-3 libraries (Frischauff et al., 1983, and from Dr. S. Orkin), a lymphoblast DNA cosmid library (from Dr. H. Vissing, Mount Sinai School of Medicine, NY), a liver DNA Charon4A (Maniatis et al., 1978), and a placental library (from Dr. E. Skolnik, NYU Medical Center, NY).

From each total genomic library about 1×10^6 clones were screened at least once. The following probes were used to screen the genomic libraries, all contain the most 5' end of our cDNA clone (Rorman and Grabowski, 1989) which begins 38 bases upstream from the ATG: 1) the 5' 1 kb *EcoR*I fragment, 2) a 198bp *EcoR*I-*PstI* fragment, which included the ATG, and 3) a PCR product of the first 78 bp of the cDNA. All probes were purified on an agarose gel (Sambrook *et al.*, 1989) and labeled. The DNA from positive clones was purified. Restriction maps were developed by Southern blot analysis (Sambrook *et al.*, 1989), using radioactive labeled oligonucleotides corresponding to sequences covering the entire cDNA, except for 80 bases of the 5' end of the cDNA (Applied Biosystems DNA synthesizer Model 380B). Clone A (Fig. 4) was digested with *SalI* and *EcoR*I and DNA fragments were subcloned into pGEM-4Z or Bluescript. Clone A was digested with *BamHI*, a fragment overlapping the three adjacent EcoRI sites was subcloned and the sequence was confirmed. This approach, as well as sequencing of much of the intact Clone A, was used to ensure that small intronic sequences were not missed. Clone B (Fig. 4) was subcloned in one piece into pGEM-4Z. All sequence analyses were conducted by the dideoxy methods (Sanger *et al.*, 1977) adapted for double-stranded DNA template (Chen and Seeburg, 1985) using synthetic oligonucleotide primers. Most of the sequence was determined in both directions.

Two regions of the genomic sequence, not fully contained in clones A or B were obtained by PCR amplification of normal genomic DNA as follows: PCR 1 (Fig. 4) included the sequence between the first and second exons. The 5' primer was synthesized from cDNA bases 100-118 5'-CAGAATGTGAAGACGGCGT, and the 3' antisense primer cDNA bases 178-197 5'-CATATGTCGCAGGGAAGGGA. PCR 2 included the sequence containing the second and third exons. The 5' primer was synthesized from cDNA bases 226-243 5'-ATGCTGAAGGACAATGCC, and antisense primer at the 3' end was synthesized from cDNA bases 252-270 5'-CAAGTAAACAAGGATCTC. Conditions for amplification were: 94^{0}C 1 min (denaturation), 56^{0}C 30 sec (annealing), 72^{0}C 2 min (extension) for a total of 30 cycles. These amplification conditions were experimentally determined to provide one specific band, which then was cloned and subjected to sequence analysis. To avoid *Taq* polymerase introduced errors, three to five independent clones from the same PCR products were isolated and sequenced. In addition, the overlapping regions of PCR 1 with clone B and PCR 2 with clone A were compared and found to be identical, except for two isolated PCR errors. All cDNA base numbers cited in the text are referenced to the A of the ATG in the sequence reported by Rorman and Grabowski (1989).

Once the entire sequences of Clone A, PCR 1, PCR 2 and Clone B were obtained and verified, their overlapping regions were aligned by computer and verified manually. Computer DNA analyses were conducted using MacMolly and the Sequence Analysis Software of the Genetic Computer Group (University of Wisconsin, Madison, WI ; Deveax *et al.*, 1984).

Primer extension studies were conducted as described (Sambrook *et al.*, 1989). Briefly, total human RNA was prepared from lymphoblastoid lines or cultured skin fibroblasts. The cells were harvested into guanidine thiocyanate, homogenized, and fractionated in a CsCl gradient (Sambrook *et al.*, 1989). After isolation of the RNA by chloroform-phenol extraction and ethanol precipitation, RNA was run on a 1% agarose gel to confirm its integrity. The RNA was selected on a poly(T) column (Sambrook *et al.*, 1989). A 20 bp long oligonucleotide starting at base 40 of the cDNA was labeled hybridized to the Poly(A) selected mRNA and RNA extension was carried out with Reverse Transcriptase at 42^{0}C. The product was resolved on 5% polyacrylamide gels.

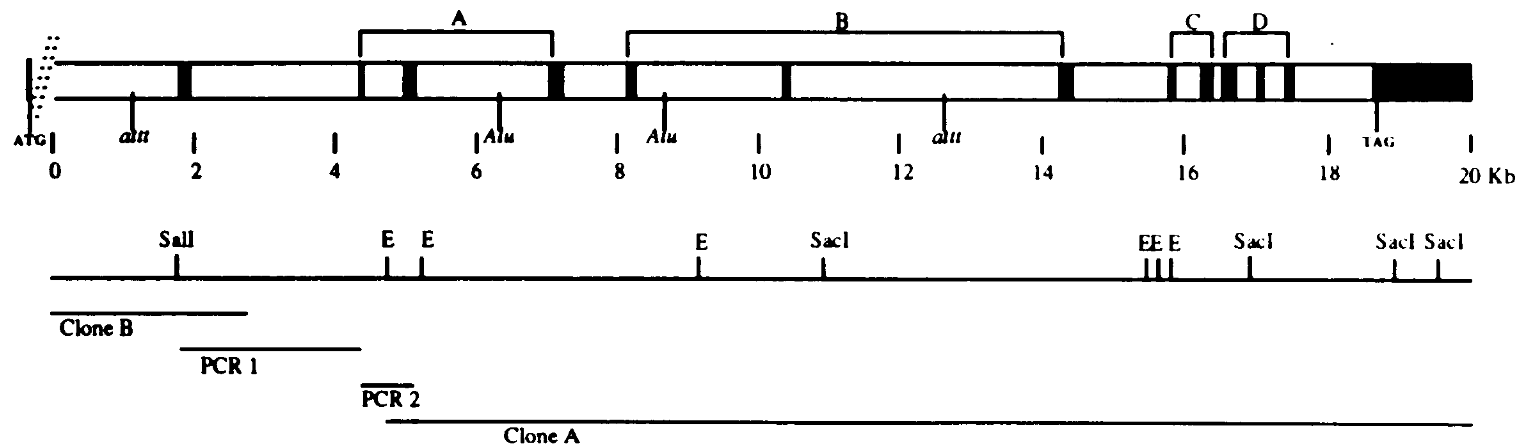


FIGURE 4.

Genomic organization, sequencing strategy and predicted restriction map for the prosaposin chromosomal gene. The open and closed regions are intronic and exonic sequences, respectively. The bracketed regions delineate the coding sequences for the mature saposins A, B, C or D. The placement of the *Alu* consensus sequences and the *attt* repeats as well as *Sal* I, *Sac* I and E (*EcoR* I) restriction sites are indicated. The scale represents kb of sequence. Clone A, clone B, PCR 1 and PCR 2 are described under Material and Methods.

RESULTS

The sequencing strategy, genomic organization, and computer-generated partial restriction map of the prosaposin locus are shown in Figure 4. The computer generated restriction map corresponded well to that obtained by Southern blot analyses of genomic DNA using the prosaposin cDNA. As expected, the intonic *EcoR* I fragment near the saposin C region was never detected. Two partial genomic clones were sequenced entirely. These clones contained about 96% of the transcribed mRNA sequence and about 99% of the coding region for the prosaposin translation product. Clone A (Fig. 4), obtained from an EMBL3 fibroblast genomic library, was about 17 kb. Its 5' end was within the first intron of the saposin A region and included all regions for saposins B, C and D as well as 3' untranslated sequence. Clone B (Fig. 4), obtained from a chromosome 10 specific library, was about 2.7 kb. Its 5' end was in the first intron of our sequence and the 3' end terminated in the second intron. To obtain the sequences between Clones A and B, PCR amplification of genomic DNA was conducted. PCR 1 (Fig. 4) spanned the most 5' exon of our sequence and the first exon of saposin A. PCR 2 (Fig. 4) spanned the first two exons of the saposin A region. A total of 19,985 bp were sequenced. From primer extension studies, two cap sites were identified at 76 and 100 bp 5' to the initiating ATG. By sequence analysis, we could identify an exon/intron boundary beginning 40 bp 3' to the initiating ATG. Together with the primer extension studies, these findings indicate that about 140 bp of exonic sequence are unaccounted for by the sequence in Figures 4 and 5. Extensive screening of five different genomic and two chromosome-10 specific libraries did not provide additional 5' sequence to that indicated in Figure 5. Screening of these libraries included the use of two short (198bp and 78bp) probes containing the most 5' end region of the cDNA. None of the clones identified with these probes contained prosaposin sequence, which was 5' to clone B. In addition, multiple PCR primers within the first intron (Fig. 4) and from the "ATG containing exon" (Fig. 5; first line) did not result in consistent products for sequence analysis. Several attempts at reverse PCR were undertaken. For these studies we used restriction endonucleases, which were predicted to have cleavage sites in the first 30 bases downstream from the ATG. Circularized DNA from such partial restriction digests yielded no sequencible products. Since we can consistently obtain PCR products of 2 to 3 kb from genomic DNA, the above results suggest that the region

between the 5' end of the first sequenced intron in Figure 5 and the ATG is likely to be greater than 2 or 3 kb. This suggests that this first intron (Figs. 4 and 5) is greater than 4 kb. These results also suggest that the 25-to 40-bp region surrounding the ATG contains additional introns and small exons.

FIGURE 5.

The sequence of the prosaposin chromosomal gene. About 99% of the coding region is contained in this sequence. About 140 bp of exonic sequence at the 5' end have not been found. The first line of the sequence shows the 5' end of the prosaposin cDNA which includes the initiating ATG (in bold). This "ATG containing exon" is likely to contain additional exons and introns. The exons are in uppercase letters and are underlined with solid lines. The two *Alu* consensus sequences are underlined by a dotted lines. The 3' stop codon (TAG) is indicated in bold.

GGGCGGGCGC ATTGCAGACT GGGGAGTCAG ACGGTGCTAT GTACG[illegible] ...

1 aagctttgt gagccatctg gggagcttgg tactgatttg catctggcac atgttttccc ctggaaacta gagtccaaac [illegible]

151 ggacagggaa ggctataatc tgggcagcag atgaataatt catttgcttg tgtctgaagg ggtgaaggtg atctggctcg agctttgtgg gactcagtct ttgtgtgagt agcttgtgga ggtagtaatg aaagaaagca [illegible]

301 taccctctcc cggagaagaa cacgtactga cctatacaac acgtgtttcc tatgtgtgtt acacacgtat gtagaattgg cttacagagg gaagcagcat acagcaagat tcctgaggtt ctaagtgtgt ttctcaggtg attgtgtttt

451 ttttttttt ttttttttc tgtctcagac ttgggtcaag tatctattgc agcaaaagtt tgttcaaccc tactgagatg cctgttggag tttgaccctc atgccaccat cctctatgtg tagtccgatg taatctcttg tttcctagct

601 aaataattgg gctagaagtg agatgtgaat ctgaggattt ttttccccat tttttggtt ctaaaatcaa ggacctgttt taagaaacat aggcactctc aaggcagttt ataaaatatt tttaggggcta ggtatggtag ct[illegible]

751 gagtgcagtg gtgagatctc ggctcactac aacctctgcc ctcctggggc tcaagctatc ctcccacctc agtctctcga gtagctggga tcataggtgc gtgtctacca tgcccggcta atttttgta gagatgaggt tttgccatgt

901 tgcccaggct ggtctcgaac tcctggactc aactgacccc acctggcctc tcaaagtgct gagattatag gcatgagcac cggcagctgg ccattgtttt atgttataag ggtatgaatt tttttaattt tatttttatt ttatttattt

1051 atttatttat ttatttattt tttgaaacgg agtctcgctc tgtcacccag gctcactgcc tagctccgcc tcccggattc acgccattct cctgcctcag cctcccaagt agctggaact acaggcgccc gccaccacac ctggctaatt

1201 gtttttttt gtatttttag tagagatggg gtttcactgt gttagccagg atggtcttga tctcctgacg tcgtgatcca cccaccttgg cctcccaaag tgctgggatt ataggcgtga gccaccgtgc ccagcctgtt tttcttttt

