Mutation Analysis, Heterologous Expression, and Characterization of Human Glucocerebrosidase

by

Graham Bernard Sinclair
B.Sc., University of British Columbia, 1995

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

DOCTOR OF PHILOSOPHY

in the Department of Biology

We accept this dissertation as conforming to the required standard

---

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

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

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

Dr. T. Pearson, Outside Member (Department of Biochemistry and Microbiology)

Dr. D. Applegarth, External Examiner (Department of Pediatrics, University of British Columbia)

Supervisor: Dr. Francis Y.M. Choy

ABSTRACT

Gaucher disease, the most common lysosomal storage disorder, results from a deficiency in the enzyme glucocerebrosidase. Inherited as an autosomal recessive disorder, Gaucher disease is clinically heterogeneous with both non-neuronopathic (Type 1) and neuronopathic (Types 2 and 3) subtypes. Although over 100 mutations in the glucocerebrosidase (GBA) gene have been identified, there still exists a poor correlation between individual genotypes and observed phenotypes, particularly for the neuronopathic subtypes. Using DNA isolated from archival tissue samples and standard molecular biology techniques, two novel and two rare mutations were identified in three individuals with neuronopathic Gaucher disease. One mutation identified only in aboriginals of Cree descent was further characterized by heterologous expression in baculovirus-infected *Sf*9 cells and displayed moderate levels of residual enzyme activity, despite the corresponding disease severity observed. Heterologous expression studies were extended to examine systems for high-level glucocerebrosidase expression for biochemical analysis and biotherapeutics. While the methylotrophic yeast *Pichia pastoris* was found to express minimal amounts of human glucocerebrosidase even when selected for high gene copy number, stable transfected *Sf*9 cells were found to produce functional glucocerebrosidase at a level of 1.0-1.3mg/L of cell culture. A subsequent analysis of synonymous codon usage bias in *Pichia pastoris* identified a significant difference in codon choice between the expression host and the GBA gene. Codon optimization studies using a 5' fragment of the GBA gene fused to a luciferase reporter gene found that alterations in both G+C content and codon bias increased expression levels 7.5 to 10 fold. This suggests that codon optimization of the entire GBA gene could significantly improve production levels of this important enzyme in *Pichia pastoris* and other expression hosts.

Examiners:

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

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

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

Dr. T. Pearson, Outside Member (Department of Biochemistry and Microbiology)

Dr. D. Applegarth, External Examiner (Department of Pediatrics, University of British Columbia)

TABLE OF CONTENTS

Title Page i
Abstract ii
Table of Contents iv
List of Tables vii
List of Figures viii
Acknowledgements x
Dedication xi
Chapter 1 – Introduction to Gaucher Disease and Glucocerebrosidase 1
1.1 – Clinical Phenotypes 1
1.2 – Glucocerebrosidase Genomics 3
1.3 – Glucocerebrosidase Mutations and Population Genetics 4
1.4 – Gluocerebrosidase Biochemistry 7
1.5 – Glucocerebrosidase Biosynthesis 9
1.6 – Therapeutics 12
1.7 – References 16
Chapter 2 – Neuronopathic Gaucher Disease Mutation Analysis 22
2.1 – Introduction 22
2.2 – Materials and Methods 25
2.2.1 – Case Reports 25
2.2.2 – Cell Culture and Fibroblast DNA Extraction 27
2.2.3 – PCR Amplification and Sequence analysis of Fibroblast DNA 27
2.2.4 – PCR Amplification and Sequence analysis of Archival DNA 29
2.2.5 – RFLP Analysis of Gaucher Mutations 30
2.2.6 – Baculovirus Expression 32
2.2.7 – Western Blotting and Activity Assay 34
2.3 – Results 35
2.4 –Discussion 47
2.4.1 – Type 2 Cases 47
2.4.2 – Type 3 Cases 52
2.5 – References 55

Chapter 3 – Heterologous Expression and Purification of Glucocerebrosidase using yeast (*Pichia pastoris*) and insect cell (*Spodoptera Frugiperda*) hosts 60
3.1 – Introduction 60
3.2 – Materials and Methods 66
3.2.1 – *Pichia* strains 66
3.2.2 – Vector Construction 67
3.2.3 – Dual-Cassette Vector 67
3.2.4 – Zeocin™ Selection 71
3.2.5 – *Pichia* Culture and Induction 72
3.2.6 – *Sf*9 Vector Construction 72
3.2.7 – *Sf*9 Cell Transfection and Stable Cell Line Selection 73
3.2.8 – Glucocerebrosidase Activity Assay 74
3.2.9 – Protein Analyses 76
3.2.10 – Southern and RNA Dot Blotting 77
3.2.11 – FPLC Purification 78
3.3 – Results 80
3.3.1 – *Pichia* Expression 80
3.3.2 – *Sf*9 Expression 90
3.4 – Discussion 100
3.4.1 – *Pichia* Expression 100
3.4.2 – *Sf*9 Expression 100
3.5 – References 105
Chapter 4 – Synonymous Codon Usage Bias and Translational Inefficiency in *Pichia pastoris* 110
4.1 – Introduction 110
4.2 – Materials and Methods 114
4.2.1 – Measures of Synonymous Codon Usage Bias 114
4.2.2 – Correspondence Analysis of Codon Usage 115
4.2.3 – Luciferase Fusion and Plasmid Construction 117
4.2.4 – Construction of Codon Optimized and G+C Altered Fragments 119
4.2.5 – *Pichia pastoris* Transformation 125

4.2.6 – *Pichia* Expression and Luciferase Assay 126
4.2.7 – Northern Blotting 127
4.3 – Results 128
4.3.1 – Correspondence Analysis of Codon Usage 128
4.3.2 – *Pichia pastoris* Expression of Luciferase Fusions 132
4.3.3 – Analysis and Expression of Codon Altered Luciferase Fusions 132
4.4 – Discussion 134
4.5 – References 141
Chapter 5 – Summary of Glucocerebrosidase Analysis and Expression 146
5.1 – References 152

LIST OF TABLES

Table 2.1– Primers for the PCR amplification of glucocerebrosidase genomic DNA and cDNA 28

Table 2.2 – Primers for RFLP analysis and mutant cDNA synthesis 31

Table 2.3- Glucocerebrosidase specific activity measurements for crude extracts of recombinant baculovirus infected *Sf*9 cells with the artificial substrate (4MUGP) 46

Table 3.1– Primers used in the construction of glucocerebrosidase expression plasmids 68

Table 3.2 – Glucocerebrosidase activity on the natural and artificial (4MUGP) substrates from *Pichia pastoris* expression clones. 84

Table 3.3 – Glucocerebrosidase activities (4MUGP) from stable *Sf*9 transformants..... 93

Table 4.1 – Primers used in the construction of GBA–luciferase fusion vectors 118

Table 4.2 – Comparison of preferred codons from *P. pastoris* and *S. cerevisiae*. 121

Table 4.3 – Primers used for the construction of codon altered GBA fragments 124

Table 4.4– Measures of codon and nucleotide bias from native and codon altered glucocerebrosidase cDNA fragments. 131

LIST OF FIGURES

Figure 2.1 – Overlap extension mutagenesis of the GBA gene 33
Figure 2.2 – Identification and delineation of the crossover mutant allele 36
Figure 2.3 – DNA sequence analysis and restrication digest of PCR-amplified genomic DNA of glucocerebrosidase exon 3 from patient 1043 38
Figure 2.4 – Sequence analysis of PCR-amplified DNA from glucocerebrosidase exon 5 of JB 39
Figure 2.5 – *Kpn*I restriction analysis of mutation P122S 41
Figure 2.6 – Western blot of baculovirus infected *Sf*9 cell crude lysates 43
Figure 2.7 – Direct sequence analysis of archival DNA from BW 45
Figure 2.8 – A mechanism for the creation of a crossover mutant 49
Figure 3.1 – *Pichia pastoris* and *Sf*9 cell expression constructs 69
Figure 3.2 – Southern blot of genomic DNA extracted from *Pichia pastoris* strains 81
Figure 3.3 – RNA dot blots of total RNA extracted from induced *Pichia pastoris* 83
Figure 3.4 – Western blot analysis of glucocerebrosidase secretion from *Pichia pastoris* 85
Figure 3.5 – SDS-PAGE and protein staining of culture medium and active FPLC fractions from glucocerebrosidase expression *Pichia* strain KM26D 87
Figure 3.6 – Elution traces from hydrophobic interaction chromatography 88
Figure 3.7 – Western blot of *Pichia* strain KM26D (GBA) culture medium and the pooled active fractions from hydrophobic interaction chromatography 89
Figure 3.8 – Western blot of transient glucocerebrosidase (GBA) expression in *Sf*9 cell lysates 91
Figure 3.9 – Western blot of culture medium from transient transfected *Sf*9 cells 92
Figure 3.10 – Elution profile from hydrophobic interaction chromatography (HIC-FPLC) of stable transfected *Sf*9 cell medium 95
Figure 3.11 – Glucocerebrosidase purified from the culture medium of a stable *Sf*9 cell line 96
Figure 3.12 – PNGase F enzymatic deglycosylation of partially purified glucocerebrosidase 97

