Gaucher Disease: Expression of Human Glucocerebrosidase in *Pichia pastoris* and Evolution of Glucocerebrosidase Gene and Pseudogene in Primates

By

Julie Rachel Wafaei
B.Sc., McMaster University, Hamilton, ON, Canada, 1997

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biology

We accept this thesis as conforming
to the required standard.

[redacted]
Dr. F.Y.M. Choy, Supervisor (Department of Biology)

[redacted]
Dr. D. B. Levin, Department Member (Department of Biology)

[redacted]
Dr. W. Kusser, Department Member (Department of Biology)

[redacted]
Dr. T.W. Pearson, External Examiner (Department of Biochemistry)

Abstract

Gaucher disease, the most common lysosomal storage disorder, is caused by insufficient levels of the enzyme glucocerebrosidase (GBA). This thesis examines two aspects of glucocerebrosidase: (1) the heterologous expression of glucocerebrosidase in *Pichia pastoris*, and (2) the molecular evolution of glucocerebrosidase functional gene and pseudogene in primate species.

The development of recombinant GBA by economical methods is of importance for treating Gaucher patients by enzyme replacement therapy. Here I explore two variables that may inhibit GBA expression in *Pichia pastoris*, proteolytic instability and gene dosage. These variables were found to have minimal impact on the consistently low levels of enzymatically and immunologically active recombinant GBA expressed in *P.pastoris*.

The GBA genes duplicated approximately 40 million years ago and subsequently diverged. Here I investigate the molecular evolution of GBA and its pseudogene by sequencing approximately 1.1 kb of the C-terminal region of nine GBA genes in chimpanzee, gorilla, orangutan, baboon, and squirrel monkey. These data indicate that gene conversion has affected the evolution of GBA and its pseudogene, as well as provide information on GBA gene copy number and the functionality of the GBA genes.

Examiners:

[redacted]

Dr. F.Y.M. Choy, Supervisor (Department of Biology)

[redacted]

Dr. D. B. Levin, Department Member (Department of Biology)

[redacted]

Dr. W. Kusser, Department Member (Department of Biology)

[redacted]

Dr. T.W. Pearson, External Examiner (Department of Biochemistry)

Table of Contents

List of Figures

List of Tables

List of Abbreviations

4MUGP	4-methylumbelliferyl-β-D-glucopyranoside
AA	amino acid
A	adenine
AOX	alcohol oxidase gene
bp	base pair
BMGY	buffered glycerol-complex medium
BMMY	buffered methanol-complex medium
C	cytosine
°C	degrees Celsius
cDNA	complementary DNA
CHO	Chinese hamster ovary
Da	Dalton
ddH_20	distilled and deionized water
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetate
Etbr	ethidium bromide
×g	relative centrifugal force
g	gram
G	guanine
GAP	glyceraldehyde 3-phosphate dehydrogenase
GBA	glucocerebrosidase
GBA	glucocerebrosidase functional gene
ψ*GBA*	glucocerebrosidase pseudogene
His	histidine
hr	hour
IgG	immunoglobul G
kb	kilobase

kDa	kilo-Dalton
lacZ	β-galactosidase gene
LB	Luria-Bertani medium
M	molar
MD	minimal dextrose
min	minute
ml	millilitre
MM	minimal medium
MMG	minimal medium with glycerol
MMM	minimal medium with methanol
mol	mole
mRNA	messenger ribonucleic acid
MTX	metaxin
Mut	methanol utilization
Mut^+	wild type for methanol utilization
Mut^s	methanol utilization slow
n	nano
nt	nucleotide
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PHO1	acid phosphatase
P. pastoris	*Pichia pastoris*
PTEF1	transcription elongation factor 1 gene promoter
RDH	regeneration dextrose histidine
rGBA	recombinant glucocerebrosidase
SDS	sodium dodecyl sulfate
Sf	*Spodoptera frugiperda*
Sh ble	*Streptoalloteichus hindustanus ble*
T	thymidine
TBS	tris-buffered saline
TTBS	tris-buffered saline with tween-20

U	unit
μ	micro
X-gal	5-bromo-4-chloro-3indolyl β-D-galactoside
YNB	yeast nitrogen base
YND	yeast nitrogen dextrose

Acknowledgments

There are an enormous number of people that I need to thank for helping me during my M.Sc. research. My supervisor, Dr. Francis Choy, has provided me with inexhaustable amount of support and encouragement. My other committee members, Dr. Wolfgang Kusser and Dr. David Levin, have helped me surmount the challenges I've encountered in my research by providing valuable guidance. The people in my lab, past and present, Graham Sinclair, Andrea Vaags, Natalie Devost, Karen Wong, Camille Wolfe, Tessa Campbell, Lisa Sharp, and ChaoWei, have not only helped me with my research but have also made the lab a very enjoyable place to be. There are many others in the Biology department, including Dr. Ben Koop, Mike Wilson, Joanne Whitehead, and Ute Rink that have given me guidance in my research and helped with various aspects of my research work. Last but certainly not least, I need to thank my boyfriend, Neil Borecky, for his continuous support and my parents for reluctantly accepting my move "out west".

Chapter 1 Introduction

1.1. Gaucher disease

Gaucher disease is an autosomal recessive disorder due to a defect in a single gene, glucocerebrosidase, which results in excessive storage of glycosphingolipids in macrophages. Gaucher disease is the most common lysosomal storage disorder,[49] as well as the most prevalent Ashkenazi Jewish genetic disease.[44] The hallmark symptoms of this disorder are splenomegaly and the appearance of unusual, large cells called "Gaucher cells" in the spleen and bone marrow. Additionally, hepatomegaly, bone abnormalities, anemia, and neuronal degradation may accompany the disorder.[7;44;48;49]

1.1.1. History of Gaucher disease

In 1882 Philippe Charles Ernest Gaucher provided the first description of a female patient with an enlarged spleen that contained abnormally large cells with a "foamy" appearance.[44] This disorder, which was the first lysosomal storage disorder discovered, became known as Gaucher disease. Other symptoms of Gaucher disease, including involvement of the liver, bone marrow, and lymph nodes, were detected in 1904.[12] The heterogeneity of symptoms in Gaucher disease and its familial linkage was already noticed by the turn of the century.[12] The large array of symptoms continued to expand with the discovery of neurological impairment in an infant in 1927. This form of the disorder became known as type 2 Gaucher disease.[12] Type 3 Gaucher disease was first noticed in 1959 when a teenager with neurological impairment was diagnosed and is characterized by later neurological onset.[12]

Since Gaucher disease was first described, considerable effort has been expended to understand the disorder. Glucosylceramide (or glucocerebroside) was identified as the accumulated lipid in organs and tissues of Gaucher patients in 1934.[12] Some 30 years later, in 1965, it was discovered that decreased activity of the enzyme glucocerebrosidase (GBA) is responsible for the accumulation of this lipid.[18] With that discovery, Gaucher disease became the second lysosomal storage disorder to have its enzymatic defect delineated.[44] Discovering the role of GBA in Gaucher disease allowed considerable progress to be made in understanding the disorder.

Two decades after the discovery of the importance of GBA in Gaucher disease, the gene coding for the enzyme was localized to the long arm of chromosome 1, on band 21.[41] The coding regions were sequenced, and from that the amino acid sequence was deduced and later confirmed by direct sequencing.[23;40;41;85;90] The genomic sequence of GBA and the highly homologous pseudogene were obtained in 1989.[54] Sequencing of glucocerebrosidase from a patient suffering from Gaucher disease led to the discovery of the gene's first mutation,[89] and since then over 110 mutations have been found.[11]

In the 1990s, Gaucher disease became the first disorder to be successfully treated by enzyme replacement therapy.[3;37] Patients saw a significant improvement in many symptoms of the disorder including splenomegaly, hepatomegaly, hematological deficiencies, and to a lesser extent bone demineralization.[48] Unfortunately the process whereby the enzyme is produced is extremely expensive.[48] Consequently the cost of the treatment prohibits many patients from benefiting from it and raises difficult ethical questions.[6;35;70] Current research is aimed at devising more economical methods of

producing glucocerebrosidase, which will make enzyme replacement therapy more widely accessible.

1.1.2. Symptoms of Gaucher disease

Gaucher disease is usually caused by a deficiency in GBA, which leads to the accumulation of its substrate glucoslyceramide. A rare exception results from a deficiency in the GBA co-factor, saposin C, which also causes a build-up of glucosylceramide.[94] There are a wide range of symptoms associated with Gaucher disease and the phenotype is extremely heterogeneous. Gaucher disease is delineated by three major types based on the absence of neurological symptoms (Type 1) and the severity of primary central nervous system involvement (Type 2 > Type 3).[49]

Type 1, the most common variant, is characterized by a lack of neurological symptoms.[44;48;49;65] The presence of enlarged lipid-laden macrophages in the various organs leads to the dysfunction of the liver and spleen, the displacement of bone marrow by these "Gaucher cells", and frequent bone infarctions and fractures.[49] Both the age of onset and phenotypes are very heterogeneous, ranging from patients with no clinical symptoms to children with splenohepatomegaly, anemia, and severe bone malformations.[44]

Type 2 is the very rare and severe infantile variant, with onset age ranging from birth to 6 months and death usually occurring before the age of two.[48;49;65] Many of the same phenotypes as those characteristic of type 1 are involved, although there is little heterogeneity in symptoms. In addition, all type 2 patients suffer from a neurological disease primarily involving oculomotor abnormalities, which often manifest themselves

in rapid head thrusts as a compensatory mechanism.[44]

Gaucher disease type 3 is also quite rare and its severity is intermediate between types 1 and 2.[48;49;65] The median age of onset is 1 year and symptoms resemble those in type 2 Gaucher disease, except they are less severe. Early symptoms usually involve splenohepatomegaly and eye movement dysfunction with other neurological manifestations developing later on.[65] The phenotypes can also vary considerably between individuals, although not to the same degree as in type 1 Gaucher disease.[16;44]

1.1.3. Incidence rates of Gaucher disease

Gaucher disease is the most common lysosomal storage disorder.[47] The disorder is pan-ethnic, although in some ethnic populations, such as the Ashkenazi Jewish population,[47] Gaucher disease is much more common. In the general population one out of 40,000 to 60,000 individuals are affected with Type 1 Gaucher disease.[44] In Ashkenazi Jews this rate increases to 1 out of 400 to 600.[13;97] Type 2 Gaucher disease occurs in fewer than 1 in 100,000 live births.[44] While type 3 is also quite rare and has an incidence of 1 in 50,000 to 100,000,[44] although in Norrbottnian Swedes the incidence of type 3 Gaucher disease is elevated.[44]

1.1.4. Mutations causing Gaucher disease

To date over 110 mutations causing Gaucher disease have been discovered.[11] The vast majority of these mutations are rare. Only five mutations are responsible for 97% of Gaucher alleles in Ashkenazi Jews and 75% in the non-Jewish population population.[13] The types of mutations in GBA include single nucleotide substitutions, deletions,

insertions, splice site abnormalities and complex or recombinant alleles resulting from interactions with the pseudogene.[11] Gene conversion and unequal crossing over between the GBA gene and its highly homologous pseudogene is responsible for approximately 10% of the mutations detected in Gaucher patients.[11;49]

The extensive elucidation of the GBA gene mutations stems from the need to explain and understand the heterogeneity of Gaucher symptoms. Differences in mutation affect whether the mutated GBA is catalytically unstable, poorly activated by its co-activator and/or phospholipids, or less thermostable.[2] Mutations can also result in a null allele, which cannot produce any GBA.[11;12;47] Despite the documentation of biochemical differences between mutations, the correlation between genotype and phenotype is not exact. The level of residual GBA activity in patients can sometimes be linked to the severity of the disorder, however, identical mutations have been found to be responsible for a wide spectrum of symptoms and varying levels of severity.[44;67]

1.1.5. Animal models for Gaucher disease

There are no adequate animal models for Gaucher disease. Gaucher disease was once detected in a dog, an Australian Sidney Silkie terrier, but has never been found to occur naturally in any other animals.[92] A mouse model has been created, wherein mice are homozygous for a null glucocerebrosidase allele. Unfortunately, this mutation causes the mice to die within 24 hours of birth.[91] The mice possess an unusual translucency of skin that is comparable to a condition seen in infants with Gaucher disease who die early in infancy.[91] This suggests that without any GBA production, neither mice nor humans

can live long. Consequently attempts are being made to create a mouse model which contains a less severely mutated GBA gene.[2]

1.2. Treatment of Gaucher disease

The future for Gaucher patients is considerably brighter now than a decade earlier. In the past only some of the symptoms of Gaucher disease could be treated, using supportive therapies such as splenectomy, bone marrow transplant, iron/vitamin supplementation, joint replacement, and pain management.[48] Since 1991, enzyme replacement therapy has been commercially available to treat type 1 Gaucher disease and has proven very successful at reversing many symptoms of the disorder.[7] In the future, gene therapy may provide a cure for Gaucher disease.

1.2.1. Diagnosis of Gaucher disease

Gaucher disease is primarily diagnosed through enzymatic tests and DNA analysis.[5;12;44;65] Morphological diagnosis based on identification of Gaucher cells is also possible, although the invasive nature of the procedure and nonspecific results makes this less desirable.[65] Enzymatic diagnosis is accomplished by assaying for the level of glucocerebrosidase activity using the natural sphingolipid substrate, N-palmitoyldihydroglucocerebroside, or more commonly the artificial fluorogenic substrate 4-methlyumbelliferyl β-D-glucopyranoside (4MUGP).[49] The patient sample can be from peripheral blood leukocytes or cultured skin fibroblasts. Prenatal diagnosis from amniotic fluid cells or chorionic villae is also possible.[49] The results from the enzymatic assay can, however, be somewhat inaccurate due to the inability of the *in vitro*

environment to perfectly mimic that of the host cell. Therefore the assay cannot differentiate between individuals affected with type 1 Gaucher disease from those with the more severe type 2 or type 3, or accurately identify heterozygotes.[44] Approximately 10-20% of heterozygotes have relatively normal GBA activity levels, which further complicates identification using this method.[17] The most conclusive diagnosis analyzes the patient's DNA using PCR based techniques, including restriction fragment length polymorphism, direct sequence analysis, and use of oligonucleotide probes.[6] Although there are a number of mutations that cause Gaucher disease, most affected individuals carry two of the same 5 mutations. By screening for these mutations the chance that both mutations will be identified in Ashkenazi Jews is 93.5%, while in other populations it is only 45.9%.[10;13] Knowledge of the specific mutation may also provide an indication of the severity of the progression of the disease, which will help in determining the appropriate course of treatment.

1.2.2. Therapy for Gaucher disease

Enzyme therapy was first proposed as a likely treatment for Gaucher disease in 1966.[18] By 1974, enzyme replacement therapy with placental purified GBA had been used to treat the first Gaucher patient. The therapy resulted in some improvements in liver and blood glucoslyceramide levels, however it was not until the mid-1980s that larger clinical trials were attempted.[5] The results were unsatisfactory, primarily due to the prompt hepatocytic clearance of the infused enzyme.[86] Work ensued on modifying the enzyme so that it was more effectively targeted to the macrophages. When a mannose-terminated variant of GBA was used for enzyme replacement therapy in clinical

trials, the results were very good.[3;83] In 1991, the therapy was approved for treatment of type 1 Gaucher disease. Now some 2,000 Gaucher patients world-wide are able to benefit from enzyme replacement therapy.[39] Unfortunately, the number of patients that receive this therapy is limited by the excessive cost of the treatment. At $350,000 US each year for an average weight individual, glucocerebrosidase replacement therapy has been dubbed "the most expensive drug on the planet" by the press.[30]

Enzyme therapy is a life long treatment that is intravenously administered on a weekly basis. The FDA recommended dosage is 60 U/kg of body weight every two weeks, although the controversial use of decreased dosage to reduce the expense of the treatment is common.[6;45;48] Within the first year of treatment, patients generally notice a significant decrease in splenic volume and those with hepatomegaly also experience a normalization in liver volume.[7;48] Anemic patients experience an increase in their hemoglobin levels. [7;48] After two years of treatment the blood platelet levels increase so that those suffering from mild thrombocytopaenia will have normalized counts.[7;48] The improvements in skeletal abnormalities are less apparent, although it has been reported that the frequency of bone crisis decreases and the mid-femoral cortical increases in thickness.[48] Unfortunately improvement of neurological symptoms is limited with enzyme therapy, likely resulting from the enzyme's inability to be transported across the blood-brain barrier. This restricts the effectiveness of the treatment for patients with types 2 and 3 Gaucher disease.[79] Other complexities in recovery arise in patients where the disorder has progressed to a level where permanent damage has been done to the tissues in the liver, spleen, and bone marrow.[48] Additionally, there are some adverse side effects to the treatment which may include weight gain, depression, menstrual

disorders, and false positive pregnancy tests.[48] Furthermore, 15% of patients develop IgG antibodies and 46% of these individuals display hypersensitivity symptoms.[39]

Although enzyme replacement therapy is presently the preferred course of treatment for type 1 Gaucher patients, other treatments may be used concurrently or in its place. Removal of an enlarged spleen through splenectomy increases the level of blood platelets, however, patients are at an increased risk of hepatomegaly and hepatic dysfunction after this procedure.[16] Joint replacement is available for those suffering from osteonecrosis.[16] Bone marrow transplants can provide very positive results for Gaucher patients suffering from non-neurological symptoms. It is, however, difficult to find a matching bone marrow donor and the procedure is extremely risky, carrying a 10% mortality rate.[6] The positive results from successful bone marrow transplants suggest that in the future correction of patients' own hematopoietic stem cells through gene transfer will provide a cure for some types of Gaucher disease.[6] Presently three Gaucher patients are being treated with gene therapy in phase 1 clinical trials.[1] The results so far suggest that their GBA activity has been increased to that of a heterozygous individual.[1]

1.3. Glucocerebrosidase

1.3.1. Biochemistry of glucocerebrosidase

Glucocerebrosidase is a monomeric glycoprotein that is normally bound to the inner membrane of the lysosome. It is 497 a.a. in length and generally has a molecular mass of between 55.6 kDa to 65 kDa, depending on its level of glycosylation.[49] Although the tertiary structure of the protein has not been determined, many additional features about the structure of the enzyme are known. Glucocerebrosidase has three disulfide

bridges and one free cysteine. The disulfide bonds are located at $Cys_4 - Cys_{16}$, $Cys_{23} - Cys_{342}$, $Cys_{126} - Cys_{248}$ and the free cysteine is positioned at amino acid 18.[61] It is likely that the bonds increase the stability of the enzyme. Disulfide exchange does not normally occur, however at a higher pH it might be responsible for the enzyme's inactivation.[61] The active site is located near the carboxy terminus and consists of three domains, a hydrophilic catalytic site that recognizes β-glucosyl moieties, a hydrophobic region that binds the alkyl chains of the substrate glucosylceramide, and a hydrophobic site that interacts with negatively charged lipids.[33]

Glucocerebrosidase possesses a 19 a.a. leader sequence that transports the enzyme across the ER and enroute is proteolytically cleaved off.[48] The enzyme is further modified by co-translational glycosylation. There are five predicted N-glycosylation sites and the first four are normally occupied.[4] Glycosylation is crucial for creating a functional enzyme, and glycosylation of the first site is particularly important for catalytic activity.[4] Unlike other lysosomal enzymes, glucocerebrosidase is not targeted to the lysosome by 'high mannose' type oligosaccharide chains. It is likely that GBA is transported by membrane flux, which is similar to the targeting system integral membrane proteins use.[2] Once the enzyme is transported to the lysosome it is retained by its hydrophobic sequences.[2]

1.3.2. Activators of glucocerebrosidase

For GBA to be active it requires a co-activator protein, saposin C. The importance of saposin C is illustrated by reports of individuals with saposin C deficiencies presenting identical symptoms as Gaucher patients with GBA deficiencies.[2]

Mice with saposin C deficiencies also demonstrate increased levels of glucosylceramide.[34] In Gaucher patients, saposin C accumulates in the spleen at levels up to 37 times greater than normal.[15] Two saposin C proteins complex to one GBA enzyme. This causes a change in the enzyme that brings about the dimerization of two glucocerebrosidase molecules. The fully active GBA molecule likely consists of four saposin C proteins, two GBA enzymes and four negative acids.[94]

Negatively charged acidic phospholipids, such as phosphatidylserine or gangliosides, are required for glucocerebrosidase to be active. Glucocerebrosidase's *in vivo* membrane association suggests that specific membrane derived lipids may be required and it has been shown that negatively charged lipids enhance the enzyme's activity.[32] The length of the lipid's fatty acid chain is not critical, although a negative charge must be present on the lipid for it to affect the enzyme's activation.[43] These negatively charged phospholipids bind to GBA, although it is not known whether this occurs before or after dimerization.[72]

1.3.3. Substrate of glucocerebrosidase

The substrate of GBA is N-acyl-sphingosyl-1-O-β-D-glucoside, which is also known as glucosylceramide, ceramide β-glucoside, glucosylceramide, or glucocerebroside. Glucosylceramide is a metabolic intermediate in the synthesis or catabolism of complex glycosphingolipids such as gangliosides or globosides. When the glycolipids are no longer needed, they are exported to the lysosomes where they are catabolized in a series of steps and glucosylceramide is the penultimate intermediate in this pathway.[47;49] Glucosylceramide is distributed widely in mammalian tissues in small

quantities, where it is mostly localized in cellular membranes.[44] Different forms of glucosylceramide exist depending on tissue location. Primarily, the variants from visceral sources differ from those in the brain in terms of the fatty acid acyl and sphingosyl moieties.[47]

Glucocerebrosidase cleaves its substrate, glucosylceramide, into ceramide and glucose. This cleavage occurs in the lysosome, after which ceramide is transported to the golgi complex where it is further degraded.[63] Reduced glucosylceramide degradation in Gaucher patients causes this intermediate, glucosylceramide, to accumulate in the lysosomes of visceral tissues at levels approaching 100 times higher than normal.[44;63]

1.3.4. Genomic organization of glucocerebrosidase

The complete genomic sequence of nucleotides for the functional glucocerebrosidase gene and its pseudogene have been elucidated.[54] The functional gene has 11 exons and 10 introns contained within a 7604 bp genomic fragment on chromosome 1q21.[54] Genetic fine localization using pairwise and multi-point analysis has identified the 3.2 cM interval between markers D1s305 and D1S2624 as being the most probable location for the GBA gene.[29] The GBA pseudogene is located 16 kb 3' to the functional gene and is slightly smaller.[54]

Variations in the nucleotide sequence of the GBA gene also exist as evidenced by the 12 polymorphic sites and over 110 different mutations that have been identified in the gene.[11;14] Thus far several different types of mutations have been found, including point mutations in the exonic regions, splice site mutation, nucleotide insertion or deletion, crossovers between the gene and the pseudogene, and gene conversions.[5] The close

proximity of the functional gene and pseudogene, and their marked similarity results in numerous recombination events between the two. Approximately 10% of the mutations in GBA are caused by this type of interaction between the functional gene and pseudogene.[11]

Structural analysis of the GBA genomic sequence indicates the existence of promotor elements in the form of two TATA boxes and two CAAT-like boxes situated respectively between nucleotides (-23) to (-27), (-33) to (-38), (-90) to (-94), and (-96) to (-99).[47] Promotor elements containing TATA and CAAT boxes are typical of highly regulated genes.[47] There are two in frame ATG initiator codons for glucocerebrosidase, the upstream ATG results in a 39 amino acid leader sequence while the downstream ATG produces a more hydrophobic 19 amino acid leader sequence.[84] Although both start codons can be used, only the downstream ATG is used.[44]

1.3.5. Glucocerebrosidase pseudogene

The GBA pseudogene is 96% homologous to the functional gene and located 16kb downstream from it on chromosome 1q21.[54] The differences in the coding sequence of the pseudogene compared to that of the functional gene are due to nucleotide substitutions and deletions. There are two exonic deletions, 5 bp and 55 bp in exons 4 and 9 respectively, which render the pseudogene unable to produce functional enzyme. Increased divergence occurs in the intronic regions of the pseudogene, resulting in a ~1.6 kb size difference between genes. The functional gene contains four large regions in introns 2, 4, 6, 7 sized 313 bp, 626 bp, 320 bp, and 277 bp respectively that are absent in

the pseudogene. These intronic deletions in the pseudogene are represented in the structural gene by *Alu* sequences flanked by direct repeats.[54]

Both the GBA functional gene and pseudogene have promoter regions, although the promoter of the functional gene is at least eight times more transcriptionally active.[83] Even so, transcription of the pseudogene in human cells has been reported.[83] These pseudogene transcripts are not translated as a result of incorrect splicing and a lack of long open reading frames.[83] Translation of the pseudogene *in vitro* produces a 30kDa polypeptide.[55]

1.4. Goals and objectives

The objective of this thesis is two fold: (1) to assess the impact that gene copy number has on the expression of heterologous GBA in *Pichia pastoris*, and (2) to compare the sequences of conserved regions of the GBA gene and pseudogene between primate species to elucidate aspects of glucocerebrosidase's molecular evolution.

Chapter 2 Expression of human glucocerebrosidase in protease deficient *Pichia pastoris* using multiple gene copies

2.1. Abstract

Gaucher disease is a lysosomal storage disorder which is caused by insufficient levels of the enzyme glucocerebrosidase (GBA). The disorder can be effectively treated by enzyme replacement therapy but at a considerable cost. Thus the development of recombinant glucocerebrosidase (rGBA) by more economical methods is of importance. Here I explore two variables that may inhibit rGBA expression in *Pichia pastoris*, proteolytic instability and gene dosage. The complete coding region for human GBA, excluding the native targeting sequence, was introduced into the genome of a protease deficient strain of *Pichia pastoris* downstream of the *AOX*1 promoter. Plasmids with multiple copies of GBA were created to introduce a controlled number of GBA genes into the *Pichia pastoris* genome. Low levels of enzymatically active and immunologically reactive rGBA were produced. Gene dosage and reduction of host proteases did not, however, significantly increase rGBA expression levels. These results demonstrate that while protease deficient *Pichia pastoris* can be used to express enzymatically and immunologically active rGBA, some intrinsic biochemical properties of the recombinant enzyme prevent it from being expressed at high levels in this organism.

2.2. Introduction

Glucocerebrosidase (GBA) is used for enzyme replacement therapy to treat Gaucher disease. Over 119.6 million units of GBA are administered to patients each

year, at a staggering cost of $478.5 million in 1999.[39] Currently there are only two sources of commercial glucocerebrosidase, human placentae and recombinant Chinese Hamster Ovary (CHO) cells.[39] However both of these sources have considerable drawbacks associated with them. The purification of enzymes from blood products is potentially dangerous due to blood borne pathogens and contaminants. Additionally, GBA is found naturally at such low levels that 50,000 placentae are required to treat a single patient for 1 year.[46] The use of recombinant enzyme solves these difficulties, although the use of CHO cells for heterologous protein expression is labour intensive and expensive. Furthermore, both the placentae purified enzyme and the recombinant enzyme require carbohydrate modification prior to therapeutic use.[53] These factors make it worthwhile to investigate producing rGBA in alternative expression systems, such as *Pichia pastoris (P. pastoris)*.

P. pastoris is a methylotrophic yeast that was initially studied and characterized by Phillips Petroleum Company in the 1970s and now has gained popularity as a heterologous protein expression system.[19] The yeast was relatively easy to develop as an expression system since the techniques used to manipulate the well characterized yeast *Saccharomyces cerevisiae* were mostly transferable to *P. pastoris*. With further investigation it become apparent that *P. pastoris* could produce high levels of properly folded, secreted enzymes with many of the post-translational modifications of mammalian cells, such as proteolytic processing, folding, disulfide bond formation, and glycosylation.[1;31;38] To date, over 223 heterologous proteins, mostly human or mammalian, have been successfully produced in *P. pastoris*.[22] *P. pastoris* is especially well suited for producing pharmaceuticals as yeasts do not possess any known pathogens,

pyrogens, or viral inclusions.[51] Heterologous insulin like growth factor-1 (IGF-1) and human serum albumin (HSA) produced in *P. pastoris* have completed clinical trials for treatment of amyotrophic lateral sclerosis and as a serum replacement product, respectively, and are pending final approval for therapeutic use.[22] Hepatitis B surface antigen also expressed in *P. pastoris* is currently being used as a vaccine against the hepatitis B virus.[22]

P. pastoris's strong inducible alcohol oxidase (AOX) promoter is one of the features that makes it an attractive expression system. When methanol is the sole carbon source, *P. pastoris* produces alcohol oxidase to utilize the methanol. Alcohol oxidase catalyzes the oxidation of methanol to formaldehyde, which is the first step in the methanol utilization pathway.[51] There are two alcohol oxidase genes (*AOX*1 and *AOX*2), although the majority of the enzyme is produced by *AOX*1.[51] The *AOX*1 promoter can be used to drive the expression of heterologous protein and has been shown to produce exceptionally high levels of recombinant protein, up to 30% of the total soluble cellular protein.[75] The tightly regulated inducibility of this promoter is advantageous since the yeast can be cultured to a high density prior to inducing heterologous protein expression, which can inhibit the growth of the cells.[76]

As a single-celled microorganism, yeast cells are easy to manipulate and culture. *P. pastoris* can also be easily scaled up from small shake flask cultures to large fermentors where the cells can be grown at even higher densities (60-100g dry cell weight/liter).[51] High cell density is especially important for secreted proteins where the level of heterologous protein expression is directly linked to the cell density.[31] *P. pastoris*'s strong preference for respiratory growth allows the cultures to achieve these

high cell densities at a rapid rate using simple inexpensive medium.[31] The recombinant protein yield in these cultures can be extremely high. The highest yield of an intracellular heterologous protein in *P. pastoris* is 12 g/litre,[27] while that for a secreted protein yield is slightly lower at 4 g/litre.[75] Furthermore, secreted proteins have the added benefit of being expressed into an almost protein free medium.[75]

One of the most important advantages of *P. pastoris* is its capacity for post-translational modifications that closely resemble mammalian processing. The processing of some foreign glycoproteins in *P. pastoris* is very similar to that in mammalian cells.[68] The N-linked oligosaccharide chains added to glycoproteins are shorter than those of other yeasts and hyperglycosylation occurs to a lower degree. *P. pastoris* has a tendency to affix high mannose chains with between 8 and 14 mannose units.[75] This is particularly useful for the expression of heterologous GBA since exposed mannose structures and preferably Man_3–Man_6 structures have a high affinity for mannose receptors and can thus be effectively targeted to the macrophages of Gaucher patients.[73] Additionally, *P. pastoris* does not use Man(α1,3) linkages, which are highly antigenic and would reduce the possibility of using the expressed protein as pharmaceutical agent.[75]

Although *P. pastoris* can be very effective as a heterologous expression system, there are also some problems that can arise. Effective secretion is complex and can depend on gene dosage, methanol utilization (Mut) phenotype, signal sequence, proteolysis, and glycosylation.[51] Heterologous secretion is more demanding than intracellular expression and some larger (>30 kDa), complex foreign proteins are inefficiently secreted.[51;76] Proteolysis of secreted polypeptides can dramatically reduce the yield of secreted proteins.[31] Genes with high AT content may experience truncated

mRNA due to yeast transcriptional terminators. [76] Glycoproteins may be inappropriately glycosylated and some proteins may prove toxic to the yeast cells.[38] Nonetheless, the literature contains examples of over 223 heterologous proteins that have been successfully produced in *P. pastoris* and it is difficult to predict how well a recombinant protein will be expressed without conducting the experiments.[22]

A number of the factors affecting low expression levels can be optimized, including protease degradation and gene dosage. Many recombinant proteins are susceptible to proteolysis by the host organism during expression or in the first stages of purification.[51] Since most protease activity occurs within membrane-bound vacuoles, exposure to proteases occurs when cells are lysed to release the intracellularly expressed recombinant protein or when cells lyse naturally, which is especially prevalent during fermentation.[51] *P. pastoris* has three important vacuole proteases, proteinase A, carboxypeptide Y, and proteinase B. Proteinase A is capable of self activation and subsequently activates carboxypeptidase Y and proteinase B.[4] Proteolytic instability can be overcome by several methods, including supplementing the growth medium with amino acids or peptides, reducing the pH to a level where degradation is minimized (e.g. pH 3.0), or using a protease deficient host strain.[51]

Optimizing gene copy number is another technique used to increase recombinant protein expression. mRNA levels increase progressively with copy number and thus transcription as a limiting factor can be eliminated by increasing copy number.[81] While some studies show that increasing copy number dramatically increases recombinant protein expression, others demonstrate a lack of effect or even a reduction in expression. For example, increasing gene copy number has resulted in an increase of up to thirteen-

fold in expression levels of some proteins,[28] while with other proteins, such as bovine lysozyme and HIVgp120, it has decreased secretion levels and caused an accumulation of intracellular products.[76;80]

Additional factors that may affect heterologous protein expression include methanol utilization (Mut) phenotype and histidine auxotrophy.[51] The Mut phenotype can have a dramatic effect on the level of heterologous expression. Which Mut phenotype is better for secreted proteins is a contentious issue, as there is evidence for both and consequently proteins need to be tested on an individual basis to determine which is optimal.[75] Methanol utilization plus (Mut^+) strains may be preferential for fermentation because they grow quicker, although they also require more methanol for induction. Methanol utilization slow (Mut^s) strains are more challenging to ferment as they are more likely to suffer from methanol toxicity due to their impaired methanol utilization.[77] Histidine auxotrophs have been shown to fare more poorly than their prototrophic counterparts during fermentation, likely as a result from histidine becoming a limiting factor.[51]

There are ten different vectors, for either intracellular expression or extracellular secretion, that can be used to express heterologous gene products in *P. pastoris*.[51] The majority of the vectors use the *AOX*1 promoter to drive expression, although recently use of a constitutive *GAP* (glyceraldehyde 3-phosphate dehydrogenase) promoter has been developed. Secretion vectors commonly use the *S. cerevisiae* α-MF prepro secretion signal, although the *P. pastoris* PHO1 secretion signal is used by the pHIL-S1 vector.[51] Some vectors, such as pPIC9K, use ampicillin resistance for selection in *E. coli*, and geneticin resistance and histidine for selection in *P. pastoris*.[56] A smaller sized vector,

pPICZα, uses Zeocin™ resistance as the sole selection marker for both *E. coli* and *P. pastoris*.[57] The substitution of Zeocin™ resistance for the 2 to 3 genes otherwise required for selection reduces the size of the vector by approximately 3.5 kb. High concentrations of Zeocin™ can be used to select hyperresistant transformants that have multiple copies of the integrated vector.[51]

Glucocerebrosidase is a challenging protein to produce in heterologous expression systems because it needs to be glycosylated and it is rather large at 55.6 kDa prior to glycosylation. Nonetheless, rGBA has been successfully produced in several expression systems, including CHO and SF9.[42;73] Recombinant GBA from CHO cells, which is used for enzyme replacement therapy to treat Gaucher disease, is only expressed at levels which are 10-20% of those achieved for other recombinant proteins in CHO cells.[53;73] Recombinant GBA has also been expressed in monkey COS and mouse L cells at low levels (200,000 U/l), as well as in SF9 cells at levels higher than in the other systems (2,400,000 U/l after 3 days).[24;42]

This thesis chapter explores improving the level of rGBA expression in *P. pastoris* by addressing two variables that may suppress it, proteolytic instability and gene dosage. Proteolytic instability can be overcome by using a protease deficient host strain. The SMD1168 *P. pastoris* strain is deficient in pep4, which encodes proteinase A, a vacuolar aspartyl protease required for activation of other vacuolar proteases such as carboxypeptidase Y and proteinase B. The gene dosage can be controlled by creating vectors that contain increasing copies of GBA, which can be used to introduce a specific number of GBA genes into *P. pastoris*. Transformants with higher gene copy numbers can be selected by using increasing concentrations of the selective agent Zeocin™.