1351 aaaaaaaccc cttcattcta cagttgttat ccattttaca gaagtcaaaa tcttccctgt atttgaattt gttagtcttt tctcctatgg tttcttttgt tacttatgga aacattttaa ggttatgatg ataccttatt acttatggaa

1501 acattttaag gttaaagttc tgatactccc tgatactgtt taaaggtagt tagagaggat gaacgtgaac cagatacatg agtcccttct agccctgtga cagacaagcc agctttcttg accgggacag gcgtgtatct gtgtggctgg

1651 tgtcccatac agcttggtga gggggtgagt agcagagtcc agctcaagga cagtcgccct cagccctggg gaaataagtc aggtcgacct ggtttgctac aaaatgggcc acctcctcac gttggttttc attttagCTC TACCC[illegible]

1801 GGTCCTTGGA CTGAAAGAAT GCACCAGGGG CTCGGCAGTG TGGTGCCAGA ATGTGAAGAC GGCGTCCGAC TGCGGGGCAG TGAAGCACTG CCTGCAGACC GTTTGGAACA AGCCAACAGT Ggtgagtgcc gctgccctcc agtcccagcc

1951 gcagcactcc tcttgggtcc ccttagaaca gctgtaactc actgccatat tgacagagca ctttttcttg ccttgttttg tttacttctc actgtgactc tgtgtgggag gcaggaccta ttggtgaatg tacctatttt acagcagagg

2101 agatggaggc tcagagaggt taaggtttca ctctcctgag gctacctggt ttcttgtggc agacaggatg aggtccccca tgcctgcggc tccaggtgcc tgccccatgc tgcctctggg cctctgtgca gatcccactt ctcaccttag

2251 ataagaggcc agtgggttgg cagtgccatt cccaccactg ggcacatggg cgcggctcta tccttggtca ctcactccaa aaagtaaatg ctgagctttt caggtacttt ctgcctgttt catatttgt tctctcctca ctcctctcaa

2401 acaggcaaac tgctgattga gatcctggag tgcagttgcc atttaaaatt ggtttccagc tgtaaccaca ctgaggaagc tagcactgat ttttcatcta aataatgcta aatgtagcgt attaatgtgg ggcatggtct aaggcgacaa

2551 gcatgggtgt gttcctcacg gtcccaagcc ctctcaggag atgctctgct gaccatgcca gctcaacaca gagctgctgc cctgggctca cttggggctg tgccggccat catgtggtac gggctgtgtt tgactgagtg gactttgggt

2701 tctaaaaaac ttgtcagtgg ccactactgc tccttcagta agagaagaaa gcttagcacc attcttttag aaacagtttg tgaattattc actggaaaac tgccttctgt ggccatactt gcccttttt ttggaggcag ggtctggctc

2851 tgtcacctag gctggagtgc agtggcacaa tcatggctca ctgcagcctc aacctcctag gctcaagcaa tcctcccacc tcagcctcct gagtactgga accacagatg tggaccagta tgcctggcta atttttgtat tttttgtaga

3001 gatagtgttt ttaccatgtt gcccaggctg gtcttgaact cctgggctaa ggtcatcctc ctgcctccca tagtgctgtg attacaggcg tgggctacca cgataggcgc tgaaactact ttcagatttg gaaactaatt tg[illegible]aat

3151 cttgtttta atactacttc ctgtttccag gtttaatgtg tgtcccgtaa ggtctttaaa acttgattgg ctggtcttgt catggcccgg ccttctgttt ttgtttttg tttttttt ttgagatgga ctctcactct gttgcctagg

3301 ctggatcagt ggcacgatct tggctcactg caacctccgc tcccgggttc aagcaattct tctgcctcag cctgccaagt agctgggatt ataggtgtac gccaccatgc ccggctaatt tttgtatttt tagtagagat ggggtttca

3451 acatgctggt caggctggtc tcgaactcct gacctcgtga tctgcccgcc tcggcctccc aagtgagcca ccacgcccgg cccggccctc tcttgatatg gttccatgca cagtaccctt gaccctctgt ttccacaaat gttgaggg[illegible]

3601 gtctgtggaa cttgagcatc ctggggtttg ggtgtccatg gggtcctgg aattgatacc ccctcagata ccaaaggatg actgtattag cctgactagg ttcaagttgg tttctctaaa tgcagctaag ttcataggtg ttcctcagt[illegible]

3751 tcacccagac atgtttcttc tccatggatg gaaagtcagt tgacatcaag aagggagcct gggcattctt ccccaagggc cgcagaacca attagaattt ccatggcctc tggtccaggt gggggcttg gtgcaatcag aagattggct

3901 ttggtgtggt cttgccttgt taacatgcct gggaggcagg ggatgcgcag gtgaacggtt cagtgtgggt gtgccttcag aggtgggtgg ggcgtttgtt tctctgtccc acccacctct ctgccatact gatgtgactc tgatttcac[illegible]

4051 tgagcctcag tttcctcaga ggtgttgtga gcattgatgt gaaattgacc tgtgccacat ggggtgctgt gtgggtatga gcaaaggttc tgtgtcctgt aactcacagt gtgttctggt gagtcacacc tcttccctcc [illegible]

4201 ccagcgcgct catgcagggt ggagaaccag gatgtgacag gagcgcggtc atgatcatat ccctggttag agagaatagc gaaaaggcca tggggagttc cagctgtgtg gggtgtgaat ttgtgtcctg caaatacaaa [illegible]

4351 tctttctctt agAAATCCCT TCCCTGCGAC ATATGCAAAG ACGTTGTCAC CGCAGCTGGT GATATGCTGA AGGACAATGC CACTGAGgtg agcggcggtg agtggggatg gtgcttgagc cacgggcacg tgcactggtt cctaagagag

4501 ctgtgattcc gggtgtctaa agaggaatgg tgtaggcctg agggtggcaa atgtcttcca caaaggatca aggagaaaag attttagggt atatgagccg tatactgtcc ctgtcacaca gtctcttttt aaaaatgtaa aaagaatgta

4651 aacaccattc cttgagttct gggttgtaca gaaataagct gaagggccta ctttgctgac cccgggacta aatgcccaaa gttaattgag atgttgtatt gcatgccact aatacctagg tccatttgg ccatctctgc ttagaatt[illegible]

4801 ctgtgcctgc gagactgcag gtccagggat tttcatttg aatgctcatg tctgctttg cctgattgac ctaatagagc cataggtaac gtggccaagt ccttttccct gagggctggc tgtttccag gcttggttt tctgggag[illegible]

4951 attgagggcc tcttgccttc agcccattta tttccaggga ggaaagccta tcgtgattct gaccctggtg tcctctcgta gGAGGACATC CTTGTTTACT TGGAGAAGAC CTGTGACTGG CTTCCGAAAC CGAACATCTC [illegible]

5101 AAGGAGATAG TGGACTCTCA CCTCCCTGTC ATCCTGGACA TCATTAAAGG AGAAATGgta agtgatgggg aataggcttt tccttgaagg tgcttctggg gcccaggaga ggcagtggcc cagagtagat tgattaaaat ggaaagtctg

5251 ttcttagctc tggaattcct caacagcatt ttgctacaat ttgacatcca gaatagctcc tttggcacag caggccactc tgtatgtaac ttgccctggg gttagcgagg gtgcggctgc tggggactgg tcggactgct gaaggaatg[illegible]

5401 aaaccggcat aaccttcaga gatgtgagtt ggatacatca tgtcaatctt taactgaccc ctttgagcag gcacaaaacc tcacttctgt gaatttattt ctgtgaactt ataatcaagg aataattcag ggaatgtatt gggt[illegible]

5551 attcatcaag tatgtattgt acacagacac aaaccctgtg tcacagactg ggtttttagc aaagaacacg gagctagctt tgtggaattt ggaaaggcct tgacaaagaa gaggtgttta agtaactgaa gagggccagt gggga[illegible]

5701 ttcccaagac aggaccaagg ttctggtgag gaaggcaggg gttgtgtcat acacagcctc ctaggccatc ttcaaggatt gcaagtgcag ccttggagtg tcatgattgt aagggcattt gtcaaagggt ttggtctagg tctaggt[illegible]

5851 attgagggaa gctgcaagtc cggaagctga caggaccctc aggaggtcat tgctttgtga gtgagatgtg atgtggacag tggaaatggc agatgcagat tctcttttgg aggattaaat ggactggact tgctcatgag gggatgtat

6001 aggagaggaa gagctataaa tgtctccctg gtttctgaga agataaaggt ggagttattt gctaaaatgg agctagttgg gcaggataga atgaggggta agtttgagat gtctgcaggg tgggcatgat tacagaagtt aaacacttgg

6151 gagccatcag catatggaag gtgttgaaag ccatcgtgtt agctcatttt gtggggaagt gacttagaat tgctgactct tctcctccag caagaagctc ctattcc[illegible]

6301 tctgtcgccc agcctggagt gcaatggcgc gatcttgact gacagcacct ctgcctcccg ggttcaagtg attctcccat ctcagcctcc caagtagctg ggatcacagg cacccaccat catgcccggc taatttttgt atttttgtag

6451 agatggggtt ttaccacatt ggccaggatg gtcttgaact cctgacctca gcgatccgcc cacctcggcc tcacaaagtg ctgggattac aggtgtgagc cactgcgcct ggcctcctgt tcctttaaag aagagctgag ggnnaagagg