Figure 3.13 – pH optimum on the artificial substrate (4MUGP) 99
Figure 4.1 – Glucocerebrosidase and luciferase fusion vectors constructed for the expression analysis of codon usage bias 120
Figure 4.2 – Sequence alignment of the first 200bp of the native glucocerebrosidase cDNA with the Pichia pastoris codon optimized (GBOPT) and G+C altered (GBGCA) constructs 123
Figure 4.3 – Correspondence analysis of codon usage 129
Figure 4.4 – Peak luciferase activities for variable length glucocerebrosidase (GBA) fragments 133
Figure 4.5 – Peak luciferase activities for codon altered glucocerebrosidase (GBA) fragments 135
Figure 4.6 – Northern blot of total RNA extracted from GBA-luciferase fustion transformed *Pichia pastoris* clones 136

ACKNOWLEDGMENTS

I would like to thank my supervisor, Dr. Francis Choy, for his unwavering support and faith in my abilities. To my labmates over the years, I thank you for helping make the process so enjoyable and for finding my presence tolerable. Brett Poulis deserves high praise for giving so much of his time to the pursuit of my protein, and Josh Eades, Michael Bridge, Mike Wilson and Kris von Schalberg for advice on all topics, scientific or not. To my family and friends (including Carter's graphics service) thanks for putting it all into perspective. Finally to Jody, thank you for understanding that a half-hour of lab work actually means three, and for letting me share this all with you.

Dedication

To Be Anounced, *in utero* 2001

Chapter 1 – Introduction to Gaucher Disease and Glucocerebrosidase

1.1 – Clinical Phenotypes

The maintenance of homeostasis in the human body requires the constant turnover of cells and the clearance or recycling of various cellular components. Much of this work occurs in the lysosomes of reticuloendothelial cells and over thirty human disorders are associated with a breakdown in lysosomal function due to inherited enzyme deficiencies. Of the human lysosomal storage disorders identified to date, Gaucher disease is by far the most common (Beutler and Grabowski, 2001). Originally misidentified as a hepatic neoplasm in 1882 by Dr. Phillipe Ernest Gaucher, this autosomal recessive disorder results from a systemic accumulation of the sphingolipid glucocerebroside (glucosylceramide) (Beutler and Grabowski, 2001). This lipid accumulation results from the dysfunction of the enzyme glucocerebrosidase (EC 3.2.1.45, acid β-glucosidase, glucosylceramidase) which normally acts as the penultimate step in the degradation of membrane glucosphingolipids to convert glucocerebroside to glucose and ceramide (Beutler and Grabowski, 2001). Accumulated glucocerebroside is taken-up by circulating macrophages and subsequently deposited in the liver, spleen, and bone marrow of affected individuals. Histological investigation of affected tissues tends to reveal lipid engorged macrophages (Gaucher Cells) which lead to a disruption of normal organ function. The resultant hepatosplenomegaly, pancytopenia, and bone crises observed are the hallmark symptoms of Gaucher disease (Balicki and Beutler, 1995).

There has been a great deal of molecular and clinical heterogeneity seen in Gaucher disease patients and the disorder has been divided into three main sub-groups. Type 1 Gaucher disease (OMIM 230800) is the mildest form, with an onset of symptoms anywhere from late childhood through adulthood. Although consistent in the absence of neurological complications, the visceral severity seen in type 1 Gaucher patients can range from nearly asymptomatic octogenarians to children who die from complications in the second decade of life. Some level of splenomegaly is observed in all symptomatic patients with the spleen itself accounting for up to 25% of total body weight in some patients (Grabowski, 1993).

Type 2 (acute neuronopathic) Gaucher disease (OMIM 230900) is the most severe form of the disorder and is characterized by debilitating neuronopathy and early mortality. Although Gaucher cells have been shown to accumulate in the sub-cortical white matter of type 2 patients, the exact cause of neuronal loss remains to be elucidated (Balicki and Beutler, 1995). The onset of symptoms tends to occur perinatally, with death of the infant occurring by two years of age. There is a distinct perinatal lethal form of the disorder contained within the type 2 phenotype defined by hydrops fetalis, icthyosis, and death within hours of birth (Sidransky et al., 1992; Strasberg et al., 1994).

The final clinical sub-group, type 3 Gaucher disease (OMIM 231000), has recently been further sub-divided into types 3a, 3b, and 3c to reflect the large degree of clinical heterogeneity within this group (Beutler and Grabowski, 2001). The type 3 group as a whole is comprised of the sub-acute neuronopathic Gaucher patients, those who show neurodegenerative symptoms but who are able to survive through childhood and into adulthood. Type 3a patients are a distinct sub-group from the Norrbottnia region of

Sweden who share homozygosity for the L444P glucocerebrosidase mutation and a consistently severe neuronopathic and visceral symptomology (Beutler and Grabowski, 2001). Type 3c Gaucher disease patients all share homozygosity for the D409H mutation leading to the presence of cardiac valve calcification and corneal opacities not observed in any other Gaucher sub-type (Chabas et al., 1995). Finally, type 3b encompasses all juvenile neuronopathic cases not previously categorized and displays a great range of severity in visceral and neurodegenerative symptoms (Beutler and Grabowski, 2001).

1.2 – Glucocerebrosidase Genomics

The full-length cDNA (Sorge et al., 1985; Tsuji et al., 1986) and genomic DNA (Horowitz et al., 1989) sequences for the glucocerebrosidase (GBA) gene have been delineated. The 7.6kb gene contains 11 exons and consensus TATA and CAAT sequences have been identified in the promoter region 250bp upstream of the translation start site (Horowitz et al., 1989). The GBA gene is unique in that it contains two functional ATG start sites, both of which appear to be functional *in vitro* and *in vivo* although at different efficiencies (Pasmanik-chor et al., 1996; Sorge et al., 1987). A standard hydrophobic 19 amino acid leader sequence or hydrophilic 39 amino acid leader (depending on the start site used) is encoded by the second exon of the GBA gene (Sorge et al., 1987). Located in a tight cluster of genes on the long arm of chromosome 1 (1q21), there is a pseudogene (psGBA) with 96% sequence similarity 16kb downstream of the GBA gene (Horowitz et al., 1989). This pseudogene is transcribed but not successfully translated as it contains numerous missense and nonsense mutations and a major deletion in exon 9 (Zimran et al., 1990). The metaxin gene (MTX) and

pseudogene (psMTX) are contiguous with the GBA gene and pseudogene in a reverse orientation at 1q21. It appears a historical tandem duplication event involving GBA and MTX gave rise to the two corresponding pseudogenes in the lineage leading to modern *Homo sapiens* (Winfield et al., 1997). This duplication event is a relatively recent event from an evolutionary perspective as it is present in the Great Apes and Old World Monkeys but is absent from the New World Primates and other mammals (Bornstein et al., 1995). Five other genes have been identified within 75kb of the GBA gene; however, none have been directly or indirectly implicated in the etiology of Gaucher disease (Winfield et al., 1997).