2.3. Materials and methods

2.3.1. Chemicals and reagents

The following were obtained from commercial sources. 2-mercaptoethanol (ACP, Montreal, Quebec); 4-methlyumbelliferyl 1-β-D-glucopyranoside (4MUGP)(Sigma, St. Louis, MO); Agar (Difco laboratories, Detroit, MI); Agarose (EM Science, Gibbstown, NJ); AlkPhos direct labeling kit (Amersham Life Sciences, Buckinghamshire, England); Anti c-myc antibody (Invitrogen, Carlsbad, CA); Biotin (Fisher Biotech, Pittsburgh, PA); Bromophenol blue (BDH Inc., Toronto, ON); Calf intestine alkaline phosphatase (New England Biolabs, Beverly, MA); Casamino acids (Difco laboratories, Detroit, MI); Chloroform (ACP, Montreal, Quebec); Citric acid (Caledon Laboratories, Georgetown, ON); Cloned pfu DNA polymerase (Stratagene, La Jolla, CA); Deoxynucleotides (GIBCO/BRL, Grand Island, NY); Dextrose (ACP, Montreal, Quebec); DNA 1kb ladder (GIBCO/BRL, Grand Island, NY); dNTPs (GIBCO/BRL, Grans Island, NY); Dry milk (Becton Dickenson, Cockeysville, MD); DTT (Invitrogen, Carlsbad, CA); ECL^{+}Plus Western blotting system (Amersham Pharmacia Biotech UK, England); EDTA (ACP, Montreal, Quebec); Ethanol (ACP, Montreal, Quebec); Ethidium bromide (Fisher Scientific, Fair Lawn, NJ); Glucose (HK) 50 (Sigma Diagnostics, St. Louis, MO); Glycerol (BDH Inc., Toronto, ON); Glycine (BDH Inc., Toronto, ON); Hybond™-P and Hybond™-N (Amersham Pharmacia Biotech, Quebec); L-glutamic acid, L-isoleucine, L-leucine, L-lysine, and L-methionine (Sigma, St.louis, MO); Low melting point agarose (Eclipse, Mississaugua, ON); Magnesium chloride ($MgCl_2 \cdot 6H_20$) (BDH Inc., Toronto, ON); Magnesium sulfide ($MgS0_4 \cdot 7H_20$) (ACP, Montreal, Quebec); Methanol (ACP, Montreal, Quebec); N-palmitoyl-dihydroglucocerebroside (Sigma, Diagnostics, St. Louis,

MO); Peptone BACTO (Difco laboratories, Detroit, MI); Phenol (Anachemia, Rosese Point, NY); Potassium acetate, Potassium chloride (KCl), and Potassium phosphate (BDH Inc., Toronto, ON); Protein marker, broad range (New England BioLabs, Beverly, MA); Restriction endonulceases (New England Biolabs, Beverly, MA); RNAaseA (GIBCO/BRL, Grand Island, NY); SDS (BDH Inc., Toronto, ON); SDS-PAGE gel loading buffer (GIBCO/BRL, Grand Island, NY); Sodium chloride (NaCl), Sodium hydroxide (NaOH) and Sorbitol (BDH Inc., Toronto, ON); Sucrose (ACP, Montreal, Quebec); T4 DNA ligase (New England Biolabs, Beverly, MA); Taq DNA polymerase (GIBCO/BRL, Grand Island, NY); Tris base (Boehringer Mannheim, Indianapolis, IN); Tryptone peptone (Difco laboratories, Detroit, MI); Tween-20 (Fisher Scientific, Fair Lawn, NJ); Xylene cyanol (Sigma, St.louis, MO); Yeast extract (Difco laboratories, Detroit, MI); Yeast nitrogen base (Difco laboratories, Detroit, MI); Zeocin™ (Invitrogen, Carlsbad, CA); Zymolase (Seikagaku Corp., Tokyo, Japan).

Monoclonal antibodies against glucocerebrosidase were prepared by Dr. Wei Chao (University of Victoria, Victoria, BC) and additional monoclonal antibodies were graciously received as a gift from Dr. Beutler (The Scripps Research Institute, La Jolla, CA). The pPICZα1-GBA recombinant vector containing the GBA gene in the pPICZαA vector (Invitrogen, Carlsbad, CA) was prepared by Dr. Wei Chao (University of Victoria, Victoria, BC).

2.3.2. Media and prepared solutions

LB (Luria-Bertani) media for *E. coli*: 1% tryptone, 0.5% yeast extract, 1% NaCl and adjusted to a pH of 7.0. For low salt LB, 0.5% of NaCl added instead of 1%.

Sterilized by autoclaving. For solid media 2% agar was added prior to autoclaving and media was poured into 150 mm plates. **SOC media for *E. coli***: 2% tryptone, 0.5% yeast extract, 10 mM NaCl. Sterilized by autoclaving. 2.5 mM KCl, 10 mM $MgCl_2 \cdot 6H_20$, 10 mM $MgS0_4 \cdot 7H_20$, 20 mM 2M glucose added (all solutions filter sterilized) and pH adjusted to 7. **RDH (regeneration dextrose histidine) agar for *P. pastoris***: 1M sorbitol, 2% agar. Sterilized by autoclaving. 1% dextrose, 1.34% YNB, 4×10^{-5}% biotin, 0.005% AA added. **YPD (yeast extract dextrose) media for *P. pastoris***: 1% yeast extract, 2% peptone. Sterilized by autoclaving. 2% dextrose and variable concentration of Zeocin™ added once solution cooled to 55°C. For solid media, 2% agar was added prior to autoclaving and media was poured into 150 mm plates. **YPDS (yeast extract dextrose sorbitol) agar for *P. pastoris***: 1% yeast extract, 2% peptone, 1M sorbitol, 2% agar. Sterilized by autoclaving. 2% dextrose and variable concentration of Zeocin™ added once solution cooled to 55°C. Poured into 150 mm plates. **MM (minimal medium) for *P. pastoris***: 20 mM potassium phosphate (pH 6.0), 10% casamino acids, 10% yeast nitrogen base, 0.4 μg/L biotin. Sterilized by autoclaving and then supplemented with 2% glycerol or 2% methanol as a carbon source to create MMG or MMM media respectively. **MD (minimal dextrose) medium for *P. pastoris***: 1.34% YNB, 4×10^{-5}% biotin, 1% dextrose. Sterilized by autoclaving. **10×YNB (yeast nitrogen base)**: 13.4% yeast nitrogen base with ammonium sulfate and without amino acids (filter sterilized). **BMGY (buffered glycerol-complex medium):** 1% yeast extract, 2% peptone. Sterilized by autoclaving. 100 mM potassium phosphate (pH 6.0), 1.34% YNB, 4×10^{-5}% biotin, 1% glycerol added. **BMMY (buffered methanol-complex medium)**: Same as BMGY except 0.5% methanol added instead of 1% glycerol. **100×AA (amino acids)**: 0.5% L-

glutamic acid, 0.5% L-methionine, 0.5% L-lysine, 0.5% L-leucine, and 0.5% L-isoleucine. Filter sterilized. **20× SSC buffer**: 3M NaCl, 0.3M Na_3 Citrate•$2H_20$. pH adjusted to 7.0. **TE buffer (pH 8.0)**: 10 mM Tris•HCl (pH8.0), 1 mM EDTA (pH 8.0). **Cracking buffer**: 0.05 M Tris-HCL (pH 7.5), 2 M EDTA, 1% SDS, 14% sucrose, 0.01% bromophenol blue, 0.01% xylene cyanol. **TBS (Tris-buffered saline)**: 0.8% NaCl, 0.02% KCl, 0.3% Tris base, 0.0015% phenol red. pH adjusted to 7.4 and sterilized by autoclaving. **TTBS (Tween tris-buffered saline)**: 0.05% Tween-20 added to TBS. **Blocking solution**: 7.5% dry milk added to TTBS. **Transfer buffer**: 39 mM glycine, 48 mM tris base, 0.037% SDS, 20% methanol. **Cell resuspension solution**: 50 mM Tris-HCl (pH 7.5), 10 mM EDTA, 100μg/ml RNAase A. **Cell lysis solution**: 0.2 M NaOH, 1% SDS. **Neutralization solution:** 1.32 M potassium acetate (pH 4.8). **Column wash solution:** 80 mM potassium acetate, 8.3 mM Tris-HCl (pH 7.5), 40 μM EDTA, 55% ethanol, 3 M sodium acetate, 2×SSC. **40× TAE Buffer**: 1.6 M Tris base, 0.8 M Na acetate•$3H_20$, 40 mM EDTA-Na_2•$2H_20$. Adjust pH to 7.2. **Gel loading buffer**: 50% glycerol, 1×TAE buffer, 1% bromophenol blue, 1% xylene cyanol. **10× PCR buffer**: 200 mM $(NH_4)_2SO_4$, 750 mM Tris-HCl (pH 8.8), 0.1% Tween-20.

2.3.3. *Escherichia coli* and *Pichia pastoris* strains

The *Escherichia coli* cell line *Top10F'* used to create the recombinant multi-insert vector was obtained from Invitrogen, Carlsbad, CA. The *Pichia pastoris* SMD 1168 cell line used to express the heterologous form of glucocerebrosidase was also obtained from Invitrogen, Carlsbad, CA. The *Pichia pastoris* SMD1168 cell line is histidine auxotrophic and protease deficient, lacking the gene *PEP*4 which encodes

proteinase A. Proteinase A is capable of self-activation and subsequent activation of additional vacuolar proteases carboxypeptidase Y and proteinase B. The *Pichia pastoris* SMD(his$^+$)(Mut$^+$) and SMD(his$^+$)(Mut^s) cell lines are identical to SMD1168 except they both possess the histidine gene and SMD(his$^+$)(Mut^s) is methanol utilization slow (Mut^s). The Mut^s phenotype occurs because the *AOX*1 gene is non-functional and thus the cell utilizes methanol at a slower rate. I engineered the SMD(his$^+$)(Mut$^+$) and SMD(his$^+$)(Mut^s) cell strains from SMD1168.

2.3.4. Construction of recombinant vector with multiple copies of GBA

i. Preparation of the GBA cassette

The GBA gene, as well as *AOX*1 promoter, *AOX*1 transcription termination, c-*myc* epitope was isolated using PCR from pPICZα1-GBA vector as a 3171 bp fragment. pPICZα1-GBA was previously constructed by inserting the cDNA for GBA (amplified using primers TATGAATTCGCCCGCCCCTGCATCCCT and GTCAGCTAGCTGGCGACGCCACAGTA, and digested using *EcoRI* and *NheI*) into the multiple cloning site of the pPICZαA vector (*EcoRI* and *XbaI* digested).[93] The GBA cassette was amplified from pPICZα1-GBA using the primers TCATGAGATCTAACATCCAAAGACGAAAGG (PICZ5) and AACGAAGGTCTCTGATCATCTTCTGTACTC (PICZBSA). The components of the PCR included 1× cloned *Pfu* DNA polymerase reaction buffer (Stratagene, La Jolla, CA), 200 μM of each dNTP, 200 ng of each PICZ5 and PICBSA primers, 10 ng template, and 5 U of cloned *Pfu* DNA polymerase in a 50 μl reaction. The conditions of PCR using the GenAmp PCR System 2400 (Perkin Elmer) were 0.5 min at 94°C, 0.5 min at 58°C, 1

min. at 72°C with an initial incubation of 5 min. at 94°C and a final incubation of 7 min. at 72°C. The number of cycles for amplification was 35. Resultant DNAs were visualized through gel electrophoresis according to the protocol described in section 3.3.5, and the 3171 bp GBA cassette was direct purified using Wizard® Plus Mini-prep purification System (Promega, Madison, WI) as outlined in section 3.3.6 (i).

The purified GBA cassette was double digested with *Bgl*II and *Bsa*I restriction enzymes to produce sticky ends for insertion into the vector. In brief, 2 µg GBA cassette, 10 U of each restriction enzyme, 1× NEB3 reaction buffer (New England Biolabs, Beverly, MA) in 20 µl total volume was incubated for 3 hrs. at 37°C, followed by 3 hrs. at 50°C. The digest was halted with a 20 min. incubation at 65°C. The digested GBA cassette was electrophoresed on 0.5% agarose gel and the band was excised for purification using Qiaex II Gel Extraction kit (Qiagen, Santa Clarita, CA) according to the protocol outlined in section 3.3.6 (ii).

ii. Preparation of the vector and ligation with insert

The pPICZα1-GBA, pPICZα2-GBA, or pPICZα3-GBA vectors were linearized with *Bgl*II. The restriction digest contained 2 µg plasmid, 10 U *Bgl*II, 1× NEB3 buffer (New England Biolabs, Beverly, MA) in 20 µl total volume and was allowed to proceed for 16 hours at 37°C. Complete linearization was verified using gel electrophoresis as outlined in section 3.3.5. Purification of the linearized vector in a final volume of 17 µl was performed by phenol/chloroform extraction and ethanol precipitation as described in Sambrook *et al.* (1989).[78] In brief, one volume of phenol/chloroform was added, vortexed for 1 min., and centrifuged at 12,000×g in an Eppendorf centrifuge 5415C

(Brinkman Instruments, Westbury, NY) for 5 minutes. The upper aqueous layer was removed, one volume chloroform was added, vortexed and centrifuged as before. The upper aqueous layer was then mixed with $1/10^{th}$ volume 3 M sodium acetate and 2 volumes ice-cold 100% ethanol. This solution was precipitated at -20°C for at least an hour, centrifuged at 12,000×g for 5 min., and the pelleted DNA was resuspended in 17 µl dH_20. Dephosphorylation of the vector was accomplished by adding 1 µl calf intestine alkaline phosphatase (CIAP) and 2 µl 10× NEB1 reaction buffer (New England Biolabs, Beverly, MA) to 17 µl of purified linearized vector. The reaction was allowed to proceed for 1 hr. at 37°C and was inactivated by adding 1 µl 0.1 M EDTA and incubating at 65°C for 5 minutes.

The digested GBA cassette and linearized dephosphorylated vector were ligated in a 3:1 ratio respectively with approximately 5 µg total DNA. The ligation reaction used 1 unit T4 DNA ligase, 1× buffer for T4 DNA ligase (New England Biolabs, Beverly, MA) in 10 µl total volume and proceeded for 16 hours at 16°C.

iii. Transformation of Top10F′ *E. coli* with recombinant plasmid

The ligation reaction was used to transform competent *Top10F′* cells according to the electroporation or $CaCl_2$ protocol modified from the *Pichia* Expression Kit Manual. In brief, *Top10F′* cells were transformed by electroporation as follows. The cells were grown overnight from a single colony in 2 ml LB medium with moderate shaking (300rpm) at 37°C. One hundred ml of LB was inoculated with 0.5 ml of the overnight culture and was grown at 37°C, with moderate shaking, to an OD_{600} of between 0.5 to 0.6. The cells were chilled on ice for 15 min. and centrifuged at 3000×g at 2°C for 20

minutes. The medium was decanted and the cells were resuspended in 1 ml ice-cold water. An additional 100 ml of cold water was added, centrifuged as before, and decanted. This was repeated once and then the cells were resuspended in 100 µl cold water. Fifty µl of cells were mixed with 1 µl of ligation reaction and transferred to a 0.1 cm cuvette (Invitrogen, Carlsbad, CA). The cuvette was pulsed with 1.5 kV at R5 resistance in Electro Cell Manipulator® 600 (BTX, San Diego, CA). One ml SOC medium was immediately added and the cells were grown for 30-60 min. at 37°C with moderate shaking. The cells were plated in 50-200 µl aliquots on low salt LB plates with 50 µg/ml Zeocin™. The plates were incubated at 37°C for 1 day.

The *Top10F'* cells were transformed using $CaCl_2$ as follows. The cells were grown overnight from a single colony in 2 ml LB medium with moderate shaking (300 rpm) at 37°C. Twenty five ml of LB was inoculated with 0.5 ml of the overnight culture and grown at 37°C with moderate shaking to an OD_{600} of 0.55. The cells were incubated on ice for 10 min., centrifuged at 3000×g at 4°C for 10 min., and then the supernatant was decanted. The cell pellet was resuspended in 10 ml 0.1 M $CaCl_2$, incubated on ice for 10 min., and centrifuged as before. The cells were resuspended in 1 ml 0.1 M $CaCl_2$ and incubated for 10 min. on ice. An aliquot of 200 µl competent cells was combined with 10 µl of ligation reaction. The tube was incubated on ice for 30 min., heat shocked at 42°C for 90 sec., with constant shaking, and immediately incubated on ice for 2 minutes. An aliquot of fresh 800 µl LB containing 0.4% glucose was added and the cells were allowed to recover by during a 1 hour incubation at 37°C with moderate shaking. Aliquots of 200 µl were spread on low salt LB plates with 50 µg/ml Zeocin™. The plates were incubated at 37°C for 1 day.

iv. Verification of transformants

E. coli transformants were analyzed for the presence and size of recombinant vectors by isolation and electrophoresis of crude bacterial DNA. In brief, the transformants were master-plated on low salt LB with 50 μg/ml Zeocin™ and incubated at 37°C for 1 day. A pinpoint portion of each colony was mixed with 50 μl cracking buffer in an Eppendorf tube. The tube was vortexed for 20 sec. and incubated at room temperature for 20 minutes. The tube was then centrifuged at 14,000×g at room temperature for 20 minutes. The entire soluble portion of the sample was electrophoresed on 0.7% agarose gel according to the procedure in section 3.3.5.

v. Screening for correct orientation and number of GBA cassettes

Transformants containing vectors corresponding to the appropriate size were grown in 2 ml low salt LB with 50 μg/ml Zeocin™ at 37°C in a shaking water bath for 20 hours. The vector was isolated for restriction digest analysis. In brief, 1.5 ml of the culture was centrifuged in an Eppendorf tube at 12,000×g for 10 minutes. The supernatant was discarded and the cell pellet was resuspended in 200 μl cell resuspension solution. Two hundred μl cell lysis solution was added and mixed by inversion, followed by 200 μl cell neutralization solution, which was also mixed by inversion. The tube was centrifuged at 12,000×g for 5 minutes at room temperature and the supernatant was transferred to clean tube for purification. The plasmids were purified using the using Wizard® Plus Mini-prep purification System according to the protocol described in the Promega product literature. The protocol is identical to that described in section 3.3.6 (i) except the DNA purification resin used was specific for plasmids.

The orientation of the GBA cassette in the vector was determined using double digest with *Bgl*II and *Eco*RV. The restriction digest reaction contained 5 µl of the vector, 10 units *Bgl*II, 10 units *Eco*RV, and 1× NEB3 reaction buffer (New England Biolabs, Beverly, MA) in 30 µl total volume. The reaction was allowed to proceed for 1 hr. at 37°C. The digested products were electrophoresed on 0.7% agarose according to the protocol in section 3.3.5.

2.3.5. Creation of *P. pastoris* SMD(His$^+$ Mut^s) and SMD(His$^+$ Mut$^+$) cell strains

i. Transformation of SMD1168

The *P. pastoris* SMD1168 cells were made competent by electroporation using a procedure modified from Invitrogen. In brief, electroporation was performed as follows. Two hundred ml of YPD was inoculated with 1 ml of an overnight culture of *P. pastoris* in YPD. The *P. pastoris* cells were grown at 30°C in a baffled 500 ml flask covered with 2 layers of cheesecloth in an orbital shaker at 250 rpm. The culture was grown to an OD_{600} of approximately 1.3 to 1.5. The cells were then collected by centrifugal force (1500xg for 20 minutes) at 4°C using a Sorvall®RC26 Plus (DuPont, Newtown, CT) centrifuge and the supernatant was discarded. The cells were resuspended in 200 ml ice cold water. The centrifugation procedure was repeated three times, with the first resuspension in 100 ml ice cold water, the second in 10 ml of ice-cold 1 M sorbitol, and the final in 400µl of ice-cold 1 M sorbitol. A 0.2 cm electroporation cuvette (Invitrogen, Carlsbad, CA) was used to hold 80 µl of electrocompetent cells and 10 µg of linearized or circular pPIC9K vector in less than 10 µl of TE buffer. Plasmid linearization was achieved by incubating 60 µg of plasmid with 60 units *Bgl*II, 1x NEB3 buffer (New

England Biolabs, Beverly, MA) in 30 µl total volume for 16 hours at 37°C. An electric pulse was applied to the cuvette using Electro Cell Manipulator® 600 (BTX, San Diego, CA) set at R6 resistance and 1.5 kV. Immediately afterwards, 1 ml ice-cold 1 M sorbitol was added. The cells were plated in 200-600 µl aliquots on RDB media and incubated at 30°C for 3-5 days. His^+ colonies that appeared in 3-5 days were resuspended in 1 ml water per plate and plated on YPD plates containing 0.5 mg/ml G418 at a final 10^5 cells/plate. The transformants that grew were confirmed for the presence of the pPIC9K vector using PCR and tested for the *AOX*1 phenotype, Mut^+ or Mut^s, as described in the following two sections.

ii. Test for Mut^+ or Mut^s phenotype

A single *P. pastoris* colony was transferred to MMM plates with a toothpick that was then used to transfer a similar sized colony onto the MMD plates. Plates were incubated for 2-5 days at 30°C and scored for density of colony growth.

iii. Confirmation of gene integration using PCR

Crude genomic DNA was isolated from the *P. pastoris* transformants according to a modified protocol from the Invitrogen product literature. A pinpoint sized part of the colony was suspended in 10 µl sterile water in a 1.5 ml Eppendorf tube. Forty five µg Zymolase was added and the tube was incubated at 30°C for 60 minutes. The sample was then immersed in liquid nitrogen for 1 minute. Five µl of the sample was used as DNA template to amplify the *AOX*1 gene with PCR using primers GACTGGTTCCAATTGACAAGC (5'AOX1) and GCAAATGGCATTCTGACATCC

(3'AOX1). The PCR also included 1× PCR reaction buffer, 2.5 mM $MgCl_2$, 250 µM of each dNTP, 0.3 µM of each primer, and 1 U *taq* polymerase in a total volume of 50 µl. The conditions of PCR using the GenAmp PCR System 2400 (Perkin Elmer) were 0.5 min. at 94°C, 0.5 min. at 58°C, 1 min. at 72°C with an initial incubation of 5 min. at 94°C and a final incubation of 7 min. at 72°C. The number of cycles for amplification was 35. Resultant DNAs were electrophoresed on 1.5% agarose according to the protocol in section 3.3.5.

2.3.6. *P. pastoris* transformation with multimer vectors and clone selection

i. *P. pastoris* transformation with multimer vectors and Zeocin™ selection

The *P. pastoris* SMD1168, SMD (his^+)(Mut^+), and SMD (his^+)(mut^s) cell lines were transformed with 50-100 µg of pPICZα1-GBA, pPICZα2-GBA, or pPICZα3-GBA plasmids using the electroporation procedure previously described in section 2.3.5. Following electroporation, an out growth procedure described in the *Pichia* expression manual by Invitrogen was followed. In brief, after 1 M sorbitol was added, the cells were incubated at 30°C without shaking for one to two hours. Then 1 ml YPD was added to the tubes and they were incubated at 30°C with shaking (~250 rpm) for another one to four hours. The cells were spread in 50-200 µl aliquots on YPDS plates containing 100 µg/ml Zeocin™. Plates were incubated at 30°C for 2 to 4 days.

ii. PCR confirmation of gene integration in *P. pastoris* genome

P. pastoris crude genomic DNA was isolated as previously described in section 2.3.5 (iii) and used as a template for PCR with primers

GACTGGTTCCAATTGACAAGC (5'AOX1) and ATAGGTGTAGGTGCGGATGGA (G9). The PCR also included 1× PCR reaction buffer, 2.5 mM $MgCl_2$, 250 µM of each dNTP, 0.3 µM of each primer, and 1 U taq polymerase in a total volume of 50 µl. The conditions of PCR using the GenAmp PCR System 2400 (Perkin Elmer) were 0.5 min. at 94°C, 0.5 min. at 58°C, 1 min. at 72°C with an initial incubation of 5 min. at 94°C and a final incubation of 7 min. at 72°C. The number of cycles for amplification was 35. Resultant DNAs were electrophoresed on 1.5% agarose as outlined in section 3.3.5.

iii. Selection using increasing concentrations of Zeocin™

To determine if transformants were hyperresistant to Zeocin™, transformants that were initially selected using 0.1 mg/ml Zeocin™ were transferred onto YPDS plates with 0.25, 0.5, 0.75, and 1 mg/ml Zeocin™. Plates were incubated at 30°C for 2-4 days and scored for density of colony growth. Hyperresistent Zeocin™ transformants were also selected by directly plating cells following the transformation onto YPDS plates with 0.1, 0.5, 1, and 2 mg/ml Zeocin™.

2.3.7. Expression and analysis of recombinant GBA.

i. Small scale induction of recombinant *P. pastoris*

P. pastoris transformants were screened for their level of recombinant GBA production by growth and methanol induction in baffled flasks. The procedure used was modified from that presented in the *Pichia* Expression Manual by Invitrogen (Carlsbad, CA). A single colony was used to inoculate 25 ml BMGY in a 250 ml baffled flask, which was grown at 30°C with shaking (~300 rpm) for 1 day or until the culture reached

an OD_{600} of 2-6. The cells were harvested by centrifugation at 1500-3000×g for 5 minutes at room temperature. The cells were resuspended in BMMY to a final OD_{600} of 1.0. The culture was transferred to a larger, 1 L baffled flask and grown with shaking at 30°C for 96 hours. Every 24 hours 200 mM DTT and 100% methanol to a final concentration of 0.5% were added. The methanol served to maintain induction, while the DTT helped retain reduced sulfhydryl groups in the highly oxygenated environment. The pH was adjusted to 7.5 with citrate buffer (pH 4.0) every 6 to 12 hours. Aliquots of 1 ml were removed at 0, 6, 12, 24, 36, 48, 60, 72, 84, and 96 hour time points. These 1 ml samples were transferred to 1.5 ml microcentrifuge tubes and centrifuged at 12,000×g for 3 minutes at room temperature. The supernatant was transferred into a clean microfuge tube and both samples were flash frozen in liquid nitrogen and stored at -20°C until required for further analysis.

ii. SDS-PAGE analysis with silver stain

Supernatant samples were thawed and 26µl of the supernatant were mixed with 13 µl of SDS-PAGE gel loading buffer with 10% DTT. The sample was boiled for 5 minutes and then centrifuged at 12,000×g for 5 minutes. The supernatant was transferred to a microfuge tube and 30 µl of the sample was loaded onto a 10% APS gel and electrophoresed. The gel was prepared and silver stained according to the protocol in Sambrook *et al.* (1989).

iii. Western blot analysis

Supernatant samples were prepared as those for the SDS-PAGE analysis (section 2.3.7 (ii) and electrophoresed on a 10% resolving and 4% stacking gel with a protein marker. The gel was prepared according to methods described in Sambrook *et al.* (1989). The proteins on the electrophoresed gel were transferred to nitrocellulose Hybond™-N membrane using a mini trans-blot cell (Bio-Rad), as outlined in Sambrook *et al.* (1989). Alternatively, methanol pre-treated Hybond™-P PVDF membrane was used as described in the Amersham Pharmacia literature. The transfer was allowed to proceed for 16 hours at 20 volts. The membrane was washed in TTBS for 10 minutes, blocked with 7.5% blocking solution for 1 hour, washed twice in TTBS for 5 minutes, incubated in primary antibody for 1 hour, washed six times in TTBS for 5 minutes, incubated in the secondary antibody (RAM-IgG) for 16 hours, washed four times in TTBS for 5 minutes, and washed twice in TBS for 5 minutes. The primary antibody was either the anti-GBA Ab, which was added at a 400× dilution, or the anti-c-*myc* antibody, which was added at a 5000× dilution. The secondary antibody was added at a 6000× dilution when used with the GBA antibody and at a 8000× dilution when used with the anti c-*myc* antibody. Colour development was performed using ECL+Plus Western blotting system according to the protocol supplied in the Amersham Pharmacia product literature.

iv. Western dot blot

Hybond™-P PVDF membrane was washed once in 100% methanol and twice in ddH_20. The membrane was placed on filter paper wetted with water so that it would maintain its moisture throughout the blotting procedure. Two and a half µl of supernatant

from induced *P. pastoris* cultures was blotted onto the membrane and allowed to air dry. The membrane was washed once in 100% methanol, twice in ddH_20, and then incubated in TTBS for 5 min.. The membrane was then incubated with antibodies and developed in a procedure identical to that outlined in the Western blot analysis procedure in section 2.3.7 (iii).

v. Southern blot analysis

The DNA probe for Southern blot analysis was created by amplification of a GBA DNA fragment using primers CTGCTGCTCTCAACATCCTT (G4) and GAAGGGGTATCCACTCAACA (G6). The PCR included 10 ng pPICZα1-GBA as template, 1× PCR reaction buffer, 2.5 mM $MgCl_2$, 250 μM of each dNTP, 10 μM of each primer, and 1 U *taq* polymerase in a total volume of 50 μl. The conditions of PCR using the GenAmp 2400 PCR were 0.5 min. at 94°C, 0.5 min. at 60°C, 0.5 min. at 72°C with an initial incubation of 5 min. at 94°C and a final incubation of 7 min. at 72°C. This PCR product was purified using Wizard Miniprep DNA purification system and diluted to 10 ng/μl. The probe was labeled using AlkPhos Direct labeling kit according to the protocol in the product literature provided by Amersham Life Science.

The recombinant and control *P. pastoris* genomic DNA was isolated according to the protocol provided by Invitrogen in the *Pichia* expression kit manual. The DNA was digested using *Eco*RI and separated by gel electrophoresis on a 1% agarose gel, which was prepared with TBE buffer and without ethidium bromide. Approximately 10 μg *P. pastoris* DNA in 50 μl was combined with 10 μl of DNA loading buffer and placed into each well. Five μl of 1 kb DNA ladder with 2 μl loading dye was run as a size reference.

The gel was run for 16 hours at 40 volts. The gel was incubated in 500 ml TBE buffer containing 0.5 μg/ml ethidium bromide for 20 minutes. The gel was visualized and photographed using an Eagleeye imager (Stratagene, La Jolla, CA) with a ruler adjacent to it. The gel was incubated in several volumes 1.5 M NaCl and 0.5 N NaOH for 30 minutes at room temperature to denature the DNA and then transferred to 1 M Tris for 30 min. to neutralize the reaction.

The DNA ladder was cut away from the rest of the gel and the electrophoresed DNAs were transferred to nitrocellulose N™-Hybond membrane using capillary transfer, according to the procedure outlined in Sambrook *et al.* (1989). The membrane was hybridized with the labeled DNA probe according to the protocol provided in the Amersham product literature. In brief, the membrane was incubated in 0.125 ml/cm^2 AlkPhos direct hybridization buffer at 55°C for 15 minutes. The labeled DNA probe was added at a concentration of 10 ng/ml buffer and incubated at 55°C overnight. The membrane was transferred to 5 ml/cm^2 primary wash buffer at 55°C for 10 min., and then washed twice in secondary wash buffer for 5 min. at room temperature. The filter was air dried and baked at 80°C for 2 hours.