6601 taacatttag gaagccctgt attagtccgt tttcatgcta ctaataaaga catacc[illegible]gag actgggtaat ttataaaaaa agaaaaaaag [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]
6751 aggaggagca aagccacatc ttacatggca gcaggcaaga gagcatgtgc aggggaactg ccctttataa aaccatcaga tcatgagact tactatcact aaacagcatg ggaaaacccg [illegible] [illegible] [illegible]
6901 ccctcacatg acacgtgggg actatggcag ctacagttca agatgagatt tgggtgggga cacagccaaa ctgcatcatt gcctaatagg agaattgagg aagcttaggg gcactgtgtg gttaggcact cggattgggt aattttg[illegible]
7051 ttcttggtat gctgagcagt ttgaatttag acaacaggtt tgtgtttctt ttaagggatt taaagagagg gactaattca gaggcactag atggcctgtc ttattttccc ctggggagag gatttt[illegible] aggacttctt [illegible]
7201 cagAGCCGTC CTGGGGAGGT GTGCTCTGCT CTCAACCTCT GCGAGTCTCT CCAGAAGCAC CTAGCAGAGC TGAATCACCA GAA[illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]
7351 ATCCCTCTCC TCCTCTACCC TCAGGACGGC CCCCGCAGCA AGCCCCAGCC AAAGgtaaga caacgggctg gccactgagg acccattttt ggttgcactg gagatgagag accaaaggga cagag[illegible] acgtgtggtt cttaaactgg
7501 ggcttatccc tgtgcgtgac tctgcagtca catgagccag cctttaaatg gttgtaccat gaagtttcta aagaatggga tatttaaaat atgcaaagat taat[illegible]cttt aaaaaaattt acttctttta tttttgagac aggg[illegible]
7651 tctgtcaccc aagctggagt gcagtggtgt gatcatggtt cactgcagcc tcaacctctg aggcttaagt gatcctccca ccttagcctc ctaagtagct gggagtacaa atgcacacca ccacacctgg ctaatttttg tattttt[illegible]
7801 tttgccatgt tgctcagact ggtctcaaac tcctgggttc aagcgatctg cctgcctcgg cttcccaaag tgctgggatt acaggtgtga gccactgtgc ctggccaacc tttttattta tactaaa[illegible] c[illegible]tttttata ccatact[illegible]
7951 actataatga acacttaatc gattttaagc ctatcagtaa tttgagagcc tgtaaagcat ctaggaactg gtcagcaagc tgcccctaat gctgcagccc tgagtatata aggacgtttt cctaggaatt tgagtcctct taagttgcaa
8101 acctaactgc ctcttctgta gGATAATGGG GACGTTTGCC AGGACTGCAT TCAGATGGTG ACTGACATCC AGACTGCTGT ACGGACCAAC TCCACCTTTG TCCAGCCCTT GGTGCAACAT GTCAA[illegible] [illegible] [illegible]
8251 GGCATGGCCG ACATAgtgag ccttgtgtcc ttgtgcattg ggatgttct gggcctggga gttgggctgt ggaggctgga gtagggcgtt cagacaacat ccttctcttt tttcccagaa agcagcttga ataggtagtg cagacttgaa
8401 tctattttgg ttaatggcat ctggatatgt aacaaaactg ggaacaaaaa gtaggagact gtttctggga agattttgtt tcctgccagg caccctgtgt attaggaaca cctgtagaaa tacctttttn c[illegible]cctagtt tttttttttt
8551 ttttttgaga ctgagtgttg ctctgttgcc caggctggag tgcagtggcg tgatctcagc tcactgcaac ctctacctca tgggttccag ggattctcct gccttagcct cccgagtagc tgggattaca ggtgtgtgca atcacgccca
8701 gctaattttt gtatttttag tagagatggg gtttcaccag tgatttgtca ggctggtctc aaactcctga tctcaagtga tctgcccgct tcagcctccc aaagtgctag gattacaggc gtgagtctct gcgcccggcc tttaccccat
8851 tttttctttt tcttttcttt tttttttttt tcttttttga gacggagtct cactctgtgc ccaggctgaa atgcagtggc gcgatctcgg ctcactgcaa cgtccgcttc ccaggttcacg ccattctcct gcctcagcct cctgagtagc
9001 tgggactaca ggtgcccgcc accatgcccg gctaattatt tttgtatttt ttagtggaga cggggtttca ctgtgttagc caggatggtc tcgatctcct gacctcatga tccgcctgcc tcggcctccc aaagtgctgg gattacaggc
9151 gtgagccacc gcgcccggcc atttacccca tttttgaaag cttggctaat gtatgtgaca ctgagaattc cagctgtgct cctgcagcaa gcatggtcgc tgaccatggt tcacctggaa gcctccctta gtcctgtggg ctcagaaact
9301 cattttcttg gccccttact tagatcgttt cattctggga ccctattcca agattattaa tcacactcag gtggatgtgt aaaaccaaaa tgccattttt agtacctgac aactgaagaa acctgctaga ctggcagctg agaag[illegible]tggt
9451 tgtaatcact ctgtcctttg ggatctagtc gaggcttttt gggtggagtg gccggctcta ggagatgagg ccttctccat tttttttttt tttcttgaga cagggtttag ctctgtcacc caggctggag tgcggtggca caatcatagc
9601 tcactgcagc ctttacctcc tgggctcaaa caatcctctt ctcagcctcc caagtagctg ggattacagg agtgagccac cacgcccagc tgagggtttc tctctgactt tccgttagga atgttaatca ggtgaacatt cattcatttt
9751 gaaagcctag gaatcctgta gcctcatttt aagtgtcttt gcataacacc atttaaacag gtttccagtt gaaccctctc tgagaggctg tatcatgtta cctttggtcc tctctctgta gaagctcccc aaatacaagg agcaatt[illegible]
9901 tggagtctac tggaagtatc tagattgctg aaaataagca gttaaaaatc attataactt gactctactg gatttaagct ttataatgcc atctgagatt gtgtaaatgg ctttgaaagg taagctccag ttcctaagca gtgtat[illegible]
10051 tagctgcctt ctccccaaat tgccactgaa tagttgctgt gtgatcctgg ggcacgtgag gttataagat aaaggaaaca ggagggtgga atctttaagg tcccttaaga aacattttca aaacgtcttt tgcacacatt tcctataaat
10201 gatagagtaa ccaaattgtc attttcatct tattaaagct ggcccagagc agacattcat aaaattctac tttagaaact tgcttgatag tcaccagttg acactgttgg aaatgctggg gggtttttgt tggttgggtt ttggtgca[illegible]
10351 gttgcagtga tttgtttagc aatctgtttc tctctcctca gTGCAAGAAC TATATCAGCC AGTATTCTGA AATTGCTATC CAGATCATGA TGCATATGgt aggaggcacg gaggcccttg ttagcatttt tctttgtttt tcaggttt[illegible]
10501 tcatttctag gaatcttgaa aattatatgg tttagtttcc cttttttacc ttagagtgct gaattgggca aatcgacccc tctggctagg agactctcca cttgcaatct cagta[illegible] cctgttacgg gagaggcagg [illegible]
10651 ttgatcccag gactgccctg gtggcttcct ggcaatttgt ttgcctctgc ccgttcaggc cagccatcag gtttctaggt ctgggatctg tagaagtcaa gtgctctcta actaagcag[illegible] gcctgagaat gggtgtgttg [illegible]
10801 cggggggaat actgcctccc ggagtttgtc ccgtgcatgg cagcagcgtc cgtgaagctg ctgcttgagc tggtaactcc tgagctggct ttggatgcat tcaggtggca ggggcattc[illegible] cttcattaga atcaagaata [illegible]
10951 aaatgagctc gctagtccac gcagagcagg gcggcctgct cacgttggga acagcactgc ctccttagat ctgattggca gggctgagtg ggtggggggc acctttt[illegible] tgtctgtcat gctttctgac tacttgagat [illegible]
11101 gcgtttgagt ttgttctctg atccaggggt cagcaaactg gcccatcatg ttttgtactt aaagggtatt tttttggaac acagcgacgc ctgttttgtg ctacagcagc acagttgagt agttgcagcc agatcttatg [illegible]
11251 ctgaaagatt gactctctga cctgctgtaa ccactcaagc aagctgagga ggagtagttt atttgggatt attactcaca ccctaccatg tttaaaaac agttaaaagg cagcttacag agacatgcac ctgaagtagg taaagataag
11401 cattgacggg atttaatctc gtgaagtact tcctcatagc taacttaaaa tattttttc ttatataaag ttatactgtg tagaaatgtg ttgaaaatat ggacaatact [illegible] aaatcatttg gagtttaaga [illegible]
11551 ggggtcatat atactacctt gtactttgta gtttgcattt tttgcctaaa aatatgttgg ccattgaata ttcttgtaca aatgactaaa tgcttgcaga atagtacct gtgtacatag aagagaattt aataattgg [illegible]
11701 cggatgtatc taggttgttt tcaggtttgc ctttgggaag atgaatatcc ttatagccaa atgcatacat atatgcatat ccctgccctt tcccttatt atagagaact aaagctggaa ctgctgaatg gaagcatgag gtttggata
11851 cattgttaaa ttggcttcca ggttggcaat ttggccacct gcagtagatt gccctagatc ccttaccctc aggatccctt acccccctta cttgggcagc cctgtgtctt gttctgtggt gctgctggtc actaactcgt gggaaag[illegible]
12001 tgtgggcccg gggcagtgtg ggctcctgag tatggagacc cttcatcaag catttctgg ggctcttgct acacagtgtt ttgagtgtga tttagtgtgt aagacctgga gatggtgcac acagagcagg acagcccca aatggt[illegible]
12151 cctgcagggc caggaaagct ttcagggccc ccttgtccgc aagatagcag gaacaaggta ggttgctaag atctgattgc aagagcattt ccctgaacc tcttgggaca tccttacacc tggttaactg cttgattgtt gttttaat[illegible]
12301 ttcttaactg ttgtgtcctt tcctttctct tattttcacc tccgatttcc tgtttttttg ttcaacagca ggatcaggta tgtgtcgggg aatctgcttg cagtagtggt ctgctgcttc ctctagttac tttgcttttt tttttttt[illegible]
12451 ggtatggcat gattagctct gacagataca aggaatggtt acgagaggca gggtccctct ctgttaatag gctggggagg gcttctgcat ctctcccctg taattcagcc cttgaggttt tggggtctc atctaccct tcggtccagt
12601 aaaatgaagc cagcagagcc ttgtgctctg gtcacagtcc acatacagca tacgtacatg caggccacca gtgtgacaat ggggaaccat gggtggccca ggaggtgcct cagcctcctt caacgtaggc cagttaggaa gtccagtgta
12751 ggtctggctt ccccagggca ccttgtgtct gtcacaacgc aagtaggcat tgcatgtctt tagtgctgtt tttcataaac acagaaatgc ttctgctgca tgagcaaaca tttgcagaca gccagcagac tgttcctggc tgttggat
12901 agagaattgg aaacaatctc ttgattgtgg cttctgcccc tctgggtaac agggtccttt gggatctaga gccacatcag atggcatgct acaagaacca aatgcccagt ccatacttgc tctgagccag tgtccttcgg ggccgtcctg
13051 aacactgcca cctctgagcg ttggcatcca tctgctagga ttagcattgg agcttttttt gaaggtattt tgaagtctaa tgggagagga ctttggttga aatacctgaa tttttaccta ggcctcctca aggagttgga acaaagctga
13201 tcgaggcaga agcagcccta tacctcctga atattggcgt cttttttttt tttctaaaaa gcttgctgta tcttcagctg cgttcttatt atttatttat ttatttattt atttttcatt cttaaggtgt ggcagagaga gggtatttgg