1.3 - Glucocerebrosidase Mutations and Population Genetics

Although Gaucher disease is a rare panethnic disorder, incidence rates are significantly elevated in Ashkenazi Jewish populations, predominantly for type 1 disease (Beutler and Grabowski, 2001). Population estimates and direct mutational screening have assessed the disease allele frequency for Gaucher disease within this population at approximately 0.03-0.04 (Beutler et al., 1993; DeMarchi et al., 1996; Eng et al., 1997) This is in contrast to global disease allele frequency estimates of 0.004-0.006 for non-Jewish populations (Beutler and Grabowski, 2001; Meikle et al., 1999) Over 100 disease causing mutations have been identified in the GBA gene (Beutler and Gelbart, 1997) and while most appear to be rare or private mutation, the elevated incidence of Gaucher disease in Ashkenazi Jews is due to four common mutations (Koprivica et al., 2000). Following standard nomenclature (Beutler et al., 1996) mutations in the glucocerebrosidase gene will be identified in this thesis by the amino acid substitution

caused (i.e. N370S as a substitution of serine for asparagine at amino acid 370 of the mature polypeptide). Where appropriate, the cDNA nucleotide substitution involved (i.e. A1226G as an A to G transition at cDNA nucleotide 1226) will also be included. Mutations N370S, L444P, ins84GG (insertion mutation causing a frameshift) and IVS2(+1) (a splice donor site variant in intron 2) account for >93% of the mutations identified in the Ashkenazi Gaucher population (Beutler et al., 1992; Koprivica et al., 2000). Comparably, these four mutations account for only 49% of the errors identified in non-Jewish patients and although complex alleles arising from interactions with the pseudogene account for 17% of non-Jewish mutations, a more heterogeneous mixture of rare, private, or currently unidentified mutations account for the rest of the Gaucher disease cases on a global level (Koprivica et al., 2000). The classes of mutations identified in the GBA gene include single basepair transitions and transversions, small deletions and insertions, triplet insertions, long tract deletions, splice site variants, and even whole gene deletions (Beutler and Gelbart, 1997; Koprivica et al., 2000).

The elevated incidence of Gaucher disease (and other lysosomal storage disorders) in the Ashkenazi Jewish population has lead to a great deal of investigation and speculation from the research community to identify a causal factor. The most attractive explanations for increased disease allele frequencies within this population are either founder effects or a heterozygote advantage. The history of the Ashkenazim does suggest that the introduction of deleterious alleles in a small founder population leads to increased disease allele frequency through subsequent population expansion. The best estimates of founding Ashkenazi Jewish populations in Poland and Lithuania in the 13th century range

from 10,000-15,000 members (Diamond, 1994; Motulsky, 1995). This population has subsequently expanded and migrated globally and 21st century estimates of those with Ashkenazi heritage have reached approximately 8 million (Motulsky, 1995). Correlating with this population structure are recent estimates of the origin of the common N370S mutation in Ashkenazi Jews in the 13th century based on fine structure linkage mapping of haplotype data (Colombo, 2000; Diaz et al., 2000). Population modeling using this same haplotype data, however, suggests that the N370S mutation appears to have arisen too recently in this population to reach its current allele frequency as the result of genetic drift alone (Boas, 2000). Also, the increased allele frequency of four mutations in Ashkenazi Gaucher patients as the chance occurrence of multiple mutations within a few founding populations is statistically improbable (Diamond, 1994; Diaz et al., 2000; Motulsky, 1995). Further confounding this founder effect model is the high incidence of two other lysosomal storage disorders (Tay-Sachs and Neimann-Pick) in this population (Diamond, 1994; Motulsky, 1995).

It has been argued that if multiple alleles for several biochemically (but not genetically) associated disorders are elevated in frequency then selection, rather than founder affects, may be playing a role in the Ashkenazim (Beutler and Grabowski, 1994; Motulsky, 1995; Peleg et al., 1998). Although heterozygote advantage has been clearly delineated as increasing disease allele frequencies for a number of human disorders, a causal factor remains to be elucidated for Gaucher disease. Early suggestions that Gaucher carriers may display some resistance to tuberculosis was attractive due to the historical marginalization of Ashkenazi Jews, but this hypothesis has not been substantiated by

GBA mutation screening in modern Jewish tuberculosis patients (Beutler and Grabowski, 1994). Several other heterozygote advantage theories have been put forward for Gaucher disease but none have been supported experimentally and it remains impossible to tease apart the possible combined effects of population dynamics and selective pressures on disease allele frequencies within this group (Colombo, 2000; Diaz et al., 2000; Peleg et al., 1998).

1.4 – Glucocerebrosidase Biochemistry

Glucocerebrosidase is a lysosomal membrane-associated hydrolytic enzyme that requires association with negatively charged phospholipids and an activator protein, Saposin C, for *in vivo* activity (Glew et al., 1988; Qi and Grabowski, 1998). The mature polypeptide is 515 amino acids in length excluding a 19 amino acid (or 39 amino acid) leader sequence that is cleaved from the nascent polypeptide during transit through the endoplasmic reticulum (ER) membrane (Erickson et al., 1985). The apparent molecular weight of the polypeptide is 55,000 but the mature protein is N-glycosylated at 4 of 5 potential glycosylation sites to produce a mature glycoprotein with a molecular weight of 62,000 to 67,000 (Choy and Woo, 1991; Erickson et al., 1985; Takasaki et al., 1984). The substrate specificity of glucocerebrosidase is directed towards glucosphingolipids with medium to long fatty-acyl chains but the identity of the substrate sugar moiety is more important than the aglycan portion of the molecule (Glew et al., 1988). This characteristic has been exploited for *in vitro* assays utilizing fluorogenic compounds such as 4-methyl-umbelliferyl-β-D-glucopyranoside (4MUGP) for measuring glucocerebrosidase activity (Daniels et al., 1980). Glucocerebrosidase contains 5

disulfide bridges and two reduced cysteines and is optimally active at a pH of 5.5 (Glew et al., 1988). The active site of the protein is encoded by exons 9 and 10 of the GBA gene and the residues putatively involved in substrate cleavage have been delineated based on comparisons to other β-glucosidase family members (Beutler and Grabowski, 2001; Dinur et al., 1986).

The crystal structure of the protein has not been determined but modeling algorithms do not suggest any transmembrane domains or large tracts of local hydrophobicity in the protein (Beutler and Grabowski, 2001). The protein is, however, tightly membrane associated and requires the presence of detergents to be solublized from the lysosomal membrane (Choy and Woo, 1991; Furbish et al., 1977; Murray et al., 1985). The regions of the protein responsible for this highly hydrophobic behavior have not been identified but deglycosylation studies suggest it is a product of the polypeptide itself rather than associated glycans (Aerts et al., 1986; Aerts et al., 1988). Glycosylation is, however; crucial to the function of the mature enzyme (Grace and Grabowski, 1990) (Berg-Fussman et al., 1993). The polypeptide contains 5 possible N-glycosylation sites (Asn-X-Ser/Thr) at amino acid positions 19, 59, 146, 270, and 462 of the mature protein (Berg-Fussman et al., 1993). Glycan site occupancy has been studied for the placental human enzyme and while the first four sites are glycosylated *in vivo,* only glycosylation at the first site (Asn 19) is required for recombinant enzyme activity *in vitro* (Berg-Fussman et al., 1993). The placental enzyme is approximately 7% carbohydrate by mass with the predominant species including terminally sialylated tri- and bi-antennary complex N-glycans and a smaller population of simple high mannose N-glycans (Takasaki et al.,

1984). Interestingly, it is the presence of N-glycosylation rather than the composition of the oligosaccharide chains themselves that is crucial, as extensive restructuring of native glycans has little impact on enzyme function (Berg-Fussman et al., 1993; Takasaki et al., 1984). With published molecular weights for the mature human glycoprotein ranging from 62,000-67,000 depending upon the cell type used, it is clear that a great deal of microheterogeneity exists in the N-glycans of glucocerebrosidase (Furbish et al., 1977; Pasmanik-chor et al., 1997) (Choy, 1986; Choy and Woo, 1991; Erickson et al., 1985).