The chemifluorescent signal generation and detection was done using ECF substrate (Amersham) according to the protocol provided in their product literature. In brief, the membrane was incubated with 25 μl/cm^2 ECF substrate in a sealed detection bag. The membrane was imaged at 1 hour and 24 hours afterwards using a Storm PhosphorImager® (Molecular Dynamics, Sunnyvale, CA).

vi. Dot blot Southern analysis

The protocol for large scale southern dot blot analysis of *P. pastoris* clones was modified from that provided by Romanos *et al.* (1991). The *P. pastoris* transformant cultures were started from single colonies in 200 µl YPD medium in 96 well plates and grown for 48 hours at 30°C. Five µl of each culture were transferred to 200 µl of fresh YPD medium in a clean 96 well plate and incubated at 30°C for 24 hours. Cells were resuspended and 5 µl were transferred onto N™-Hybond nitrocellulose membrane. The membrane was air dried. The membrane was incubated in solution by placing it on filter-paper saturated with solutions as follows: 15 min. in solution 1 (50 mM EDTA, 2.5% 2-mercaptoethanol (pH 9.0)), 4 hrs. in 3 mg/ml zymolase at 37°C, 5 min. in solution 3 (0.1 M NaOH, 1.5 M NaCl), and two incubations of 5 min. in 2× SSC for 5 min.. The filter was air dried and baked at 80°C for 2 hours. The membrane was hybridized with the labeled DNA probe and the chemifluorescence signal was detected as previously described in section 2.3.7 (v).

vii. Northern dot blot

RNA was extracted from *P. pastoris* according to the protocol outlined by Sambrook *et al.* (1989). Two and a half µl of each sample were blotted onto N™-Hybond nitrocellulose membrane and the membrane was baked at 80°C for 30 minutes to fix the RNA. The membrane was hybridized with the probe and visualized according to the same procedures used for Southern blot analysis described in section 2.3.7 (v).

viii. Enzyme activity analysis

Recombinant enzyme activity was measured using two assays; the sphingolipid substrate, N-palmitoyldihydroglucocerebroside, assay and the fluorogenic substrate, 4-methlyumbelliferyl β-D-glucopyranoside (4MUGP), assay. The assay using the sphingolipid substrate, N-palmitoyldihydroglucocerebroside, was modified from that previously described by Choy and Davidson (1980).[26] In brief, 25 µl of supernatant were added to 100 µl of a mixture of 1mM N-palmitoyldihydroglucocerebroside, 1% sodium taurocholate, 0.1% Triton X-100, and 40 mM citrate buffer (pH 4.0). A control tube containing all components except for N-palmitoyldihydroglucocerebroside was used to measure endogenous glucose levels. The reaction was incubated at 37°C with shaking for 4 hours and stopped by immersion in boiling water for 5 minutes. The tubes were centrifuged at 12,000× g for 15 minutes at 4°C. Sixty µl of the supernatant were added to 540 µl of glucose assay reagent and incubated for 5 minutes at room temperature. The hexokinase in the glucose HK converts any glucose into glucose-6-phosphate in the presence of ATP, which then reduces NADP. NADPH was then measured at 340 nm using Spectronic Genesys 5 Spectrophotometer (Milton Roy). The glucocerebrosidase activity was determined by calculating the amount of glucose present using a glucose standard curve (0-60nmol) calibrated under identical pH and detergent conditions. The endogenous glucose activity was determined from the negative control and subtracted from the sample.

The assay for measuring GBA enzyme activity using the fluorogenic substrate 4MUGP was modified from the procedure described in Choy and Davidson (1978).[25] In brief, 10 µl supernatant was added to 70 µl 5 mM 4MUGP, 15 µl 200 mM citrate buffer

(pH 5.5), and 5 µl 2% sodium taurocholate. The reaction was incubated at 37°C for 30 minutes and then stopped by the addition of 1.4 ml 0.2 M glycine buffer (pH 10.5). The fluorescence was measured with a Model 2400 Fluorometer (Abbott Diagnostics, Abbott Park, IL) and the enzyme activity was calculated based on a standard curve.

2.4. Results

2.4.1. Construction of multi-copy GBA expression vector

The pPICZαA expression vector was selected for the creation of multimers, primarily due to its small size of 3.6 kb. The vector includes: transcription elongation factor 1 gene promoter (*PTEF1*), which drives expression of the *Streptoalloteichus hindustanus ble* (*Sh ble*) gene in *P. pastoris*; *EM7* promoter (*PEM7*) which drives expression of the *Sh ble* gene in *E. coli*; *Sh ble* gene, which confers resistance to Zeocin™ in both *E. coli* and *P. pastoris*; *CYC1* transcription termination region (*CYC1*TT), which allows efficient 3′ processing of the *Sh ble* gene; ColE1 origin (ColE1) for replication and maintenance of the plasmid in *E. coli*; *AOX*1 promoter and transcription termination regions.

The sequence of the glucocerebrosidase (GBA) gene used for heterologous expression begins after the native targeting sequence and ends just before the translation termination codon, and all non-coding sequences have been removed for a total size of 1517 bp. A 3171 bp fragment containing GBA as well as the flanking *AOX*1 regions was PCR amplified from pPICZα1-GBA (Figure 2.1). The amplified GBA cassette begins at the 5′*AOX*1 promoter and ends at the *AOX1* transcription terminator in pPICZα1-GBA. The 3171 bp fragment (Figure 2.1) contains the *AOX1* promoter and α-

factor secretion signal upstream of the GBA gene, as well as the *myc* epitope tag, polyhistidine tag, and *AOX1* transcriptional terminating sequence downstream of the GBA gene.

This GBA cassette was cloned into the *Bgl*II linearized vector using overhanging complementary ends that had been created with *Bsa*I and *Bgl*II. The BsaI recognition site was created in the GBA cassette using mismatch PCR mutagenesis with the PICZBSA primer (Figure 2.2). Although, pPICZαA possesses a *Bam*HI site 3′ of the *AOX*1 transcriptional termination, which creates a complementary overhang to *Bgl*II when cleaved, *Bam*HI could not be used to digest the multimer plasmid as GBA also has a recognition site for this enzyme. The *Bsa*I site was created 23 bp upstream of the *Bam*HI cut site. When the GBA cassette integrated into the vector, the *Bgl*II vector overhang and *Bsa*I cassette overhang combined and abolished the *Bgl*II recognition site. This enabled *Bgl*II to be used to linearize the recombinant plasmid for insertion of additional GBA cassettes.

The GBA cassette was cloned into the pPICZα1-GBA vector to create a two cassette vector, pPICZα2-GBA, and this was used to create a three GBA cassette vector, pPICZα3-GBA (Figure 2.3). DNA sequencing confirmed the integrity of the vectors by demonstrating that mutations were not introduced into the vectors during the engineering process. This step-wise approach to cloning was used because the size of the GBA cassette (3.2 kb) was too large to easily allow two or more GBA cassettes to be simultaneously inserted into the vector. Even this approach was difficult and it took nine attempts to create the pPICZα3-GBA vector. Frequent results from the ligation reactions and subsequent transformation into *E. coli* included low transformation efficiencies,

recombination events involving the vector, and religation of the vector. $CaCl_2$ was initially used to bring the *E. coli* cells to competency but due to low transformation efficiencies, electroporation was substituted and eventually proved successful. Two attempts at inserting a fourth GBA cassette in the pPICZα3-GBA plasmid were made but were unsuccessful.

The *E. coli* transformants were screened using Zeocin™ resistance, and the presence and size of the vector were confirmed by electrophoresis. The 3171 bp size difference in the plasmids containing an additional insert was readily apparent in the electrophoresed DNAs (Figure 2.4). The plasmid containing two GBA cassettes is 8281 bp while that with three GBA cassettes is 11,452 bp. Linearization of the vector confirmed the size of the vector, and the orientation of the GBA cassettes was determined using a double digest with *Bgl*II and *Eco*RV. *Eco*RV cuts in the transcription termination region for the Zeocin™ resistance gene (*CYC*TT) and *Bgl*II cleaves at the junction where the *AOX1* promoter of the GBA cassette has either joined the vector, in the case of correct orientation, or another GBA cassette, in the case of incorrect orientation. The pPICZα3-GBA vector with all three GBA cassettes in the correct orientation showed two bands sized 10,459 bp and 993 bp, while the vector with two GBA cassettes in the correct orientation had a 7288 bp and a 993 bp band (Figure 2.5).

2.4.2. Construction of *P. pastoris* SMD Mut^{s}His$^+$ and *P. pastoris* SMD Mut$^+$His$^+$ cell lines

It was necessary to convert the SMD1168 cell strain to a histidine prototroph and convert the methanol utilization (Mut) phenotype from methanol utilization plus (Mut$^+$)

to methanol utilization slow (Mut^s) prior to transformation with pPICZα1-GBA, pPICZα2-GBA, or pPICZα3-GBA. This is because the pPICZαA vector does not carry a histidine gene and thus it is unable to affect the histidine auxotrophy. Additionally, pPICZα2-GBA or pPICZα3-GBA can only integrate into the *P. pastoris* genome by gene insertion, not by gene replacement, which is required to alter the Mut phenotype. This is because gene replacement requires linearization of the vector in the *AOX*1 region, and neither pPICZα2-GBA or pPICZα3-GBA can be linearized in this region. Cleavage of these vectors, which contain two or three copies of the GBA cassette, in the *AOX*1 region would fragment the plasmid into more than one piece.

The SMD(His^+ Mut^s) and SMD(His^+ Mut^+) cell lines were created by transforming SMD1168 cells with the pPIC9K vector. The pPIC9K vector is 9.3 kb, contains the gene for histidine, and the kanamycin resistance gene as a selection marker in *P. pastoris*. The plasmid can integrate either by gene replacement, which is favoured by a linear plasmid, or gene insertion, which is favoured by a circular plasmid (Figure 2.6).

pPIC9K was linearized with *Bgl*II, which cuts at both the 5'*AOX1* and the 3'*AOX1* regions, producing two fragments (Figure 2.7). The 6873 bp fragment contains the *AOX*1 regions of homology flanking the α-factor secretion signal, 3' *AOX1* transcription termination, histidine gene, and kanamycin resistance gene. This fragment was integrated into the *P. pastoris* genome by gene replacement at the *AOX1* locus, altering the Mut^+ phenotype to Mut^s. Gene replacement occurs by a double crossover event between the 5'and 3' *AOX1* regions, replacing the functional *P. pastoris AOX1* gene with a portion of the pPIC9K vector (Figure 2.6a). This hinders the ability of *P. pastoris* to

utilize methanol, resulting in a methanol utilization slow (Mut^s) phenotype. The smaller, 2427 bp fragment does not integrate into the *P. pastoris* genome as it only contains the ampicillin resistance gene and the ColE1 origin.

Circular pPIC9K was used to integrate the vector into the *P. pastoris* genome by gene insertion. Gene insertion occurs by a single crossover event at either the 5'and 3' *AOX*1 regions (Figure 2.6b). The *AOX*1 gene was not disrupted and the *P. pastoris* strain retains its Mut^+ phenotype, although the cell was converted into a histidine prototroph.

i. Transformation efficiency

The transformation efficiency of integrating the pPIC9K vector into the *P. pastoris* SMD cell line ranged from 7 to 13 transformants per μg *Bgl*II linearized plasmid and 1 transformant per μg circular plasmid. These transformants were selected using histidine and then further screened using G418, which was necessary since histidine selection can result in false positives.

ii. PCR confirmation

The transformants were confirmed for the presence of the pPIC9K vector by PCR using 5'*AOX1* and 3'*AOX1* primers that flank the *AOX1* region. If the vector integrated into the *P. pastoris* genome, a 492 bp band corresponding to the plasmid's *AOX1* promoter and transcription termination region was expected. If the plasmid integrated by gene insertion an additional band of 2.2 kb corresponding to the size of the uninterrupted *AOX1* gene in the *P. pastoris* genome was expected. If the pPIC9K plasmid integrated by gene replacement, only the 492 bp PCR fragment was expected since the *P. pastoris*

AOX1 gene had been replaced by the plasmid's incomplete *AOX*1 gene. The transformants created with circular pPIC9K display two bands sized 492 bp and ~2.2 kb, suggesting that pPIC9K integrated by gene insertion. In contrast, the transformants created with linear pPIC9K demonstrate only one band, sized 492 bp, which is consistent with integration via gene replacement (Figure 2.8).

iii. Phenotypic selection for *AOX*1

To verify the Mut^+ or Mut^s phenotype, the pPIC9K transformants were grown on plates with only methanol as a carbon source (MMM plates) and replica plated on agar with dextrose as a carbon source (MMD plates). Most of those which were transformed with *Bgl*II digested pPIC9K grew poorly on MMM media and well on MMD media. This Mut^s (methanol utilization slow) phenotype suggests that the vector was integrated into the genome by a double cross over event that replaced the *AOX*1gene. In contrast, the transformants that resulted from the transformation with circular pPIC9K grew well on both MMM and MMD media suggesting that their phenotype is Mut^+ (methanol utilization plus). The results from the phenotypic selection correspond well with those from the direct yeast PCR, with Mut^+ strains demonstrating 492 bp and 2.2 kb bands, and Mut^s strains only showing 492 bp bands.

2.4.3. Insertion of multiple copies of the GBA gene in *P. pastoris* and confirmation using PCR and Southern blot analysis

The plasmids pPICZα1-GBA, pPICZα2-GBA, and pPICZα3-GBA were transformed into SMD1168, SMD(His^+ Mut^s) and/or SMD(His^+ Mut^+) *P. pastoris* strains

and transformants were selected using Zeocin™. To compensate for reduced transformation efficiencies due to the large size and circular nature of the vector, electroporation and an out growth procedure were used to transform the cells. Despite this, the transformation efficiency when selected with 0.1 mg/ml Zeocin™ remained low, ranged from 2-18 transformants/μg plasmid (data not shown).

The transformants were confirmed by PCR analysis with primers that amplified a portion of the alcohol oxidase gene and glucocerebrosidase gene which was 762 bp in size (Figure 2.9). To ensure that the amplified products were from the *P. pastoris* DNA and not artifacts or contaminants, a negative control identical to the other samples except without *P. pastoris* DNA was included. The absence of amplified products in the negative control strongly suggests that amplification could only occur if the plasmid had integrated into the yeast genome. The size of the amplified fragment is the expected size of 762 bp and is identical in size to the product amplified from the pPICZα1-GBA vector, which served as the positive control (Figure 2.9).

The presence of the GBA gene in the *P. pastoris* genome was also confirmed using Southern blot analysis. A ~3.2 kb fragment was detected in the transformants and was absent in the negative control (Figure 2.10). The *P. pastoris* transformants depicted in Figure 2.10 were transformed with pPICZα3-GBA, which would cause at least three copies of the GBA cassette to be inserted in tandem into the yeast genome. *Eco*RI cuts at the start of each GBA cassette, which creates fragments sized approximately 3.3 kb containing the GBA gene. To ensure this band was not an artifact, a negative control containing DNA of non-recombinant SMD1168 and identical to the other reactions in all

other aspects was included. The lack of bands in the negative control suggests that the results seen are due to the integration of GBA into *P. pastoris*. A positive control of *P. pastoris* transformed with a different GBA vector was also included and demonstrated a band sized approximately 1.5 kb. This illustrates that the conditions of the Southern blot were adequate to accurately detect GBA. The size difference between the positive control and the transformants results from the use of a different vector to integrate GBA into *P. pastoris*. Southern dot blot analysis was also used to screen high numbers of *P. pastoris* transformants for high GBA gene copy number (data not shown).

The Mut phenotype of the *P. pastoris* transformants was confirmed by comparing the growth of transformants on MMM agar to that on MMD agar. The phenotype of the *P. pastoris* strains transformed with the pPICZαA vectors did not change, Mut^s strains remained Mut^s and Mut^+ strains continued to be able to utilize methanol (data not shown). This stability in Mut phenotype was expected since the circular vector preferentially integrates via gene insertion, which does not interrupt the *AOX*1 gene.

2.4.4. Selection of hyperresistant transformants

There is a direct relationship between the number of Zeocin™ resistance genes in the *P. pastoris* genome and resistance to increasing concentrations of Zeocin™. Zeocin functions stoichiometrically, which allows the copy number of Zeocin™ resistance genes to be fairly accurately predicted by the transformant's level of Zeocin™ resistance.[51] PCR confirmed transformants were further analyzed for multiple pPICZα3-GBA integration events by assessing their ability to grow on increasing concentrations of Zeocin™. Of the transformants that were able to grow at 0.1 mg/ml Zeocin™, only 22%

were able to grow well at 1 mg/ml Zeocin™ (data not shown). The differences in colony growth was slight, with all growing at the highest Zeocin™ concentrations but some colonies were denser and larger.

Hyperresistant pPICZα1-GBA pPICZα2-GBA, and pPICZα3-GBA transformants were also selected by directly plating transformed cells onto medium with increasing concentrations of Zeocin™. The lowest concentrations of Zeocin™ (0.1 mg/ml) contained approximately 35 transformations per plate while at the highest Zeocin™ concentration (1 mg/ml), only some plates had a single transformant (Table 2.1). The number of transformants that grew on increased Zeocin™ concentrations and the transformation efficiency did not vary significantly with the size of the transforming vector. These values were approximately equal for pPICZα1-GBA pPICZα2-GBA, and pPICZα3-GBA transformants (Table 2.1).

2.5.4. Recombinant GBA expression and enzyme analysis

i. Induction of *P. pastoris*

It was observed that following induction there was a greater pH change in the GBA transformants when compared to the *P. pastoris* control strains. The average pH change every 12 hours following induction was from 5.5 to 8.0 for most SMD transformants although some had an even greater change to above 9.0. The pH of the control SMD strains did not vary much from the initial pH of 5.5. The growth of the transformant and control strains following induction was comparable. Northern blot analysis demonstrated increasing mRNA expression with longer induction periods but this trend was also apparent in the negative control (data not shown).

ii. SDS-PAGE with silver stain and Western blot analysis

Comparison of the supernatant of the GBA recombinant *P. pastoris* strains with that of the control strains indicated differences in proteins expressed under methanol induction. SDS-PAGE separation followed by silver staining did not demonstrate significant differences in the protein production of the recombinant strains compared to the control (data not shown). The multitude of proteins detected by silver staining in all strains tested contributed to the difficulty in detecting slight differences in expression. Western blot analysis of the supernatant proteins demonstrated bands sized 65-77 kDa that were immunoreactive with the anti-GBA antibody (Figure 2.11). While some non-specific binding activity was detected in the negative control, this was consistently absent in the 65-77 kDa range.

iii. Enzyme activity assays

The activity of GBA from ten supernatant samples was compared to that of the control SMD and SMD(his$^+$) cell lines. The highest enzyme activity using the sphingolipid substrate N-palmitoyldihydroglucocerebroside assay was 10 nM/hr/mg (data not shown). The highest enzyme activity using the fluorogenic substrate 4-methlyumbelliferyl β-D-glucopyranoside (4MUGP) assay was found to be much higher at 562 nM/hr/mg(data not shown). The 4MUGP assay, however, also demonstrated high enzyme activity for the control strain.

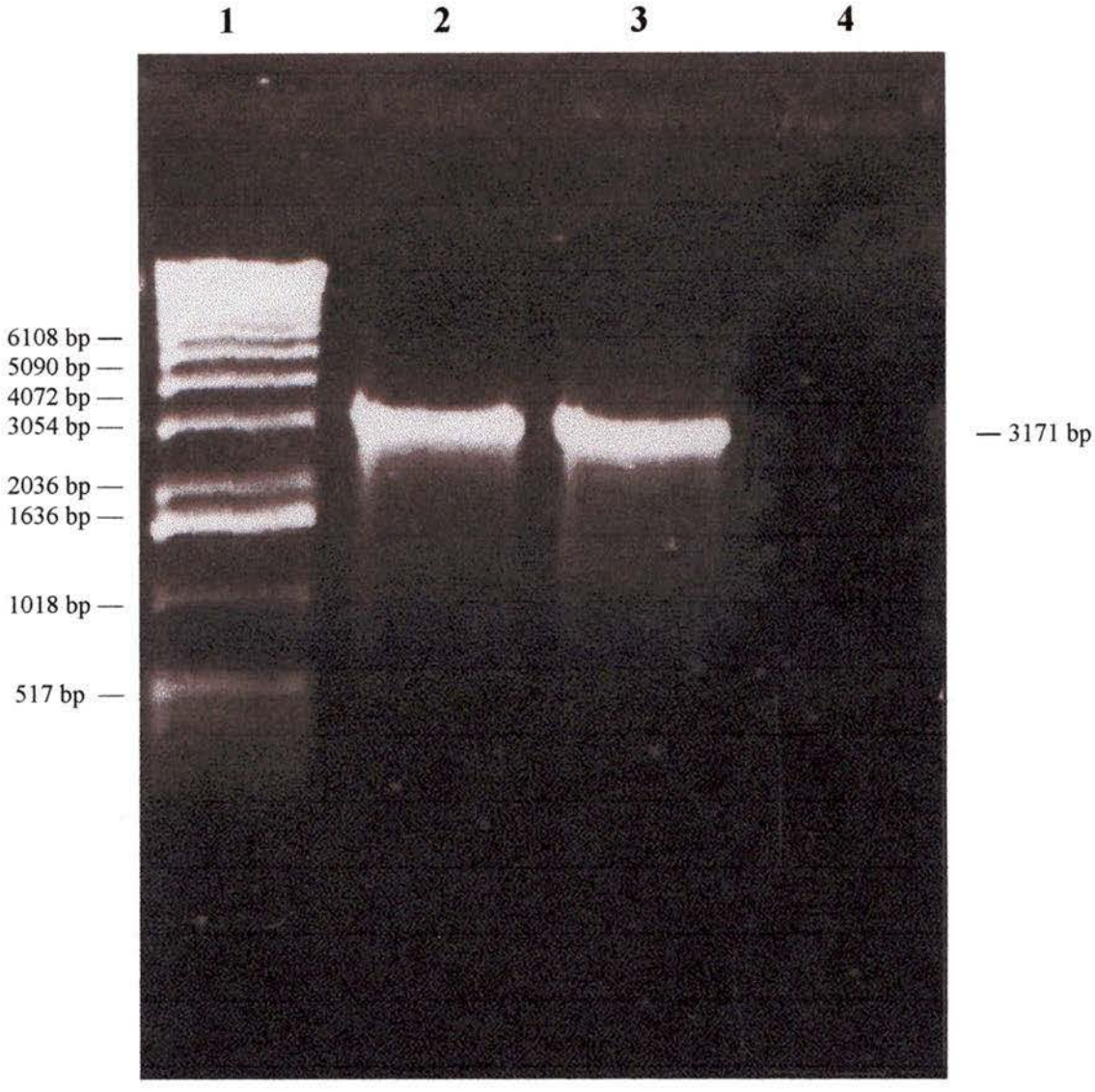


Figure 2.1. Electrophoresis of DNAs from PCR of GBA cassette using primers PICZ5 and PICBSA with the pPICZα1-GBA plasmid as template. The GBA cassette is 3171 bp and contains the *AOX1* promoter, α-factor secretion signal, coding region of the GBA gene, *myc* epitope tag, polyhistidine tag, and *AOX1* transcriptional terminating sequence. Lane 1) DNA 1kb ladder; lanes 2 and 3) GBA cassette amplified from pPICZα1-GBA cassette; lane 4) negative control, which is identical to the other reactions except lacking pPICZα1-GBA as DNA template.

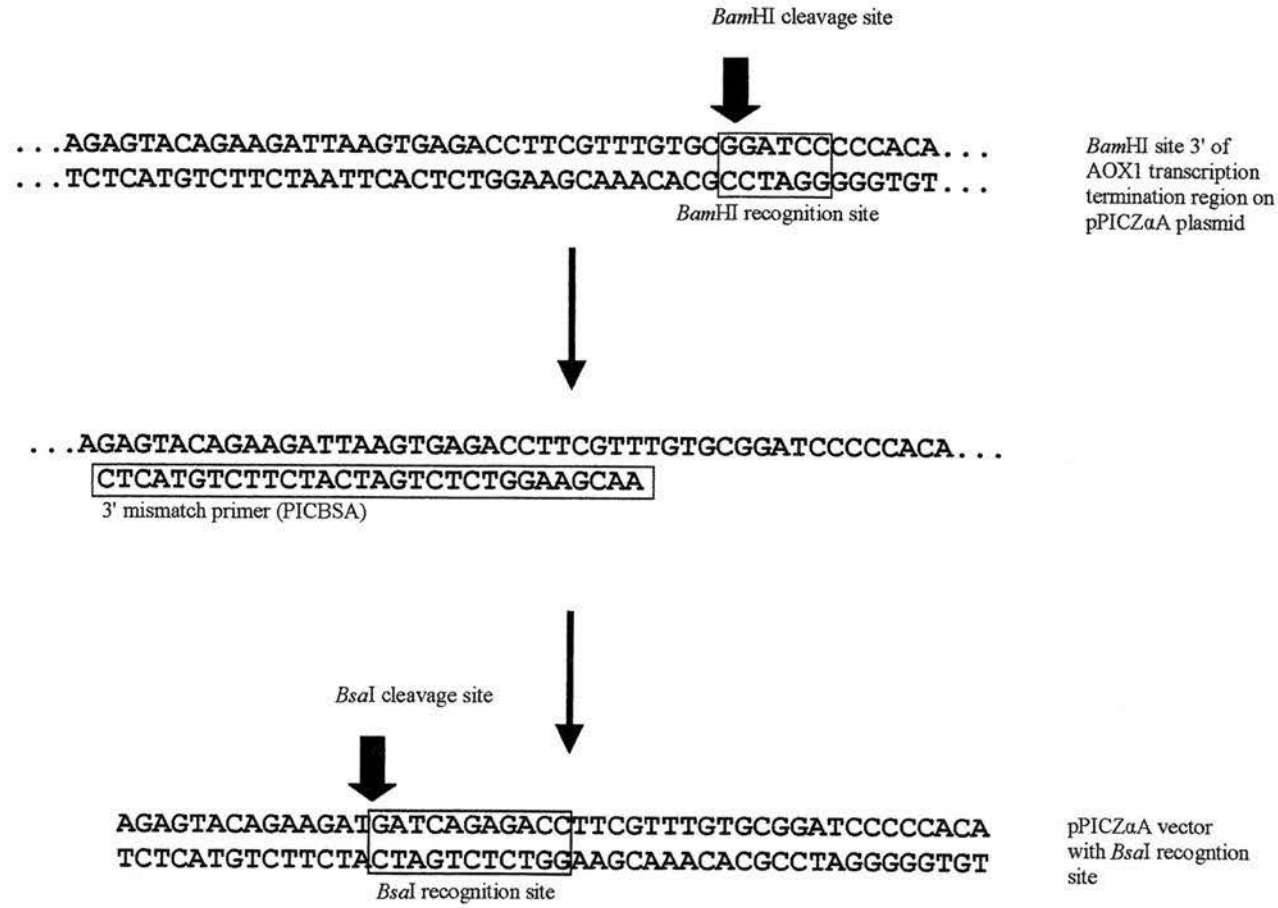


Figure 2.2. Site directed mutagenesis of GBA cassette using mismatch PCR to create a *Bsa*I recognition site 3′ of the *AOX1* transcription termination region. The PICBSA primer is the reverse primer used to amplify the GBA cassette from the pPICZα1-GBA vector. Restriction digest with *Bsa*I creates overhangs which are complementary to those created by *Bgl*II. The vector already contains a *Bam*HI site, which also creates an overhang that is complementary to *Bgl*II, however, the GBA gene also contains a *Bam*HI recognition site and consequently the enzyme cannot be used. The *Bsa*I site is 23bp upstream of the *Bam*HI site.

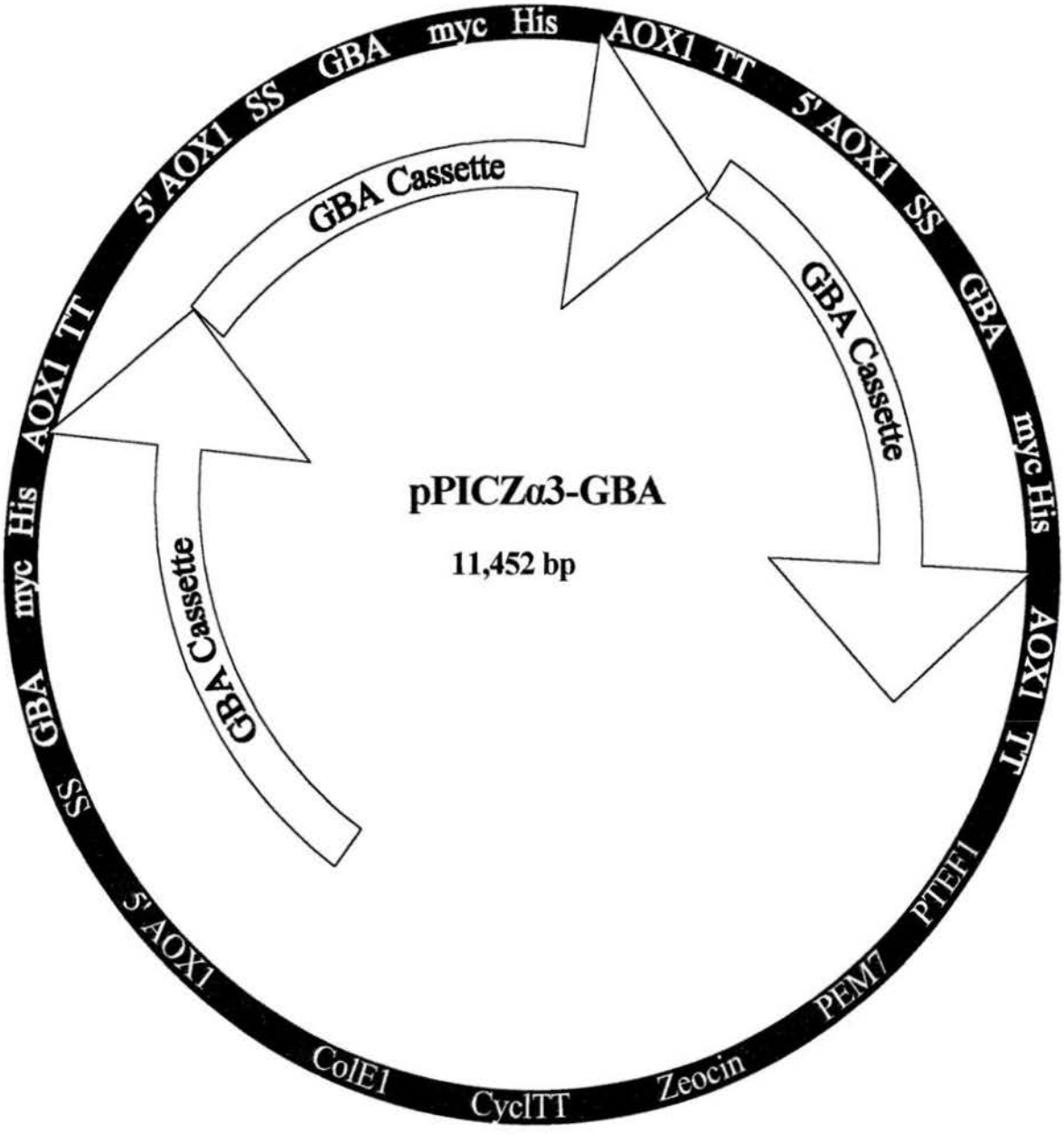


Figure 2.3. Vector pPICZα3-GBA constructed from pPICZαA to carry three copies of the coding region of glucocerebrosidase (GBA). Each copy of GBA also has the *AOX1* promoter (*5' AOX1*) and α-factor secretion signal (SS) upstream of it as well as the C-terminal *myc* epitope tag (*myc*), C-terminal polyhistidine tag (His), native transcription termination and polyadenylation signal from the *AOX1* gene (*AOX1* TT) downstream. The vector also contains single copies of the transcription elongation factor 1 gene promoter (PTEF1) which drives expression of the *Streptoalloteichus hindustanus ble* (*Sh ble*) gene in *P. pastoris*, EM7 promoter (PEM7) which drives expression of the *Sh ble* gene in *E. coli*, *Sh ble* gene which confers resistance to Zeocin™, *CYC1* transcription termination region (*CYC1*TT) which allows efficient 3' processing of the *Sh ble* gene, and the ColE1 origin (ColE1) for replication and maintenance of the plasmid in *E. coli*.

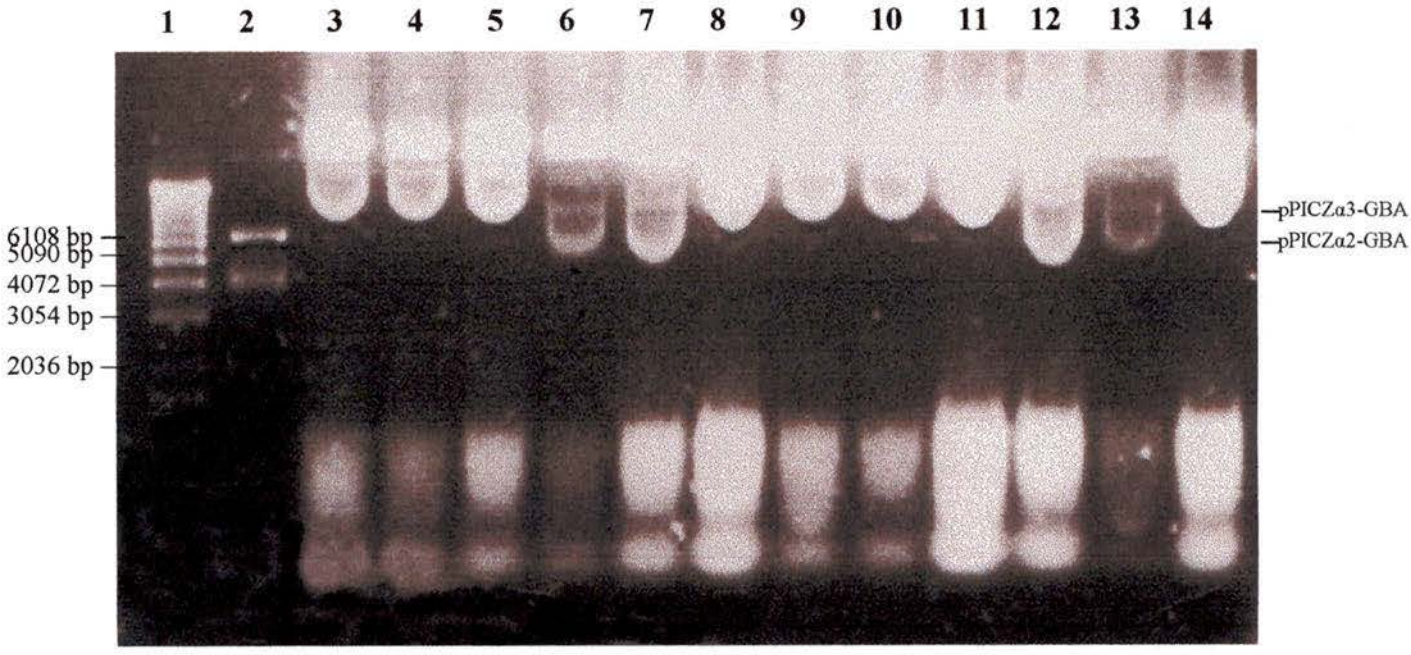


Figure 2.4. Agarose gel of DNA isolated from recombinant *E. coli* transformed with a ligation product consisting of linearized pPICZα2-GBA and a GBA cassette. Lane 1) 1 kb DNA ladder; lane 2) pPICZα2-GBA as a control; lanes 3 to 14) DNAs from the *E. coli* transformants. Lanes 3, 4, 5, 8, 9, 10, 11, 14 contain the pPICZαA plasmid with three copies of the GBA cassette, while lanes 6, 7, 12, 13 contain the pPICZα2-GBA plasmid.