13351 gatatacttt tcagactccc ccggtcttat taattctcga accttcttga tattaaaaca agatgagcca tagtcatggt [illegible]
13501 aaggcaaatg cactcaggct cagagctagt tgctggttgc atttatttcc acaatataca gagtctatac acttgtctca ctcttggcca tcagagtgtc atggtcagga tgtctagact atctagtgct ctgtgtttgg tgtctaaatg
13651 gtgaaagggg atcaatctgg agaccttgtg gcttaaaatg attcatccat ttggcgggtt gggttttcat gtgcactgtg tgcagcttga acatagaatg tggtttttta aaaagccagc atttcccctt aatgcatcat tttaaatggt
13801 cctgccgctg aaaaatctgg tgtggtttta tttctaacac atttggcttt ctgttagttt atgggcatgg tagttccccc tctagaaaaa agcatagctt agttcctacc tagagtgacc tgaaaagaca cattctagaa cagaccaggt
13951 gtccttggag atgtggaagg tgctggggag tctccctcgt ggtggccctg gatggagact cctctgtttg atgttcctta ggagctttgt gtggtctcct ccccgagatg taactggcac atatgcactg cttcagacac gagccccttta
14101 acatggaaac tggaatctgc cccgttgtgg gagggggaga gggaggtagc cttgacctgg gcccagggtg ctgtccagct ccctgtcctc tggctggatt gggactgagg agcagtgggc gtaggcttgc tgctgactgg ttgttctgtg
14251 ctctcttcag CAACCCAAGG AGATCTGTGC GCTGGTTGGG TTCTGTGATG AGGTGAAAGA GATGCCCATG CACACTCTGG TCCCCGCCAA AGTGGCCTCC AAGAATGTCA TCCCTGCCCT GGAACTGGTG GACCCCATTA AGgtacctgg
14401 tgtcactttg tgtcctggtc agggctgtgc aatgccacac tactccctgc tgtgtcacat cacatcttac gagtcacttg tttctttcgg ttccctgcct ggctcctaaa ggtctttgag tttgtgatca ctggtttatc ctatccccta
14551 tactaatgca gcccccgacc ctgagttctt ttgaaaacat tttatttcct tggttgtaaa ccctaatgtt cgtgggatca ttagcctgtg tattttcagc atgccattca aaggtgactt tgcagtgcct ggcttcagct caccttctca
14701 catgcactgt ctgggtcctt ccccatcctg tggccctgac agactggatg ttcccagcac acccctcaga gtgtacggaa gcctcactgg gcctgcagcc tgtatactaa gtacatgtgg gtgtctattt tatttaacta aaacacctca
14851 cttggtgaag attagactac ttcatcccat aataaatttg tatttggcat ttattatggg tcaggcagca tgttaggggc tgtggtgaac agatgatgtc attgtgccgt ggggcttaca ttctagttct tcaaggcctg ggctgtgtga
15001 tgccttaaca cgtggcacat gtacacgctt cttgttctag aacatcccac ggcacttgac tgccagctgt tggtctctct ctgctgtctc ctctctgcat gtgcagagtg tgtgacccag gggccaggcc gaagtgtcag ctggaatcca
15151 ggaagcactg cttcttgcct tcctctccac cgggtgctcc tcagctccag cctgctgacc aaagtgacgg ccagagaaag tgaagagtta tagctcgcac tccaaccaca tgtgactact taaattttaa tgaaatacca tggagcaaaa
15301 ttaaaaattc agccccttgg ctacaccagc cacattgcag gtgctcagtg cacaggtagc tggagcttat agtggacagc agagatagag aacatttcca ttgccacaga acatcctctg taacagtatt gccttattaa aggtgggttc
15451 cagaagagca ttctcaagtg ctaatgagac atgggagaca cacgagcaga cctgaagtgt gttcaacatt ctgtcttcta gatagtggac cgatagtgag ccccttcact ttgcttttgg tccttcatcc tcctccccaa aatgaaccct
15601 gaatgatgaa atgattgtaa ccgaacacat catcaggaca gctcttcggg gtggggaagg gactttctgt ctcctgaact gtgaattcct gggtttgtt ttggtggggt taggggaggg tgggggtggt agaggattgg ggaagggggt
15751 gaatgtaaaa atagggaatt aacttggata ttagaagcca cacgccattt ggaattcagg aggcagtggc ccttgatgtg ctaacccacc ctggtctctg tgtccccttt cctttagggg agcaagacca agggctgtga gtcccagcat
15901 tgaaagtact ctccttctac tctttccaca gAAGCACGAG GTCCCAGCAA AGTCTGATGT TTACTGTGAG GTGTGTGAAT TCCTGGTGAA GGAGGTGACC AAGCTGATTG ACAACAACAA GACTGAGgta tgctgtcctc ctggcagagg
16051 gagggcctct cctgcccccc cgagcctgaa tgtcccagtg ggaaagtccc acaggagaga gggcttggcc agccttgaca tgtgtacagc catgtcccca tctcgtggaa ggcaatcctc accagtggga cctctaggct ggcgggggtc
16201 cggggaggtg gacaccattg ccctgctttg tagtttggta gtttttgtg ggtggaaacg gcacccacca ttgactcatt tcctggccat gttagtattt tcagaatgtc ccccaacccct tgtccggcct cccgtcttct cacacgatcc
16351 tctcgtgtgt ttcagAAAGA AATACTCGAC GCTTTTGACA AAATGTGCTC GAAGCTGCCG AAGTCCCTGT CGGAAGAGTG CCAGGAGGTG GTGGACACGT ACGGCAGCTC CATCCTGTCC ATCCTGCTGG AGGAGGTCAG CCCTGAGCTG
16501 GTGTGCAGCA TGCTGCACCT CTGCTCTGGC ACGGGGCTGC CTGCACTGAC CGgtgagcgc tgggtgtggt gcgggagagg atggtgcccc tcttgggtct gggagggccc tgccagcgtt gagccccgctg ccctcttatg tagTTCAGCT
16651 GACTCAGCCA AAGGACGGTG GCTTCTGCGA AGTGTGCAAG AAGCTGGTGG GTTATTTGGA TCGCAACCTG GAGAAAAACA GCACCAAGCA GGAGATCCTG GCTGCTCTTG AGAAAGGCTG CAGCTTCCTG CCAGACCCTT ACCAGAAGCA
16801 Ggtacgcccc gggctgggct gcggtaggcc agcatgcgag aacctgatga agtatgccaa ggctggggta cattttgatg gaaaaccagc tgtttctgcc tccgggcatg gagtccaatg aagttttact tggaggtgat ttgttggctt
16951 aagatgagcc actgggggaa gcaacctgaa gtagagaggc caaatgtaga tagggtgtct tggaaaggga ggccccagga cccagtgggc tgggaatgcc aggagctctc agggaacagt gtatccaggc ctgcctggcc ttttgtgccc
17101 cactatgctc ggacaccccc agaagctgga cctgtgtccc cactttcctt tatagTGTGA TCAGTTTGTG GCAGAGTACG AGCCCGTGCT GATCGAGATC CTGGTGGAGG TGATGGATCC TTCCTTCGTG TGCTTGgtga gcctcactgg
17251 gtggtgtggt cctctcgggc agctgtaacc tggggggctg cagagacccg gaaggttgct ctgagacttg gccgtggggt ggagtctctg ggttctgctg cgggagagcc aagaaggtag ggagccagcc ccccaagccc catcgctttc
17401 ctcagcaacc ctgatgttct gaaagagcgt ggtggagctg tgatcctagc tccttctgga aactgttttc ttagtttga aagcgataca tcgggcagct ctcattctcc cccttttcttc tttccttttc tcaccatata gAAAATTGGA
17551 GCCTGCCCCT CGGCCCATAA GCCCTTGTTG GGAACTGAGA AGTGTATATG GGGCCCAAGC TACTGGTGCC AGAACACAGA GACAGCAGCC CAGTGCAATg tgagtagctc agcagacagt gtcaccagcc tgtcttgccc tggggtagcc
17701 tgtgtgagaa tcctgcttta atgggaaccc agtgatcaac tctccaccct gcttttatgt tatcagaaag ctgaggccca gagaggtggt taggcttact catggtcccc tagaaaagac agggtgaact acactcttgt agtgaagggt
17851 ggggtggatt ttatttattt ttttttttta gtattttttt gcttctgctt tgttgctttt gtcttcattt cccttgtgac agtgtcagac ttgtacaagg tcactccagg gattgtacaa ggtctgtggg attcattgcc cagagctttg
18001 cccaggtcct gtcccccggg aacatgtttg cccttgctga agccgggccc gagtcctcac tcaaggccct gagaactggc ccctgcccag cagccctagg gcactgtggc cctggctggc ctggcagcca ccagttgggt gacttgtgct
18151 gcttccctgg gtggggatgt ttggttcggt ccctgttact taatgttgct ttgtgcaact tttccttcct caccgggctt ccctgtgacc atgtcagtgg cggttgtgtc atttcggacc actgcaggga ggggcagggg gccattctgg
18301 ccatgtccag ctctttcctc cagctaggct ccagcccatc tccaccgcca gccccacccc acagctggtt ctgctgctgt cctggggaaa tgtgggcctg cttctcaagg cccgagctgg tgtgtagcct aactcaggca aggaggttct
18451 gcctgggagt tagtgtgcag gctgctagca ggacgctggg tgccgggtgc aggtagatga agggtcctga gccctgcctg tgtgggtact catctgcaaa ctgctgcagg catctgctgc tctggactcg tgccttagaa gttcccaaaa
18601 ctctgatggc cacgtgttcc ctgtttttaa gtgagatcct gtcctgggtc ttcagcatct ggtgccatcc agctgaccta ggacaagaga acaggtcctg ttttggggtg atatctgata cattgatggg ccgggaacgg agtttttcat
18751 ccgaagacta aactcagctt cctctctcct accagGCTGT CGAGCATTGC AAACGCCATG TGTGGAACTA GGAGGAGGAA TATTCCATCT TGGCAGAAAC CACAGCATTG GTTTTTTTCT ACTTGTGTGT CTGAAAGAAT GAAGCAGAAA
18901 ATCTGTTTGA CTTTGTTATA AAAATAGGGC TCCCCCACCT CCCCCATTTC TGTGTCCTTT ATTGTAGCAT TGCTGTCTGC AAGGGAGCCC CTAGCCCCTA GCCCCTGGCA GACATAGCTG CTTCAGTGCC CCTTTTCTCT CTGCTAGAAG
19051 GATGTTGATG CACTGGAGGT CTTTTAGCCT GCCCTTGCAT GGCGCCTGCT GGAGGAGGAG AGAGCTCTGC TGGCATGAGC CACAGTTTCT TGACTGGAGG CCATCAACCC TCTTGGTTGA GGCCTTGTTC TGAGCCCTGA CATGTGCTTG
19201 GGCACTGGTG GGCCTGGGCT TCTGAGGTGG CCTCCTGCCC TGATCAGGGA CCCTCCCCGC TTTCCTGGGC CTCTCAGTTG AACAAAGCAG CAAAACAAAG GCAGTTTTAT ATGAAAGATT AGAAGCCTGG AATAATCAGG CTTTTTAAAT
19351 GATGTAATTC CCACTGTAAT AGCATAGGGA TTTTGGAAGC AGCTGCTGGT GGCTTGGGAC ATCAGTGGGG CCAAGGGTTC TCTGTCCCTG GTTCAACTGT GATTTGGCTT TCCCGTGTCT TTCTGGTGGA TGGCTTGTTT GGGGTTCTGT
19501 GGGTTTGGGT GGGAAGAGGG CCATCTGCCT GAATGTAACC TGCTAGCTCT CCGAAGCCCT GCGGGCCTGG CTTGTGTGAG CGTGTGGACA GTGGTGGCCG CGGTGTGCCT GCTCGTGTTG CCTACATGTC CCTGGCTGTT GAGGCCATCA
19651 TTCAGCCTGC ACCCCTCCCT TGTCTCATAG ATGCTCCTTT TGACCTTTTC AAATAAATAT GGATGGCGAG CTCCTAGGCC TCTGGCTTCC TGGTAGAGGG CGGCATGCCG AAGGGTCTGT CTGGGTGTGG ATTGGATGCT GGGGGTGTGC
19801 GGGTTGGAAG CTGTCTGTGG CCCACTTGGG CACCCACGCT TCTGTCCACT TCTGGTTGCC AGGAGACAGC AAGCAAAGCC AGCAGGACAT GAAGTTGCTA TTAAATGGAC TTCGGTGATT TTGTTTTGCA CTAAAGTTTC TGTGAATTAA
19951 CAATAAAATT CTGTTAGCCA GATCTGGGCC TAATG

Figure 4 shows the genomic organization of the prosaposin locus derived from the complete sequence data (Fig. 5). The sequenced region contained 13 exons varying in size from 57 to 119 bp. No differences were found between the sequences of the exons in Figure 5 and those reported for the corresponding cDNA sequences. The introns were from 91 to 3812 bp in size, with the largest introns (2126 and 3812 bp) in the saposin B region and the smallest (91 bp) in the sequence between the saposin C and D regions. None of the intron/exon boundaries were located exactly at the beginning or end of the respective mature saposin regions, although only one or two codons of non-saposin A or non-saposin C sequence, respectively, were present at the 5' end of these regions. The regions containing the exons for saposins A, B and D each had two introns, whereas that for saposin C had only one. Each of the intron/exon boundaries conformed to the gt/ag rules (Table 1).

To search for related sequences in the entire prosaposin, the Compare program in the Wisconsin package and the MacMolly program were used. Except for repetitive sequences and the saposin exonic regions, no convincing duplications were found. In addition, the entire prosaposin sequence contained only two *Alu* consensus sequences and these were in the reverse orientation (Figs. 4 and 5). One *Alu* was in the second intron of the saposin A region and had 89% homology to the consensus *Alu* sequence (Jurka and Smith, 1988; Britten *et al.*, 1988). Similarly the other *Alu* sequence was in the first intron of the saposin B region and had 92% homology to the consensus *Alu* sequence. For a sequence of ~20 kb, about 4 to 6 *Alu* sequences would be expected. Short tandem sequences of *attt* repeats were found in the first, fourth and seventh introns (Fig.4).

Table 1: Intron/Exon Boundries of Prosaposin Gene

cDNA Base #	EXON	Intron	Intron	Exon	size
40			acctcctcacgttggttttcatttt**ag**	CTCTA	>1787
174	CAGTG	**gt**gagt..	acaaacttctcgtattctttctctt**ag**	AAATC	2441
249	CTGAG	**gt**gagc..	gattctgaccctggtgtcctctcgt**ag**	GAGGA	594
375	AAATG	**gt**aagt..	cagcaggacttctttccatcccacc**ag**	AGCCG	2046
576	CAAAG	**gt**aaga..	ttgcaaacctaactgcctcttctgt**ag**	GATAA	[illegible]
720	ACATA	**gt**gagc..	tttagcaatctgtttctctctcctc**ag**	TGCAA	2126
777	ACATG	**gt**agga..	tgactggttgttctgtgctctcttc**ag**	CAACC	3812
909	TTAAG	**gt**acct..	agtactctccttctactctttccac**ag**	AAGCA	1539
1005	CTGAG	**gt**atgc..	ctcacacgatcctctcgtgtgtttc**ag**	AAAGA	338
1192	GACCG	**gt**gagc..	cgttgagcccgctgccctcttatgt**ag**	TTCAC	91
1350	AGCAG	**gt**acgc..	gacctgtgtccccactttcctttat**ag**	TGTGA	354
1431	GCTTG	**gt**gagc..	ttcttctttccttttctcaccatat**ag**	AAAAT	305
1539	GCAAT	**gt**gagt..	taaactcagcttcctctctcctacc**ag**	GCTGT	1136

A QTVWNKPTVKSLPCDICKDVVTAAGDMLKDNATEEEILVYLEKT.CDWLPKPNMSASCKEIVDSYLPVILDIIKGEMSRPGEVCSALNLCES

B PRSKPQPKDNGDVCQDCIQMVTDIQTAVRTNSTFVQALVEHVKEECDRLGPG.MADICKNYISQYSEIAIQMMMH.M.QPKEICALVGFCDE

C IKKHEVPAKSDVYCEVCEFLVKEVTKLIDNNKTEKEILDAFDKM.CSKLPKS.LSEECQEVVDTYGSSILSILLEEVS.PELVCSMLHLCSG

D LIVHVTQPKDGGFCEVCKKLVGYLDRNLEKNSTKQEILAALEKG.CSFLPDP.YQKQCDQFVAEYEPVLIEILVEVMD.PSFVCLKIGACPS

FIGURE 6.