1.5 – Glucocerebrosidase Biosynthesis

The unique requirement of glycosylation for activity of the glucocerebrosidase protein has lead to a thorough investigation of the biosynthesis of this enzyme and roles for glycans in both protein folding and lysosomal targeting have been discussed. Glucocerebrosidase is translated as a 55,000 molecular weight core polypeptide that is cotranslationally transported across the ER and glycosylated by the *en bloc* transfer of four high mannose glycans from their dolichol precursors (Erikson and Wahlberg, 1985). Accordingly, the core polypeptide is only visualized in the presence of tunicamycin, an antibiotic that blocks glycosylation (Aerts et al., 1986). The only proteolytic processing of the protein occurs at this point with the removal of the small leader peptide responsible for ER localization. Again, the 19 amino acid or 39 amino acid leader resulting from either consensus ATG both function to appropriately target the nascent polypeptide (Sorge et al., 1987). The initial cotranslational glycosylation and peptide cleavage gives rise to a 65,000-68,000 molecular weight species which then has its glycans trimmed to core penta- or tri-mannosyl structures (Erikson and Wahlberg, 1985). This 60,000

molecular weight intermediate finally has its glycans converted to complex type bi- and tri-antenary structures on the *cis* face of the Golgi (Erikson and Wahlberg, 1985). The final mature protein now appears as a 62,000-67,000 molecular weight form dependent upon cell type as discussed above. The transport of glucocerebrosidase from the *trans* Golgi complex to the lysosomal membrane, however, can currently be defined only by the mechanisms that are not involved, rather than those that are.

Most lysosomal hydrolases are sorted to the lysosome at the *trans* face of the Golgi through mannose-6-phospate receptor (MPR) mediated mechanisms. High mannose glycans on these proteins are phosphorylated in the *trans* Golgi network and bound by a specific MPR that facilitates segregation of lysosomal proteins from those targeted for secretory vesicles (Glickman and Kornfeld, 1993; Lemansky et al., 1985). The hydrolase/receptor complex is then shuttled to endosomes via clathrin-coated vesicles for final transport to the lumen of the lysosome (Glickman and Kornfeld, 1993; Lemansky et al., 1985). Glucocerebrosidase, however, is unique in that it appears to be transported by a mannose-6-phosphate independent mechanism. Placental and fibroblast glucocerebrosidase contain a low proportion of high mannose glycans that appear not to be phosphorylated (Aerts et al., 1988). More convincingly, an inherited deficiency in phosphotransferase known as I-Cell disease, results in the absence of mannose phosphorylation and thus improper targeting of most lysosomal proteins. Glucocerebrosidase, however, is found at normal levels in the lysosomes of I-Cell patients confirming an alternative mechanism of targeting for this protein (Glickman and Kornfeld, 1993; Lemansky et al., 1985).

Recent studies on other membrane-bound lysosomal proteins have identified two non-MPR mediated mechanisms for lysosomal targeting utilizing specific cytoplasmic tail sequences. One model involves the direct transfer of proteins to the lysosome through a mechanism similar to mannose-6-phosphate receptor localization, while the second mechanism involves internalization in clathrin-coated pits following transfer of the proteins to the cell surface (Guarnier et al., 1993; Mathews et al., 1992). Both of these mechanisms, however, appear to be mediated by specific tyrosine-based or dileucine containing cytoplasmic tails that differentially control protein internalization and lysosomal sorting. Computer analysis of the glucocerebrosidase protein core, however; fails to identify any definitive transmembrane domains outside of the short ER targeting signal that is cleaved following transposition to the lumen of the ER (Sorge et al., 1985). Accordingly, the mature glucocerebrosidase lacks the appropriate cytoplasmic tail required by such lysosomal targeting mechanisms.

As neither high mannose phosphorylation nor a cytoplasmic tail appears to play a role in lysosomal targeting of glucocerebrosidase, there must be another alternative mechanism for the trafficking of this enzyme. The role of the N-glycans themselves have been investigated using inhibitors of glycan formation and while some indirect evidence suggests a role for glycosylation in lysosomal targeting, the protein will still become membrane associated in the absence of glycosylation (Aerts et al., 1986; Aerts et al., 1988; Rijnboutt et al., 1991). Glycosylation is, however, a clear requirement for appropriate folding and stability of the protein as active glucocerebrosidase will not be

produced in the presence of glycosylation inhibitors (Berg-Fussman et al., 1993; Grace and Grabowski, 1990). Glucocerebrosidase has also been shown to interact with several lysosomal membrane proteins (i.e. LAMP-1 and LAMP-2) prior to lysosomal targeting and a glycan-protein or protein-protein association with these other species may be involved in lysosomal sorting at the *trans*-face of the Golgi apparatus (Qi and Grabowski, 2000; Zimmer et al., 1999).

1.6 – Therapeutics

Numerous theories for the treatment of Gaucher disease have been presented since the biochemical defect in glucocerebrosidase activity was first identified. As many of the symptoms associated with this disorder result directly or indirectly from lipid accumulation in circulating macrophages, these cells have been the main targets for therapeutics. Hematopoetic cell repopulation by bone marrow transplantation (BMT) was the first effective treatment available for Gaucher disease and remains a viable option where an appropriate HLA-match donor is available (Ringden et al., 1995). However, the relatively high mortality rate associated with allogenic transplantation from a poor HLA match (20-25%) and an inability to regress skeletal and neurological symptoms has limited the use of BMT particularly with the availability of enzyme replacement therapy (Hoogerbrugge et al., 1995). The accessibility of the target cell population and relative simplicity as a single enzyme deficiency means that intravenous enzyme replacement therapy (ERT) can be an effective option for the treatment of Gaucher disease. Early attempts with unmodified human placental glucocerebrosidase met with minimal success as the administered protein was inefficiently targeted to

macrophages (Barton et al., 1990). An alteration of the glycans present on the placental enzyme to remove terminal sialic acid, galactose and N-acetylglucosamine residues and reveal the inner mannose residues has allowed efficient targeting of glucocerebroside *in vivo* to circulating macrophages and liver cells (Barton et al., 1991; Barton et al., 1990; Mistry et al., 1996). This targeting appears to be mediated by a mannose receptor on the surface of macrophages and liver Kupffer and endothelial cells, which leads to the internalization of the modified glucocerebrosidase and subsequent transport to the lysosome (Grabowski et al., 1998; Willemsen et al., 1995). Although the reversal of visceral symptoms in type 1 Gaucher patients on ERT has been significant, the relatively large dose required for clinical response suggests that the efficiency of targeting for the replacement enzyme remains low (Mistry et al., 1996; Sato and Beutler, 1993).

Regardless of inefficient targeting of the protein, two available commercial enzymes have become the standard for Gaucher disease treatment. *Ceredase*™, glycan modified human placental glucocerebrosidase, and *Cerezyme*™, recombinant glucocerebrosidase produced in Chinese Hamster Ovary (CHO) cells, have both been shown effective in reducing hepatic and splenic volume and reversing hematological complications (Grabowski et al., 1998). Unfortunately, ERT has been less effective in reducing the painful skeletal complications experienced by most Gaucher disease patients (Grabowski et al., 1998). Enzyme replacement therapy has also been shown to be ineffective in significantly affecting the progression of neurodegeneration in type 2 and type 3 Gaucher patients, and is currently approved only on an limited trial basis for those individuals showing neuronopathy (Elstein et al., 1998; Erikson et al., 1993). These limitations in efficacy for

some sub-types and symptoms of Gaucher disease are compounded by the prohibitive cost of development and production of both the placental and recombinant enzymes. Based on dose regimes for an adult patient, annual treatment costs average $150,000 USD per individual (Grabowski et al., 1998). A great deal of discussion has emerged regarding the most affective dose regimes for these treatments, centered on decreasing the amount, and thus the cost of the enzyme utilized (Altarescu et al., 2000; Brady and Barton, 1994; Moscicki and Taunton-Rigby, 1993).

While Gaucher disease has become a model disorder for the investigation of enzyme replacement therapies, it has also been touted as a promising model for the development of gene therapies. As early as 1987, retroviral-mediated transfer of the glucocerebrosidase cDNA into cultured type 1 Gaucher fibroblasts was shown to effectively correct the inherent enzyme deficiency (Sorge et al., 1987). Mononuclear cells isolated from a type 3 Gaucher patient and transduced with the human GBA cDNA were able to produce glucocerebrosidase in culture suggesting the possibility of an *ex vivo* autologous stem cell gene therapy (Nolta et al., 1992). To allow repopulation of the patient's hematopoetic cells however, full bone marrow ablation would theoretically be required, leading to a high risk of mortality and morbidity (Grabowski, 1993). The most promising proposed therapy to date involves the transformation of myoblasts isolated from the patient for long term expression of mannose terminated glucocerebrosidase (Liu et al., 1998a). Initial studies have shown that recombinant myoblasts overexpressing glucocerebrosidase will secrete the high mannose precursor of the glycoprotein and that this expression is maintained over a period of several months *in vitro* (Liu et al., 1998a).