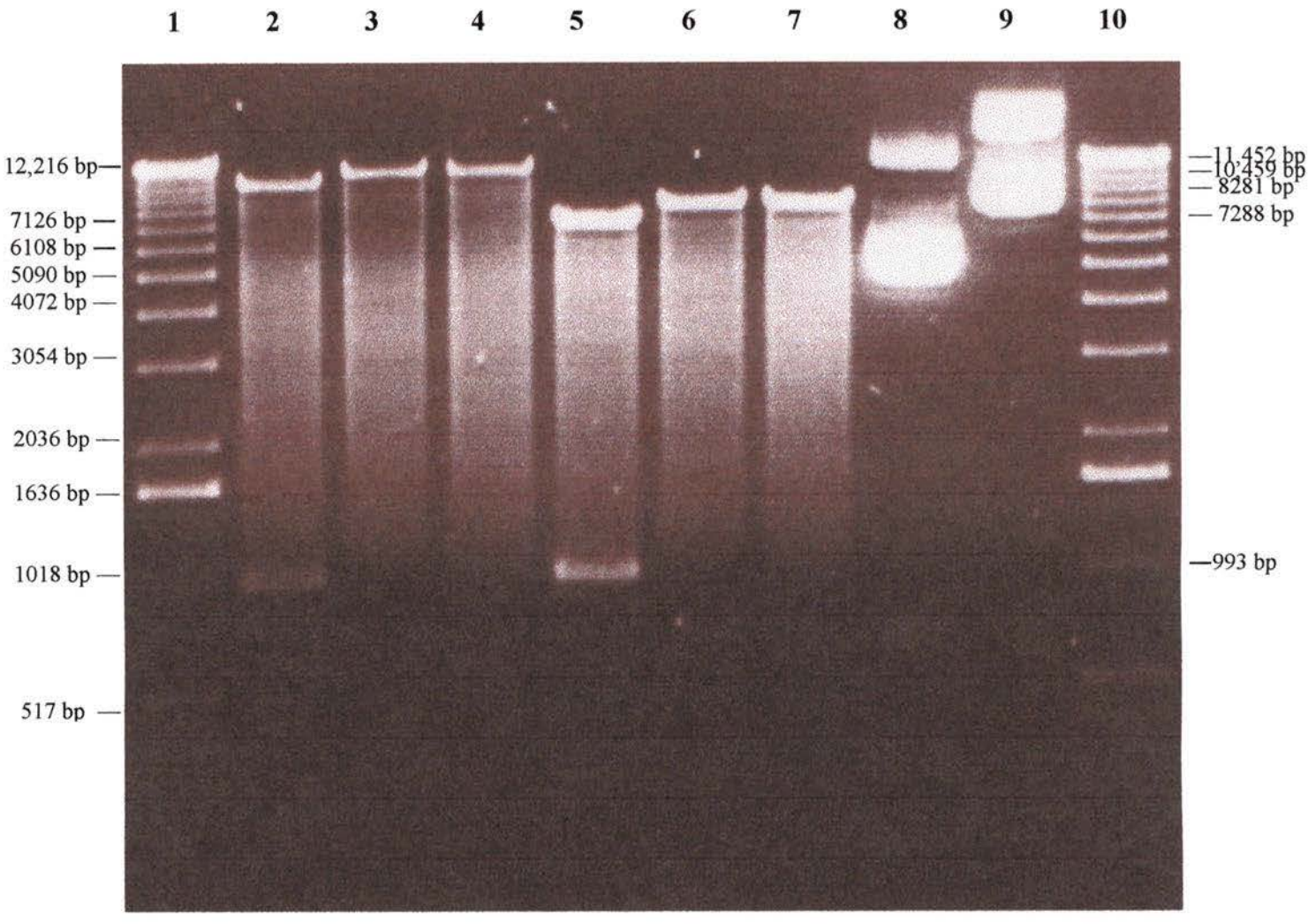


Figure 2.5. Agarose gel of restriction digested pPICZαA with three copies of GBA cassette to determine the orientation of GBA cassettes and pPICZα2-GBA as a control. The plasmids were isolated from *E. coli* transformants and double digested with *Eco*RV and *Bgl*II. Lanes 1 and 10) 1 kb ladder; lane 2) pPICZαA vector with three GBA cassettes (pPICZα3-GBA) digested with *Bgl*II and *Evo*RV; lane 3 pPICZα3-GBA digested with *Bgl*II; lane 4 pPICZα3-GBA digested with *Eco*RV; lane 5) pPICZαA vector with two GBA cassettes (pPICZα2-GBA) digested with *Bgl*II and *Evo*RV; lane 6) pPICZα2-GBA digested with *Bgl*II; lane 7) pPICZα2-GBA digested with *Eco*RV; lane 8) circular pPICZα2-GBA; lane 9) circular pPICZα3-GBA.

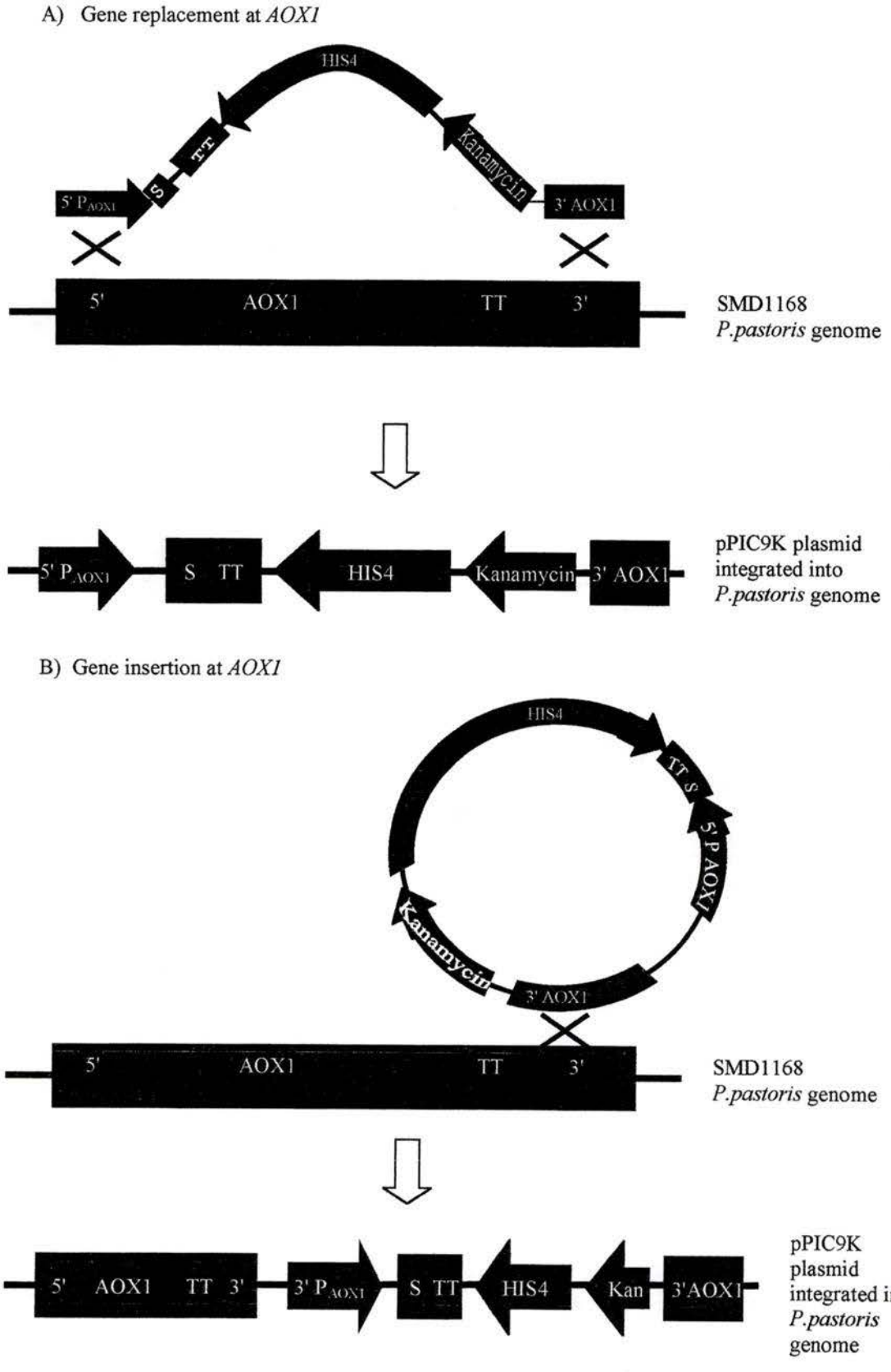


Figure 2.6. A) Integration of *BglII* linearized pPIC9K into *P. pastoris* SMD genome by gene replacement. B) Integration of circular pPIC9K into *P. pastoris* SMD genome by gene insertion.

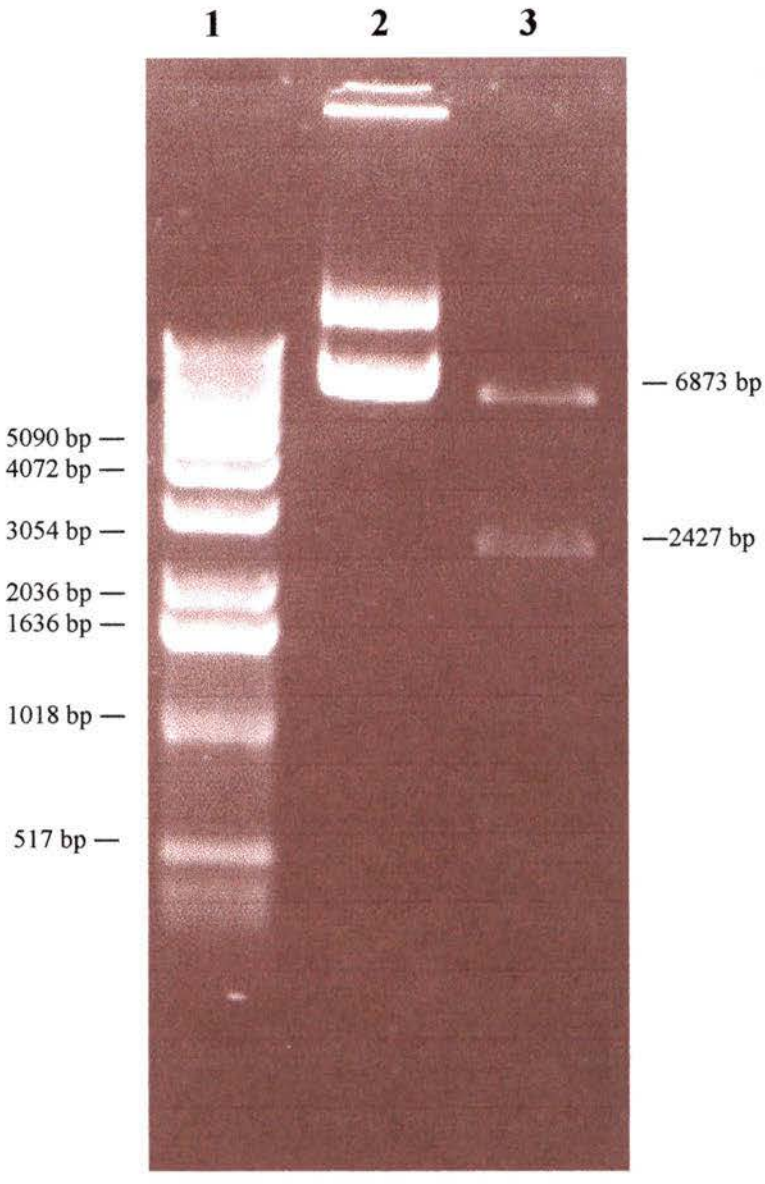


Figure 2.7. Confirmation of complete *Bgl*II cleavage of pPIC9K vector, which is required for transformation of SMD1168. Lane 1 contains the 1 kb ladder, lane 2 contains the circular pPIC9K vector, and lane 3 contains the *Bgl*II digested vector. *Bgl*II digests the pPIC9K vector into two fragments, sized 6873 bp and 2427 bp.

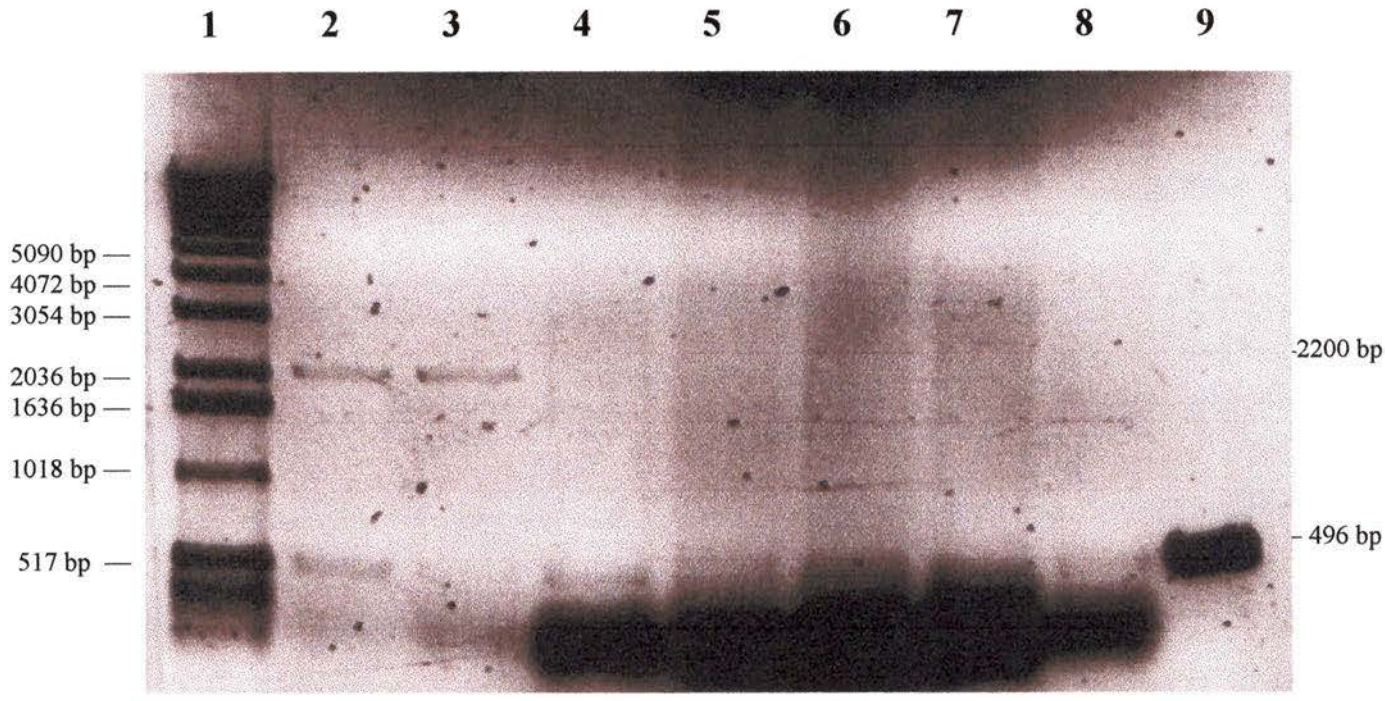


Figure 2.8. PCR confirmation of pPIC9K plasmid integration in the genome of *P. pastoris* SMD1168. PCR amplification was performed using genomic DNA of the *P. pastoris* transformants as template in lanes 2 through 8 and pPIC9K as template in lane 9 for the positive control. A reaction with non-transformed cells as template was included as a negative control and another without template as a negative control for contamination (data not shown). The 5'AOX1 (GACTGGTTCCAATTGACAAGC) and 3'AOX1 (GCAAATGGCATTCTGACATCC) primers were used to amplify the AOX1 locus. The 496 bp band in lanes 2 to 9 represents the locus from the pPIC9K vector and the 2200 bp band in lanes 2 and 3 represents the complete *AOX*1 gene from the *P. pastoris* genome.

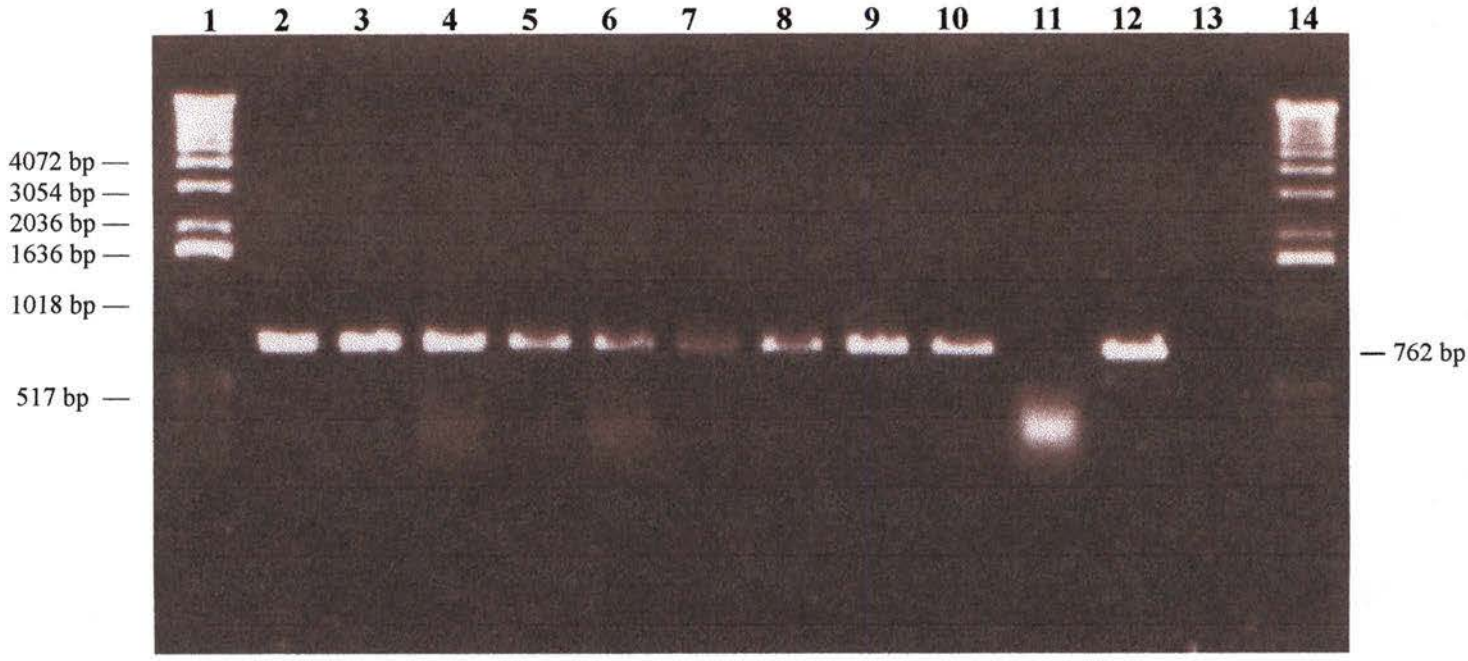


Figure 2.9. PCR confirmation of pPICZα3-GBA integration in *P. pastoris* using primers 5′AOX1 and G9. Lanes 1 and 14) 1 kb ladder; lanes 2 to 10) recombinant *P. pastoris*; lane 11) non-recombinant *P. pastoris*; lane 12) pPICZα3; and lane 13) negative control, which lacks template DNA. The 762 bp fragment confirms the integration of pPICZα3-GBA in *P. pastoris*.

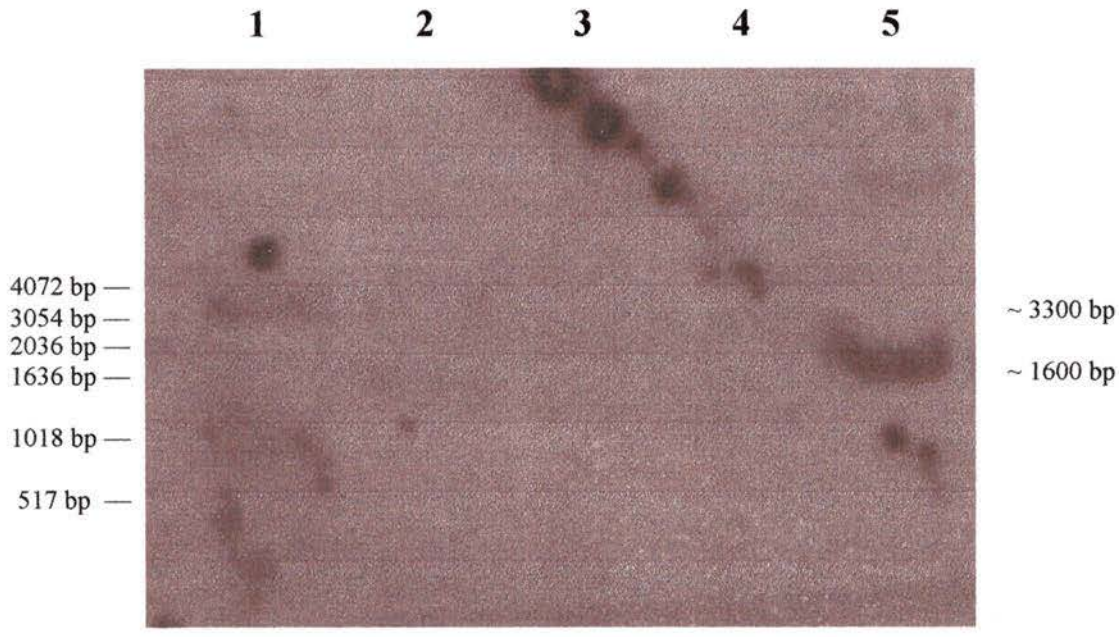


Figure 2.10. Southern blot analysis of DNA from *P. pastoris* transformed with pPICZα3-GBA. DNA was isolated from transformants, digested with *Eco*RI, separated using electrophoresis and probed with an oligonucleotide specific to GBA. Lanes 1 and 2) SMD1168 transformants; lanes 3 and 4) non-transformed SMD1168 as a negative control; and lane 5) KM71 transformant confirmed for expression of heterologous GBA as a positive control. The ~3.3 kb fragment in lane 1 corresponds to the expected size of the GBA fragment in SMD1168, while the 1.6 kb band which is visible in lane 5 corresponds to the expected size of the GBA fragment in KM71. The KM71 and SMD1168 transformants were created using different vectors resulting in the size difference seen in this Southern blot.

Table 2.1. Number of *P. pastoris* colonies transformed with pPICZα1-GBA, pPICZα2-GBA and pPICZα3-GBA that grew on plates containing various concentrations of Zeocin™. *P. pastoris* was transformed with each of the three vectors and allowed to recover using an outgrowth procedure. The cells were plated in 200 µl aliquots onto YPD plates containing Zeocin™ in concentrations ranging from 100 µg/ml to 2000 µg/ml. The average number of transformants that appeared after 2-4 days is indicated.

Plasmid	Zeocin conc. (µg/ml)	Average number of colonies
pPICZα1-GBA	100	5
pPICZα1-GBA	500	2
pPICZα1-GBA	1000	2
pPICZα1-GBA	2000	0
pPICZα2-GBA	100	56
pPICZα2-GBA	500	9
pPICZα2-GBA	1000	1
pPICZα2-GBA	2000	1
pPICZα3-GBA	100	44
pPICZα3-GBA	500	2
pPICZα3-GBA	1000	2
pPICZα3-GBA	2000	0

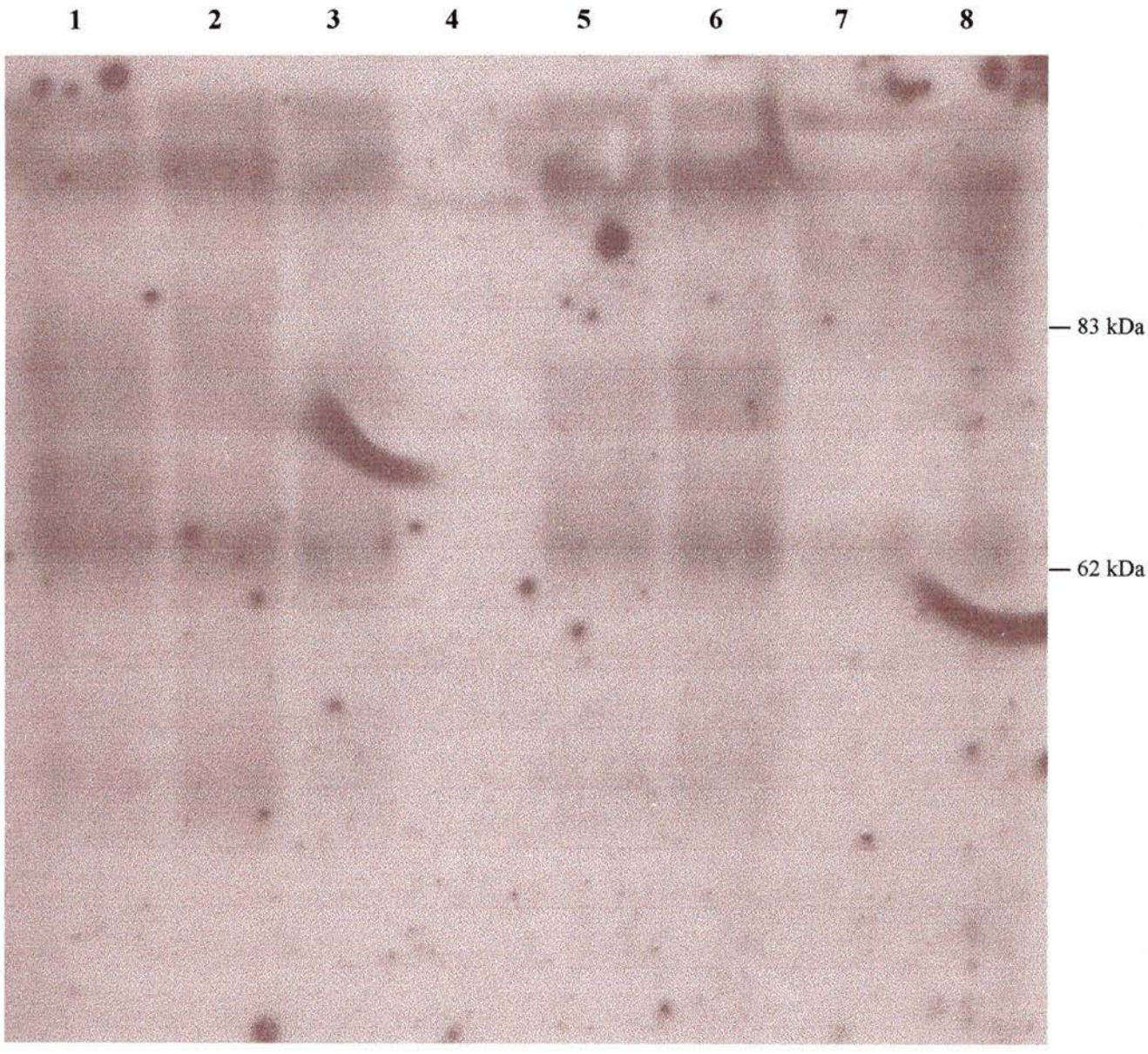


Figure 2.11. SDS-PAGE and Western blot analysis of glucocerebrosidase from supernatant of recombinant clones at 80 hours following induction. *P. pastoris* transformants were induced to express recombinant GBA and their supernatant was collected at 80 hours following induction. The supernatant was electrophoresed and the proteins were transferred to a membrane. Antibodies against GBA were used to identify the recombinant GBA enzyme. Lanes 1 through 6 contain recombinant *P. pastoris* clones. Lanes 7 and 8 contain SMD1168 and SMD(His$^+$) as the negative controls.

2.5. Discussion

I have demonstrated recombinant glucocerebrosidase (rGBA) expression in protease deficient *P. pastoris*. The heterologous glucocerebrosidase is enzymatically active and immunologically reactive, although the level of expression is low. Comparison of rGBA expression between protease deficient and other strains suggests that at the shake flask level proteolytic instability does not affect recombinant glucocerebrosidase. Gene dosage does not measurably increase or decrease the level of rGBA expression. Furthermore, the methanol utilization phenotype (Mut^+ or Mut^s) and His phenotype (His^+ or His^-) of the *P. pastoris* strain also does not affect rGBA expression.

2.5.1. Vector construction

Secreted expression of GBA, rather than intracellular expression, was chosen based on the findings of Wei (1998), which demonstrated that intracellular expression of GBA in *P. pastoris* did not produce enzymatically active recombinant protein. Secretion has the additional benefit of simplifying purification of the recombinant protein.[51] The α-factor signal from *Saccharomyces cerevisiae* was used to secrete rGBA instead of the native leader sequence of GBA. Wei (1998) demonstrated the GBA native leader sequence resulted in the production of a truncated rGBA polypeptide that was not properly secreted, while the α- factor signal allowed the complete protein to be secreted. In general, yeast signal sequences have proved to be more effective than native signal peptides, and in some cases have increased the yield two to three fold.[9;76]

The pPICZα vector was chosen to create the multimer plasmid containing several copies of GBA primarily due to the vector's small size, ~3.3 kb. Its small size is a result of the presence of only one selectable marker, Zeocin™ resistance, which has a reading frame 375 bp and is effective in both *E. coli* and *P. pastoris*. Most other vectors contain ampicillin resistance for selection in *E. coli*, as well as histidine and kanamycin for selection in *P. pastoris*.[51] The use of Zeocin™ resistance reduces the size of the vector by at least 2.5 kb. The pPICZα vector is designed so that once a recombinant vector with a single copy of the desired gene has been constructed, *Bam*HI and *Bgl*II can be used remove the gene and surrounding regions. This cassette can then be religated into a *Bgl*II linearized recombinant plasmid to produce a 2 copy multimer vector. The presence, however, of a *Bam*HI site in the GBA gene and the lack of another restriction site that would produce complementary sticky ends, necessitated the creation of an alternate restriction enzyme recognition site using site directed mutagenesis. The creation of a *Bsa*I site adjacent to the *Bam*HI site enabled the GBA cassette to be digested with *Bgl*II and *Bsa*I and inserted into the linearized vector using complementary sticky ends (Figure 2.2).

With the addition of a single GBA cassette into pPICZαA, the vector was 5.1 kb in size. This size increased to 8.3 kb with two cassettes and to 11.5 kb with three (Figure 2.3). Multimers with more than three GBA cassettes were not successfully created. This is likely a result of the large size of the three insert vector and the considerable size increase, ~3.3 kb, which would be required to insert a fourth GBA cassette. This is supported by the observation that a higher frequency of recombination events occurred when the creation of this large vector was attempted. Nonetheless it is possible to create

larger multimer vectors as demonstrated by the construction of a 17 kb plasmid containing six copies of IGF-1.[20]

2.5.2. Construction of recombinant *P. pastoris*

The glucocerebrosidase gene was inserted into the *P. pastoris* genome using the pPICZα1-GBA, pPICZα2-GBA, or pPICZα3-GBA vectors. The use of a circular plasmid in the transformation favours integration via gene insertion, while the use of a linearized vector favours integration via gene replacement (Figure 2.1).[56] The crossover event(s) between the vector and *P. pastoris* genome occur at a regions of homology between the two. With pPICZα, crossovers could occur between the *AOX1* regions on the vector (*AOX1* promotor and *AOX1* transcription termination region) and any of the three *AOX*1 regions in the yeast (*AOX1* promotor, *AOX1* transcription termination region, or the region downstream of *AOX1*). I confirmed the integration of the glucocerebrosidase into *P. pastoris* using PCR and Southern blot analysis (Figures 2.9 and 2.10). PCR was used to amplify a portion of DNA from the *AOX1* promotor to exon 4 of glucocerebrosidase.

2.5.3 Protein expression and analysis

P. pastoris transformants that were confirmed for the presence of GBA cassettes by PCR and Southern blot analysis, were induced to express rGBA. The *P. pastoris* cultures were grown to optimal cell density and then transferred from glycerol to methanol as a carbon source. The methanol induces *P. pastoris* to express alcohol oxidase which is required for the cell to metabolize methanol.[75] Since GBA is downstream of the alcohol oxidase promotor it should be expressed when the cells are

grown on methanol. The ability to grow *P. pastoris* to the optimal density prior to expressing rGBA is beneficial as production of rGBA appears to affect cell growth. Recombinant *P. pastoris* grow more slowly and have more extreme pH fluctuations than the negative control when grown with methanol as the sole carbon source. It is likely that the production, secretion, and glycosylation of GBA is taxing to *P. pastoris* and consequently affects its growth.