The alignment of the intron positions in the human saposin A, B, C and D amino acid sequences.

The introns positions are indicated by double ended arrows. The mature saposin sequences are contained in the stippled area.

DISCUSSION

Several laboratories have characterized cDNAs encoding the multifunctional human prosaposin from fibroblast and hepatic sources (O'Brien *et al.*. 1988; Nakano *et al.* 1989; Rorman and Grabowski 1989; Reiner *et al. 1989).* These cDNAs differed by the insertion of 9 bp (three codons) within the saposin B exonic regions (Nakano *et al.*, 1989) (Fig 5). In an effort to gain insight into the evolutionary relationships of the human and murine analogues of prosaposin, into the gene organization, and into the origin of the above 9 bp insertion, the bulk of the human prosaposin genomic sequence was determined. A cryptic splice site was indicated in the center of the second intron of the saposin B region by: 1) analysis of the present sequence, 2) the partial DNA sequence saposin B (Holtschmidt *et al.*, 1991) and 3) a 33 bases insertion, that included the above 9bp, in the cDNA of a patient with saposin B deficiency (Zhang *et al.*, 1990). This splice site has low frequency usage during normal mRNA processing leading to cDNAs with a 9 pb (in-frame) insertion. Activation of this cryptic splice site also accounts for a 33 bp insertion leading to the aberrant function of the mature, albeit mutant, saposin B in a metachromatic leukodystrophy-like disease (Zhang *et al.* 1990).

Comparison of the human and murine cDNAs for prosaposin or sulfated glycoprotein 1, SGP-1, demonstrated extensive regions of homology, which preserved the four-region structure for saposins A, B, C and D (Collard et al., 1989, Rorman and Grabowski, 1989) (Fig. 3). The major difference in the sequences was a 90-bp region encoding 30 amino acids in the murine Sertoli cell SGP-1, which was absent in the human cDNA from fibroblast or hepatic sources. If present, this 90-bp region would be localized to the intron between the saposin C and D regions in the human sequence (Fig. 3). Although there is a 91 bp intron in this region, examination of this sequence did not reveal homology to the murine cDNA, which would account for the absence of codons encoding this "proline-rich" domain (Collard *et al.*, 1988). Consequently, the absence of this domain in the human prosaposin protein encoded by cDNAs from fibroblasts or hepatic sources (O'Brien *et al.*, 1988; Nakano *et al.*, 1989; Rorman and Grabowski, 1989; Reiner *et al.*, 1989) cannot be the result of tissue-specific alternative splicing, but rather the loss of the sequences that encode this domain.

Since the mature saposins have a very high degree of amino acid similarity (70 to 95%), it is likely that they arose by gene duplication from an ancestral gene. If this duplication arose after the introduction of introns, the sequence and/or position of the intron insertion within the exonic sequences would be expected to have been preserved. Comparison of the sequences surrounding the intron/exon boundaries (Table 2 and Fig. 5) did not reveal any extensive regions of sequence similarity. However, the positions of the introns relative to the amino acid sequences could be aligned between the different saposins (Fig. 6). For example, the first intron positions in the saposin A and B directly correspond, whereas the second intron position of saposin A aligns with the analogous intron in the saposin C region. Figure 6 indicates the alignment of the other intron positions in saposins A, B, C and D. A scheme to explain how the saposins may have arisen from a single ancestral gene (e.g. saposin B) is shown in Figure 7. This scheme requires the ancestral gene to have been duplicated prior to the introduction of introns. Following the insertion of introns (Fig. 7) into Saposin B and C predecessors, the entire gene would have been duplicated in tandem to produce the final order of the saposin regions. A gene rearrangement resulting from a double crossover would have produced the alignment of intron positions (Fig. 6), if the crossovers occurred between the first and second and between the second and third intron positions of the saposin B- and C-like regions. These results provide support for the contention that the saposins arose by gene duplication(s) of a very ancient ancestor gene.

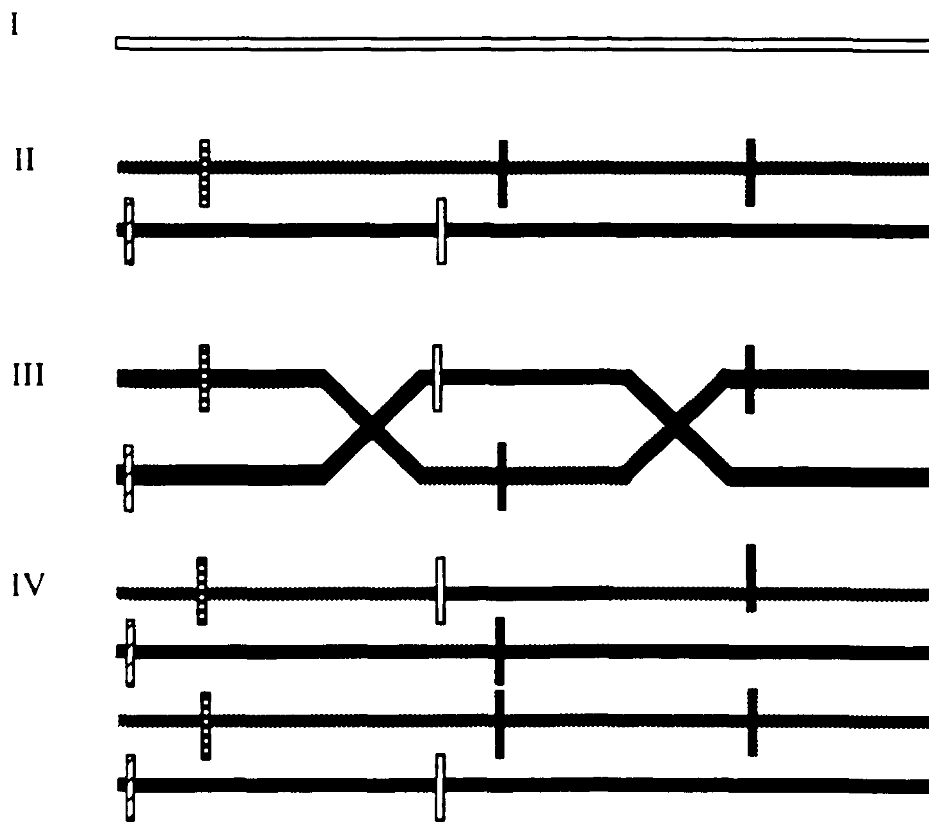


FIGURE 7.
An evolutionary model to account for the maintenance of the intronic positions in the amino acid sequences of the saposins. (I) An ancestral gene that was a precursor of the saposin B region. (II) A saposin B (upper, stippled) and C (lower, solid) tandem duplicated version of (I) into which introns (vertical boxes) have been inserted. (III) and (IV) indicate a rearrangement mechanism that reshuffles the intronic positions to obtain the alignment in Figure 5. The sequences derived from either the saposin B or C precursors are indicated by shading. (IV) Requires duplication of the tandem structure in (II) to produce the intron insertion alignment shown in Figure 6.

ACKNOWLEDGEMENTS

This work was supported by grants from the National Institutes of Health (DK 36729), the National Gaucher Foundation (NGF 19) and from the Division of Research Resources of the National Institutes of Health (RR-71). EGR is the recipient of a March of Dimes-National Birth Defects Foundation Predoctoral fellowship.

REFERENCES

1.BERNIER, I., AND JOLLES, P. (1987). A survey on cytosolic non-enzyme proteins involved in the metabolism of lipophilic compounds: from organic anion binders to new protein families. *Biochimie.* **69**:1127-1152.

2.BRADY, R.O., KANFER, J.N., AND SHAPIRO, D. (1965). Metabolism of glucocerebrosides.II. evidence of an enzymatic deficiency in Gaucher's disease. *Biochem. Biophys. Res. Commun.* **18**:221-225.

3.BRITTEN, R.J., BARON, W.F., STOUT, D.B., AND DAVIDSON, E.H. (1988). Sources and evolution of human *Alu* repeated sequences. *Proc. Natl. Acad. Sci. USA* **85**:4770-4774.

4.CHEN, E.Y., AND SEEBURG, P.H. (1985). Supercoil sequencing: A fast and simple method of sequencing plasmid DNA. *DNA* **4**:165-170.

5.CHRISTOMANOU, H., AIGNESBERGER, A., AND LINKE, R. (1986). Immunological characterization of two activator proteins stimulating enzymic sphingomyelin degradation in vitro: Absence of one of them in a human Gaucher disease variant. *Biol. Chem. Hoppe-Seyler* **367**:879-890.

6.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/GM_1 activator. *Biochemistry* **27**:4557-4564.

7.CONZELMAN, E., AND SANDHOFF, K. (1979). Purification and characterization of an activator protein for the degradation of glycolipids GM2 and GA2 by hexosaminidase A. *Hoppe-Seyler's Z. Physiol. Chem.* **360**:1837-1849.

8.DEWJI, N.N., WENGER, D.A., AND O'BRIEN, J.S. (1987). Nucleotide sequence of cloned cDNA from human sphingolipid activator 1 precursor. Proc. Natl. Acad. Sci. USA **86**:8652-8656.

9.DEVEUX, J., HAEBERLI, P., AND SMITHIES, O. (1984). A comprehensive set of sequence analysis program for the VAX. *Nucleic Acids Res.* **12**:387-395.

10.FABBRO, D., AND GRABOWSKI, G.A. (1991). Human acid β-glucosidase: Use of inhibitory and activating monoclonal antibodies to investigate the enzyme's catalytic mechanism and saposin A and C binding sites. *J. Biol. Chem.* **266**:15021-15027.

11.FISCHER, G., AND JATZKEWITZ, H. (1975). The activator of cerebroside sulfatase. Purification from human liver and identification as a protein. *Hoppe-Seyler's Z. Physiol. Chem.* **356**:605-613.

12.FRISCHAUF, A.-M., LEHRACH, H., POUSTKA, A., AND MURRAY, N. (1983). Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* **170**:827-842.

13.FUJIBAYASHI, S., AND WENGER, D.A. (1986a). Synthesis and processing of sphingolipid activator protein-2 in cultured human fibroblasts. *J. Biol. Chem.* **261**:15336-15343.

14.FUJIBAYASHI, S., AND WENGER, D.A. (1986b). Biosynthesis of the sulfatide/GM1 activator protein in control and mutant cultured skin fibroblasts. *Biochim. Biophys. Acta* **875**:554-562.

15.FUJIBAYASHI, S., AND KAO, F.-T., JONES, C., MORSE,.H., LAW, M., AND WENGER, D.A. (1985). Assignment of the gene for human sphingolipid activator protein-2 (SAP-2) to chromosome 10. *Amer. J. Hum. Genet.* **37**:741-748.

16. 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**: 317-328.

17.GRABOWSKI, G.A., GATT, S., AND HOROWITZ, M. (1990). Acid β-Glucosidase: Enzymology and molecular biology of Gaucher Disease. *CRC Critical Reviews of Biochemistry and Molecular Biology* **25**:385-414.

18.HO, M.W., AND O'BRIEN, J.S. (1971). Gaucher disease: deficiency of acid-β-glucosidase and reconstitution of enzyme activity in vitro. *Proc. Natl. Acad. Sci. USA* **68**:2810-2813.

19.HOLTSCHMIDT, H., SANDHOFF,K., FURST, W., KWON, H-Y., SCHNABEL, D., AND SUZUKI., K. (1991). The organization of the gene for the human cerebroside sulfate activator protein. *FEBS.* **280**:267-270.