Myoblasts have the potential to divide, differentiate, and associate with myofibres *in vivo,* allowing the glucocerebrosidase producing cells to populate the circulatory system and act as a long-term source of recombinant enzyme (Liu et al., 1998b). While the aforementioned *in vitro* studies appear promising, and two phase-I clinical trials for *ex vivo* hematopoetic stem cell transduction have been initiated (U.S. National Institutes of Health Study ID Numbers 88-N-0019 and 199/11727), unfortunately, the only literature available regarding these studies suggest that only one patient of four displayed initial glucocerebrosidase expression following *ex vivo* therapy, but has not sustained that protein production over the longterm (Mosciki, 2000).

1.7 - **References**

Aerts, J.M., Brul, S., Donker-Koopman, W.E., et al. Efficient routing of glucocerebrosidase to lysosomes requires complex oligosaccharide chain formation. *Biochemical and Biophysical Research Communications* 141:452-458, 1986.

Aerts, J.M., Schram, A.W., Strijland, A., et al. Glucocerebrosidase, a lysosomal enzyme that does not undergo oligosaccharide phosphorylation. *Biochimica et Biophysica Acta* 964:303-308, 1988.

Aerts, J.M.F.G., Brul, S., Donker-Koopman, W.E., et al. Efficient routing of glucocerebrosidase to lysosomes requires complex oligosaccharide chain formation. *Biochemical and Biophysical Research Communications* 141:452-458, 1986.

Aerts, J.M.F.G., Schram, A.W., Strijland, A., et al. Glucocerebrosidase, a lysosomal enzyme that does not undergo oligosaccharide phosphorylation. *Biochimica et Biophysica Acta* 964:303-308, 1988.

Altarescu, G., Schiffmann, R., Parker, C.C., et al. Comparative efficacy of dose regimes in enzyme replacement therapy of type I Gaucher disease. *Blood Cells, Molecules, and Diseases* 26:285-290, 2000.

Balicki, D. and Beutler, E. Gaucher disease. *Medicine* 71:305-323, 1995.

Barton, N.W., Brady, R.O., Dambrosia, J.M., et al. Replacement therapy for inherited enzyme deficiency - Macrophage targetted glucocerebrosidase for Gaucher's Disease. *New England Journal of Medicine* 324:1464-1470, 1991.

Barton, N.W., Furbish, F.S., Murray, G.J., Garfield, M. and Brady, R.O. Therapeutic response to intravenous infusions of glucocerebroisdase in a patient with Gaucher disease. *Proceedings of the National Academy of Sciences, U.S.A* 87:1913-1916, 1990.

Berg-Fussman, A., Grace, M.E., Ioannou, Y. and Grabowski, G.A. Human acid β-glucosidase. *Journal of Biological Chemistry* 268:14861-14866, 1993.

Beutler, E. and Gelbart, T. Hematologically important mutations: Gaucher disease. *Blood Cells, Molecules, and Diseases* 23:2-7, 1997.

Beutler, E., Gelbart, T., Kuhl, W., Zimran, A. and West, C. Mutations in Jewish patients with Gaucher disease. *Blood* 79:1662-1666, 1992.

Beutler, E. and Grabowski, G.A. Gaucher disease. In: *Molecular and Biochemical Disease*, Anonymous1994, p. 2641-2653.

Beutler, E. and Grabowski, G.A. Gaucher disease. In: *The Metabolic and Molecular Bases of Inherited Disease*, edited by Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D. New York: McGraw-Hill, 2001, p. 3645-3668.

Beutler, E., McKusick, V.A., Motulsky, A.G. and Scriver, C.R. Mutation nomenclature: Nicknames, systemic names, and unique identifiers. *Human Mutation* 8:203-206, 1996.

Beutler, E., Nguyen, N.J., Henneberger, M.W., et al. Gaucher disease: gene frequencies in the Ashkenazi Jewish population. *American Journal of Human Genetics* 52:85-88, 1993.

Boas, F.E. Linkage to Gaucher mutations in the Ashkenazi population: Effect of drift on decay of linkage disequilibrium and evidence for heterozygote selection. *Blood Cells, Molecules, and Diseases* 26:348-359, 2000.

Bornstein, P., McKinney, C.E., LaMarca, M.E., et al. Metaxin, a gene contiguous to both thrombospondin 3 and glucocerebrosidase, is required for embryonic development in the mouse: implications for Gaucher disease. *Proceedings of the National Academy of Science, U.S.A.* 92:4547-4551, 1995.

Brady, R.O. and Barton, N.W. Enzyme replacement therapy for Gaucher disease: Critical investigations beyond demonstration of clinical efficacy. *Biochemical Medicine and Metabolic Biology* 52:1-9, 1994.

Chabas, A., Cormand, B., Grinberg, D., et al. Unusual expression of Gaucher's disease: cardiovascular calcifications in three sibs homozygous for the D409H mutation. *Journal of Medical Genetics* 32:740-742, 1995.

Choy, F.Y.M. Purification of human placental glucoceerebrosidase using a two-step high-performance hydrophobic and gel permeation column chromatography method. *Analytical Biochemistry* 156:515-520, 1986.

Choy, F.Y.M. and Woo, M. Purification and the effect of peptide N-glycosidase F on lysosomal membrane-bound glucocerebrosidase from human cultured fibroblasts. *Biochemistry and Cell Biology* 69:551-556, 1991.

Colombo, R. Age estimate of the N370S mutation causing Gaucher disease in Ashkenazi Jews and european populations: A reappraisal of haplotype data. *American Journal of Human Genetics* 66:692-697, 2000.

Daniels, L.B., Glew, R.H., Radin, N.S. and Vunnam, R.R. A revised fluorometric assay for Gaucher's disease using conduritol-β-epoxide with liver as the source of β-glucosidase. *Clinica Chimica Acta* 106:155-163, 1980.

DeMarchi, J.M., Caskey, C.T. and Richards, C.S. Population-specific screening by mutation analysis for diseases frequent in Ashkenazi Jews. *Human Mutation* 8:116-125, 1996.

Diamond, J.M. Jewish lysosomes. *Nature* 368:291-292, 1994.

Diaz, G.A., Gelb, B.D., Risch, N., et al. Gaucher disease: The origins of the Ashkenazi Jewish N370S and 84GG acid β-glucosidase mutations. *American Journal of Human Genetics* 66:1821-1832, 2000.

Dinur, T., Osiecki, K.M., Legler, G., Gatt, S., Desnick, R.J. and Grabowski, G.A. Human acid beta-glucosidase: isolation and amino acid sequence of a peptide containing the catalytic site. *Proceedings of the National Academy of Science, U.S.A.* 83:1660-1664, 1986.

Elstein, D., Abrahamov, A. and Zimran, A. Ethical considerations for enzyme replacement therapy in neuronopathic Gaucher disease. *Clincal Genetics* 54:179-184, 1998.

Eng, C.M., Schechter, C., Robinowitz, J., et al. Prenatal genetic carrier testing using triple disease screening. *Journal of the American Medical Association* 278:1268-1272, 1997.

Erickson, A.H., Ginns, E.I. and Barranger, J.A. Biosynthesis of the Lysosomal Enzyme Glucocerebrosidase. *Journal of Biological Chemistry* 260:14319-14324, 1985.

Erikson, A., Johansson, K., Mansson, J.E. and Svennerholm, L. Enzyme replacement therapy of infantile Gaucher disease. *Neuropediatrics.* 24:237-238, 1993.

Erikson, A. and Wahlberg, I. Gaucher disease--Norrbottnian type. Ocular abnormalities. *Acta Ophthalmology (Copenhagen)* 63:221-225, 1985.

Furbish, F.S., Blair, H.E., Shiloach, J., Pentchev, P.G. and Brady, R.O. Enzyme replacement therapy in Gaucher's disease: large-scale purification of glucocerebrosidase suitable for human administration. *Proceedings of the National Academy of Science, U.S.A.* 74:3560-3563, 1977.