The fluctuation in pH of the induced cultures was problematic, and the pH had to be maintained at 5.5 for both the health of *P. pastoris* and the enzymatic functioning of rGBA. *P. pastoris* are grown optimally at pH 4.8 to 5.2 although they are able to survive in pH conditions ranging from 3.0 to 7.0.[51] The optimal pH of glucocerebrosidase is approximately 5.5 and at higher pHs (>7.4) the enzyme is highly unstable.[48] The pH of induced cultures, however, increased to a high of 7.0 to 9.0 in 12-24 hours. This is likely due to the pH-dependent proteases and has been observed in the expression of other foreign proteins.[28] Although the SMD1168 strain of *P. pastoris* is deficient in proteinase A, it still possess partially active proteinase B. This dramatic increase in pH was not observed in the non-transformed strain, which suggests that the protease activity is associated with foreign gene expression. It is also likely that exposing the secreted rGBA to such high pH conditions will negatively affect its activity and yields. Unfortunately, in shake flask cultures it is infeasible to maintain a constant pH as it is in fermentation runs.

The proteins expressed by recombinant *P. pastoris* were evaluated by silver staining and Western blot analysis. Silver staining indicated the presence of a multitude of proteins in both the control and recombinant *P. pastoris* strains (data not shown). It

was not, however, possible to detect any distinct bands that were present in recombinant *P. pastoris* and absent in the control. Although silver staining can detect as little as 0.1-1.0 ng of polypeptide in a single band,[78] the number of proteins detected made it was difficult to determine if any bands were exclusively present in recombinant *P. pastoris*. Additionally, from the silver stained gel it is apparent that if rGBA is expressed, it is not at very high levels.

Western blot analysis using monoclonal antibodies specific to glucocerebrosidase indicated that proteins sized 65-77 kDa were immunoreactive (Figure 2.11). However the monoclonal antibody also crossreacted with proteins other than GBA as seen by the immunoreactivity with proteins in the negative control. Nonetheless, the recombinant transformants had crossreactive bands sized 65-77 kDa which were not present in the negative control. The lack of specificity of the monoclonal antibodies is not due to the antibodies themselves, as two different anti-GBA antibody preparations were tested, both of which had been shown to react with glucocerebrosidase from non-*P. pastoris* sources.[93] Part of the non-specificity is likely due to cross-reactivity of proteins naturally expressed by *P. pastoris*. To decrease the amount of non-specific reactivity the stringency of the conditions of the Western blot were increased. This, however, obscured the detection of rGBA, which suggests that the levels of immunologically reactive rGBA expressed by *P. pastoris* are low. This could result from one of three possibilities; very low levels of immunologically reactive rGBA, rGBA which is expressed at higher levels but is not immunoreactive, or rGBA which is expressed at low levels and is poorly immunoreactive.

To circumvent the difficulty of poor reactivity of the rGBA to the anti-GBA monoclonal antibody, rGBA was expressed as a fusion protein with *myc* epitope and polyhistidine tag at its carboxy terminus. This enabled the anti-c-*myc* antibody to be used to detect the c-*myc* tag fused to rGBA. The use of anti-c-*myc* antibody for Southern blot analysis, however, did not provide an improved signal (data not shown). It is possible that incomplete translation of rGBA truncated the c-*myc* tag, which is fused to the carboxy terminus of rGBA, and consequently the fusion protein could not be detected. Alternatively it suggests that the levels of rGBA expression are very low. Furthermore, the conditions of this Southern blot were not as rigorously tested as those using the anti-GBA monoclonal antibody, and further optimization may provide improved signal.

The Western blot analysis suggests that induced recombinant *P. pastoris* cultures express proteins of variable sizes (65-77 kDa) that are immunoreactive against the anti-GBA Ab. If these proteins are rGBA, a possible explanation for the heterogeneity of the expressed GBA involves variable glycosylation. Mature non-glycosylated GBA is 56 kDa and its glycosylated form is 65 kDa, although various glycosylated forms of GBA, ranging from 60-68 kDa, have been isolated from human placenta and fibroblasts.[42] Glucocerebrosidase has five N-glycosylation sites, of which the first four are normally occupied. The extent of glycosylation required for enzymatic activity is unknown, except that the glycosylation at the first site is essential.[4] The high mannose glycosylation abilities of *P. pastoris* resemble those of higher eukaryotic systems, and some heterologous glycoproteins expressed in *P. pastoris* are very similar to those produced in mammalian cells.[68] Nonetheless, expression of glycoproteins in a foreign system is expected to produce forms variable from those expressed in the native host and there are

instances where *P. pastoris* has hyperglycosylated glycoproteins or added oligosaccharide chains that significantly differ from those normally found on the protein.[80]

Most of the immunoreactive proteins are larger (65-77 kDa) than human GBA, which can range in size from 60 to 68 kDa.[48] Glucocerebrosidase was expressed as a fusion protein with *myc* and His c-tag, which increases the size of the recombinant protein by 21 amino acids. [57] The remaining size difference could be due to improper glycosylation, although alternative explanations also exist. It is possible that the α-secretion signal is improperly proteolytically cleaved. The α-factor prosequence is 10 kDa and contains three putative glycosylation sites which can increase the mass by another 7-10 kDa. This is a distinct possibility as improper proteolytic processing has been observed in the expression of other heterologous proteins in *P. pastoris*.[95]

Enzyme activity assays suggest that there are low levels of rGBA activity (10 nmol/hr/mg) in the supernatant of induced *P. pastoris* transformants. The enzyme activity of the recombinant glucocerebrosidase was analyzed using both the sphingolipid N-palmitoyldihydroglucocerebroside substrate assay and the fluorogenic 4-methlyumbelliferyl β-D-glucopyranoside (4MUGP) substrate assay. Only the sphingolipid substrate assay, however, was found to be effective. Although the artificial substrate has been used to effectively measure natural GBA as well as rGBA from several sources, in *P. pastoris* the background activity with 4MUGP is too high to accurately determine the level of activity that is due to rGBA.[58;73] Even though the conditions of the assay (low pH and presence of taurocholate, which inhibits other non-specific glucosidases) favour detection of GBA activity, the 4MUGP substrate can still be cleaved

by other glucosidases. The sphingolipid substrate N-palmitoyldihydroglucocerebroside, which is specific for GBA, provided a much more accurate indication of GBA activity.[26] Unfortunately, the high cost and limited availability of the natural substrate, restricts its usefulness as a measure of rGBA enzymatic activity.

2.5.4. Gene dosage

I selected an *in vitro* approach to increasing gene dosage over an *in vivo* method. Although constructing the multiple GBA cassette vector was more labour intensive it allowed precision in determining the number of glucocerebrosidase copies possessed by *P. pastoris*. It also allowed a step-wise analysis of the effects of copy number on protein expression. Additionally it was possible to screen the transformants for multiple plasmid insertions, which would greatly increase the number of gene copies with the 2 or 3 insert multimer. Each integration that occurs via a single crossover event would cause an additional 1, 2, or 3 copies of GBA to be incorporated into the *P. pastoris* genome, depending on the vector used.

The transformants were selected for high copy number by assessing their hyperresistancy to Zeocin™. With each additional copy of the Zeocin™ resistance gene incorporated into the genome, an additional 1-3 copies of GBA were also integrated. Initially transformants were tested by patching cells onto increasing concentrations of Zeocin™. It was found, however, that all cells grew on the high Zeocin™ concentrations, although the size and density of the patch was variable. It is probable that the high number of cells transferred in such a procedure may result in false positives for Zeocin™ hyperresistance.[51] The likelihood of false positives was decreased by directly

plating the transformants on increasing concentrations of Zeocin™ (Table 3.1). The number of GBA genes in the *P. pastoris* transformants was further evaluated using Southern dot blot (data not shown).

Increasing the number of GBA genes should maximize the level of mRNA transcripts, which would be expected to overcome difficulties associated with suboptimal 5'-untranslated sequence, mRNA secondary structure, or protein instability.[51] Variations in the level of RNA expression between transformants were detected by RNA dot blot at various time points following methanol induction (data not shown). Much of the signal, however, is non-specific as demonstrated by the presence of signal in the negative control. This high non-specific activity was also present in Northern blot analysis of other strains of *P. pastoris* transformants expressing GBA (Sinclair, G., unpublished results). Therefore, at this level of transcription it is not possible to deduce from the Northern dot blot that increasing copy number of GBA in *P. pastoris* results in higher levels of mRNA production.

Heterologous GBA expression in transformants with various copies of GBA was fairly uniform. This was not unexpected since there are numerous examples where increasing gene dosage for secreted proteins does not have a positive effect or is detrimental to secretion.[51] Although maximal gene dosage is preferred for intracellular proteins, high levels of mRNA can complicate the secretion pathway and reduce the yield of secreted proteins. It has generally been found that secreted proteins require an optimal gene dosage.[20] In some instances, increasing the gene dosage above 1 has a detrimental effect on protein expression.[80] Nonetheless, in the literature there are some examples of maximal gene dosage producing the highest levels of a secreted protein. For example

increasing gene copy number to 19 increased the expression of Murine epidermal growth factor 13 fold.[73] Thill *et al.* (1990) suggested that secreted proteins which take longer to transverse the cell (>1 hour) would be less likely to benefit from high copy number.[88] Glucocerebrosidase may be slow to transverse the cell due to its large size (55 kDa) and required glycosylation. Pulse chase experiments in human fibroblasts indicate that GBA is made within 30 minutes and continues to be processed for 15 hours afterwards.[36]

2.5.5. Effect of histidine auxotrophy or prototrophy

Histidine prototrophy was found to have no significant effect on heterologous GBA expression in protease deficient *P. pastoris* at the shake flask level. In larger scale inductions, histidine becomes a limiting factor and histidine auxotroph strains generally do not perform as well.[51]

2.5.6. Protease instability

Protease instability can be problematic in the expression of heterologous proteins. This problem can be overcome by expressing the protein in a protease deficient cell line, adding amino acids or peptide supplements to the growth media, or reducing the pH. For the expression of GBA, use of a protease deficient cell line is the best choice. Reducing the pH is not possible as GBA is not enzymatically active at low pH values, while adding additional supplements would increase the cost of producing the large volumes that would be required if this recombinant enzyme were to be produced as a pharmaceutical agent. Of the available protease deficient cell lines, SMD1168 was selected. SMD1168 is more robust than SMD1163, which was found to be sensitive to the high

concentrations of the selective agent G418 needed to screen for high copy number transformants.[95] Even so, SMD1168 is less robust than the wildtype strains and requires greater care in growth and storage.[51]

Protease degradation does not seem to significantly effect glucocerebrosidase at the shake flask level of growth. A comparison of the rGBA expression levels in protease deficient *P. pastoris* to those previously achieved in cell strains wildtype for protease activity indicates that expression is consistently low across all cell strains.[93] Although protease instability does not appear to be problematic at this small scale growth, in fermentation it is generally worsened due to the higher concentration of proteases in the denser culture or in the presence of different media.[51]

2.5.7. Conclusions and future considerations

Considerable effort has been invested in optimizing the expression of glucocerebrosidase in *P. pastoris,* but achieving the high levels of expression seen with other heterologous proteins remains elusive. There still remain several optimizations that may result in increased expression. Scaling up the culture to fermentation size could result in dramatic improvement in the level of expression since the conditions, such as oxygen supply, pH, and methanol induction can be tightly controlled. Switching from shake-flask to fermentation expression has shown to increase the level of expression of some proteins 10 fold.[27] For secreted proteins fermentation is essential, since protein yields are directly correlated with cell density, and only through fermentation can *P. pastoris* achieve very high cell densities.[51] Optimization of the GBA coding sequence for expression in *P. pastoris* can be used to overcome any difficulties associated with poor

translation due to rare codons. This is particularly relevant to GBA as the gene has a high number of codons that are rare in *P. pastoris*. For GBA to be expressed in CHO cells, the sequence near the translation start had to be altered to match the consensus sequence for optimal translation in mammalian cells.[73]

It would also be of value to further analyse the post-translational modifications of recombinant glucocerebrosidase. This research suggests the enzyme is heterogeneously glycosylated, and some forms are larger than the expected size. This difference in size could be entirely due to the addition of glycans or the α-factor secretion signal may not have been properly cleaved. Removal of the oligosaccharide chains would allow the mass of unglycosylated GBA to be determined, which would be expected to be 55kDa for GBA and 65kDa for GBA with the α-factor signal. Furthermore it would be of interest to study the types of oligosaccharides that were added onto GBA to determine how similar *P. pastoris* glycosylation of rGBA is compared to mammalian.

Unfortunately, the possibility exists that heterologous glucocerebrosidase will never be expressed at high levels in *P. pastoris*. This yeast has been shown to inappropriately glycosylate glycoproteins and inefficiently secrete some larger (>30kDa) proteins.[38] For example, the HIV gp120 glycoprotein, which is similar in size to GBA, was very poorly secreted and improperly glycosylated in *P. pastoris*. Nonetheless, investigating the expression of rGBA in *P. pastoris* has contributed to the understanding of complex protein expression in this yeast.

Chapter 3 Evolution of the glucocerebrosidase gene and pseudogene in primates

3.1. Abstract

Humans have one functional GBA gene (*GBA*) and one GBA pseudogene (ψ*GBA*) that share 96% overall sequence similarity. In order to investigate the molecular evolution of the *GBA* and ψ*GBA*, I have sequenced approximately 1.1 kb of the C-terminal region of nine glucocerebrosidase (GBA) genes from orangutan, gorilla, baboon, squirrel monkey, and chimpanzee. *GBA and ψGBA* were PCR amplified from genomic primate DNA, cloned, sequenced, and analysed using DNAstar, BCM, and PAUP4.0 software. These data indicate that orangutan, gorilla, baboon, and chimpanzee each have two copies of GBA, while the squirrel monkey only possesses a single copy. Analysis of the 115 a.a. sequence indicates that several of these genes are non-functional. The two GBA genes in each species are least similar in gorilla and chimpanzee (97.8%) and most similar in orangutan (99.5%) and baboon (99.4%). The increased similarity of the GBA genes in baboon and orangutan, as well as phylogenetic tree analysis indicates that gene conversion caused the evolution of *GBA* and ψ*GBA* to be concerted.

3.2. Introduction

The evolution of glucocerebrosidase is of particular interest because its duplication has contributed to human pathology through interaction between the functional gene and pseudogene.[60;87;98] Approximately 10% of *GBA* mutations, causing Gaucher disease in the homozygous state, are believed to be caused by recombination or gene conversion between the human *GBA* and ψ*GBA* genes.[11;48] Nonetheless little is known about the molecular evolution of glucocerebrosidase and its pseudogene. In 1989

the complete genomic sequence of human *GBA* and ψ*GBA* was delineated by Horowitz *et al.*.[54] That same year the cDNA of the murine *GBA* gene was characterized and localized to chromosome 3E3-F1.[71] The gene rich area surrounding human glucocerebrosidase on chromosome 1q21 has recently been investigated in greater detail and the precise location of the duplicated region leading to the development of the pseudogene has been identified.[96]

It is believed that ψ*GBA* arose from the duplication of an ancestral *GBA* gene and the subsequent divergent evolution of both genes, which resulted in a loss of function in one gene.[54] Gene duplication is an essential part of genome evolution and has been implicated as being the primary source of 'new' functional genes.[50] However, deleterious mutations occur more frequently than advantageous mutations, and consequently it is far more likely for a duplicated gene to evolve into a pseudogene.[50] Whether the duplicated gene becomes a pseudogene or gains a new function, it will likely be more similar to the other member(s) of its gene family than would be expected through independent evolution. This is due to concerted evolution, which occurs through the exchange of information between members of the gene family, either by unequal crossing over or gene conversion.[50]

Gene conversion is a nonreciprocal recombination process between two DNA strands where by sequences within one gene are "converted" to sequences of the other gene. The area altered in gene conversion can range from a few nucleotides to several thousand base pairs, although the size is limited as generally only the heteroduplex region is involved. Gene conversion between duplicated genes has been found to occur in every

species,[50] and has been suggested as a mechanism of gene homogenization in human α- and γ-globin gene families.[82]

The duplication of GBA has been identified as a tandem duplication event and the end points of the duplication have been defined (Figure 3.1).[54;96] The duplicated area stretches from 5.5 kb upstream of *GBA* to within exon 2 of metaxin (*MTX*), which is contiguous to, and transcribed convergently with *GBA*.[62;96] While the duplicated *MTX* gene is non-functional, as it is missing all of exon 1 and part of exon 2, the entire *GBA* gene was duplicated. Thus, immediately after the duplication event, two functional *GBA* genes existed. Analysis of the human GBA genes, however, indicates that through subsequent evolution the ancestral GBA gene became non-functional, while the duplicated *GBA* remained functional.[54]

Comparison of human *GBA* and ψ*GBA* indicates that they share 96% similarity in their genomic sequence.[54] The lack of functionality in the pseudogene is due in part to two exonic deletions, 5 bp and 55bp in exons 4 and 9, respectively. The difference in size of the *GBA* and ψ*GBA* genes results from the insertion of *Alu* sequences primarily into the intronic regions of *GBA* (Figure 3.2).[54] The *Alu* family is one of the most abundant families of retrosequence transposible elements, which are characterized by their ability to insert themselves into the genome. *GBA* contains partial, single, or multiple *Alu* inserts in introns 2, 4, 6, 7 sized 313 bp, 626 bp, 320 bp, and 277 bp respectively that are absent in the pseudogene.[54] Excluding the intron 7 *Alu* insert, all these transposible elements were introduced into *GBA* following duplication. This suggests that in terms of size, the pseudogene bears a closer resemblance to the ancestral gene.[54]

The presence of *Alu* inserts helps to date the duplication event of *GBA* at approximately 36-40 million years ago (mya). The *Alu* repeat in intron 7 shared by *GBA* and ψ*GBA* belongs to the *Sx* family, which is 40 million years old.[64] Since this repeat is in both GBA genes it must have been inserted into ancestral *GBA* prior to the duplication event, suggesting that the duplication event could not have occurred more than 40 mya. This date of *GBA* duplication is further supported by the analysis of gene copy number in various species. Human and rhesus macaque have two *GBA* genes, while mouse, chicken and rat only have one.[21;54;71] This suggests that the duplication event occurred prior to the divergence of old world monkeys, which was approximately 36-48 mya.[59;69]

The relatively recent duplication event hypothesized is supported by the high homology between *GBA* and ψ*GBA* (96%) including conserved splice sites, as well as similar sized coding, intronic (excluding *Alu* sequences) and untranslated regions.[54] Additionally the adjacent *MTX* and ψ*MTX* genes show 98.9% homology in the coding regions, excluding the truncated 5' end in exon 2 of the pseudogene. The ψ*GBA* promoter still retains some activity as evidenced by the detection of ψ*GBA* mRNA in human cell lines[55;83] and by effectively attaching the promotor to a reporter gene.[74] ψ*GBA* was translated *in vitro* to produce a 30 kDa polypeptide, which is considerably smaller than the 65 kDa enzyme produced by *GBA*.[55]

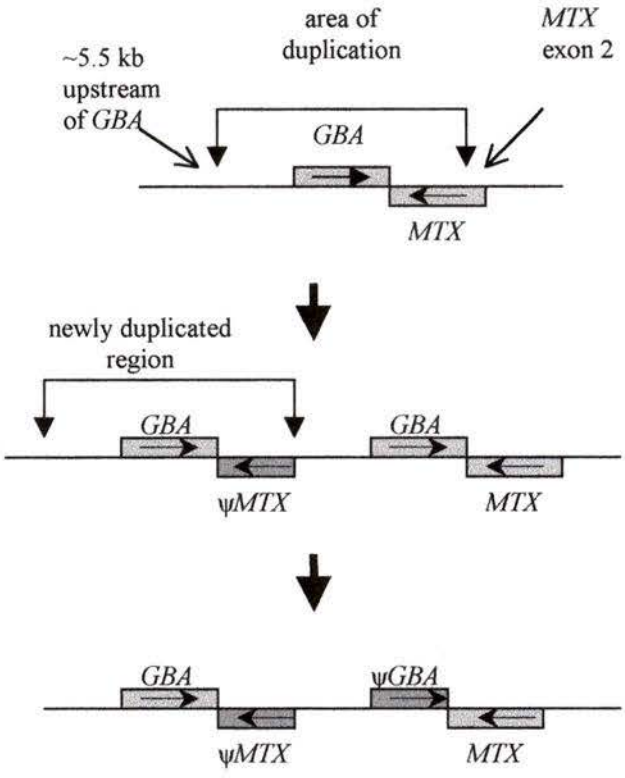
GBA is a highly conserved gene as evidenced by comparing the cDNA for mouse and human *GBA*.[71] Murine *GBA* is located on chromosome 3E3-F1 and shares 84% similarity with the exonic regions of human *GBA* in the protein coding regions, although the similarity drops to 53% in the 5' noncoding region and to 78% in the 3' noncoding regions.[71] The deduced amino acid sequence of the murine *GBA* is 86% similar to human

GBA and only a single amino acid deletion at position 273 is predicted.[71] The C-terminal is more highly conserved, while the N-terminal has the highest rate of variation.[71] Since the structural domains for glucocerebrosidase and active site are located in the 3' end, it is likely that mutations in this area negatively affect the enzyme and thus are less likely to become established.[33] High conservation is also maintained in the positions of all the intron/exon splicing regions, which are identical in both mouse and human *GBA*.[71]

A more detailed account of the molecular evolution of *GBA* and ψ*GBA* can be obtained by comparing *GBA* genes from several species. In order to examine both the *GBA* and ψ*GBA* it is necessary to analyze species that have diverged from their common ancestor with humans less than 40 mya, since that is the earliest the *GBA* duplication event is hypothesized to have occurred.[54] This limits the analysis to great apes and old world monkeys (Figure 3.3). The great apes (*Pongidae*) include chimpanzee (*Pan troglodytes* and *Pan paniscus*), gorilla (*Gorilla gorilla*), and orangutan (*Pongo pygmaeus*). The chimpanzee is the closest living relative of humans, having diverged approximately 4.7-7.1 mya, while the gorilla is slightly more distant at 5.3-8.0 mya.[66] The orangutan is thought to have diverged from the common ancestor to African apes and humans, approximately 15 mya.[69] Old world monkeys are the next closest relatives to humans, having diverged approximately 30 mya. There are over sixty species that are considered old world monkeys, including the rhesus monkey (*Macaca mulatta*) and baboon (genus *Papio*).

This chapter of my thesis examines the changes that *GBA* and ψ*GBA* have undergone through evolution by analyzing a contiguous 1059-1149 bp region starting in exon 8 and ending near the end of exon 10. I have sequenced this portion of the

glucocerebrosidase genes in gorilla, chimpanzee, baboon, orangutan, and squirrel monkey, and deduced the 115 a.a. sequence in this region. I analyzed these sequences using DNAstar and PAUP4.0 to obtain insight into how GBA and ψGBA diverged following duplication. The orangutan, gorilla, baboon, and chimpanzee each have two copies of the gene, while the squirrel monkey only possesses a single copy. Analysis of the genes indicates that in gorilla and chimpanzee, one copy is functional while the other is a pseudogene, whereas in the orangutan, both GBA genes may be functional. The two GBA genes in each species are least similar in gorilla and chimpanzee (97.8%) and most similar in orangutan (99.5%) and baboon (99.4%). The increased similarity of the GBA genes in baboon and orangutan, as well as phylogenetic tree analysis indicates that gene conversion caused the evolution of *GBA* and ψ*GBA* to be concerted. Furthermore, the distribution of nucleotide variation is not equal throughout this region of the gene.



The ancestral *GBA* gene was duplicated through a tandem duplication event encompassing the area approximately 5.5 kb upstream of *GBA* to exon 2 of *MTX*, for a total of about 16 kb.

The newly duplicated region was positioned directly adjacent to the original *GBA* and *MTX* genes. However the duplicated *MTX* gene was missing all of exon 1 and part of exon 2 and therefore was a pseudogene.

The genes accumulated mutations and in humans the original *GBA* gene has become a pseudogene.

Figure 3.1. Events surrounding the duplication of *GBA* and *MTX*.

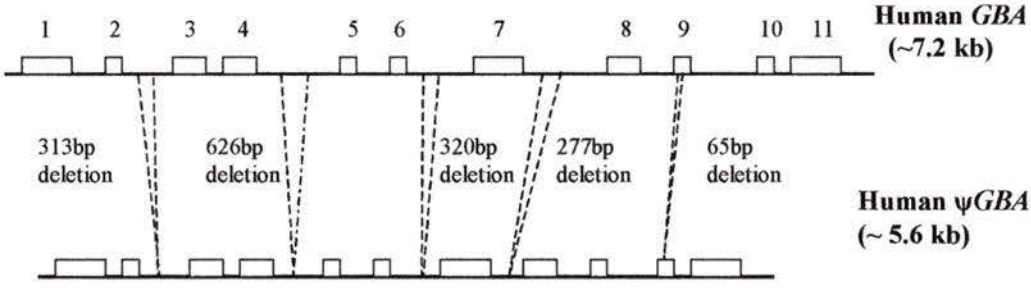


Figure 3.2. Comparison of human glucocerebrosidase functional gene (*GBA*) and pseudogene (ψ*GBA*). There are 11 exons in each of the genes and the the exon/intron boundaries are identical. The exons are boxed and numbered. The pseudogene is smaller in size compared to the functional gene as a result of 4 large deletions in introns 2, 4, 6, and 7, and one smaller deletion in exon 9 totalling ~1.6 kb. The deletions are marked by dashed lines and their size is indicated in bp. Most of the intronic deletions are thought to be due to insertion of *Alu* sequences in the functional gene as well as short (~10 bp) repetitive sequences flanking the *Alu* sequence that are associated with the *Alu* integration event.

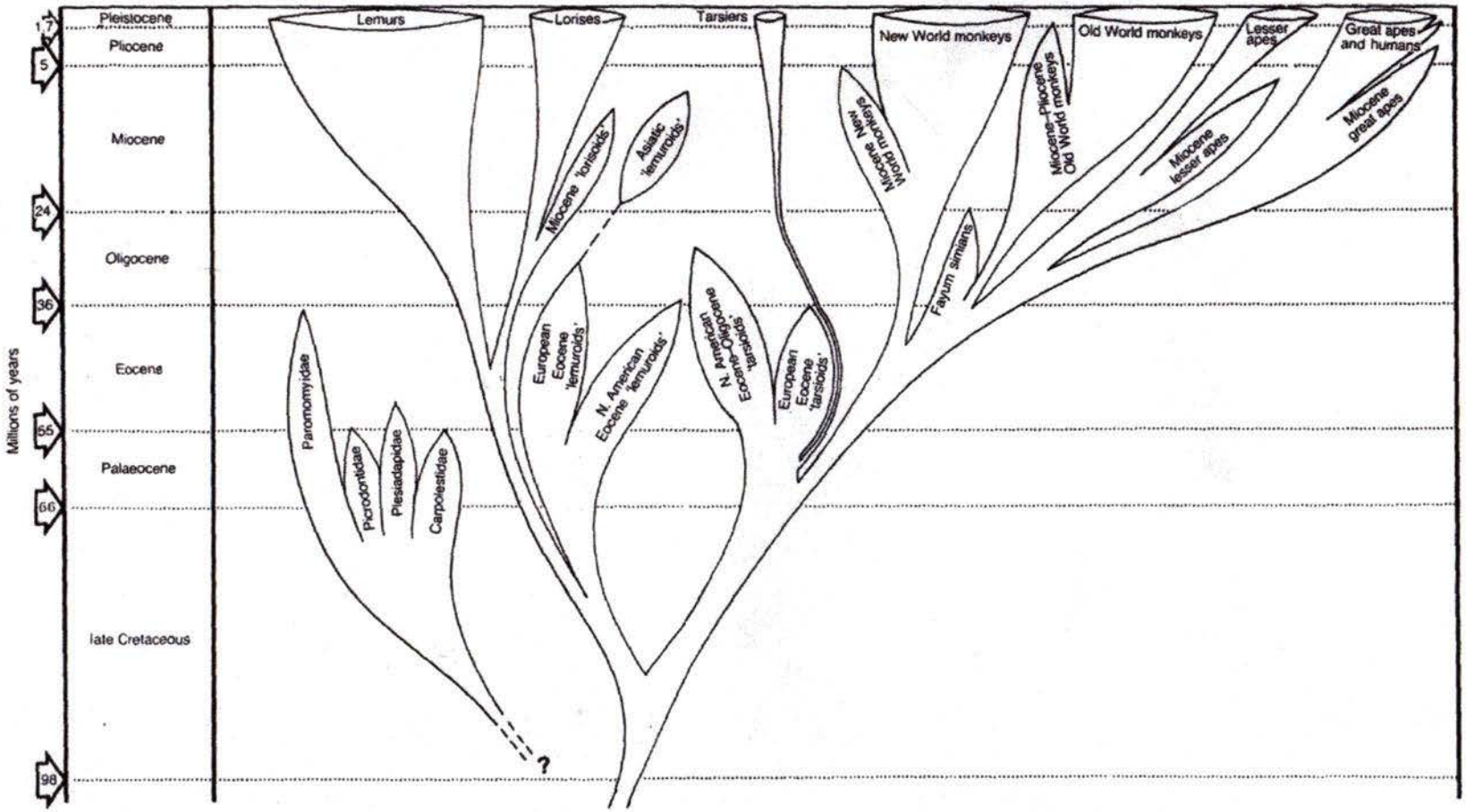


Figure 3.3. Phylogenetic tree for primates. Figure is from Martin (1990).

3.3. Materials and methods

3.3.1. Chemicals and reagents

The following were obtained from commercial sources. 2-mercaptoethanol (ACP, Montreal, Quebec); Agar (Difco laboratories, Detroit, MI); Agarose (EM Science, Gibbstown, NJ); AlkPhos direct labeling kit (Amersham Life Sciences, Buckinghamshire, England); Ampicillin (Sigma, St.louis, MO); Boric Acid (ACP, Montreal, Quebec); Bromophenol blue (BDH Inc., Toronto, ON); Calf intestine alkaline phosphatase (New England Biolabs, Beverly, MA); Chloroform (ACP, Montreal, Quebec); Deoxynucleotides (GIBCO/BRL, Grans Island, NY); Dextrose (ACP, Montreal, Quebec); DNA 1kb ladder (GIBCO/BRL, Grans Island, NY); DNAzol (GIBCO/BRL, Grans Island, NY); dNTPs (GIBCO/BRL, Grans Island, NY); Dry milk (Becton Dickenson, Cockeysville, MD); EDTA (ACP, Montreal, Quebec); Ethanol (ACP, Montreal, Quebec); Ethidium bromide (Fisher Scientific, Fair Lawn, NJ); Glycerol (BDH Inc., Toronto, ON); Hybond™-P and Hybond™-N (Amersham Pharmacia Biotech, Quebec); IPTG (GibcoBRL, Grans Island, NY); Low melting point agarose (Eclipse, Mississaugua, ON); Magnesium chloride ($MgCl_2 \cdot 6H_20$) (BDH Inc., Toronto, ON); Magnesium sulfide ($MgS0_4 \cdot 7H_20$) (ACP, Montreal, Quebec); Potassium acetate, Potassium chloride (KCl), and Potassium phosphate (BDH Inc., Toronto, ON); Restriction endonulceases (New England Biolabs, Beverly, MA); RNAaseA (GIBCO/BRL, Grans Island, NY); SDS (BDH Inc., Toronto, ON); Sodium acetate, Sodium chloride (NaCl) and Sodium hydroxide (NaOH) (BDH Inc., Toronto, ON); Sucrose (ACP, Montreal, Quebec); *Taq* DNA polymerase (GIBCO/BRL, Grans Island, NY); Tris base (Boehringer Mannheum, Indianapolis, IN); Tryptone peptone (Difco

laboratories, Detroit, MI); Tween-20 (Fisher Scientific, Fair Lawn, NJ); X-gal (GibcoBRL); Xylene cyanol (Sigma, St.louis, MO); Yeast extract (Difco laboratories, Detroit, MI)

3.3.2. Media and prepared solutions

LB media for *E. coli*: 1% tryptone, 0.5% yeast extract, 1% NaCl and adjusted to a pH of 7.0. For low salt LB, 0.5% of NaCl added instead of 1%. Sterilized by autoclaving. For solid media 2% agar was added prior to autoclaving and media was poured in 150 mm plates. **SOC media for *E. coli***: 2% tryptone, 0.5% yeast extract, 10 mM NaCl. Sterilize by autoclaving. Add 2.5 mM KCl, 10 mM $MgCl_2 \cdot 6H_20$, 10 mM $MgS0_4 \cdot 7H_20$, 20 mM 2 M dextrose (all solutions sterilized by filter) and adjust pH to 7. **TE buffer (pH 8.0)** : 10 mM Tris•HCl (pH8.0), 1 mM EDTA (pH 8.0). **Cracking buffer**: 0.05 M Tris-HCL (pH 7.5), 2 M EDTA, 1% SDS, 14% sucrose, 0.01% bromophenol blue, 0.01% xylene cyanol. **Cell resuspension solution**: 50 mM Tris-HCl (pH 7.5), 10 mM EDTA, 100 µg/ml NAase A. **Cell Lysis Solution:** 0.2 M NaOH, 1% SDS. **Neutralization Solution:** 1.32 M Potassium acetate (pH 4.8). **Column Wash Solution:** 80 mM potassium acetate, 8.3 mM Tris-HCl (pH 7.5), 40 µM EDTA, 55% ethanol. **40× TAE Buffer:** 1.6 M Tris base, 0.8 M Na acetate•$3H_20$, 40 mM EDTA-$Na_2 \cdot 2H_20$. Adjust pH to 7.2. **20×TBE buffer:** 1 M Tris base, 1 M Boric Acid, 20 mM EDTA-$Na_2 \cdot 2H_20$. **Gel Loading Buffer:** 50% glycerol, 1×TAE buffer, 1% Bromophenol blue, 1% Xylene cyanol. **20× SSC Buffer:** 3 M NaCl, 0.3 M Na_3 Citrate•$2H_20$. Adjust pH to 7.0.