20.INUI, K., EMMETT, M., AND WENGER, D.A. (1983).Immunological evidence for deficiency in an activator protein for sulfatide sulfatase in a variant form of metachromatic leukodystrophy. *Proc. Natl. Acad. Sci. USA* **68**:2810-2813.

21.INUI, K., KAO, F.-T., FUJIBAYASHI, S., JONES, C., MORSE, H. G., LAW, M.L., AND WENGER, D.A. (1985). The gene coding for a sphingolipid activator protein, SAP-2, is on chromosome 10. *Hum. Genet.* **69**:197-200.

22.JATZEWITZ, H. AND MEHL, E. (1969). Cerebroside sulfatase and arylsulfatase A deficiency in metachromatic leukodystrophy. *J. Neurochem.* **16**:19-27.

23.JURKA, J., AND SMITH,T. (1988). A fundamental division in the *Alu* family of repeated sequences. *Proc. Natl. Acad. Sci. USA* **85**:4775-4778.

24.KAO, F.-T., LAW, M. L., HARTZ, J., JONES., ZHANG, X.-L., DEWJI, N.N., O'BRIEN, J.S., AND WENGER, D.A. (1987). Regional localization of the gene coding for sphingolipid activator protein SAP-1 on human chromosome 10. *Somatic Cell Mol. Genet.* **13**: 685-693.

25.KLEINSCHMIDT, T., CHRISTOMANOU, H., AND BRAUNITZER, G. (1987). Complete amino-acid sequence and carbohydrate content of the naturally occurring glucosyl ceramide activator protein (A_1 activator) absent from a new human Gaucher disease variant. *Biol. Chem. Hoppe-Seyler* **368**:1571-1578.

26.LI, S.-C., KIHARA, H., SERIZAWA, S., LI,Y.-T., FLUHARTY, A.L., MAYER, J.S., AND SHAPIRO, L.J.(1985). Activator protein required for the enzymatic hydrolysis of cerebroside sulfate. *J. Biol. Chem.* **260**:1867-1871.

27.MANIATIS, T., ROSS, C. H., LACY, E., LAUER, J., O'CONNELL, C., AND QUON, D. (1978). The isolation of structural genes from libraries of Eucaryotic DNA. *Cell* **15**:667-701.

28.MORIMOTO, S., MARTIN, B.M., KISHIMOTO, Y., AND O'BRIEN, J.S. (1988). Saposin D: A sphingomyelinase activator. *Biochem. Biophys. Res. Commun.* **156**:1098-1101.

29.MORIMOTO, S., KISHIMOTO, Y., TOMICH, J., WEILER, S., OHASHI, T., BARRANGER, J.A., KRETZ, K.A., AND O'BRIEN, J.S. (1989). Interaction of saposins, acidic lipids, and glucosylceramidase. *Proc. Natl. Acad. Sci. USA.* **86**:3389-3393.

30.NAKANO, T., SANDHOFF, K., STUPER, 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 alternative forms of the sulfatide activator. *J. Biochem.* **165**: 152-154.

31.O'BRIEN, J.S., OKADA, S., CHEN, A., FILLERUP, D.L. (1970). Tay-Sachs disease: Detection of heterozygotes and homozygotes by serum hexosaminidase assay. *N. Engl. J. Med.* **283**:15-20.

32.O'BRIEN, J. S., KRETZ, K.A., DEWJI, N.N., WENGER, D.A., ESCH, F., AND FLUHARTY, A. L. (1988). Coding of two sphingolipid activator proteins (SAP-1 and SAP-2) by the same genetic locus. *Science* **214**:1098-1101.

33.REINER, O., DAGAN, O., AND HOROWITZ, M. (1989). Human sphingolipid activator protein-1 and sphingolipid protein-2 are encoded by the same gene. *J. Mol. Neurosci.* **1**:225-233.

34.RORMAN, E.G., AND GRABOWSKI, G.A. (1989). Molecular cloning of a human co-β-glucosidase cDNA: Evidence that the four sphingolipid hydrolase activator proteins are encoded by single genes in human and rats. *Genomics* **5**: 486-492.

35.SAMBROOK, J., MANIATIS, T., AND FRITSCH, E.F. (1989) "Molecular Cloning: A laboratory Manual", 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

36.SANGER, F., NICKLEN, S., AND COULSON, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.

37.ZHANG, X.-L., MOHAMMAD A.R., 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. USA* 87:1426-1430.

Conclusions

Studies of the small, heat stable activator protein, saposin C, using pulse-chase labeling gave the first hint that the locus on chromosome 10 encodes a large precursor protein [81,82]. When the amino acid sequence for saposin C and partial amino acid sequence for saposin B became available, a comparison of the two showed high homology, especially in the location of the cysteine residues and N-glycosylation sites [89]. These observations suggested a possible evolutionary relationship between these two proteins involved in sphingolipid degradation. Using molecular biology techniques to isolate the cDNA of the saposin C confirmed that one open reading frame encodes four, highly homologous protein domains [60-63].

In this chapter the genomic organization and a model for the evolution of the prosaposin gene will be discussed first, followed by the different steps from DNA to the mature proteins. The last part will discuss the known and more speculative functions of the different proteins encoded by this unique domain.

A. Genomic Organization and Evolution of the Prosaposin Gene

1. Genomic Organization

The prosaposin gene was localized to chromosome 10 q21-q22 [83,84]. The genomic organization of the 19,985 bp-long region coding the four protein domains and the carboxyl-terminus of the signal peptide includes 13 exons ranging in size from 59 to 119 bp, with the size of the introns ranges from 91 to 3812 bp. The four protein domains encoded on prosaposin are highly homologous; in each one there are 6 cysteine residues, N-glycosylation sites and proline residues which are placed in similar locations. In addition, 80% identity was found between the human cDNA and a rat cDNA encoding for sulfated glycoprotein-1. These finding suggest one anscetral gene which underwent duplication through evolution. The location of the only two *Alu* repetitive sequences cannot explain the duplication of the four domains, although gene duplication followed by various DNA manipulation such as nucleic bases substitutions or crossovers have an important contribution to evolution of the genetic system. It is estimated that only 1,000-

10,000 primordial domains, e.g. exons, may have been sufficient for the structure of all existing proteins [93,94]. It is therefore highly unlikely, that the four saposin coding regions arose independently. While the sequence and size of the introns do not show any similarity, examining of the location of the introns in respect to the exonic sequence suggest an explanation for the evolutionary evolvement of this locus.

2. Model for the Evolution of Prosaposin

Intron location in respect to exonic sequence is conserved in many gene families such as the globin gene family [95,96], as well as in closely related genes. For example, the α and β chains of β-hexosaminidase have overall structural and functional similarities suggesting a common evolutionary origin, which is supported by their genomic structure. The exonic sequence of both genes reveals that 12 out of the 14 exons have 63% identity, moreover, 12 of the 13 introns ;errupt the coding regions at homologous positions [97]. Another example of two lysosomal glycohydrolases which have similar function and are highly conserved in their structure and intron placement is that of α-galactosidase A and α-N-acetylgalactosaminidase, which show an overall 46.9% amino acid identity. The structural organization of the genes reveals that all six introns of α-galactosidase A are located in identical positions on the exonic sequence as do their counterparts in the α-N-acetylgalactosaminidase gene, the latter containing two additional introns in the 5' and the 3' ends of the sequence [98]. When all four domains of prosaposin are optimally aligned, the location of the prosaposin introns reveals an interesting pattern. As shown in figure 6, the introns 5' to saposin A and B, and the introns 5' of saposin C and D, are in identical exonic location, as well as the first intron of saposin A and the only intron of saposin C. In addition, the first introns of saposin B and D as well as the last introns of saposin A and B, have identical locations with respect to the coding sequence. Using this pattern as a guide, the following evolutionary model is suggested (figure 7); One ancestral gene was duplicated followed by insertion of introns in a pattern similar to the one in saposin B and C. Then, another duplication of the two domains in tandem created all four domains followed by a double crossover which brought the organization of the gene to its present one.

It is now clear that the gene, in its present form, is a single transcription and translation unit.

B. Post-Translational Processing

1. Targeting to the Lysosome and Lysosomal Processing

As prosaposin is translated and inserted through the endoplasmic reticulum membrane, its signal peptide is clipped. The 16 amino acid long signal peptide was predicted from the cDNA sequence [61,62], and its cleavage site was recently confirmed by N-terminal amino acid sequence of prosaposin obtained from human milk which starts at Gly^{17} [99]. Addition of the mannose-6-phosphate group in the Golgi compartment directs the precursor protein into the lysosome via the mannose-6-phosphate receptor pathway. As with many lysosomal proteins, final maturation of the saposins requires proteolytic processing in the lysosome. Supporting data for the path described above come from finding of the precursor protein in the media of I-cell disease fibroblasts and normal fibroblasts treated with NH_4Cl [81,85]. The order and exact sites of the proteolytic clipping is not known yet, but it is likely that endoprotease activity occurs around the basic amino acid residues that are present between saposin A and B, saposin B and C and saposin C and D followed by exoprotease trimming to obtain each protein. An example of a well studied post-translational processing where a similar sequence of events was shown, is that of the human β-hexosaminidase. The enzyme is composed of two subunits; the α subunit comprised a major polypeptide (Mr=56,000) and a minor polypeptide (Mr=6,000) connected through a disulfide bond. The β subunit comprised three polypeptides, Mr = 11,000, 19,000 to 26,000, and 28,000, which are also connected through disulfide bonds. Both the α and the β subunits are synthesized as precursor polypeptides, pro-α and pro-β, respectively. After clipping of the signal peptide in the endoplasmic reticulum, extensive proteolytic processing in the lysosome takes place. The pro-α chain undergoes an internal proteolytic cleavage followed by a carboxy-terminal trimming in some tissue, e.g., placenta. The pro-β subunit undergoes two internal cleavages followed by minimal or no amino-terminal processing and carboxyl-terminal trimming, the latter process is tissue specific as studies of the placental enzyme indicate [100,101]. Another example of post-translational processing in the lysosome is that of lysosomal

endopeptidases, which are synthesized as inactive proenzymes. A single-chain active enzyme is then generated which can undergo a tissue-specific asymmetric cleavage in the lysosome to yield a two-chain active form. The ratio of single-chain enzyme to two-chain form varies between different tissues and species [102].

2. Secretion of Prosaposin

Another interesting aspect of the prosaposin processing is the fact that the precursor protein is secreted from various cells and is found in body fluids and different tissues, i.e., plasma, semen fluid, bile, pancreatic juice, milk, testis, cerebral gray matter, liver, and spleen [86-88] where it may have distinct function(s). This implies that either some of the precursor form does not follow the route into the lysosome but instead is secreted out, or that after reaching the lysosome some of the precursor protein escapes the last maturation step and is secreted out from the lysosome. It is well documented that all lysosomal enzymes are secreted to some extent, the secreted form is the proenzyme that has lost its signal peptide and bears the mannose-6-phosphate group [for Review: 103,104]. Therefore, it is thought that some percentage of lysosomal enzymes which use the mannose-6-phosphate receptor path escapes it by an unknown mechanism. Since there are mannose-6-phosphate receptors on the cell surface, the secreted enzymes can return into the lysosome by endocytosis. Such an example is the precursor form of β-hexosaminidase which is found in cell medium, and may even be active in this early biosynthetic form [105]. Another example is the lysosomal endopeptidases which are also secreted during biosynthesis, in most cells the proenzyme form is secreted. These inactive proenzymes forms, under certain physiologic conditions, may undergo conformational changes and become activated [102]. The prosaposin form found in various tissues and body fluids has a very similar tertiary structure as the one sent to the lysosome [87,99]. This implies that the prosaposin, in analogy to other lysosomal proteins, escapes the mannose-6-phosphate receptor path. More detailed studies are needed to determine the nature of the carbohydrate and phosphate groups that are attached to the secreted polypeptide backbone. At this stage of the research, the physiological function of prosaposin is not known and therefore its activity cannot be assessed (see Discussion section C4). Since the saposins are probably needed to activate lysosomal hydrolases in all tissues and prosaposin is found only in a few, than secretion of prosaposin is an additional,

regulated, tissue-specific process. As was described above for β-hexosaminidase and endoproteases, tissue-specific processing occurs in lysosomal proteins.