Glew, R.H., Basu, A., LaMarco, K.L. and Prence, E.M. Mammalian glucoerebrosdiase: Implications for Gaucher's Disease. *Journal of Laboratory Investigation* 58:5-25, 1988.

Glickman, J.N. and Kornfeld, S. Mannose 6-phosphate-independent targeting of lysosomal enzymes in I-cell disease B lymphoblasts. *Journal of Cell Biology* 123:99-108, 1993.

Grabowski, G.A. Gaucher disease: Enzymology, genetics, and treatment. In: *Advances in Human Genetics*, edited by Harris, H. and Hirschhorn, K. New York: Plenum Press, 1993, p. 377-439.

Grabowski, G.A., Leslie, N. and Wenstrup, R. Enzyme therapy for Gaucher disease: the first 5 years. *Blood Reviews* 12:115-133, 1998.

Grace, M.E. and Grabowski, G.A. Human acid β-glucosidase: glycosylation is required for catalytic activity. *Biochemical and Biophysical Research Communications* 168:771-777, 1990.

Guarnier, F.G., Arteburn, L.M., Penno, M.B., Cha, Y. and August, J.T. The motif Tyr-X-X-hydrophobic residue mediates lysosomal membrane targeting of lysosome-associated membrane protein 1. *Journal of Biological Chemistry* 268:1941-1946, 1993.

Hoogerbrugge, P.M., Brouwer, O.F., Bordigoni, P., et al. Allogeneic bone marrow transplantation for lysosomal storage diseases. The European Group for Bone Marrow Transplantation. *Lancet* 345:1398-1402, 1995.

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

Koprivica, V., Stone, D.L., Park, J.K., et al. Analysis and classification of 304 mutant alleles in patients with type 1 and type 3 Gaucher disease. *American Journal of Human Genetics* 66:1777-1786, 2000.

Lemansky, P., Gieselmann, V., Hasilik, A. and von, F.K. Synthesis and transport of lysosomal acid phosphatase in normal and I-cell fibroblasts. *Journal of Biological Chemistry* 260:9023-9030, 1985.

Liu, C., Bahnson, A.B., Dunigan, J.T., Watkins, S.C. and Barranger, J.A. Long-term expression and secretion of human glucocerebrosidase by primary murine and human myoblasts and differentiated myotubes. *Journal of Molecular Medicine* 76:773-781, 1998a.

Liu, C., Dunigan, J.T., Watkins, S.C., Bahnson, A.B. and Barranger, J.A. Long-term expression, systemic delivery, and macrophage uptake of recombinant human glucocerebrosidase in mice transplanted with genetically modified primary myoblasts. *Human Gene Therepy* 9:2375-2384, 1998b.

Mathews, P.M., Martinie, J.B. and Fambrough, D.M. The pathway and targeting signal for delivery of the integral membrane glycoprotein LEP100 to lysosomes. *Journal of Cell Biology* 118:1027-1040, 1992.

Meikle, P.J., Hopwood, J.J., Clague, A.E. and Carey, W.F. Prevalence of lysosomal storage disorders. *Journal of the American Medical Association* 281:249-254, 1999.

Mistry, P.K., Wraight, E.P. and Cox, T.M. Therapeutic delivery of proteins to macrophages: implications for treatment of Gaucher's disease. *Lancet* 348:1555-1559, 1996.

Moscicki, R.A. and Taunton-Rigby, A. Treatment of Gaucher's Disease. *New England Journal of Medicine* 328:1564-1568, 1993.

Moscicki, R.A. New approaches in lysosomal disorders – Research and developments Gaucher disease, Fabry Disease, MPS Disease, Pompe Disease. *Proceedings of the Canadian Gaucher Club Meeting* :2-5, 2000.

Motulsky, A.G. Jewish diseases and origins. *Nature Genetics* 9:99-101, 1995.

Murray, G.J., Youle, R.J., Gandy, S.E., Zirzow, G.C. and Barranger, J.A. Purification of beta-glucocerebrosidase by preparative-scale high-performance liquid chromatography: the use of ethylene glycol-containing buffers for chromatography of hydrophobic glycoprotein enzymes. *Analytical Biochemistry* 147:301-310, 1985.

Nolta, J.A., Yu, X.J., Bahner, I. and Kohn, D.B. Retroviral-mediated transfer of the human glucocerebrosidase gene into cultured Gaucher bone marrow. *Journal of Clinical Investigation* 90:342-348, 1992.

Pasmanik-chor, M., Elroy-Stein, O., Aerts, H., Agmon, V., Gatt, S. and Horowitz, M. Overexpression of human glucocerebrosidase containing different-sized leaders. *Biochemical Journal* 317:81-88, 1996.

Pasmanik-chor, M., Madar-shapiro, L., Stein, O.E., Aerts, H., Gatt, S. and Horowitz, M. Expression of mutated glucocerebrosidase alleles in human cells. *Human Molecular Genetics* 6:887-895, 1997.

Peleg, L., Frisch, A., Goldman, B., et al. Lower frequency of Gaucher disease carriers among Tay-Sachs disease carriers. *European Journal of Human Genetics* 6:185-186, 1998.

Qi, X. and Grabowski, G.A. Acid beta-glucosidase: intrinsic fluorescence and conformational changes induced by phospholipids and saposin C. *Biochemistry* 37:11544-11554, 1998.

Qi, X. and Grabowski, G.A. Molecular and cell biology of acid beta-glucosidase and prosaposin. *Progress in Nucleic Acid Research and Molecular Biology* 66:203-39.:203-239, 2000.

Rijnboutt, S., Aerts, H., Gueze, H.J., Tager, J.M. and Strous, G.J. Mannose 6-phosphate-indepentent menbrane association of cathespin D, glucoerebrosidase, and sphingolipid-activation protein in HepG2 cells. *Journal of Biological Chemistry* 266:4862-4868, 1991.

Ringden, O., Groth, C.G., Erikson, A., Granqvist, S., Mansson, J.E. and Sparrelid, E. Ten years' experience of bone marrow transplantation for Gaucher disease. *Transplantation* 59:864-870, 1995.

Sato, Y. and Beutler, E. Binding, internalization, and degradation of mannose-terminated glucocerebrosidase by macrophages. *Journal of Clinical Investigation* 91:1909-1917, 1993.

Sidransky, E., Sherer, D.M. and Ginns, E.I. Gaucher disease in the neonate: A distinct Gaucher phenotype is analogous to a mouse model created by targeted disruption of the glucocerebrosidase gene. *Pediatric Research* 32:494-498, 1992.

Sorge, J., West, C., Westwood, B. and Beutler, E. Molecular cloning and nucleotide sequence of human glucocerebrosidase cDNA. *Proceedings of the National Academy of Science, U.S.A.* 82:7289-7293, 1985.

Sorge, J.A., West, C., Kuhl, W., Treger, L. and Beutler, E. The human glucocerebrosidase gene has two functional ATG initiator codons. *American Journal of Human Genetics* 41:1016-1024, 1987.

Strasberg, P.M., Skomorowski, M.A., Warren, I.B., Hilson, W.L., Callahan, J.W. and Clarke J.T.R. Homozygous presence of the crossover (fusion gene) mutation identified in a type II Gaucher disease fetus: Is this analogous to the Gaucher knock-out mouse model? *Biochemical Medicine and Metabolic Biology* 53:16-21, 1994.

Takasaki, S., Murray, G.J., Furbish, S., Brady, R.O., Barranger, J.A. and Kobata, A. Structure of the N-asparagine-linked oligosaccharide units of human placental β-glucocerebrosidase. *Journal of Biological Chemistry* 259:10112-10117, 1984.

Tsuji, S., Choudary, P.V., Martin, B.M., Winfield, S., Barranger, J.A. and Ginns, E.I. Nucleotide sequence of cDNA containing the complete coding sequence for human lysosomal glucocerebrosidase. *Journal of Biological Chemistry* 261:50-53, 1986.

Willemsen, R., Tibbe, J.J., Kroos, M.A., Martin, B.M., Reuser, A.J. and Ginns, E.I. A biochemical and immunocytochemical study on the targeting of alglucerase in murine liver. *Histochemical Journal* 27:639-646, 1995.