3.3.3. Primate DNA and bacterial cell line

Chimpanzee (*Pan paniscus*), orangutan (*Pongo pygmaeus*), squirrel monkey (*Saimiri sciureus*), baboon (*Papio papio*) and gorilla (*Gorilla gorilla*) genomic DNA were graciously provided by Dr. Ben Koop (University of Victoria). *Escherichia coli* cell line *Top10F'* was obtained from Invitrogen, Carlsbad, CA.

3.3.4. PCR

i. Amplification of DNA fragments greater than 3 kb

Elongase enzyme mix (Gibco) was used to amplify up to 5.5 kb fragments from genomic primate DNA. The PCR included 1× PCR reaction buffer (Gibco), 2.5 mM $MgCl_2$, 200 µM of each dNTP, 10 µM of each primer, and 1 U elongase polymerase in a total volume of 50µl. The conditions of PCR using the GenAmp PCR System 2400 (Perkin Elmer) were 0.5 min. at 94°C, 0.5 min. at 56°C, 5 min.. at 68°C with an initial incubation of 5 min. at 94°C and a final incubation of 7 min. at 72°C. The number of cycles for amplification was 35.

ii. Amplification of DNA fragments less than 3 kb

Taq polymerase (Gibco) was used to amplify fragments which were smaller than 3kb. The PCR included 1× reaction buffer (Gibco), 2.5 mM $MgCl_2$, 250 µM of each dNTP, 10 µM of each primer, and 1 U taq polymerase in a total volume of 50 µl. The conditions using the GenAmp PCR System 2400 were 0.5 min. at 94°C, 0.5 min. at 56°C, 1 to 3 min. at 68°C with an initial incubation of 5 min. at 94°C and a final incubation of 7

min. at 72°C. The time allowed for DNA elongation was 1 min. for 1kb fragments and an additional 1 min. for each 1kb increase in size. The number of cycles for amplification was 35.

iii. Nested PCR

The primary DNA fragment was amplified using *taq* polymerase or elongase enzyme mix and then gel or direct purified using the techniques described in section 3.3.7. The secondary PCR included approximately 10 ng of this PCR template (1 µl of a 1 in 10,000 or 100,000 dilution), 1× reaction buffer, 2.5 mM $MgCl_2$, 250 µM of each dNTP, 10 µM of each primer, and 1 U taq polymerase in a total volume of 50 µl. The conditions using the GenAmp PCR System 2400 were 0.5 min. at 94°C, 0.5 min. at 56°C, 1 min. at 68°C with an initial incubation of 5 min. at 94°C and a final incubation of 7 min. at 72°C. The number of cycles for amplification was 35.

iv. Core PCR

The primary DNA fragment was amplified using *taq* polymerase or elongase enzyme mix and then electrophoresed on 0.5 to 0.7% agarose gel. The desired band was excised and transferred to a 1.5 ml microfuge tube. The sample was then heated at 50°C for 10 min. to melt the gel and 5µl of this sample was used as template in the PCR. The PCR also included 1× PCR reaction buffer (Gibco), 2.5 mM $MgCl_2$, 250 µM of each dNTP, 10 µM of each primer, and 1 U *taq* polymerase (Gibco) in a total volume of 50µl. The conditions using the GenAmp PCR System 2400 were 0.5 min. at 94°C, 0.5 min. at

56°C, 1 min. at 68°C with an initial incubation of 5 min. at 94°C and a final incubation of 7 min. at 72°C. The number of cycles for amplification was 35.

v. Touch down PCR

PCR components varied depending on whether *taq* polymerase or elongase enzyme mix was used. The conditions in the GenAmp PCR System 2400 were as follows, 5 min. initial incubation at 94°C, followed by 30 cycles of 0.5 min. at 94°C, 0.5 min. at 65°C with a 0.5°C decrease in temperature every cycle, 1 to 5 min. at 72°C (1 min. for approximately 1 kb and an additional minute for each 1 kb extra). This was followed by 15 cycles of 30s at 94°C, 30s at 50°C and 1 to 5 min. at 72°C. A final incubation of 7 min. at 72°C was implemented.

3.3.5. Gel electrophoresis of DNA

Products from PCR were electrophoresed on 0.5-1.5% agarose containing 0.5 µg/ml EtBr and visualized using UV light. 0.5% agarose was used to separate DNA fragments greater than 3 kb in size, 0.7% agarose for DNAs 1-3 kb, and 1.5% for DNAs less than 1kb. Agarose gels were prepared according to Sambrook *et al.*(1989). In brief, agarose was melted in 1× TAE buffer and 0.5 µg/ml ethidium bromide was added. The solution was poured into a mold and allowed to solidify. The gel was placed in a horizontal electrophoresis apparatus (Fisher Biotech) containing 1× TAE buffer. 10 µl of each of the PCR products was combined with 2 µl gel loading buffer and loaded into the wells. A 1 kb ladder (Gibco) was used as a size reference with DNA fragments greater than 1 k b and a 0.1 kb ladder was used with DNA fragments less than 1 kb in size. 0.5 µl

of loading buffer was added to 2 µl of ladder prior to be loaded into the gel. The DNA was electrophoresed at approximately 80 volts for 45 minutes. The DNA was visualized with UV light using EagleEye (Stratagene) and photographed using the EagleEye digital camera.

3.3.6. DNA purification of PCR products

i. Direct purification

Purification of clean PCR samples was accomplished using Wizard® Plus Mini-prep purification System. (Promega, Madison, WI) as outlined in the Promega literature. In brief, 100 µl of direct purification buffer (Promega) was added to the PCR product and mixed by inversion, 1 ml of purification resin (Promega) was then added and also mixed by inversion. This solution was pushed through a minicolumn using a 3 ml syringe. Two ml of 80% isopropanol was then pushed through the minicolumn (Promega) using a 3 ml syringe. The minicolumn was placed in 1.5 ml microfuge tube and spun for 20 sec. at 12,000×g to elute the remaining solution. The minicolumn was placed into a clean 1.5 ml microfuge tube and the DNA was eluted by adding 50 µl ddH_20 to the minicolumn, incubating it for 1 minute and centrifuging it at 12,000×g for 20 seconds.

ii. Gel purification

PCR samples with additional bands were gel purified using Wizard® Plus Mini-prep purification system (Promega, Madison, WI) for DNA fragments less than 3 kb in size or Qiaex II Gel Extraction kit (Qiagen) for DNA fragments greater than 3 kb. Gel purifications of PCR products less than 3 kb were performed using the Wizard® Plus Mini-prep purification system according to the procedure outlined in their literature. In

brief, the DNA samples were electrophoresed on 1.5% low melting point agarose gel and the desired bands were excised. Low melting point agarose gels were prepared using the same procedure outlined previously for agarose gels in section 3.3.5. The excised band was placed in a 1.5 ml microfuge tube and melted in a 70°C waterbath. The remainder of the procedure is identical to that for direct purification described earlier in section 3.3.6(i) except the direct purification buffer is omitted.

PCR samples greater than 3 kb were gel purified using Qiaex II Gel Extraction kit according to the protocol outlined in the Qiagen product literature. In brief, the DNA samples were electrophoresed on 1.0% agarose gel as described in section 3.3.5 and the desired bands were excised. The excised band was placed in a 1.5 ml microfuge tube, with 3 volumes of Buffer QX1(Qiagen) and 30µl of Qiaex II (Qiagen). These volumes of these two solutions varied depending on the size of the DNA fragment and the concentration of DNA. The tube was incubated in a 50°C waterbath for 10 min. to solubilize the agarose, centrifuged for 0.5 min., and the supernatant was discarded. The pellet was washed once with 500 µl QX1 Buffer (Qiagen), twice with PE Buffer (Qiagen), and air dried for 15-30 min.. The DNA was eluted into the supernatant by adding 20 µl TE buffer, incubating for 5 min. at room temperature, and centrifuging for 0.5 min. to pellet the debris.

3.3.7. Determination of DNA concentration

DNA concentrations of PCR products were determined by spectrophotometry or by comparison to a DNA ladder reference following gel electrophoresis. The optical density at 260 nm of 100 µl of a $1/50^{th}$ dilution was determined using a Spectronic

Genesys 5 Spectrophotometer (Milton Roy). This value was multiplied by 2500 to factor in the dilution and convert the value to ng/µl. Alternatively, the DNA concentration was determined by visual comparison of the PCR product bands with DNAs of known concentrations. Varying dilutions (1, $1/10^{th}$, $1/100^{th}$) of the PCR sample were subjected to electrophoresis beside DNAs of known concentrations (0.1 kb ladder) on an agarose gel. Following visualization, the intensity of the PCR bands were compared to bands in the DNA reference ladder. The concentration of the PCR product was estimated by calculating the concentration of the specific band within the 100 bp ladder that was closest in range to the PCR product.

3.3.8. Cloning of PCR fragments

The pGEM®-T -vector System (Promega) was used to clone the PCR fragments for sequence analysis. The protocol was obtained from the product literature provided by Promega. In brief, desired fragments from PCR were purified and ligated with pGEM®-T vector. The ligation reaction included 50 ng of vector, 50 ng PCR product, 3 units T4 DNA ligase (Promega), and 1× rapid ligation buffer (Promega) in a total volume of 10 µl. The ligation reaction was incubated at 4°C for 16 hours. The ligation reaction was transformed into Top10F' *E. coli* cells as outlined in section 2.3.3 (iii), except the transformed cells were plated onto LB containing 100 µl/ml ampicillin, 0.5 mM IPTG, and 80 µg/ml X-gal.

Following a 24 hour incubation at 37°C the white colonies were selected for further analysis. This included isolating the DNA of the transformed bacterial colony and analyzing it by gel electrophoresis as outline in section 2.3.4(iv). Once the presence of an

insert in the plasmid was confirmed by its size, the plasmid was purified from a small scale growth of the culture according to the protocol described in section 2.3.4(v). As well the plasmid was linearized with 5 units of *Bsa*I and gel electrophoresed to confirm the size of the plasmid+insert. These procedures are also outlined in section 2.3.4(v).

3.3.9. DNA sequencing of primate DNA

DNA sequencing was kindly performed by the laboratory of Dr. Ben Koop. Both PCR products and cloned PCR fragments were sequenced using fluorescently labeled dye terminator chemistries (Amersham and ABI) with specific primers (M13R, M13F, F31, G15, S10, E85). Sequencing reactions were run on an ABI 377 automated DNA sequencer. PCR products were directly sequenced primarily with forward primer E85 and reverse primer S10, although primers within this region (G15, F31) were also used occasionally (primer sequences are given in Table 3.1 and primer locations relative to human *GBA* are given in Figure 3.4). PCR fragments cloned into pGEM®-T -vector were sequenced with pUC/ M13 forward primer (GTTTTCCCAGTCACGAC) and pUC/ M13 reverse primer (CAGGAAACAGCTATGAC).

3.3.10. DNA sequencing strategy and sequence data analysis

The contiguous stretch of *GBA* and ψ*GBA* that was sequenced stretched from primer location G15 to F31(Figure 3.4). The forward primer was located at 5332 bp and the reverse primer was at 6488 bp in human *GBA*, for a total size of 1116 bp between the primers in human *GBA* and ranging from 1059-1149 bp in the other primates analyzed. The forward and reverse sequences for a single PCR product or clone usually had a region of overlap which could be used to align the sequences. This enabled the 1059-

1149 bp stretch of DNA to be sequenced in its entirety.

DNAstar programs, including Seqman, and Megalign were used to analyse and manipulate the sequence data. Using Seqman, the forward and reverse sequences for *GBA* and ψ*GBA* genes were aligned together in each primate species, for a total of 5 alignment files. Once all the sequences were aligned, the sequence data for *GBA* and ψ*GBA* was separated into individual files. This enabled the creation of a single consensus sequence for each *GBA* and ψ*GBA* gene. Each *GBA* and ψ*GBA* consensus sequence was compiled of at least 6 sequences so that the entire 1059-1149 bp region had been sequenced at least three times. The sequence data for each primate *GBA* and ψ*GBA* gene was obtained from at least three different PCR amplifications that had been sequenced directly from PCR products and from PCR fragments cloned into plasmid. In Seqman the sequence trace data was viewed to ensure the quality of the sequences was acceptable. The consensus sequence for each *GBA* and ψ*GBA* gene was compiled in Megalign and patterns of nucleotide substitutions, deletions, and insertions, as well as the percent differences between the primate *GBA* and ψ*GBA* were analyzed. The primate *GBA* and ψ*GBA* consensus sequences were aligned with human *GBA* and ψ*GBA* using Boston College of Medicine alignment.

Maximum parsimony analysis of the DNA sequences was done by PAUP 4.0b4a (Swafford, 1993) using a heuristic search. Gaps were considered as single events. This enabled the creation of an evolutionary tree for the GBA genes. Bootstrap values were obtained by 1000 replications with 10 shufflings per replication.

3.3.11. Southern blot analysis of genomic DNA or PCR fragments

i. Creation of DNA probe

PCR fragments corresponding to region of the GBA gene were amplified using primer combinations E105 and PRE11, as well as G5 and E75 (primer sequence given in Figure 3.1 and primer location relative to human *GBA* given in Figure 3.4). These PCR products were gel purified using Wizard® Plus Mini-prep purification system (Promega, Madison, WI) according to the protocols previously described in section 3.3.7. The probe was labeled with ECF (Amersham) for chemiluminescent detection as previously described in section 2.3.7. (vi).

ii. Digestion of primate genomic DNA

Primate genomic DNA was digested with a variety of combinations of restriction enzymes: i) *Eco*RI, ii) *Bam*HI, iii) *Bgl*II, iv) *Bam*HI and *Bgl*II, v) *Eco*RI and *Bgl*II, vi) *Eco*RI and *Bam*HI, vii) *Eco*RI, *Bgl*II and *Bam*HI, viii) *Pst*I, ix) *Pvu*II and x) *Pvu*II and *Pst*I. Each restriction digest contained 10 µg of genomic DNA, 50 U of each restriction enzyme in 1× buffer (New England Biolabs, Beverly, MA) for a total volume of 50 µl. The restriction digest was allowed to proceed at 37°C for 16 hours. The success of the restriction digest was verified by electrophoresing 2 µl of the sample on a 0.5% agarose gel as outlined in section 3.3.5.

iii. Electrophoresis of DNA

A 10.5cm × 20 cm 1% agarose gel was prepared with TBE buffer and without ethidium bromide as described in section 3.3.6. Fifty µl of DNA (PCR products or

digested primate genomic DNA) was combined with 10 µl of loading dye and loaded into a well. 5 µl of 1 kb ladder (Gibco) with 2 µl loading dye was run as a size reference. The gel was run for 16 hours at 40 volts. The gel was incubated in 500 ml TBE buffer containing 0.5 µg/ml ethidium bromide for 20 minutes. The gel was visualized and photographed using Eagleeye (Stratagene) with a ruler adjacent to it.

iv. Denaturation of electrophoresed DNAs

The gel was dried for 2 hours at 80°C. Then it was incubated at room temperature in denaturation solution (5 M NaOH, 1 M NaCl) for 30 min. and in neutralization solution (1 M Tris-HCl Ph 7.2, 1 M NaCl) for 30 minutes.

v. Transfer of electrophoresed DNA to nitrocellulose membrane

The DNA ladder was removed from the gel and the electrophoresed DNAs were transferred to nitrocellulose Hybond™-N using capillary transfer. The procedure used corresponds to that in Sambrook *et al.* (1989). Following the transfer, the gel was stained with 0.5 µg/µl ethidium bromide in TBE buffer and visualized using UV to ensure that none of the DNA remained.

vi. Membrane hybridization and visualization

The membrane was hybridized with the labeled probe according to the protocol provided in the Amersham product literature. In brief, the membrane was incubated in 0.125 ml/cm^2 AlkPhos direct hybridization buffer (Amersham) at 55°C for 15 minutes. The labeled DNA probe was added at a concentration of 10 ng/ml buffer and incubated at

55°C for 16 hours. The membrane was transferred to 5 ml/cm^2 primary wash buffer (Amersham) at 55°C for 10 min., and then washed twice in secondary wash buffer (Amersham) for 5 min. at room temperature. The membrane was air dried and baked at 80°C for 2 hours.

The chemifluorescent signal generation and detection was done using the ECF substrate according to the protocol provided in the Amersham product literature. In brief, the membrane was incubated with 25 µl/cm^2 ECF substrate in a sealed detection bag. The membrane was imaged at 1 hour and 24 hours afterwards using a Storm PhosphorImager® (Molecular Dynamics, Sunnyvale, CA) with the excitation at 430 nm and emission at 560 nm.

3.4. Results

3.4.1. PCR amplification and DNA sequencing of primate *GBA* and ψ*GBA*

To amplify primate *GBA* and ψ*GBA*, I constructed five degenerate primers corresponding to highly conserved areas of *GBA* (Table 3.1 and Figure 3.4). The areas of high conservation were determined by comparing the murine *GBA* cDNA sequence to that of human *GBA* and ψ*GBA*.[54;71] Additionally, primers designed for human *GBA* were used (Table 3.1 and Figure 3.4). In total, the primers used spanned the entire *GBA* gene from exons 1-11 and corresponded to both intronic and exonic regions. The use of such a variety of primers was necessitated by the variable success achieved by the primer combinations in different primate species.

The primer combinations and PCR conditions were optimized so that both *GBA* and ψ*GBA* could be amplified in a single reaction. Primers G3 and S10 were used to

amplify exons 3-10 in *GBA* and ψ*GBA* (Figure 3.5). In orangutan, gorilla, and chimpanzee this produced two PCR products sized approximately 3.5 kb and 4.8 kb. These PCR products correspond to the expected size of those from human *GBA* and ψ*GBA* (Figure 3.4). In baboon a single fragment size approximately 3.5 kb was amplified with these primers, while in squirrel monkey a single band of about 4.8 kb was amplified (Figure 3.6). The PCR products are not due to contamination as the negative control, which was identical to the other reaction tubes except lacking primate genomic DNA, did not include any PCR products (data not shown). Furthermore, Southern blot analysis using a *GBA* specific probe confirmed that these PCR products were GBA (data not shown).

In chimpanzee, gorilla, and orangutan *GBA* and ψ*GBA* were isolated from each other and used as template in a nested PCR. The genes were separated using gel electrophoresis and then the purified product was used as template in a second PCR or the DNA fragment was excised from the gel and used directly in a "core" PCR. In the nested PCR, primers G15 and F31 were used to amplify a 1078-1168 bp portion of *GBA* and ψ*GBA* from gorilla, orangutan, and chimpanzee (Figure 3.7). The primers G15 and F31 flank the region of *GBA* and ψ*GBA* that is analyzed by sequencing. This PCR fragment was cloned into pGEM T-vector and sequenced using forward and reverse pUC/M13 primers. Alternatively, primers E85 and S10, which flank the region amplified by G15 and F31, were used in the nested PCR and the PCR products were directly sequenced using E85 and S10 primers. When PCR products are directly sequenced it is necessary to use primers that flank the region for sequencing by at least 30 bp as the first part of the sequence is usually of poor quality. When cloned PCR fragments are

sequenced, however, the sequencing primers (forward and reverse pUC/M13) correspond to the vector and the first portion of the sequence is derived from the vector.

In baboon and squirrel monkey, PCR amplification of GBA and ψGBA from genomic DNA only produced one size of fragment. This fragment was approximately 3.5 kb in baboon and about 4.8 kb in squirrel monkey. Nested PCR using G15 and F31 amplified a 1129 bp fragment in baboon and a 1111 bp fragment in squirrel monkey (Figure 3.7), which were cloned and sequenced. Additionally primers E85 and S10 were used in the nested PCR to amplify PCR products for direct sequencing.

PCR products were sequenced using the same primers with which they were amplified or the DNAs were cloned into a plasmid and sequenced with the pUC/M13 primers corresponding to the plasmid. Direct sequencing of PCR fragments worked well in gorilla, chimpanzee, and orangutan where *GBA* and ψ*GBA* were isolated and used as template in a nested PCR, as well as in squirrel monkey were only one *GBA* sequence was obtained. For the baboon, however, direct sequencing resulted in simultaneously sequencing of two different *GBA* sequences and thus cloning was necessary. To maintain consistency, the *GBA* genes in all primates were sequenced directly and from plasmids. Each *GBA* and ψ*GBA* gene was sequenced from PCR products obtained in least three different amplifications. These sequences were compiled in Seqman to create a consensus sequence for *GBA* and ψ*GBA* from each primate. The consensus sequence was compiled from at least 3 forward and 3 reverse sequences so that the entire 1059-1149 bp region had been sequenced at least three times. In total, 73 sequences were obtained for *GBA* and ψ*GBA* in the five primates studied.

3.4.2. Differentiation of primate *GBA* and ψ*GBA* based on size

PCR amplification from exon 3 to intron 10 produced two fragments approximately sized 3.6 kb and 4.8 kb in human, orangutan, gorilla, and chimpanzee (Figure 3.6). The same PCR amplification, however, produced a single fragment of approximately 3.6kb in baboon and about 4.8kb in squirrel monkey (Figure 3.6). This discrepancy in size between *GBA* and ψ*GBA* in gorilla, chimpanzee, and orangutan is also seen in the PCR products from other amplifications. For example amplification from exons 4-11 (using primers PrFOR and PrREV) in human, orangutan, gorilla, and chimpanzee produces two bands sized ~3.2 kb and ~4.5 kb (data not shown).

3.4.3. Determination of gene copy number through sequence analysis and Southern blot analysis

DNA sequencing of ~1.1 kb (1059-1149 bp) in *GBA* and ψ*GBA* produced two distinct sequences in chimpanzee, gorilla, orangutan, and baboon (Figure 3.8). Sequencing revealed that the *GBA* sequences in these primates possessed differing nucleotides and in some instances deletions or insertions existed (Figure 3.8). Whereas only one *GBA* sequence was detected in squirrel monkey when this region of the gene was sequenced (Figure 3.8). Southern blot analysis was used to verify the number of glucocerebrosidase genes in the analyzed primate species (data not shown). However, the signal obtained with Southern blot analysis was never sufficiently above background to accurately determine *GBA* copy number. The Southern blot analysis was performed over ten times to try to optimize the procedure. Conditions which were altered included concentration of genomic DNA, restriction enzymes used to digest the genomic DNA,

concentration of probe, size of probe, area to which probe annealed, hybridization temperature, method of adhering DNA to the membrane, and length of time allowed for signal to develop.

3.4.4. Sequence analysis of ~1.1 kb in primate *GBA* and ψ*GBA*

Nine sequences from *GBA* and ψ*GBA* in gorilla, chimpanzee, orangutan, baboon, and squirrel monkey were analyzed and compared to human *GBA* and ψ*GBA* using DNAstar, PAUP 4.0, and BCM. All the investigated primates, except for squirrel monkey, had two sequences for *GBA*. The analyzed DNA sequence stretched from exon 8 to near the end of exon 10 and its size ranged from 1059-1149 bp, depending on whether the gene was *GBA* or ψ*GBA* and which primate species it came from (Figure 3.8).

The variation in size of the sequenced portion of *GBA* and ψ*GBA* is due to several deletions and insertions (Figure 3.8). Gorilla and chimpanzee ψ*GBA* has a 55bp deletion in exon 9 that is also present in human ψ*GBA*. There are two other exonic deletions: a 2 nucleotide deletion in exon 8 of both baboon *GBA* genes; and a single nucleotide deletion in exon 8 of gorilla ψ*GBA*. The number of deletions and insertions increases dramatically in the intronic regions. In the five primate species analyzed, there are a total of nineteen intronic insertions and deletions, twelve in intron 8 and seven in intron 9, ranging in size from a single nucleotide to 33 bp.

The similarity between *GBA* and ψ*GBA* in the different primate species varies considerably. The chimpanzee and gorilla *GBA* and ψ*GBA* are least similar at 97.8% and 97.9%, while baboon and orangutan *GBA*1 and *GBA*2 are most homologous at 99.4% and

99.5% similarity (Table 3.3). Furthermore, the homology between GBA genes in different species varies even more. Chimpanzee *GBA* is the most similar to human *GBA* (99.5%), while squirrel monkey *GBA* is least similar, sharing only 91.7% similarity (Table 3.2). The variations in nucleotides are not evenly distributed throughout all regions of *GBA* and ψ*GBA*. In the exonic regions 12.2% of sites are variable, while 19.7% are variable in the intronic regions. Furthermore the variation is lower is exons 9 and 10 at approximately 10.5 %, while in exon 8 it is 15.7% (Figure 3.9).

3.4.5. Phylogenetic analysis of primate GBA genes

A comparison of sequence data indicates that there are 173 variable characters of which 82 were informative for parsimony analysis. The phylogenetic tree for glucocerebrosidase in the five primate species analyzed was constructed using maximum parsimony analysis and the best tree was selected (Figure 3.10). The bootstrap values are consistently high, four are 100 and the lowest bootstrap value is 72. These bootstrap values indicate that the grouping in the parsimony tree are significant at $P \geq 0.05$.[52]

3.4.6. Deduction of 115 amino acids in primate GBA

A 115 amino acid sequence has been deduced for the C-terminal region of chimp *GBA*, gorilla *GBA*, orangutan *GBA*1 and *GBA*2, and squirrel monkey *GBA*, as well as a 104 a.a. sequence for baboon *GBA*1 and *GBA*2 (Figure 3.11). The chimpanzee and gorilla ψ*GBA* amino acid sequences are not shown because the 55bp deletion in exon 9 indicates they would be non-functional. Additionally, only the amino acid sequence following the frameshift mutation in exon 8 of baboon *GBA*1 and *GBA*2 is given (Figure 3.11). This

two nucleotide deletion in baboon *GBA*1 and *GBA*2 suggests that a reading frameshift would occur, but since the entire sequence of the genes is not known it is not possible to predict the extent of the effect the mutation would have. These sequences are compared to that for human *GBA* and the known human *GBA* mutations leading to Gaucher disease (Figure 3.11).[11]

The squirrel monkey *GBA* amino acid sequence is identical to human, while baboon *GBA*1, which differs from human *GBA* at six sites, is the most different from human (Figure 3.11). Furthermore baboon *GBA*1 has a stop codon introduced in exon 8. None of the amino acid changes in the primate sequences correspond to known mutations in human *GBA* (Figure 3.11).

Table 3.1. Primers used to amplify the glucocerebrosidase gene from genomic primate DNA. The location of these primers relative to the intronic/exonic regions of GBA are shown. The degenerate primers marked by an asterisk (*) were constructed by comparing the mouse and human *GBA* genes and determining the most conserved regions. The letters represent different nucleotides as follows: A=adenine, C=cytosine, G=guanine, T=thymidine, R=A+G, Y=C+T, I=deoxyinosine, H=A+T+C, S=C+G, N=A+T+C+G.

Forward primer	**Location**	**Reverse Primer**	**Location**
G0 CGG AAT TAC TTG CAG GGC TA	Exon 1	G9 TGC ATA GGT GTA GGT GCG GA	Exon 5
E25 AGG AGA GTA GTT GAG GGG TG	Intron 1	F31 GTT TAG CAC GAC CAC AAC AGC	Exon 10
G3 TTT CCT GCC CTT GGT ACC TT	Exon 3	S10 TTC AGC CCA CTT CCC AGA CC	Intron 10
G4 TGC TGC TCT CAA CAT CCT T	Exon 4	PrE11 GTG TGA ATG GAG TAG CCA GGT GA	Exon 11
PRE4 AGA AGT TCC AGA AAG TGA ARG G	Exon 4	PRE11B* RAT AGT YTC CAG RAA ICC CAC	Exon 11
PrFOR* AAR TTY CAG AAR GTG AAR GGI TTY GGI GGI GCC ATG	Exon 4	PrREV* TCA CTG ICG ICG CCA CAG GTA GGT GTG RAT	Exon 11
G13 TCC GCA CCT ACA CCT ATG CA	Exon 5		
E65 GGA TGC ACT GGT TGG GCT AG	Intron 5		
PrE6* GAY ATH TAY CAY CAR ACI TGG	Exon 6		
E75 AGC TGG TCT GGT CCA CTT TC	Intron 6		
PRE7* CAY AAY GTS CGS CTN CTI ATG	Exon 7		
G5 TAC AGT TCT GGG CAG TGA CA	Exon 7		
E85 TGT GCA AGG TCC AGG ATC AG	Intron 7		
G15 CAC CAT GCT CTT TGC CTC AG	Exon 8		

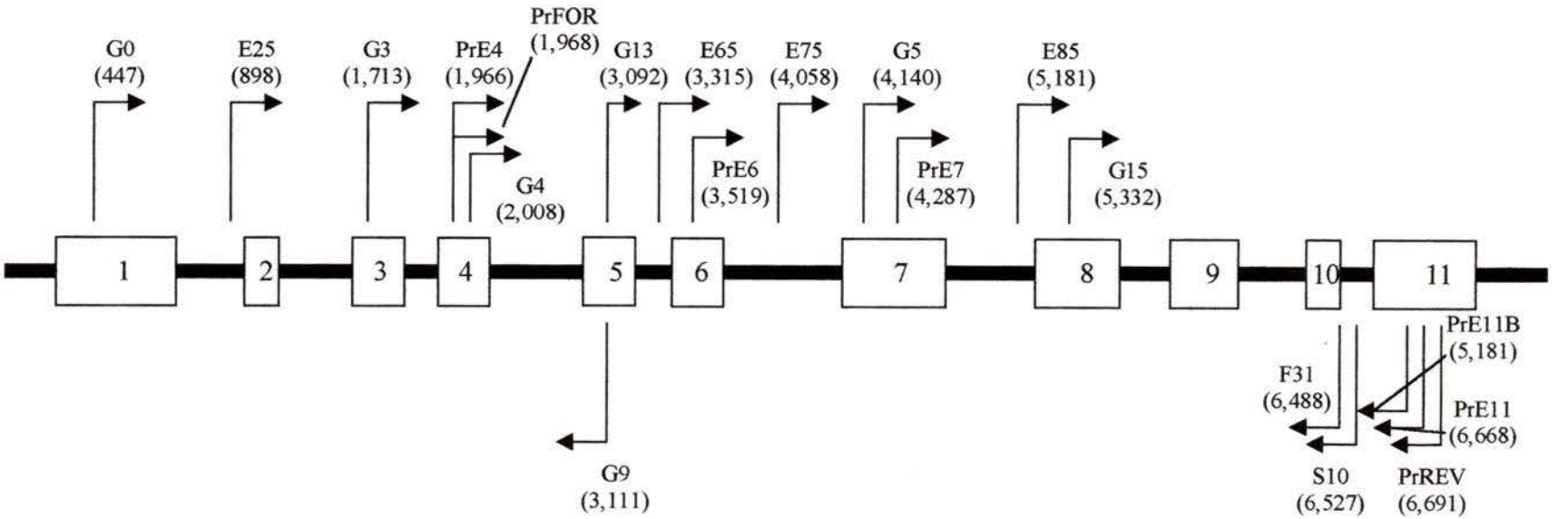


Figure 3.4. Nucleotide position relative to human GBA and direction of primers used for primate GBA and ψGBA amplification and sequencing. Arrows represent sense of amplification and numbers in parentheses indicate nucleotide position on the full genomic sequence of human GBA.

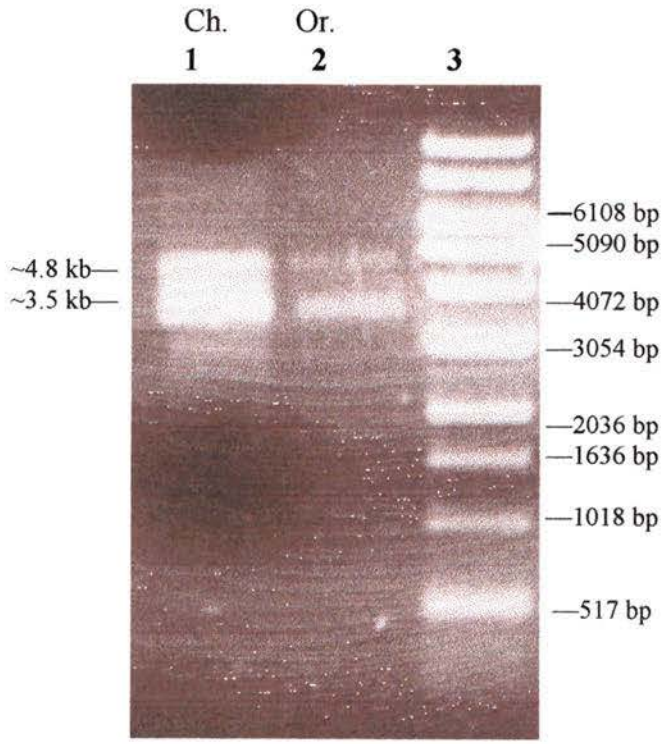


Figure 3.5. PCR of exon 3 to intron 11 of glucocerebrosidase using primers G3 and S10. The PCR products in lane 1 are from chimpanzee and those in lane 2 are from orangutan. Lane 3 contains a 1 kb ladder with the fragment sizes indicated on the right. The DNA fragments in lanes 1 and 2 are sized approximately 3.5 and 4.8 kb.