3. Post-Translational Regulation

Tissue-specific regulation was also suggested from the observation that the various activator proteins are elevated to different extents in tissues from lysosomal storage diseases. Since all four proteins are found on one transcription and translation unit, the level of each specific saposin must be regulated by a post-translational mechanism. The sequence of proteolytic events that produces the small activators might be under saposin-specific control in respect to the presence or absence of a specific lysosomal hydrolase, or to the accumulation of a specific substrate. In this case, after the formation of a mature saposin or a saposin that is not totally processed by proteases, its stability might depend on the local environment, i.e.; the tertiary structure of the protein, its accessibility and susceptibility to proteolysis, and the existence of proteases. The latter mechanism may imply that the regulation of prosaposin is a by product of regulation of expression of proteases. Since the intralysosomal level of proteases varies between tissues, as was described above for endopeptidases (see Discussion section B2 and B3), the availability of the proteases involved in prosaposin processing may vary. This will yield different quantities and quality of the products generated from the prosaposin precursor protein.

C. Function of the Four Saposins and the Precursor Protein

1. Function of Saposins as Measured *In Vitro*

In vitro assays defined the activation capacity of each saposin: Saposin A was reported to activate acid β-glucosidase and β-galactosidase when galactosyl ceramide is the substrate [59]. Saposin B was found to increase the hydrolysis of galactosyl sufatide by arylsulfatase A, G_{M1} ganglioside by acid β-galactosidase and globotriaosyl ceramide by α-galactosidase [67,106-108]. Saposin C activates acid β-glucosidase, β-galactosidase towards the hydrolysis of galactosyl ceramide, and under some assay condition, sphingomyelinase [58,65-67]. Saposin D was reported to specifically stimulate sphingomyelinase [108,109]. The mechanism of

activation is different among the saposins. While saposin A, C and D bind to the enzymes [69,77,109-111], saposin B binds to its substrates and acts as a detergent-like agent in solubilizing lipids [71-73]. Wynn [39] explained the specificity of interaction between saposin B and its substrate by a triple-binding-domain model where hydrophobic interactions, electrostatic interactions, and hydrogen bonding between sugar residues are important. Supportive data for this model came from binding studies which indicate that saposin B is a nonspecific glycosphingolipid binding protein [112]. As described in the introduction to this work (see Introduction section A1), a similar binding domain was proposed for phospholipid transfer protein which interacts with phospholipids [38], and to the active site of acid β-glucosidase [34,36].

2. Mutations Causing Saposin Deficiency Diseases

The physiologic role of saposin B and saposin C was demonstrated by their deficiency states. The deficiency of saposin B causes a variant of Metachromatic Leukodystrophy (MLD) disease [108]. The deficiency of saposin C causes a variant of Gaucher-like disease [8,9,10]. Until now, five mutations have been described as the cause of saposin deficiency diseases, the mutations are listed in table 2.

Table 2: Mutations Causing Saposin Deficiency Diseases

Deficiency of	Nucleic Acid Change base # as counted from ATG	Amino Acid Change as counted from the first Met	Ref.
Saposin B	650 C to T	Thr^{216} to Ile Loss of a glycosylation site.	113
Saposin B	33 bp insertion between 777 and 778	11 amino acid insertion	114,115
Saposin B	722 G to C	Cys^{241} to Ser	116
Saposin C	1154 G to T found in heterozygous state only.	Cys^{385} to Phe	117
Saposin C and B	1 A to T destroys the initiation codon.	No in-frame translation	118

Except for the mutation at the initiation codon, all other mutations have a unique affect on one of the four protein domains and thus impairs the function of only one of the saposins. In addition, all mutations must change the prosaposin amino acid sequence. Elimination of the glycosylation site of saposin B [113] may change the folding properties of the protein before and after its proteolytic processing in the lysosome. Insertion of 11 amino acids [114,115] causes significant changes in the

hydropathy profile and probably produces an unstable protein. The source of the nucleic acid insertion is in the second intron of the saposin B region [119,120]. It is interesting to note that a 9 base insertion, corresponding to the last nucleotides of the 33 bp insertion, occurs in normal individuals [61]. Both insertions skip the same splice site but activate two different cryptic splice sites. Two mutations changed conservative cysteines, the one in the sequence of saposin B substitutes a serine [116], and the other one in the sequence of saposin C substitutes a phenylalanine [117]. Elimination of a cysteine residue can disrupt the formation of a disulfide bridge and cause a severe change in the compact tertiary structure of the saposin creating an unstable protein. This is supported by the observation that in both reports the patients had no CRIM for the mutated saposin, but no effect on any other saposin that was tested was observed [8,116,117]. The last mutation described an elimination of the translation initiation Met [118]. This creates a situation where neither the prosaposin nor any of the activator proteins exist, which indicates that the functions of those proteins are important but not crucial for embryonic development.

3. Saposin B and C and their Potential Involvement in Other Genetic Diseases

From the above studies it is clear that saposin B has a physiological role in the hydrolysis of galactosyl sulfatide by arylsulfatase A, and saposin C in the hydrolysis of glucosyl ceramide by acid β-glucosidase. Mutations in those enzymes cause various forms of MLD and Gaucher disease. As described in the introduction to this work (see Introduction section A2), about 30 Gaucher disease alleles have been described but, there is no clear correlation between most genotypes and phenotypes. A brief introduction and update on the current research of MLD will be described here; the substrate for arylsulfatase A, galactosyl sulfatide, is found mainly as a component of the nervous system and is accumulated in MLD patients in the white matter of the central nervous system and in peripheral nerves [121]. There are three forms of the disease distinguished according to the age of onset: late infantile onset at 1 to 2 years of age, juvenile onset at 3 to 16 years of age, and adult onset between midteens to the seventh decade. There is a wide range of clinical manifestation of the disease with various neurological symptoms such as weakness, ataxia, progressive spastic tetraparesis, optic atrophy, and dementia. There are also various degrees of psychological symptoms such as

schizophrenia and depression. A correlation between the late infantile form and an allele that had three single base substitutions in the gene encoding arylsulfatase A was established [122]. Of the three changes one was a polymorphism, another one had no affect on enzymatic activity, but the third one caused the loss of a splice junction. This allele does not encode a functional arylsulfatase A and in the homozygous state it causes the most severe form of the disease, the late-infantile form. A second MLD allele is a result of a point mutation which reduces the enzymatic activity of arylsulfatase A to 3% as compared to the normal allele. Individuals who are homozygous for this allele may have either the juvenile form of the disease or the adult form. Such variability can occur even within one family [122,123]. A similar situation was observed in Gaucher disease; siblings who have the identical genotype at the acid β-glucosidase locus but very different progression rates of Gaucher disease [124]. These discrepancies between the genotype and the phenotype imply that in addition to the locus encoding the defective enzyme, another locus may play a role in modulating the enzymatic activity *in vivo*, the prosaposin locus is a natural candidate. To explain phenotypic variability between patients with identical genotype a change in the modulator protein is suggested. For example, minor changes in the amino acid sequence of saposin may alter *in vivo* enzymatic activity . Only one change in the nucleic acid sequence of saposin B was observed in cDNA from normal individuals. This is a 9 bp insertion caused by alternative splicing [61,116]. Although, the amino acid sequence for saposin B does not contain the 3 amino acid insertion [125], the cDNA insertion was found frequently enough to suggest a functional change. If the saposin B with additional 3 amino acids can function as an activator protein, it may have a different affinity and interaction with the substrate, and thus change the specific activity of arylsulfatase A.

4. Function of Prosaposin

The function of prosaposin is not known yet. According to a few reports prosaposin does not have any activator function similar to the four saposins [88,99]. As was mentioned earlier in this work, the precursor form is found at different levels in human tissues and body fluids. High amounts of prosaposin was found in human plasma. During all stages of the developing rat brain, some level of prosaposin is detected though, the ratio of precursor to mature forms increases

during development [86]. Prosaposin is most abundant during the stage of neuronal differentiation, it is also found in cerebrospinal fluid [87] which may suggest a role in differentiation of neural cells. Prosaposin is also found at high levels in human milk immediately after delivery and towards the mature milk lactating period [87]. A highly homologous rat protein, sulfated glycoprotein-1, is one of the major proteins secreted from sertoli cells but its function is not known yet [126]. Collard and coworkers proposed that since the sulfated glycoprotein-1 is secreted into the aduminal compartment of the seminiferous tubule during spermatogenesis, it may play a role in spermatogenesis by binding to glycolipids which are found on the cell surface.

The limited data from the reports on the location of prosaposin and estimation of the level of expression cannot provide a clear idea for the function of the protein. From the amino acid sequence, the structure of the protein may suggest that it contains binding sites which are also used by the small saposins. For example, the binding site of saposin B was shown to have non-specific affinity to different glycolipids, using their sugar residues. If this domain is available in the prosapsin molecule, it may play a role in the binding of different glycolipids which are found on cell surfaces and are thought to be involved in differentiation processes.

References

1. Brady RO, Kanfer JN, Shapiro D, 1965, Biochem Biophys Res Commun 18:221-225.

2. Patrick AD, 1965, Biochem J 97:17c-18c.

3. Desnick RJ, Gatt S, Grabowski GA, editors, 1982, Gaucher Disease: A Century of Delineation , Research; Alan R. Liss, New York p. 760.

4. Klibansky C, Hoffman J, Zaizov R, Matoth Y, Pinkhas J, deVries A, 1973, Biomedicine 19:345-348.

5. Wenger DA , Olson GC, 1981, in Lysosomes and Lysosomal Storage Disease; edited by Callahan JW and Lowden JA, Raven Press, New York.

6. Peters SP, Coffee CJ, Glew RH, Lee RE, Wenger DA, Li SC , Li YT, 1977, Arch Biochem Biophys 183:290-297.

7. Turner BM, Hirschhorn K, 1978, Am J Hum Genet 30:346-358.

8. Christomanou H, Aignesberger A, Linke R, 1986, Biol Chem Hoppe-Seyler 367:879-890.

9. Harzer K, Paton BC, Poulos A, Kustermann-Kuhn B, Roggendrof W, Grisar T, Popp M, 1989, Eur J Pediatr 149:31-39.

10.Christomanou H, Chabas A, Pampols T, Guardiola A,1989, Klin Wochenschr 67:999-1003.

11. Fredrickson DS, Sloan HR, 1978, in The Metabolic Basis of Inherited Disease: edited by Stanbury JS, Wyngaarden JB, Frederickson DS, McGraw-Hill, NY.

12. Zimran A, Gelbart T, Westwood B, Grabowski GA, Beutler E, 1991, Am J Hum Genet 49:855-859.

13. Beutler E, 1991, New Engl J of Med 325:1354-1360.

14. Fried K, 1958, Bull Res Council Isr, 7B:213-218.

15. Hodson P, Goldblatt J , Beighton P, 1979, Arch Dis Child, 54:707-716.

16. Dreborg S, Erikson A, Hagberg B, 1980, Eur J Pediatr 133:107-118.

17. Reiss O, Kato K, 1981, Am J Dis Child 43:365-398.

18. Maloney AFS, Cumming JN, 1960, J Neurol Neurosurg Psychiatry 23:207-216.

19. Sorge J, West C, Westwood B, Beutler E, 1985, Proc Natl Acad Sci USA 82:7289-7293.

20. Furbish FS, Blair HE, Shiloach J, Pentchev PG, Brady RO, 1977, Proc Natl Acad Sci USA 74: 3560-35

21. Erickson AH, Ginns EI, Barranger JA, 1985, J Bio Chem 260(26):14319-14324.

22. Bergmann JE, Grabowski GA, 1989, Am J Hum Genet 44:741-750.

23. Berg A, Grace ME, Grabowski GA, 1992, J Biol Chem, in review.

24. Varki A, Kornfeld A, 1981, J Biol Chem 256:11977-11980.

25. Kaplan A, Achord DT, Sly WS, 1977, Proc Natl Acad Sci USA 74:2026-2030.

26. Hickman S, Neufeld EF, 1972, Biochem Biophys Res Commun 49:992-999.

27. Takasaki S, Murray GJ, Furbish S, Brady RO, Barranger JA, Kobata A, 1984, J Biol Chem 259:10112-10117.

28. Strasberg PM, Lowden JA, Mahuran D, 1982, Can J Biochem 60:1025-2031.

29. Grabowski GA, Dagan A, 1984, Anal Biochem 141:267-279.

30. Grabowski GA, Gatt S, Horowitz M, 1990, CRC Crit Rev Biochem Mol Bio 25:385-414.

31. Pentchev PG, Brady RO, Hibbert SR, Gal AE, Shapiro P, 1973, J Biol Chem 248:5256-5261.

32. Blonder E, Klibansky C , DeVries A, 1976, Biochim Biophys Acta 431:45-48.

33. Dale GL, Villacorte DG, Beutler E, 1976, Biochem Biophys Res Commun 71:1048-1053.

34. Osiecki-Newman K, Legler G, Grace ME, Dinur T, Gatt S, Desnick RJ, Grabowski GA, 1988, Enzyme 40:173-188.

35. Dinur T, Osiecki KM, Legler G, Gatt S, Desnick RJ, Grabowski GA, 1986, Proc Natl Acad Sci USA 83:1660-1664.

36. Osiecki-Newman K, Fabbro D, Legler G, Desnick RJ, Grabowski GA, 1987, Biocim Biophys Acta 915: 87-100.

37. Wynn CH, 1986, Biochem J 240:921-924.

38. Van Paridon PA, Veisser AJ, Wirtz KW, 1987, Biochim Biophys Acta 898:172-180.

39. O'Neill R, Tokoro T, Kozak CA, Brady RO, 1989, Proc Natl Acad Sci USA 86:5049-5053.

40. Shafit-Zagardo B, Devine EA, Smith M, Arradondo-Vega F, Desnick RJ, 1981, Am J Hum Genet 33:564-575.

41. Ginns EI, Choudary PV, Tsuji S, Martin B, Stubblefield BK, Sawyer J, Hozier J, Barranger JA, 1985, Pro Natl Acad Sci USA 82: 7101-7105.