Winfield, S.L., Tayebi, N., Martin, B.M., Ginns, E.I. and Sidransky, E. Identification of three additional genes contiguous to the glucocerebrosidase locus on chromosome 1q21: implications for Gaucher disease. *Genome Research* 7:1020-1026, 1997.

Zimmer, K., le Coutre, P., Aerts, J.M.F.G., et al. Intracellular transport of acid β-glucosidase and lysosome-associated membrame proteins is affected in Gaucher's disease (G202R mutation). *Journal of Pathology* 188:407-414, 1999.

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

Chapter 2 – Neuronopathic Gaucher Disease Mutation Analysis

2.1 – Introduction

The increased incidence of Gaucher disease in the Ashkenazi Jewish population has lead to extensive analysis of the mutations associated with this group and with type 1 Gaucher disease in general. The analysis of Gaucher mutations outside of this ethnic group and clinical subtype has, in turn, remained comparatively limited due to the rarity of type 2 and type 3 disease. The mutational analysis of neuronopathic patients is further confounded by the tendency toward premature death of these patients due to the severity of disease, and the panethnic nature of these subtypes limiting population based analyses, aside from the notable Norrbottnian type 3a cluster in Sweden (Erikson and Wahlberg, 1985). Although the mutations currently identified from type 2 and 3 Gaucher patients have generally been characterized as severe or null alleles, there still exists a great deal of both molecular and clinical heterogeneity within these neuronopathic subtypes (Koprivica et al., 2000). Importantly, the presence of an allele associated with neuronopathy in one patient is not necessarily predictive of neurodegeneration in another (Koprivica et al., 2000). In fact, the biochemical basis of neurodegeneration in Gaucher disease remains unclear (Orvisky et al., 2000). Adding further complexity is the fact that the majority of mutations identified in neuronopathic patients remain private or rare alleles present as heterozygotes in only a small number of individuals, usually in the context of a second poorly characterized allele (Koprivica et al., 2000).

The study of larger, genetically homogeneous populations has been central to the prediction of disease severity for mutations identified in type 1 Ashkenazi Jews and Type

3a Norrbottnian patients. As a result of large-scale screening of Jewish populations, researchers have been able to correlate the common N370S mutation exclusively with type 1 Gaucher disease (Beutler et al., 1996; Germain et al., 1998). In all compound heterozygotes identified to date, the presence of the N370S mutation has been protective against the onset of neurogeneration and retains a relatively high degree of residual enzyme activity *in vitro* (Choy et al., 1996; Grace et al., 1990). Norrbottnian type 3a Gaucher disease represents another well characterized disease cluster due to the prevalence of the L444P mutation in an isolated community in Northern Sweden (Erikson and Wahlberg, 1985). In this group, it has been shown that L444P homozygosity will invariably lead to type 3 Gaucher disease, although there still exists a great deal of clinical heterogeneity with regards to the overall severity of symptoms seen (Dahl et al., 1990). The analysis of the L444P mutation in the general population has been much less definitive due to the occurrence of this mutation in the context of a series of mutations resulting from interaction between the glucocerebrosidase gene and pseudogene. These "complex alleles" tend to include a number of deleterious mutations corresponding to the sequence of the pseudogene in exons 9 and 10 and have often been misidentified as simple L444P point mutations in large scale screening projects (Koprivica et al., 2000). Accordingly, the literature has incorrectly attributed much of the clinical heterogeneity associated with this allele to poor genotype/phenotype correlation rather than the heterogeneous genetic context of the L444P mutation.

One of the major contraints to the investigation of non-Ashkenazi neurodegenerative sub-types has been a difficulty in obtaining the appropriate whole blood samples or fibroblast

cell lines necessary for a typical molecular study. The clinical characteristics of Gaucher disease, however, present a unique opportunity to utilize paraffin-embedded, formalin-fixed archival tissue samples as a source of genomic DNA. While classical tissue fixation techniques crosslink proteins, they have been found to leave DNA in a more functional state. DNA has been extracted from fixed tissues up to forty years in age and although somewhat fragmented, been found sufficient for PCR and Southern blotting analysis (Coleman et al., 1991). Gaucher disease has classically been diagnosed through a histological analysis of liver, spleen and bone marrow tissue to identify characteristic lipid engorged Gaucher cells (Beutler and Grabowski, 2001). Accordingly, there exists a repository of untapped genetic information from Gaucher patient samples stored in hospital pathology departments. The development of a procedure for the isolation of DNA from Gaucher archival tissues will expand the number of cases available for study, not only spacially, but also temporally as long as appropriate clinical records are available.

This chapter reports a study designed to address the current limitations in the genotype / phenotype correlation of neuronopathic Gaucher mutations through three related avenues. First, DNA available from the whole blood and fibroblast cell cultures of Gaucher patients was analysed to add to the current catalog of mutations and a novel point mutation and a novel complex allele associated with neuronopathy were identified. Second, a technique for DNA isolation from formalin-fixed archival tissue was used to identify a rare allele associated with perinatal type 2 Gaucher disease and to characterize an individual homozygous for a second rare type 3 mutation. As this mutation has only

been identified in individuals of Cree descent in northern British Columbia and Alberta, family members of the affected individuals were screened for the presence of this rare allele. Third, the nucleotide substitution identified from the archival analysis of this aboriginal allele was introduced into normal copies of the glucocerebrosidase cDNA and expressed heterologously to confirm the biochemical defect and assess residual enzyme activity levels.

2.2 - Materials and Methods

2.2.1- *Case Reports*

Patient 1043, an African-American, was diagnosed in 1976 as having type 2 Gaucher disease at the Children's Hospital in Pittsburgh, Pennsylvania. She was noted to have splenomegaly and failure to thrive. Glucocerebrosidase activity was found to be 8.9% of normal. She died at the age of 18 months.

JB was the third child born to a 23 year old First Nations mother of borderline intelligence. There was little prenatal care, and dates were uncertain. Hepatosplenomegaly was noticed at 13 months. At this stage he was obviously developmentally delayed and was rolling over, but not sitting, with moderate hypotonia. Bone marrow and liver biopsies suggested a diagnosis of Gaucher disease, which was confirmed by low leukocyte β-glucosidase activity (2.1 U/hr/mg protein, normal 5.5-12.5 U/hr/mg protein). X-rays of femurs showed characteristic "Erlenmeyer flask" deformities.

Treatment with *Ceredase*™ was initiated at 15 months, using 400 units every 2 weeks (103 U/kg/mo.). This dose was continued over the next 15 months, with considerable clinical benefit. Liver and spleen sizes declined by both clinical and objective measurement, and there was marked improvement in hematological parameters. Decreasing the dose to 150 Units every 2 weeks (26 U/Kg/mo.) over the next 8 months was inadequate to maintain the improvement, so it was raised to 400 Units every week (100 U/Kg/mo.) This dose was maintained from age 5 to 8, over which time the disease remained clinically stable. *Cerezyme*™ was substituted for *Ceredase*™ at age 6. Subsequently the dose has been increased to 800 Units weekly (135 U/Kg/mo.) because of evidence that he was developing neurological features consistent with Type 3 Gaucher disease. JB is globally developmentally delayed, without obvious regression. From age 9, oculomotor apraxia, dysarthria, and tremor have been noted. These have not so far improved on the higher dose of *Cerezyme*™.

Patient MS, a second cousin to JB, first presented at 2 years of age when hepatosplenomegaly was noted while she was being treated for impetigo. The disease was fairly rapidly progressive, with development of massive abdominal organomegaly, severe anemia and thrombocytopenia, a pathological fracture of the femur, and eventual demise at 7 years. She was also developmentally delayed, though the extent was never formally evaluated. MS received no enzyme replacement therapy and was suffering from massive hepatosplenomegaly and severe anemia (Hgb 5.5g%) when she died from these complications of untreated Gaucher disease.

Patient BW was born to parents of non-Jewish caucasian descent with hydrops fetalis and severe ichthyosis and died 4 hours after birth. Autopsy revealed severe hepatosplenomegaly and a phenotype consistent with perinatal type 2 Gaucher disease.