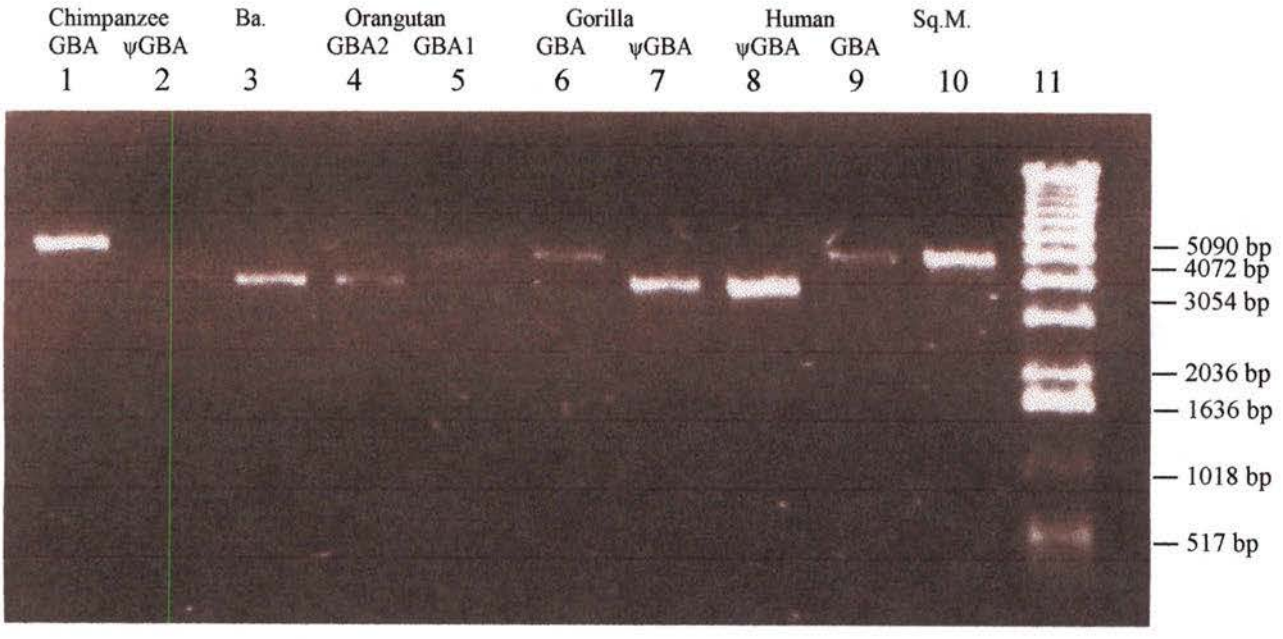


Figure 3.6. Size comparison of primate *GBA* and ψ*GBA* PCR products amplified using primers G3 and S10. Lane 1)chimpanzee *GBA*; lane 2) chimpanzee ψ*GBA*; lane 3) baboon *GBA*; lane 4) orangutan *GBA*2; lane 5) orangutan *GBA*1; lane 6) gorilla *GBA*; lane 7) gorilla ψ*GBA*; lanes 8) and 9) are human ψ*GBA* and *GBA* controls, lane 10) squirrel monkey *GBA*. Lane 11 contains a 1 kb ladder with the size of the fragments indicated on the right of the gel.

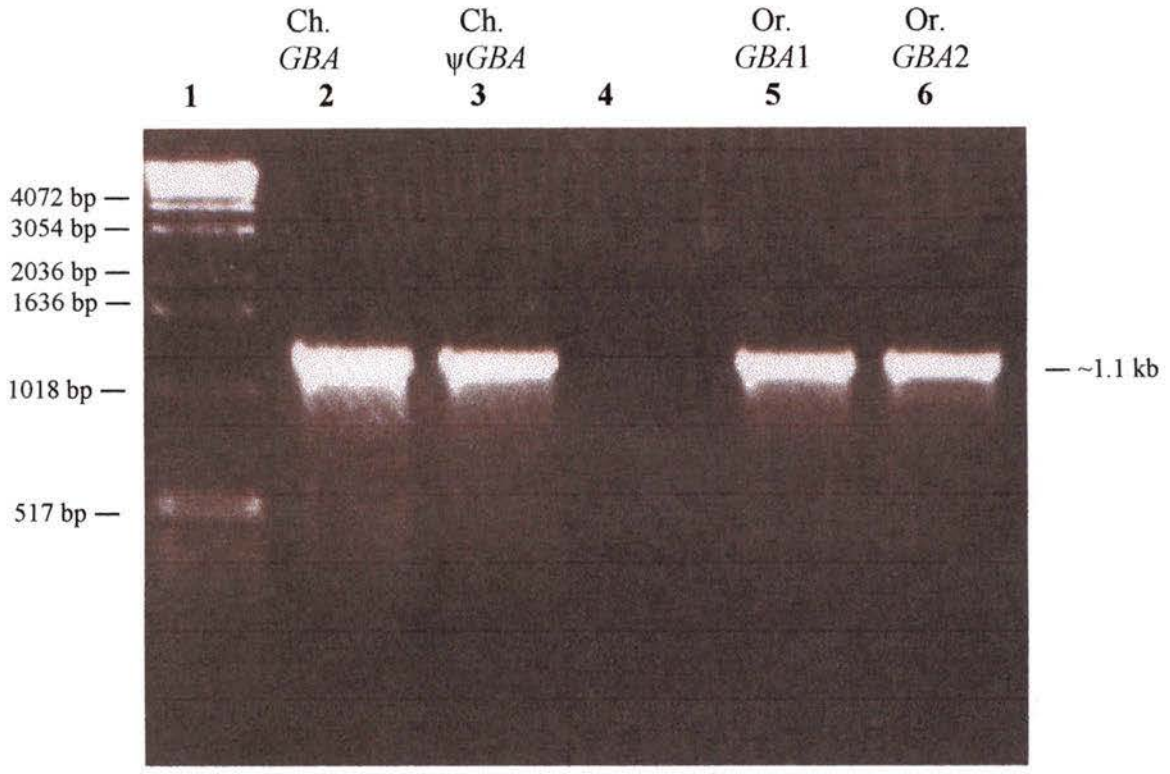


Figure 3.7. Agarose gel of PCR products from nested PCR using primers F31 and G15. The DNA template for the reaction was amplified using primers G3 and S10, and *GBA* and ψ*GBA* were isolated and purified. Lane 1 contains a 1 kb ladder with the fragment sizes indicated on the left and lane 4 is the negative control. Lane 2 contains chimpanzee GBA, lane 3 chimpanzee ψ*GBA*, lane 5 orangutan *GBA*1, and lane 6 orangutan *GBA*2. The DNA fragments in lanes 2, 3, 5, and 6 are sized approximately 1.1 kb.

```
Human GBA              1 AGGCCTGTGTGGGCTCCAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGGATCGAG
Human ψGBA             1 ............T.............................................
Chimp GBA              1 .............................................................
Chimp ψGBA             1 ............T.............................................
Gorilla GBA            1 .............................................................
Gorilla ψGBA           1 ....-.........................................................
Orangutan GBA1         1 ..............................................G..T...........
Orangutan GBA2         1 ..............................A.......A.....................
Baboon GBA1            1 ........C...................A..--.........G.........A.......
Baboon GBA2            1 ........C...................A..--.........G.....................
Squirrel Monkey        1 ..............................................T..............

Human GBA             61 GGATGCAGTACAGCCACAGCATCATCACGGTAAGCCACCCCAGTCTCCCTTCCTGCAAAG
Human ψGBA            61 ..............................A..............................
Chimp GBA             61 ..............................A..............................
Chimp ψGBA            61 ............C.............A..............................
Gorilla GBA           61 ..............................A..............................
Gorilla ψGBA          60 ..............................A..............................
Orangutan GBA1        61 .............................................................
Orangutan GBA2        61 .............................................................
Baboon GBA1           59 ..................A.............................G.............
Baboon GBA2           59 ..................A.............................G.............
Squirrel Monkey       61 ..............................................C.G.........G..

Human GBA            121 CAGACCTCAGACCTCTTACTAGTTTCACCAAAG--------------------------
Human ψGBA           121 G............CA...G....C.........--------------------------
Chimp GBA            121 ..............................--------------------------
Chimp ψGBA           121 .....A.......CA...G....C.........--------------------------
Gorilla GBA          121 ..............................CAGACCTCAGACCTCTTACTAGTTTCA
Gorilla ψGBA         120 .............CA...G....C.........--------------------------
Orangutan GBA1       121 .............C.........C.........--------------------------
Orangutan GBA2       121 .............C.........C.........--------------------------
Baboon GBA1          119 ..C..........C....G....C.........--------------------------
Baboon GBA2          119 ..C.........TC....G....C.........--------------------------
Squirrel Monkey      121 .............C...C..G..C.........--------------------------

Human GBA            154 ------ACTGACAGAAGCCCTTCCTGTCCAGCTT-TCCCCAGCTAGCCTGCCCTTTTGAG
Human ψGBA           154 ------.....T.....................-.......G..............C.
Chimp GBA            154 ------.....................................-.....................
Chimp ψGBA           154 ------.....T.....................-.....................G.
Gorilla GBA          181 CCAAAG......................................-.....................G.
Gorilla ψGBA         153 ------.....A.....................-.....................-.G.
Orangutan GBA1       154 ------.....................-.....................G.
Orangutan GBA2       154 ------.....................-.....................G.
Baboon GBA1          152 ------.........--.................-.G........T...........G.
Baboon GBA2          152 ------.........--..T.............-.G........T..C.........G.
Squirrel Monkey      154 ------..G........A.......C....C...C..............T..C....G.

Human GBA            207 CAACTCTGGGGAACCATGATTCCCTATCTTCCCTTTCCTTCACAGGTCTGCACACCTCAT
Human ψGBA           207 ........................G....G.............................
Chimp GBA            207 .............................................................
Chimp ψGBA           207 ........................G....G.............................
Gorilla GBA          240 .....................................................A.......
Gorilla ψGBA         205 ............G..........G....G.....................A.......
Orangutan GBA1       207 ........................G....G.............................T.
Orangutan GBA2       207 ........................G....G.............................T.
Baboon GBA1          203 .....--.................G....G.............................
Baboon GBA2          203 .....--.................G....G.......C.....................
Squirrel Monkey      208 ...T................TT..G....G.............................C
```

Figure 3.8. Variable nucleotide sites of aligned *GBA* and ψ*GBA* sequences (Exon 8 – Exon 10) in five primate species. Primate sequences are compared to the known human GBA sequences. Numbers indicate the nucleotide positions, dashes (-) indicate alignment gaps, nucleotide substitutions compared to the human functional gene are shown while identical nucleotides are indicated by a period (.). Exonic regions are indication in bold.

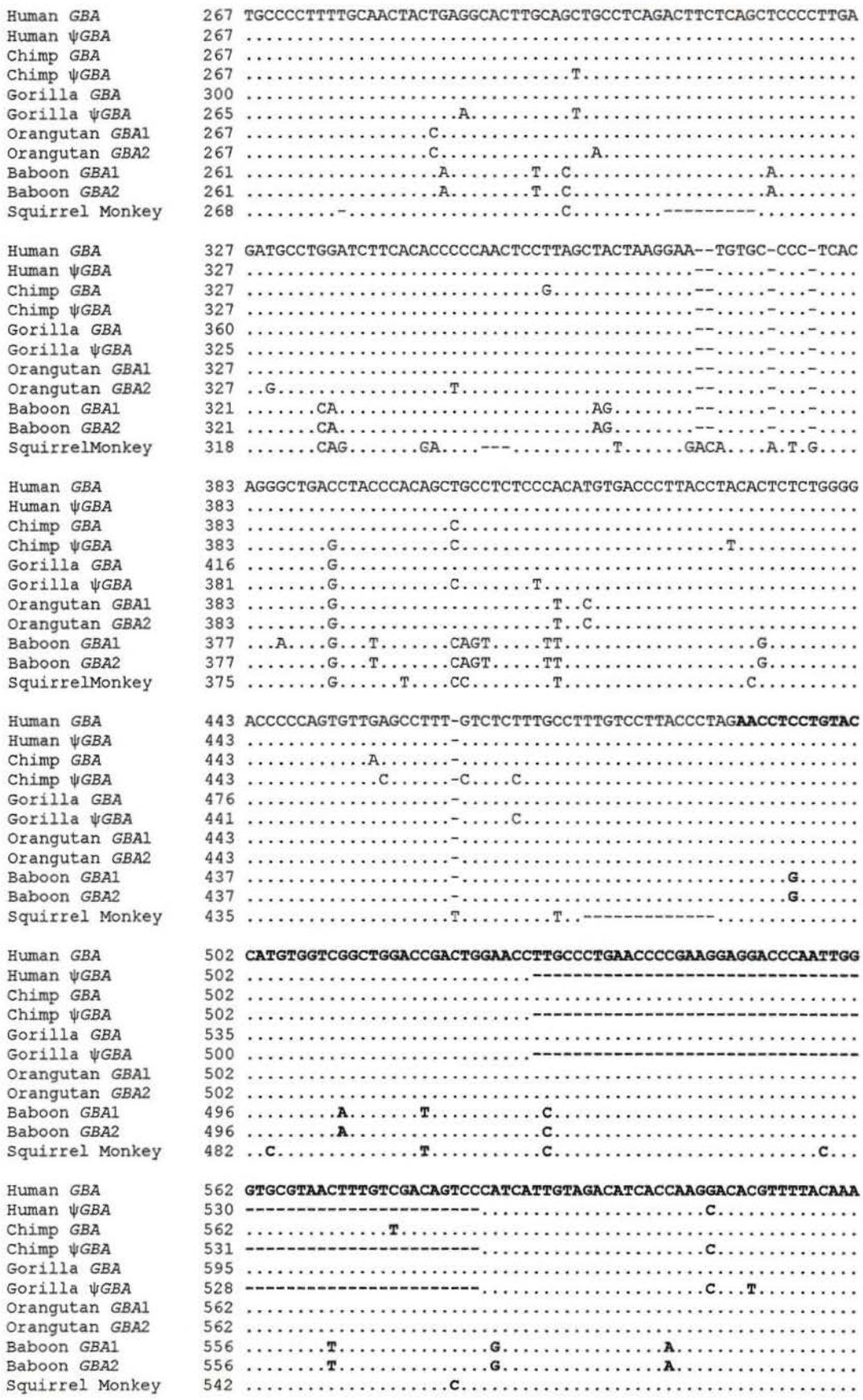


Figure 3.8. (continued)

```
Human GBA          622 CAGCCCATGTTCTACCACCTTGGCCACTTCAGGTGAGTGGAGGGCGGGCACCCCCATTCC
Human ψGBA         567 ..........................................................
Chimp GBA          622 ..........................................................
Chimp ψGBA         567 ..........................................................
Gorilla GBA        655 ..........................................................G....
Gorilla ψGBA       565 ..........................................................G....
Orangutan GBA1     622 ..........................................................
Orangutan GBA2     622 ..........................................................
Baboon GBA1        616 ......................T...........................T.....G.
Baboon GBA2        616 ......................T.....................A.....T.....G.
Squirrel Monkey    602 .....T..................................................TT......

Human GBA          682 ATACCAGGCCTATCATCTCCTACATCGGATGGCTTACATCACTCTACACCACGAGGGAGC
Human ψGBA         627 ..........................................................
Chimp GBA          682 ......................T...................................
Chimp ψGBA         627 ..........................................................
Gorilla GBA        715 ..........................................................
Gorilla ψGBA       625 ..........................................................
Orangutan GBA1     682 ..........................................................T........
Orangutan GBA2     682 ..........................................................T........
Baboon GBA1        676 ...T.....................A.....................T........
Baboon GBA2        676 ...T.....................A.....................T........
SquirrelMonkey     662 ......................----------.............T.C......

Human GBA          742 AGGAAGGTGTTCAGGGTGG-AACCTCGGAAGAGGCACACCCATCCCCTTTTGCACCATGG
Human ψGBA         687 ..................-...........................................
Chimp GBA          742 ..................-...........................................
Chimp ψGBA         687 ..................-...G.......................................
Gorilla GBA        775 ..................-...........................................
Gorilla ψGBA       685 ..................-...........................................
Orangutan GBA1     742 ...........T.......-...........................................
Orangutan GBA2     742 ...........T.......-...........................................
Baboon GBA1        736 ...........T.......-.....T..G.....................T.......
Baboon GBA2        736 ...........T.......-.....T..G.....................T.......
Squirrel Monkey    708 ...........T.......G.....T..G......-.........T.C...........

Human GBA          801 AGGCAGGAAGTGACTAGGTAGCAACAGAAAACCCCAATGCCTGAGGCTGGA-CTGCGATG
Human ψGBA         746 ..........................................................-........
Chimp GBA          801 ..........................................................-........
Chimp ψGBA         746 ..........................................................-...T....
Gorilla GBA        834 ..........................................................-........
Gorilla ψGBA       744 ..........................................................-........
Orangutan GBA1     801 ........................G..C.......C.............A.-........
Orangutan GBA2     801 ........................G..C.......C.............A.-........
Baboon GBA1        795 ............................C.......G.............-........
Baboon GBA2        795 ............................C.......G.............-........
Squirrel Monkey    767 .............G.............C.....................GG........

Human GBA          860 CAGAAAAGCAGGGTCAGTGCCCAGCAGCATGGCTCCAGGCCTAGAGAGCCAGGGCAGAGC
Human ψGBA         805 ..........................................................
Chimp GBA          860 ..........................................................
Chimp ψGBA         805 ..........................................................
Gorilla GBA        893 .......A..................................G...................
Gorilla ψGBA       803 ..........................................................
Orangutan GBA1     860 ..........................................................C.....
Orangutan GBA2     860 ..........................................................C.....
Baboon GBA1        854 T.....................................T.......C.....
Baboon GBA2        854 T.....................................T.......C.....
Squirrel Monkey    827 .G....G..........C.G..............TG........C.A.T.........

Human GBA          920 CTCTGCAGGAGTTATGGGGTGGGTCCGTGGGTGGGTGACTTCTTAGATGAGGGTTTCATG
Human ψGBA         865 ..T..............-..............C.........................
Chimp GBA          920 ..........................................................
Chimp ψGBA         865 ..........................................................
Gorilla GBA        953 ........................T.AA..............................
Gorilla ψGBA       863 ..........................................................
Orangutan GBA1     920 ......................A.A.................................
Orangutan GBA2     920 ......................A.A.................................
Baboon GBA1        914 T.....................AT.C.....................TA.........
Baboon GBA2        914 T.....................AT.C.....................TA.........
Squirrel Monkey    887 T......C..C.........A...AG..............................C.
```

Figure 3.8. (continued)

```
Human GBA              980 GGAGGTACCCCGAGGG------ACTCTGACCA--TCTGTT-CCCACATTCAGCAAGTTCA
Human ψGBA             924 ..............------..........--......-...................
Chimp GBA              980 ..............------..........--......-...................
Chimp ψGBA             925 ..............------..........--......-...................
Gorilla GBA           1013 ..............------..........--......-...................
Gorilla ψGBA           923 ..............------..........--......-...................
Orangutan GBA1         980 ..............------T.........--...T..-...................
Orangutan GBA2         980 ..............------T.........--...T..-...................
Baboon GBA1            974 ..C...........------..........--...C..-T.................
Baboon GBA2            974 ..C...........------..........--...C..-T.................
Squirrel Monkey        947 ..............CCAGAGT.C.......TG......T...................

Human GBA             1031 TTCCTGAGGGCTCCCAGAGAGTGGGGCTGGTTGCCAGTCAGAAGAACGACCTGGACGCAG
Human ψGBA             975 ...................................................C........
Chimp GBA             1031 ....................................................................
Chimp ψGBA             976 ...................................................A...
Gorilla GBA           1064 ..........GG..........................................
Gorilla ψGBA           974 ..........GG..........................................
Orangutan GBA1        1031 ...................................................A...
Orangutan GBA2        1031 ...................................................A...
Baboon GBA1           1025 ..........T.....................C...................A...
Baboon GBA2           1025 ..........T.....................C...................A...
Squirrel Monkey       1007 ....................................G..................A...

Human GBA             1091 TGGCACTGATGCATCCCGATGGCTCT
Human ψGBA            1035 .........................
Chimp GBA             1091 .........................
Chimp ψGBA            1036 ..........G..............
Gorilla GBA           1124 .........................
Gorilla ψGBA          1034 .........................
Orangutan GBA1        1091 .........................
Orangutan GBA2        1091 .........................
Baboon GBA1           1085 .........................
Baboon GBA2           1085 .........................
Squirrel Monkey       1067 .A............T..C......
```

Figure 3.8. (continued)

Table 3.2. The percent nucleotide similarity between ~1.1 kb of C-terminal region of primate GBA genes (*GBA* and ψ*GBA*) and human *GBA*. Nucleotide substitutions are equally weighted and each gap is equal to a single nucleotide substitution. Human ψ*GBA* is also included for comparison.

Primate GBA gene	**Percent similarity to human *GBA* (%)**
Human ψ*GBA*	98.4
Chimpanzee *GBA*	99.5
Chimpanzee ψ*GBA*	97.8
Gorilla *GBA*	98.8
Gorilla ψ*GBA*	97.8
Orangutan *GBA*1	97.8
Orangutan *GBA*2	97.4
Baboon *GBA*1	93.7
Baboon *GBA*2	93.5
Squirrel Monkey *GBA*	91.7

Table 3.3. The percent nucleotide similarity of ~1.1 kb of the C-terminal region between the two GBA genes in each primate species. Percent similarity determined for chimpanzee *GBA* and ψ*GBA*, gorilla *GBA* and ψ*GBA*, orangutan *GBA*1 and *GBA*2, and baboon *GBA*1 and *GBA*2. Nucleotide substitutions are equally weighted and each gap is equal to a single nucleotide substitution. Percent similarity of human *GBA* and ψ*GBA* in this analyzed region is also included for comparison.

Primate (genes compared)	**Percent similarity between GBA genes (%)**
Human (*GBA* and ψ*GBA*)	98.4
Chimpanzee (*GBA* and ψ*GBA*)	97.8
Gorilla (*GBA* and ψ*GBA*)	97.9
Orangutan (*GBA*1 and *GBA*2)	99.5
Baboon (*GBA*1 and *GBA*2)	99.4

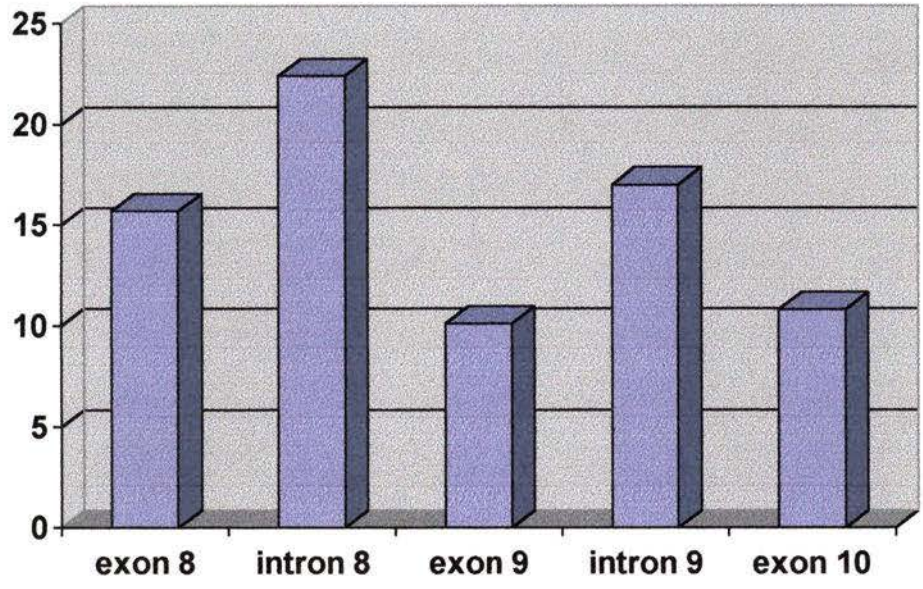


Figure 3.9. Percent nucleotide variation according to distribution in C-terminal region of *GBA* and ψ*GBA* in human, gorilla, orangutan, chimpanzee, baboon, and squirrel monkey. Percent nucleotide variation was calculated by determining the number of sites at which nucleotide variation occurred in each exon and intron and dividing that value by the number of nucleotides in human *GBA* for that region.

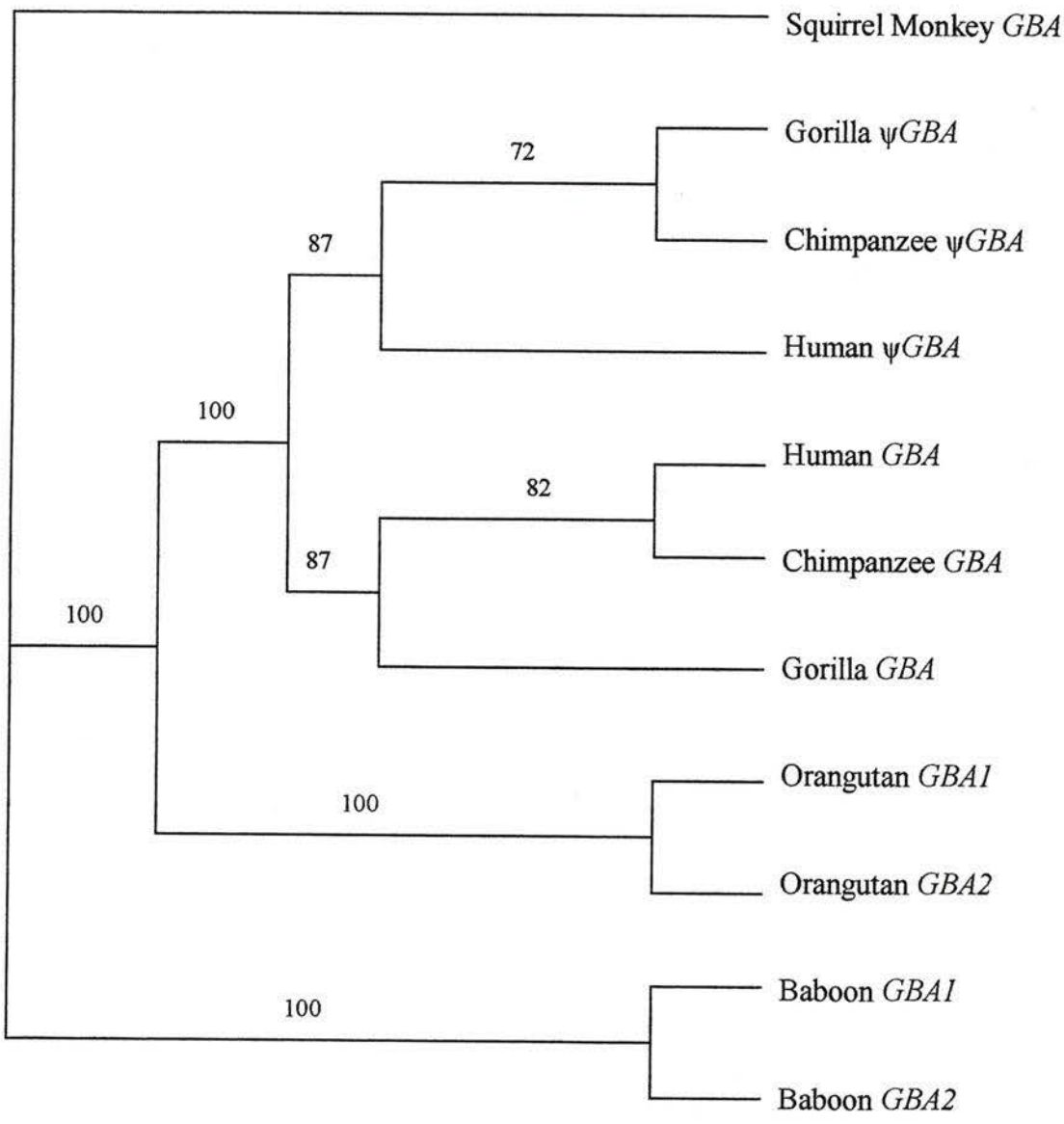


Figure 3.10. Parsimony tree obtained using PAUP 4.0 for 11 aligned *GBA* and ψ*GBA* sequences. Numbers aside nodes represent bootstrap values obtained with 1000 replicates using parsimony analysis.

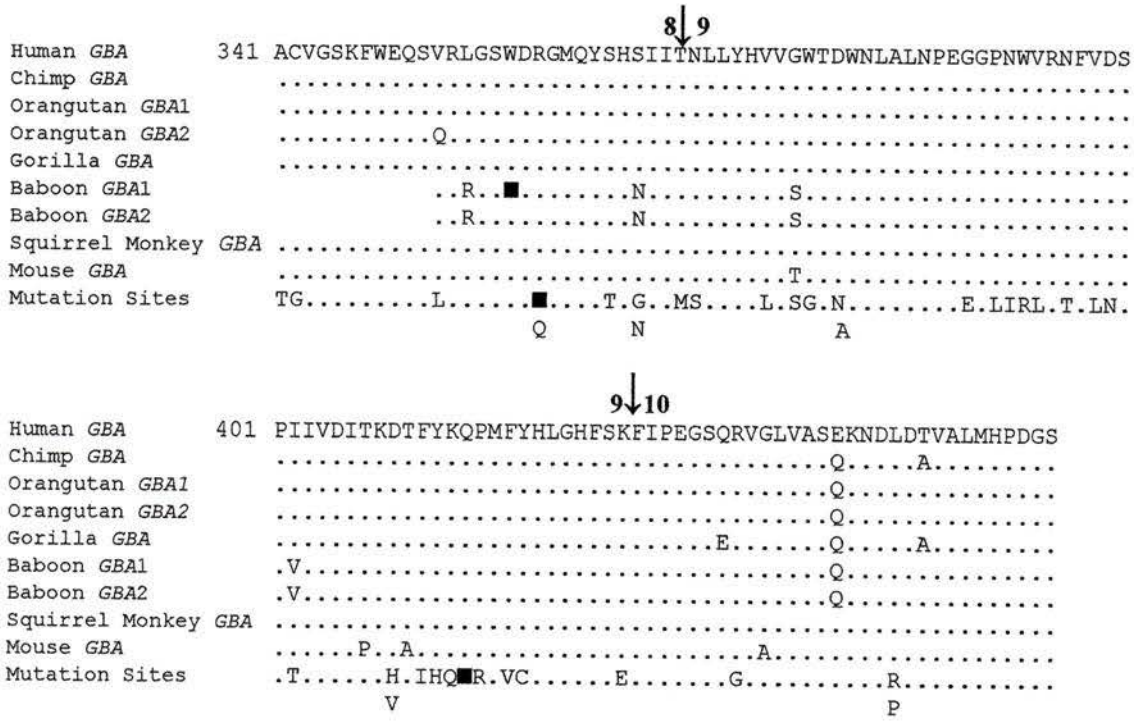


Figure 3.11. Variable amino acid sites of primate *GBA* and ψ*GBA* sequences compared to known mouse and human *GBA*. The human *GBA* sequence shown is from position 341 to 457 and approximate exon boundaries are marked. Mutations in human *GBA* leading to Gaucher disease are indicated and, where more than one mutation occurs at the same site, the additional mutation is indicated beneath the first (Beutler and Gelbart, 1998). Numbers indicate the amino acid position, dashes (-) indicate alignment gap or gaps, and squares (▪) indicate a stop codon. Amino acid substitutions compared to human GBA are shown while identical amino acids are indicated by a period (.). Amino acids are defined as follows: A= alanine, R=arginine, N= asparagine, D=aspartic acid, C=cysteine, E=glutamic acid, Q=glutamine, G=glycine, H=histidine, I=isoleucine, L=leucine, K=lysine, M=methionine, F=phenylalanine, P=proline, S=serine, T=threonine, W=tryptophan, Y=tyrosine, V=valine.

3.5. Discussion

I have sequenced 1059-1149 bp of the C-terminal region of nine glucocerebrosidase (GBA) genes in orangutan, gorilla, baboon, squirrel monkey and chimpanzee. The orangutan, gorilla, baboon, and chimpanzee each have two copies of the gene, while the squirrel monkey only possesses a single copy. Analysis of the genes indicates that in gorilla and chimpanzee, one copy is functional while the other is a pseudogene, whereas in the orangutan, both GBA genes may be functional. The two GBA genes in each species are least similar in gorilla and chimpanzee (97.8%) and most similar in orangutan (99.5%) and baboon (99.4%) (Table 3.3). The increased similarity of the GBA genes in baboon and orangutan, as well as phylogenetic tree analysis indicates that gene conversion caused the evolution of *GBA* and ψ*GBA* to be concerted.

3.5.1. Sequencing strategy

I chose to investigate the molecular evolution of glucocerebrosidase by sequencing approximately 1.1 kb in the C-terminal region of both GBA genes in five primate species (gorilla, chimpanzee, orangutan, baboon, squirrel monkey). This region of the GBA gene was selected because it contains a 55 bp exonic deletion in the human GBA pseudogene (ψ*GBA*) that contributes to the gene's lack of function.[54] Thus it is possible to differentiate the functional GBA gene (*GBA*) from the ψ*GBA* by the presence or absence of this deletion. Although other regions of variability exist between human *GBA* and ψ*GBA*, this is the most significant variation that contributes to the non-functionality of human ψ*GBA*.[54] Additionally the C-terminal region of *GBA* is more conserved, which facilitates the use of human primers to isolate this region of *GBA* and

ψ*GBA* from primate DNA.[71]

Five primates, gorilla, chimpanzee, orangutan, baboon, and squirrel monkey, were analyzed to investigate the molecular evolution of *GBA* and ψ*GBA*. Since the *GBA* duplication event occurred relatively recently, approximately 36 to 40 million years ago,[54] it was necessary to select primates which had diverged from their common ancestor with humans less than 40 million years ago (Figure 3.3). This includes the old world monkeys and the great apes, of which I analyzed gorilla, chimpanzee, orangutan, and baboon.[69] The new world monkeys diverged at least 40 million years ago, and consequently should only possess one copy of *GBA*.[69] To compare the sequence of ancestral *GBA* to that of the duplicated GBA genes, I examined one new world monkey species, the squirrel monkey.