42. Horowitz M, Wilder S, Horowitz Z, Reiner O, Gelbart T, Beutler E, 1989, Genomics 4:87-96.

43. Reiner O, Wigderson M , Horowitz M, 1988, DNA 7(2):107-116.

44. Graves PN, Grabowski GA, Ludman MD, Palese P, Smith FI, 1986, Am J Hum Genet 39:763-774.

45. Riener O, Wilder S, Givol D, Horowitz M, 1987, DNA 6(2): 101-108.

46. Fabbro D, Desnick RJ, Grabowski GA, 1987, Am J Hum Genet 40:15-31.

47. Zimran A, Sorge J, Gross E, Kubitz M, West C, Beutler E, 1990, J Clin Invest 85:219-222.

48. Theophilus B, Latham T, Grabowski GA, Smith FI, 1989, Am J Hum Genet 45:212-225.

49. Zimran A, Gross E, West C, Sorge J, Kubiz K, Beutler E,1989, Lancet 2:349-352.

50. Tsuji S, Choudary PV, Martin BM, Stubblefield BK, Major JA, Barranger JA, Ginns EI, 1987, New Eng J Med 316:570-575.

51. Grace ME, Graves PN, Smith FI, Grabowski GA, 1990, J Biol Chem 265:6827-6835.

52. Ohashi T, Mu Hong C, Weiller S, Tomich JM, Aertz JMFG, Tager JM, Barrangere JA, 1991,J Biol Chem 266:3661-3667.

53. Latham T, Grabowski GA, Theophilus BDM, Smith FI, 1990, Am J Hum Genet 47:79-86.

54. Dahl N, Lagerstrom M, Erickson A, Pettersson U, 1990, Am J Hum Genet 47:275-278.

55. Dahl N, Erickson A, Hammarstrom-Heeroma K, Pettersson U, 1988, Genomics 3:296-298.

56. He G-S, Grabowski GA, 1992, Am J Hum Genet, In Press.

57. Gelbart T, Kuhl J, Sorge J, West C, 1991,Proc Natl Acad Sci USA 88:10544-10547.

58. Ho MW, O'Brien JS, 1971, Proc Natl Acad Sci USA 68:2810-2813.

59. Morimoto S, Martin BM, Yamamoto Y, Kretz KA, O'Brien JS, Kishimoto Y, 1989, Proc Natl Acad Sci USA 86:3389-3393.

60. O'Brien JS, Knetz KA, Dewji NN, Wenger DA, Esch F, Fluharty AL, 1988, Science 241:1098-1101.

61. Nakano T, Sandhoff K, Stumper J, Christomanou H, Suzuki K, 1989, J Biochem 165:151-154.

62. Rorman EG, Grabowski GA, 1989, Genomics 5:486-492.

63. Reiner O, Dagan O, Horowitz M, 1989, J Mol Neurosci 1:225-233.

64. Sano A, Radin NS, 1988, Biochem Biophys Res Com 154(3):1197-1203.

65. Wenger DA, Sattler M, Roth S, 1981, Trans Am Soc Neurochem 12:210.

66. Christomanou H, 1980, Hoppe-Seyler's Z. Physiol Chem 361:1489-1502.

67. Wenger DA, Sattler M, Roth S, 1982, Biochim Biophys Acta 712:639-649.

68. Ho MW, Light NL, 1973, Biochem J 136:821-823.

69. Berent SL, Radin NS, 1981, Biochim Biophys Acta 664:572-582.

70. Basu A, Glew RH, Daniels LB, Clark LS, 1984, J. Neuropath Exp Neuro 41:630.

71. Mehl E, Jatzkewitz H, 1964, Hoppe-Seyler's Z Physiol Chem 339:260-276.

72. Fischer G, Jatzkewitz H, 1975, Hoppe-Seyler's Z Physiol Chem 356:605-613.

73. Conzelman E, Burg J, Stephan G, Sandhoff K, 1982, Eur J Biochem 123:455-464.

74. Ho MW, 1975, FEBS Lett 53:243-247.

75. Peters SP, Coyle P, Coffee CJ, Glew RH, Kuhlenschmidt MS, Rosenfeld L, Lee YC, 1977, J Biol Chem 252:563-573.

76. Berent SL, Radin NS, 1981, Arch Biochem Biophys 208:248-260.

77. Fabbro D, Grabowski GA, 1991, J Biol Chem 266:15021-15027.

78.Sano A, Radin NS, Johnson LL, Tarr GE,1988, J Biol Chem 263:19597-19601.

79. Prence E, Chakravorti S, Basu A, Clarck LS, Glew RH, Chambers JA, 1985, Arch Biochem Biophys, 1985, 236:98-109.

80. Weiler S, Carson WD, Ohashi T, Kishimoto Y, Morimoto S, O'Brien JM, Aerts JM, Tager M, Barranger JA, Tomich JM, 1989, Am J Hum Genet 45:A12.

81. Fujibayashi S, Wenger DA, 1986, J Biol Chem 261:15336-15343.

82. Fujibayashi S, Wenger DA, 1986, Biochim Biophys Acta 875:554-562.

83. Inui K, Kao FT, Fujibayashi S, Jones C, Morse HG, Law ML, Wanger DA, 1985, Hum Genet 69:197-200.

84. Kao F-T, Law ML, Hartz J, Jones C, Zhang X-L, Dewji NN, O'Brien JS, Wenger DA, 1987, Somatic Cell Mol Genet 13:685-693.

85. Varon R, Kleijer WJ, Thompson EJ, d'Azzo A, Hum Genet 62:66-69.

86. Sano A, Hineno T, Mizuno T, Kondoh K, Ueno S-i, Kakimo Y, Inui K, 1989, Biochem Biophys Res Commun 165:1191-1197.

87. Hineno T, Sano A, Kondoh K, Ueno S-i, Kakimoto Y, Yoshida K-i, 1991, Biochem Biophys Res Commun 176:668-674.

88. O'Brien JS, Kishimoto Y, 1991, FASEB J 5:301-308.

89. Kleinschmidt T, Christomanou H, Braunitzer G, 1987, Biol Chem Hoppe-Seyler 368:1571-1578.

90. Inui K, Wenger DA, 1983, J Clin Invest 72:1622-1625.

91. Morimoto S, Yamamoto Y, O'Brien JS, Kishimoto Y, Proc Natl Acad Sci USA1990, 87:3493-3497.

92. Datta SC, Snider RM, Radin NS, 1986, Biochimica et Biophysica Acta 877:387-398.

93. Doolittle RF, 1981, Science 214:149-159.

94. Dorit RL, Schoenbach L, Gilbert W, 1990, Science 250:1377-1382.

95. Jhiang SM, Garey JR, Riggs AF, 1988, Science 240:334-336.

96. Jhiang SM, Riggs AF, 1989, J Biol Chem 264:19003-19008.

97. Proia LR,1988, Proc Natl Acad Sci USA 85:1883-1887.

98. Wang AM, Desnick RJ, 1991, Genomics 10:133-142.

99. Kondoh K, Hineo T, Sano A, Kakimoto Y, 1991, Biochem Biophys Res Comm 181:286-292.

100. Mahuran DJ, Neote K, Klavins MH, Leung A, Gravel RA, 1988, J Biol Chem 263:4612-4618.

101. Mahuran DJ, 1990, J Biol Chem 265:6794-6799.

102. Erickson AH, 1989, J Cell Biochem 40:31-41.

103. von Figura K, Hasilik A, 1986, Ann Rev biochem 55:167-193.

104. Pfeffer SR, 1991, Curr Top Microbiol Immunol 170:43-65.

105. Stirling J, Leung A, Gravel RA, Mahuran D, 1988, FEBS-Lett 231:47-50.

106. Li SC, Kihara H, Serizawa S, Li YT, Fluharty AL, Mayes JS , Shapiro LJ, 1985, J Biol Chem 260:1867-1871.

107. Cozelman E, Sandhoff K, 1979, Hoppe-Seyler's Z Physiol Chem 360:1837-1849.

108. Inui K, Emmett M, Wenger DA, 1983, Proc Natl Acad Sci USA 80:3074-3077.

109. Morimoto S, Martin BM, Kishimoto Y, O'Brien JS, 1988, Biochem Biophys Res Commun 156(1):403-410.

110. Furst W, Machleidt W, Sandhoff K, 1988, Biol Chem Hoppe-Seyler 369:317-328.

111. Morimoto S, Kishimoto Y, Tomich J, Weiler S, Ohashi T, Barranger JA, Kretz KA, O'Brien JS, 1990, J Biol Chem 265:1933-1937.

112. Schwarzmann G, Sandhoff K, 1991, J Biochem 200:591-597

113.Rafi AM, Zhang X-L, DeGala G, Wenger DA, 1990, Biochem Biophys Res Comm 166:1017-1023.

114.Kretz KA, Carson GS, Morimoto S, Kishimoto Y, Fluharty AL, O'Brien JS, 1990, Proc Natl Acad Sci USA 87: 2541-2544.

115.Zhang X-L, Rafi MA DeGala G, Wenger DA, 1990, Proc Natl Acad Sci USA 87:1426-1430.

116.Holdschmidt H, Sandhoff K, Kwon HY, Harzer K, Nakano T, Suzuki K, 1991, J Biol Chem 266: 7556-7560.

117.Schnabel D, Schroder M, Sandhoff K, 1991, FEBS 284:57-59.

118.Schnabel D, Schroder M, Furst W, Klein A, Hurwitz R, Zenk T, Weber J, Harzer K, Patom BC, Poulus A, Suzuki K, Sandhoff K, 1992, J Biol Chem 267:3312-3315.

119. Rorman EG, Scheinker V, Grabowski GA, 1992, Genomics 13:312-318.

120. Holtschmidt H, Sandhoff K, Furst W, Kwon H-Y, Schnabel D, Suzuki K, 1991, FEBS 280:267-270.]

121. Kolodny EH, 1989, in Metabolic basis of inherited disease: edited by Scriver CR, Beaudet AL, Sly WS, Valle D, 6th edition, McGraw-Hill, NY.

122. Polten A, Fluharty AL, Fluharty CB, Kappler J, von Figura K, Gieselmann V, 1991, N Engl J Med 324:18-22.

123. Clarke JTR, Skomorowski AM, Chang PL, 1989, J Med Genet 33:10-13.

124. Latham TE, Theophilus BDM, Grabowski GA, Smith FI, 1991, DNA and Cell Biol 10(1):15-21.

125. Furst W, Schubert J, Machleidt W, Meier H, Sandhoff K, 1990, Eur J Biochem 192:709-714.

126. Collard MW, Sylvester SR, Tsuruta JK, Griswold MD, 1988, Biochemistry 27:4457-4564.