The *GBA* and ψ*GBA* genes in each primate species were isolated by PCR amplification of large regions, encompassing exons 3-10 (Figure 3.5). In gorilla and chimpanzee, this region of *GBA* was approximately 1.3 kb larger than that of ψ*GBA*, while in orangutan one GBA gene (*GBA*1) was ~1.3 kb larger than the other (*GBA*2) (Figure 3.6). In human *GBA* and ψ*GBA*, the 1223 bp size difference in this amplified region (exon 3-10) is due to the insertion of *Alu* elements into the introns of *GBA*. Given the similarity in size between these primate and human *GBA* and ψ*GBA* genes, it is likely that the *Alu* insertions are also present in gorilla and chimpanzee *GBA*, as well as orangutan *GBA*1. Isolation of *GBA* from ψ*GBA* in gorilla and chimpanzee, and separation of orangutan *GBA*1 and *GBA*2 prior to sequencing the C-terminal region ensured that the variation in sequences was due to amplification of different *GBA* genes and not allelic variation. Separation of the two baboon GBA genes (*GBA*1 and *GBA*2)

was not possible as both genes were similar in size (Figure 3.6).

3.5.2. Determination of non-functionality of primate GBA genes

Chimpanzee and gorilla have a GBA pseudogene and functional gene. This is made evident by the 55 bp deletion in exon 9 of ψ*GBA* for these species (Figure 3.8). This deletion is also present in human ψ*GBA* and contributes to its non-functionality.[54] The presence of this 55 bp deletion in human, chimpanzee, and gorilla ψ*GBA* and absence in the orangutan GBA genes indicates that the deletion was incorporated prior to the divergence of gorilla and after the divergence of orangutan from their common ancestor to human. That implies that the 55bp deletion in exon 9 occurred approximately 5.5-15 million years ago (mya).[69] This suggests that primates that have two copies of the GBA gene and are more distantly related to human than gorilla, may possess two functional GBA genes.[54]

The two orangutan GBA genes, *GBA*1 and *GBA*2, are 99.5% similar (Table 3.3). The deduced amino acid sequence of these genes varies by only one substitution event wherein glutamine replaces arginine in exon 8 (Figure 3.11). There are no known human mutations at this site in *GBA*, thus the implications on glucocerebrosidase functionality of mutation at this site are not known.[11] However in human *GBA*, a amino acid substitution of arginine→glutamine at a.a. position 359 has been found to result in mild Gaucher disease phenotype.[8] As arginine and glutamine possess similar properties it is likely that the mutation is not very severe.[50] Thus it is highly unlikely that this mutation in orangutan *GBA*2 would make the gene non-functional. It is possible, however, that other mutations exist elsewhere that would make one of the orangutan GBA genes non-

functional.

The amino acid sequence for the single squirrel monkey GBA gene is identical to that for human *GBA* (Figure 3.11), indicating that the squirrel monkey *GBA* is functional. Additionally, since the squirrel monkey only has one GBA gene, the essential nature of glucocerebrosidase suggests that the gene would have to be functional.

The baboon GBA genes, *GBA*1 and *GBA*2, are 99.4% similar in their nucleotide sequence (Table 3.3). They differ from each other at one amino acid location in exon 8. A nonsense mutation in exon 8 of baboon *GBA*1 creates a translation stop codon (Figure 3.11), which would produce a truncated polypeptide. Consequently, the baboon *GBA*1 gene is non-functional. Furthermore, both baboon *GBA*1 and *GBA*2 have a two nucleotide deletion in exon 8, which would in theory create a frameshift mutation in both genes. Frameshift mutations should have a significant impact on the expression of glucocerebrosidase, as all frameshift mutations found in human *GBA* create a null allele.[11] This result creates a conundrum, as it is unlikely that baboon lacks a functional GBA gene because glucocerebrosidase plays an essential role in lipid catabolism that both humans and mice require to survive.[48]

The region of baboon *GBA*1 and *GBA*2 that contains the AG deletion, which would create a frameshift mutation, was sequenced 7 times and every sequence showed the same deletion. Furthermore, the sequenced baboon genes were PCR amplified three separate times, suggesting that *taq* polymerase error is not responsible. The non-functionality of both sequenced baboon GBA genes can be alternatively explained. 1) It is possible that the sequencing results are inaccurate. 2) Baboon *GBA*1 and *GBA*2 could represent ψ*GBA* alleles and the sequence of the functional gene has not yet been

determined. Since I was unable to separate the two baboon GBA genes by size this is a distinct possibility. In the 1110 bp of *GBA*1 and *GBA*2 sequenced there are eight nucleotide variations (Figure 3.8). This is a high level of variation to attribute to polymorphic sites, particularly since the same region in human *GBA* contains only one polymorphic site.[14] 3) There are more than two GBA genes in baboon, and I have only been able to amplify two, both of which are pseudogenes.

3.5.3. Primate *GBA* and ψ*GBA* nucleotide variability

The C-terminal region of genomic *GBA* and ψ*GBA* in the five analyzed primates is 91.7-99.5% similar to human *GBA* (Table 3.2). Chimpanzee *GBA* (99.5%) and gorilla *GBA* (98.8%) are the most similar to human *GBA*, while the similarity drops to 91.7% in the squirrel monkey. The high level of similarity in gorilla, chimpanzee, and orangutan (97.4-99.5%) suggests that glucocerebrosidase is conserved as the average DNA similarity between these higher primates is reported to be between 96.5-98.4%.[69] This supports the murine *GBA* sequence data, which also suggests that the C-terminal region of *GBA* is more conserved.[71]

The variation in nucleotides is not evenly distributed throughout all regions of the *GBA* and ψ*GBA* C-terminus in the analyzed primates (Figure 3.9). In the exonic regions, 12.2% of sites are variable, while 19.7% of the intronic sites are variable. The nucleotide variation distribution is also variable within exons, exons 9 and 10 have the least variation (~10.5 %) and exon 8 has a higher value (15.7%) (Figure 3.9). Intronic regions do not have the same selective pressure as exons since intronic mutations, excluding splice site mutations, will not affect the functionality of the gene.[50]

The differences in the size of the primate GBA genes are due to deletions and insertions. In the five primate species, there are nineteen intronic deletions and insertions, all in introns 8 and 9 (Figure 3.8). The deletions and insertions range in size from a single nucleotide to a 33 bp insertion in gorilla *GBA*. The exonic deletions are far less numerous compared to those in introns. There are three exonic deletions, a two nucleotide deletion in exon 8 of the baboon GBA genes, a single nucleotide deletion in exon 8 of gorilla ψ*GBA*, and a 55bp deletion in exon 9 of ψ*GBA* in human, gorilla, and chimpanzee.

Allelic variation was not observed in the *GBA* and ψ*GBA* regions sequenced in any of the primates. This is not surprising since the comparable region in human *GBA* only contains one polymorphic site.[14] It is probable that if a larger portion of the primate GBA genes was sequenced some polymorphic sites would be discovered. Since there are 12 polymorphic sites in human *GBA* it likely that other primates would also have some *GBA* polymorphic sites.[14]

3.5.4. Determination of GBA gene copy number in primates

PCR amplification of GBA genes in gorilla, chimpanzee, and orangutan enabled the separation of the two genes in each primate by size (Figure 3.5). This was not possible in baboon and squirrel monkey. Sequencing in the baboon revealed two different nucleotide sequences (*GBA*1 and *GBA*2) (Figure 3.8). These sequences varied at eight sites, which is too high a level to attribute to polymorphic variation, especially since human *GBA* only has 1 polymorphic site in this region.[14] Sequencing in the squirrel monkey revealed only one *GBA* sequence (Figure 3.8). A single GBA gene in the

squirrel monkey is expected since the GBA duplication event is hypothesized to have occurred after squirrel monkey divergence.[54;69] Southern blot analysis was used to verify the gene copy number in these primate species. However the signal obtained was not sufficiently above background to determine gene copy number. Numerous approaches to optimizing the procedure were attempted without success. It is possible that the quality of the primate genomic DNA was poor and consequently other optimization techniques did not work.

3.5.5. Phylogenetic tree analysis and gene conversion

The phylogenetic tree of *GBA* and ψ*GBA* suggests that gene conversion has affected the evolution of the genes (Figure 3.10). If the GBA genes had arisen from a single duplication event and independently diverged, it would be expected the *GBA* would be grouped independently from ψ*GBA*. This is what occurs for gorilla, chimpanzee, and human *GBA* and ψ*GBA*. The orangutan *GBA*1 and *GBA*2 are, however, grouped separately, as are baboon *GBA*1 and *GBA*2 (Figure 3.10). The bootstrap values on the independent grouping of the orangutan and baboon genes are both 100, suggesting that the phylogenetic groupings are highly significant. This also suggests that (1) gene conversion homogenized *GBA*1 and *GBA*2 in orangutan and baboon or (2) the GBA genes in baboon and orangutan were recently duplicated and subsequently have had less time to diverge. It is more probable that gene conversion has masked the age of *GBA*1 and *GBA*2 in baboon and orangutan, than for three separate GBA duplication events to have occurred independently in these primate lineages. Additionally, the tree in Figure 3.10 contains a phylogenetic anomaly whereby it classifies chimpanzee and gorilla ψ*GBA*

as a clade, giving further support to the theory that concerted evolution has affected GBA. If GBA had been able to evolve independently it would be expected that the human and chimp ψ*GBA* would be grouped in a clade.

The likelihood that gene conversion will homogenize the genes in a multigene family is affected by factors such as gene number, gene size, and homology between genes. Glucocerebrosidase only has two genes in the higher primates which suggests that a single gene conversion can homogenize a considerable portion of the genes. Human ψ*GBA* is 5769 bp and *GBA* is 7604 bp, which means at most several thousand nucleotides of the gene can be affected by a single gene conversion event. Gene conversion has a higher incidence when genes are located in tandem as is the case for glucocerebrosidase.[50] Additionally, the high similarity of the duplicated GBA genes (96% between human *GBA* and ψ*GBA*) increases the likelihood of gene conversion since gene conversion relies on similarity for misalignment of repeats.[50;54]

3.3.6. Amino acid substitutions in primate GBA

I have deduced a region of 115 amino acids in chimpanzee *GBA*, gorilla *GBA*, orangutan *GBA*1 and *GBA*2, and squirrel monkey *GBA*, as well as a 104 a.a. region in baboon *GBA*1 and *GBA*2 (Figure 3.11). There are twelve variable sites in this 115 a.a. region when all six primates are compared. Baboon *GBA*1 is most dissimilar compared to human *GBA*, as well it contains a stop codon in exon 8. Baboon *GBA*1 would produce a truncated polypeptide and therefore must be a pseudogene. Interestingly none of the variable amino acid sites correspond to known human *GBA* mutations. Although this does not reveal the impact these amino acid variations have on the GBA genes, it

suggests that for primate *GBA* to be functional it cannot possess mutations which have been shown to compromise the functionality of its enzymatic product in humans.

3.5.7. Conclusions and future considerations

This research allows conclusions to be drawn on the divergence between the GBA genes, the mechanism which guided their evolution, and the functionality of the genes. Following the duplication of the glucocerebrosidase gene approximately 40 million years ago, the genes independently evolved, resulting in a functional gene and pseudogene in humans. However, the GBA genes have not diverged significantly in all primates. In baboon and orangutan, both GBA genes are highly similar suggesting that gene conversion homogenized the genes. Nonetheless these genes still differ from each other through nucleotide substitutions, deletions, and insertions. Most notable is the 55 bp deletion in exon 9 of chimpanzee and gorilla ψ*GBA*. This contributes to the non-functionality of human ψGBA and would also cause the chimpanzee and gorilla ψGBA to be non-functional. Neither of the orangutan genes possess this mutation, suggesting they are functional, as long as no other deleterious mutations exist. The C-terminal of the primate GBA genes is 91.7-99.5% similar to human GBA, while the amino acid sequence is 94.8-100% similar.

This research indicates the level of divergence that has occurred in the glucocerebrosidase genes since duplication. The lack of an exon 9 deletion in orangutan *GBA*1 and *GBA*2 suggests that both glucocerebrosidase genes may be functional. However to conclusively determine this it is necessary to sequence the genes in their entirety and/or attach the cDNA to a promoter for expression studies. The presence of a

frameshift mutation in baboon *GBA*1 and *GBA*2 indicates that both genes are non-functional. It is unlikely, however, that baboon does not require glucocerebrosidase. Therefore it would be of interest to look for an additional baboon GBA gene and sequence baboon *GBA*1 and *GBA*2 in their entirety to determine the extent of the frameshift mutations.

References

1. Barranger JA, Rice EO, Swaney WP: Gene transfer approaches to the lysosomal storage disorders. Neurochem.Res. 24: 601-615, 1999

2. Barranger JA, Tomich J, Weiler S, Sakallah S, Sansieri C, Mifflin T, Bahnson A, Wei FS, Wei JF, Vallor M: Molecular biology of glucocerebrosidase and the treatment of Gaucher disease. Cytokines Mol.Ther. 1: 149-163, 1995

3. Barton NW, Brady RO, Dambrosia JM, Di Bisceglie AM, Doppelt SH, Hill SC, Mankin HJ, Murray GJ, Parker RI, Argoff CE: Replacement therapy for inherited enzyme deficiency--macrophage-targeted glucocerebrosidase for Gaucher's disease [see comments]. N.Engl.J.Med. 324: 1464-1470, 1991

4. Berg-Fussman A, Grace ME, Ioannou Y, Grabowski GA: Human acid beta-glucosidase. N-glycosylation site occupancy and the effect of glycosylation on enzymatic activity. J.Biol.Chem. 268: 14861-14866, 1993

5. Beutler E: Gaucher disease: new molecular approaches to diagnosis and treatment. Science 256: 794-799, 1992

6. Beutler E: Gaucher disease as a paradigm of current issues regarding single gene mutations of humans. Proc.Natl.Acad.Sci.U.S.A 90: 5384-5390, 1993

7. Beutler E: Enzyme replacement therapy for Gaucher's disease. Baillieres Clinical Haematology 10: 751-763, 1997

8. Beutler E: Commentary: the natural history of Gaucher disease. Blood Cells Mol.Dis. 24: 82, 1998

9. Beutler E, Demina A, Gelbart T: Glucocerebrosidase mutations in Gaucher disease. Mol.Med. 1: 82-92, 1994

10. Beutler E, Gelbart T: Glucocerebrosidase (Gaucher disease). Hum.Mutat. 8: 207-213, 1996

11. Beutler E, Gelbart T: Hematologically important mutations: Gaucher disease. Blood Cells Mol.Dis. 24: 2-8, 1998

12. Beutler E, Grabowski GA: Gaucher Disease, in scriver (ed): The Metabolic and Molecular Basis of Inherited Disease, vol 7. New York, McGraw-Hill, 1995, pp 2641-2670

13. Beutler E, Nguyen NJ, Henneberger MW, Smolec JM, McPherson RA, West C, Gelbart T: Gaucher disease: gene frequencies in the Ashkenazi Jewish population. Am.J.Hum.Genet. 52: 85-88, 1993

14. Beutler E, West C, Gelbart T: Polymorphisms in the human glucocerebrosidase gene. Genomics 12: 795-800, 1992

15. Bierfreund U, Kolter T, Sandhoff K: Sphingolipid hydrolases and activator proteins. Methods Enzymol. 311: 255-276, 1999

16. Brady RO, Barton NW, Grabowski GA: The role of neurogenetics in Gaucher disease. Arch.Neurol. 50: 1212-1224, 1993

17. Brady RO, Johnson WG, Uhlendorf BW: Identification of heterozygous carriers of lipid storage diseases. Current status and clinical applications. Am.J.Med. 51: 423-431, 1971

18. Brady RO, Kanfer JN, Bradley RM, Shapiro D: Demonstration of a deficiency of glucocerebroside-cleaving enzyme in Gaucher's disease. J.Clin.Invest 45: 1112-1115, 1966

19. Bretthauer RK, Castellino FJ: Glycosylation of Pichia pastoris-derived proteins [see comments]. Biotechnol.Appl.Biochem. 30 (Pt 3): 193-200, 1999

20. Brierley, R, Davis, G, and Holtz, G. Production of insulin-like growth factor-1 in methylotrophic yeast cells. 23,463(5,324,639). 1994. United States.
Ref Type: Patent

21. Callahan M, Samimi R, Tayebi N, Sidransky E: Evolution of the glucocerebrosidase pseudogene. Am.J.Med.Genet.A183, 1999 (abstr)

22. Cereghino JL, Cregg JM: Heterologous protein expression in the methylotrophic yeast Pichia pastoris. FEMS Microbiol.Rev. 24: 45-66, 2000

23. Choudary PV, Ginns EI, Barranger JA: Molecular Cloning and Analysis of the Human Glucocerebrosidase Gene. DNA 4: 74, 1985

24. Choudary PV, Tsuji S, Martin BM, Guild BC, Mulligan RC, Murray GJ, Barranger JA, Ginns EI: The molecular biology of Gaucher disease and the potential for gene therapy. Cold Spring Harbor Symposia on Quantitative Biology 51 Pt 2: 1047-1052, 1986

25. Choy FY, Davidson RG: Gaucher's disease. I. Solubilization and electrophoresis of beta- glucosidase from leukocytes and cultured fibroblasts. Pediatr.Res. 12: 1115-1120, 1978

26. Choy FY, Davidson RG: Gaucher disease. III. Substrate specificity of glucocerebrosidase and the use of nonlabeled natural substrates for the investigation of patients. Am.J.Hum.Genet. 32: 670-680, 1980

27. Clare J, Rayment F, Ballantine S, Sreekrishna K, Romanos M: High level expression of tetanus toxin fragment C in Pichia Pastoris containing multiple tanden integrations of the gene. Biotechnology 9: 455-460, 1991

28. Clare JJ, Romanos MA, Rayment FB, et al: Production of mouse epidermal growth factor in yeast: high level secretion using Pichia pastoris strains containing multiple gene copies. Gene 105: 205-212, 1991

29. Cormand B, Montfort M, Chabas A, Vilageliu L, Grinberg D: Genetic fine localization of the beta-glucocerebrosidase (GBA) and prosaposin (PSAP) genes: implications for Gaucher disease. Hum.Genet. 100: 75-79, 1997

30. Cramer CL, Weissenborn DL, Oishi KK, Grabau EA, Bennett S, Ponce E, Grabowski GA, Radin DN: Bioproduction of human enzymes in transgenic tobacco. Ann.N.Y.Acad.Sci. 792: 62-71, 1996

31. Cregg JM, Vedvick TS, Raschke WC: Recent advances in the expression of foreign genes in Pichia pastoris. Biotechnology (N.Y.) 11: 905-910, 1993

32. Dale GL, Villacorte DG, Beutler E: Solubilization of glucocerebrosidase from human placenta and demonstration of a phospholipid requirement for its catalytic activity. Biochem.Biophys.Res.Commun. 71: 1048-1053, 1976

33. Dinur T, Osiecki KM, Legler G, Gatt S, Desnick RJ, Grabowski GA: Human acid beta-glucosidase: isolation and amino acid sequence of a peptide containing the catalytic site. Proc.Natl.Acad.Sci.U.S.A 83: 1660-1664, 1986

34. Doering T, Holleran WM, Potratz A, Vielhaber G, Elias PM, Suzuki K, Sandhoff K: Sphingolipid activator proteins are required for epidermal permeability barrier formation. J.Biol.Chem. 274: 11038-11045, 1999

35. Elstein D, Abrahamov A, Zimran A: Ethical considerations for enzyme replacement therapy in neuronopathic Gaucher disease. Clin.Genet. 54: 179-184, 1998

36. Erickson AH, Ginns EI, Barranger JA: Biosynthesis of the lysosomal enzyme glucocerebrosidase. J.Biol.Chem. 260: 14319-14324, 1985

37. Fallet S, Grace ME, Sibille A, Mendelson DS, Shapiro RS, Hermann G, Grabowski GA: Enzyme augmentation in moderate to life-threatening Gaucher disease. Pediatr.Res. 31: 496-502, 1992

38. Fischer R, Drossard J, Emans N, Commandeur U, Hellwig S: Towards molecular farming in the future: pichia pastoris-based production of single-chain antibody fragments [see comments]. Biotechnol.Appl.Biochem. 30 (Pt 2): 117-120, 1999

39. Genzyme. Annual Report. 1999.
Ref Type: Report

40. Ginns EI, Choudary PV, Martin BM, Winfield S, Stubblefield B, Mayor J, Merkle-Lehman D, Murray GJ, Bowers LA, Barranger JA: Isolation of cDNA clones for human beta-glucocerebrosidase using the lambda gt11 expression system. Biochem.Biophys.Res.Commun. 123: 574-580, 1984

41. Ginns EI, Choudary PV, Tsuji S, Martin B, Stubblefield B, Sawyer J, Hozier J, Barranger JA: Gene mapping and leader polypeptide sequence of human glucocerebrosidase: implications for Gaucher disease. Proc.Natl.Acad.Sci.U.S.A 82: 7101-7105, 1985

42. Ginns, E. J., Martin, B, Maysak, K, Eliason, W, and LaMarca, M. Cloned DNA sequence for synthesizing unique glucocerebrosidase. 452,398(5,879,680). 1999. United States.
Ref Type: Patent

43. Gonzales ML, Basu A, de Haas GH, Dijkman R, van Oort MG, Okolo AA, Glew RH: Activation of human spleen glucocerebrosidases by monoacylglycol sulfates and diacylglycerol sulfates. Arch.Biochem.Biophys. 262: 345-353, 1988

44. Grabowski GA: Gaucher disease. Enzymology, genetics, and treatment. Adv.Hum.Genet. 21: 377-441, 1993

45. Grabowski GA: Treatment of Gaucher's disease [letter]. N.Engl.J.Med. 328: 1565, 1993

46. Grabowski GA, Barton NW, Pastores G, Dambrosia JM, Banerjee TK, McKee MA, Parker C, Schiffmann R, Hill SC, Brady RO: Enzyme therapy in type 1 Gaucher disease: comparative efficacy of mannose-terminated glucocerebrosidase from natural and recombinant sources. Ann.Intern.Med. 122: 33-39, 1995

47. Grabowski GA, Gatt S, Horowitz M: Acid beta-glucosidase: enzymology and molecular biology of Gaucher disease. Crit Rev.Biochem.Mol.Biol. 25: 385-414, 1990

48. Grabowski GA, Leslie N, Wenstrup R: Enzyme therapy for Gaucher disease: the first 5 years. Blood Rev. 12: 115-133, 1998

49. Grabowski GA, Saal HM, Wenstrup RJ, Barton NW: Gaucher disease: a prototype for molecular medicine. Crit Rev.Oncol Hematol. 23: 25-55, 1996

50. Graur D, Li W: Fundamentals of Molecular Evolution. Sutherland, Sinauer Associates, Inc., 2000

51. Higgins DR, Cregg JM: Pichia Protocols. Totowa, Humana Press, 1998

52. Hillis DM, Bull JJ: An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst.Biol. 42: 182-192, 1993

53. Hoppe H: Cerezyme--recombinant protein treatment for Gaucher's disease. J.Biotechnol. 76: 259-261, 2000

54. Horowitz M, Wilder S, Horowitz Z, Reiner O, Gelbart T, Beutler E: The human glucocerebrosidase gene and pseudogene: structure and evolution. Genomics 4: 87-96, 1989

55. Imai K, Nakamura M, Yamada M, Asano A, Yokoyama S, Tsuji S, Ginns EI: A novel transcript from a pseudogene for human glucocerebrosidase in non-Gaucher disease cells. Gene 136: 365-368, 1993

56. Invitrogen: *Pichia* Expression Kit. SanDiego, Invitrogen, 1994

57. Invitrogen: pPICZalphaA, B, and C Instruction Manual. 1996

58. Kass DH, Batzer MA, Deininger PL: Gene conversion as a secondary mechanism of short interspersed element (SINE) evolution. Mol.Cell Biol. 15: 19-25, 1995

59. Lasker GW, Tyzzer RN: Physical Anthropology. Orlando, Saunders College Publishing, 1981

60. Latham T, Grabowski GA, Theophilus BD, Smith FI: Complex alleles of the acid beta-glucosidase gene in Gaucher disease. Am.J.Hum.Genet. 47: 79-86, 1990

61. Lee Y, Kinoshita H, Radke G, Weiler S, Barranger JA, Tomich JM: Position of the sulfhydryl group and the disulfide bonds of human glucocerebrosidase. J.Protein Chem. 14: 127-137, 1995

62. Long GL, Winfield S, Adolph KW, Ginns EI, Bornstein P: Structure and organization of the human metaxin gene (MTX) and pseudogene. Genomics 33: 177-184, 1996

63. Madar-Shapiro L, Pasmanik-Chor M, Dinur T, Dagan A, Gatt S, Horowitz M: Intracellular degradation of fluorescent glycolipids by lysosomal enzymes and their activators. J.Inherit.Metab Dis. 22: 623-637, 1999

64. Makalowski W, Mitchell GA, Labuda D: Alu sequences in the coding regions of mRNA: a source of protein variability. Trends Genet. 10: 188-193, 1994

65. Martin BM, Sidransky E, Ginns EI: Gaucher's disease: advances and challenges. Adv.Pediatr. 36: 277-306, 1989

66. Martin RD: Primate origins and evolution: a phylogenetic reconstruction. Cambridge, Princeton University Press, 1990

67. Masuno M, Tomatsu S, Sukegawa K, Orii T: Non-existence of a tight association between a 444leucine to proline mutation and phenotypes of Gaucher disease: high frequency of a NciI polymorphism in the non-neuronopathic form. Hum.Genet. 84: 203-206, 1990

68. Miele RG, Nilsen SL, Brito T, Bretthauer RK, Castellino FJ: Glycosylation properties of the Pichia pastoris-expressed recombinant kringle 2 domain of tissue-type plasminogen activator. Biotechnol.Appl.Biochem. 25 (Pt 2): 151-157, 1997

69. Miyamoto MM, Koop BF, Slightom JL, Goodman M, Tennant MR: Molecular systematics of higher primates: genealogical relations and classification. Proc.Natl.Acad.Sci.U.S.A 85: 7627-7631, 1988

70. NIH Technology Assessment Panel on Gaucher Disease: Gaucher disease. Current issues in diagnosis and treatment. JAMA 275: 548-553, 1996

71. O'Neill RR, Tokoro T, Kozak CA, Brady RO: Comparison of the chromosomal localization of murine and human glucocerebrosidase genes and of the deduced amino acid sequences. Proc.Natl.Acad.Sci.U.S.A 86: 5049-5053, 1989

72. Prence EM, Garrett KO, Panitch H, Basu A, Glew RH, Wherrett JR, Huterer S: Sulfogalactocerebroside and bis-(monoacylglyceryl)-phosphate as activators of spleen glucocerebrosidase. Clin.Chim.Acta 156: 179-189, 1986

73. Rasmussen, J. Barsomian G. and Bergman, B. Enzymatically active recombinant glucocerebrosidase. Genzyme. 748283(5236838). 1993. United States.
Ref Type: Patent

74. Reiner O, Wigderson M, Horowitz M: Structural analysis of the human glucocerebrosidase genes. DNA 7: 107-116, 1988

75. Romanos MA: Advances in the use of Pichia pastoris for high-level gene expression. Current Opinion in Biotechnology 6: 527-533, 1995

76. Romanos MA, Scorer CA, Clare JJ: Foreign gene expression in yeast: a review. Yeast 8: 423-488, 1992

77. Rosenfeld SA: Use of Pichia pastoris for expression of recombinant proteins. Methods Enzymol. 306: 154-169, 1999

78. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual. New York, Cold Spring Harbor Laboratory Press, 1989

79. Schiffmann R, Heyes MP, Aerts JM, Dambrosia JM, Patterson MC, DeGraba T, Parker CC, Zirzow GC, Oliver K, Tedeschi G, Brady RO, Barton NW: Prospective study of neurological responses to treatment with macrophage-targeted glucocerebrosidase in patients with type 3 Gaucher's disease. Ann.Neurol. 42: 613-621, 1997

80. Scorer CA, Buckholz RG, Clare JJ, Romanos MA: The intracellular production and secretion of HIV-1 envelope protein in the methylotrophic yeast Pichia pastoris. Gene 136: 111-119, 1993

81. Scorer CA, Clare JJ, McCombie WR, Romanos MA, Sreekrishna K: Rapid selection using G418 of high copy number transformants of Pichia pastoris for high-level foreign gene expression. Biotechnology (N.Y.) 12: 181-184, 1994

82. Scott A: The sequence of the gorilla fetal globin genes: Evidence for multiple gene conversion in human evolution. Mol.Biol.Evol. 1: 371-389, 1984

83. Sorge J, Gross E, West C, Beutler E: High level transcription of the glucocerebrosidase pseudogene in normal subjects and patients with Gaucher disease. J.Clin.Invest 86: 1137-1141, 1990

84. Sorge J, West C, Kuhl W, Treger L, Beutler E: The human glucocerebrosidase gene has two functional ATG initiator codons. Am.J.Hum.Genet. 41: 1016-1024, 1987

85. Sorge J, West C, Westwood B, Beutler E: Molecular cloning and nucleotide sequence of human glucocerebrosidase cDNA. Proc.Natl.Acad.Sci.U.S.A 82: 7289-7293, 1985

86. Takasaki S, Murray GJ, Furbish FS, Brady RO, Barranger JA, Kobata A: Structure of the N-asparagine-linked oligosaccharide units of human placental beta-glucocerebrosidase. J.Biol.Chem. 259: 10112-10117, 1984

87. Tayebi N, Stern H, Dymarskaia I, Herman J, Sidransky E: 55-base pair deletion in certain patients with Gaucher disease complicates screening for common Gaucher alleles. Am.J.Med.Genet. 66: 316-319, 1996

88. Thill G, Davis G, Stillman C, Holtz G, et al: Positive and negative effects of multicopy integrated expression vectors on protein expression in Pichia pastoris., in Heslot H, Davies J, Florent J, et al (eds): Proceedings of the 6th international symposium on genetics of microorganisms, vol. 2. Paris, Societe Francaise de Microbiologie, 1990, pp 477-490

89. Tsuji S, Choudary PV, Martin BM, Stubblefield BK, Mayor JA, Barranger JA, Ginns EI: A mutation in the human glucocerebrosidase gene in neuronopathic Gaucher's disease. N.Engl.J.Med. 316: 570-575, 1987

90. Tsuji S, Choudary PV, Martin BM, Winfield S, Barranger JA, Ginns EI: Nucleotide sequence of cDNA containing the complete coding sequence for human lysosomal glucocerebrosidase. J.Biol.Chem. 261: 50-53, 1986

91. Tybulewicz VL, Tremblay ML, LaMarca ME, Willemsen R, Stubblefield BK, Winfield S, Zablocka B, Sidransky E, Martin BM, Huang SP: Animal model of Gaucher's disease from targeted disruption of the mouse glucocerebrosidase gene. Nature 357: 407-410, 1992

92. Van De Water NS, Jolly RD, Farrow BR: Canine Gaucher disease--the enzymic defect. Aust.J.Exp.Biol.Med.Sci. 57: 551-554, 1979

93. Wei, C. Molecular Analysis and Expression of the Human Glucocerebrosidase Gene. 1998. University of Victoria.
Ref Type: Thesis/Dissertation

94. Weiler S, Kishimoto Y, O'Brien JS, Barranger JA, Tomich JM: Identification of the binding and activating sites of the sphingolipid activator protein, saposin C, with glucocerebrosidase. Protein Sci. 4: 756-764, 1995

95. Weiss HM, Haase W, Michel H, Reilander H: Comparative biochemical and pharmacological characterization of the mouse 5HT5A 5-hydroxytryptamine receptor and the human beta2-adrenergic receptor produced in the methylotrophic yeast Pichia pastoris. Biochem.J. 330 (Pt 3): 1137-1147, 1998

96. Winfield SL, Tayebi N, Martin BM, Ginns EI, Sidransky E: Identification of three additional genes contiguous to the glucocerebrosidase locus on chromosome 1q21: implications for Gaucher disease [letter] [see comments]. Genome Res. 7: 1020-1026, 1997

97. Zimran A, Gelbart T, Westwood B, Grabowski GA, Beutler E: High frequency of the Gaucher disease mutation at nucleotide 1226 among Ashkenazi Jews. Am.J.Hum.Genet. 49: 855-859, 1991

98. Zimran A, Sorge J, Gross E, Kubitz M, West C, Beutler E: A glucocerebrosidase fusion gene in Gaucher disease. Implications for the molecular anatomy, pathogenesis, and diagnosis of this disorder. J.Clin.Invest 85: 219-222, 1990

Partial Copyright License

I hereby grant the right to lend my thesis (or dissertation) to users of the University of Victoria Library, and to make single copies for only such users or in response to a request from the Library of any other university, or similar institution, on its behalf or for one of its users. I further agree that permission for extensive copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis:

Gaucher Disease: Expression of Human Glucocerebrosidase in *Pichia pastoris* and Evolution of Glucocerebrosidase Gene and Pseudogene in Primates

Author: [redacted]

Julie Rachel Wafaei

October 31, 2000

Curriculum Vitae

Julie Rachel Wafaei

Education:

University of Victoria	1997-2000
McMaster University, Hamilton, ON	1993-1997

Degrees Awarded:

B.Sc. (McMaster University)	1997

Honours and Awards:

BC Health Foundation Summer Scholarship	1998
McMaster Travel Award	1995
Nicholson-St.Paul's Bursary	1993-1995
Knight's of Columbus Award	1993