2.2.2 - *Cell Culture and Fibroblast DNA Extraction*

Fibroblasts from patients 1043 and JB were cultured in Eagle minimum essential medium supplemented with 10% fetal bovine serum and harvested as previously described (Choy 1994). Genomic DNA and poly-A mRNAs were isolated from harvested fibroblasts using the DNAzol™ (Life Technologies, Bethesda, MD), and Micro-FastTrack mRNA Isolation Kits [Invitrogen Corporation, San Diego, CA) as previously described (Choy et al., 1997). The cDNA of the glucocerebrosidase gene was synthesized by reverse transcription of fibroblast poly-A mRNA using primer A in the anti-sense orientation to glucocerebrosidase genomic DNA nucleotide positions 6154-6135 (Table 2.1) and the Superscript™ Preamplification System (Life Technologies, Bethesda, MD). The cDNA containing the entire coding region of the glucocerebrosidase gene was then amplified by PCR using primers A and B (Table 2.1) as previously described (Choy et al., 1994).

2.2.3 - *PCR Amplification and Sequence Analysis of Fibroblast DNA*

For genomic DNA samples, a PCR method (Choy et al., 1997) was used to selectively amplify the glucocerebrosidase functional gene, but not the pseudogene that shares more than 96% sequence similarity with the functional gene (Horowitz et al., 1989). In brief, primers C and D were used to amplify exons 1 to 3, primers E and F for exons 4 to 6, and primers G and H for exons 7 to 11 (Table 2.1). In this first stage PCR, all sense primers (C, E, G) were specific to the functional gene, while all antisense primers (D, F, H) were

Table 2.1 – Primers for the PCR amplification of glucocerebrosidase genomic DNA and cDNA

Primer[c]	Nucleotide sequences(5'>3')	Nucleotide position	Orientation
A	CTTTAATGCCCAGGCTGAGC	Genomic nt[a] 6154-6135	Anti-sense
B	CGGAATTACTTGCAGGGCTA	cDNA nt[b] minus 137-118	Sense
C	CGGAATTACTTGCAGGGCTA	Genomic nt minus137-118	Sense
D	TGCATAGGTGTAGGTGCGGA	Genomic nt 2526-2507	Anti-sense
E	GCTGGGTACTGATACCCTTA	Genomic nt 1312 -1331	Sense
F	CAACTGTGGGATCCATGGCA	Genomic nt 3915-3896	Anti-sense
G	GCCATCTTCTACTCACTGTAA	Genomic nt 3248-3267	Sense
H	CTTTAATGCCCAGGCTGAGC	Genomic nt 6154-6135	Anti-sense
I	ACTTTGTCGACAGTCCCATC	Genomic nt 5336-5355	Sense
J	CATGGAGAGGTCATCTCAGTT	Metaxin nt 2701-2721[d]	Sense

[a]The nucleotide position of glucocerebrosidase genomic DNA is numbered from the upstream initiator ATG (where A is position no. 1) according to Beutler et al. (1996).

[b]The nucleotide position of glucocerebrosidase cDNA is numbered according to Sorge et al. (1987) where A of the upstream ATG initiation codon is position no. 1.

[c]Primer A was used for the first strand cDNA synthesis and with primer B in the PCR amplification of glucocerebrosidase cDNA. Primers C through H were used for the PCR amplification of the glucocerebrosidase genomic DNA. Primers I and J were used to amplify the complex mutant allele in patient 1043.

[d]The genomic nucleotide position is numbered according to Long et al. (Long et al., 1996) EMBL/Genbank Data Libraries Accession No. U46920.

non-specific and capable of annealing to both the functional and pseudogene sequences. By utilizing a non-specific anti-sense primer in each reaction most complex alleles resulting from gene to pseudogene recombination would be amplified and detected. A nested PCR method was used in the amplification of smaller fragments for RFLP analysis. Primer J is sense to exon 2 of the metaxin gene. The reverse orientation of this gene to the glucocerebrosidase genes allowed primer J to be utilized as an antisense primer for glucocerebrosidase amplification (Long et al., 1996). Genomic DNA fragments from the first stage PCR were subjected to the dideoxynucleotide chain termination sequencing method of Sanger *et al.* (Sanger et al., 1977) using the fmol DNA Sequencing System (Promega Corporation, Madison, WI).

2.2.4 – *PCR Amplification and Sequence Analysis of Archival Tissue DNA*

Formalin-fixed bone marrow aspirate slides obtained from patient MS and paraffin-embedded liver tissue samples obtained from patient BW were irradiated under long wave UV light for 10 minutes to control for surface exogenous DNA contamination. A thin scraping of the embedded tissue from patient BW was first incubated in 100% xylene for 15 minutes to deparaffinize the sample followed by removal of the xylene and a short incubation at 58°C for 5 minutes to evolve the residual solvent. The deparaffinized liver and bone marrow aspirate samples were scraped into 1.5ml microfuge tubes containing 100ul of a proteinase K digestion buffer (1M Tris-Cl pH 8.0, 100mM MgCl2, 100ug/ml Gelatin, 0.45% Nonidet P-40, 0.45% Tween 20, and 60mg/ml Proteinase K) and incubated overnight at 55°C (Coleman et al., 1991).

The mixture was heated to 95°C for 5 minutes to denature the proteinase and centrifuged at 14,000x*g* for 5 minutes to pellet the cell debris. The supernatant was diluted 1:4 and a 10ul aliquot used directly as template for PCR amplification. The nucleotide sequences of the primers used in the PCR amplification of archival DNA are shown in Table 2.1. PCR products from patient MS were cloned directly into the pCRII vector following the TA Cloning Kit procedure (Invitrogen Corporation, San Diego, Ca.) for sequence analysis. Insert bearing clones were identified and plasmids isolated as directed using a Wizard Miniprep Kit (Promega Corporation, Madison, WI)

2.2.5 - *RFLP Analysis for Gaucher Mutations*

The five most common Jewish mutations N370S, L444P, 84insG, IVS2+1, and V394L were screened using the restriction enzymes *Xho*I, *Nci*I, *BsaB*I, *Hph*I, and *Ban*I respectively (Beutler et al., 1991; Beutler et al., 1992; Beutler et al., 1990; Theophilus et al., 1989; Tsuji et al., 1987). A mismatch PCR method using a mutagenized sense primer and an antisense primer (M and N respectively, Table 2.2) were used to create a *BsrD*I cleavage site to confirm the novel E41K mutation in patient 1043. The identity of the L444P complex allele fragment amplified by primers I and J was confirmed by *SnaB*I restriction digest. A nonrecombinant product, which could result from spurious annealing of primer J to the metaxin pseudogene rather than the functional gene, would be digested by this enzyme while a true crossover would not. *Kpn*I RFLP analysis was used to confirm the presence of mutation P122S in JB and MS and for mutation screening of suspected heterozygote family members. After amplification by the PCR using primers O and P for JB and Q and R for MS (Table 2.2), glucocerebrosidase genomic DNA or cDNA was digested with *Kpn*I endonuclease.

Table 2.2 –Primers for RFLP analysis and mutant cDNA synthesis

Primer[b]	Nucleotide sequences (5'>3')[a]	Nucleotide position	Orientation
K	GATGTGTCCATTCTCCATGT	Genomic 974-993	Sense
L	CCAAGGGCAGGAAAGGACGG	Genomic 1140-1121	Anti-sense
M	TTGGTACCTTCAGCCGCAAT	Genomic nt 1139-1158	Sense
N	TAAGGGTATCAGTACCCAGC	Genomic nt 1331-1312	Anti-sense
O	CGGAATTACTTGCAGGGCTA	cDNA nt minus137-118	Sense
P	TGGGTGACAGAGAGAGAGACT	Genomic nt 3229-3209	Anti-Sense
Q	GCAGAGTCCCATACTCTCCT	Genomic nt 2395-2414	Sense
R	CTCAACCCCCAGACATCAGG	Genomic nt 2642-2623	Anti-sense
S	**GCTGAATTCC**TTCATCTAATG ACCCTGA	cDNA nt minus 23-5	Sense
T	CTGGCCATGGATACCCGGAT	cDNA nt 491-469	Anti-sense
U	ATCCGGGTATCCATGGCCAG	cDNA nt 469-491	Sense
V	**GCTGAATTC**TTTAATGCCCAG GCTG	cDNA nt1658-1642	Anti-sense

[a]Mutagenic mismatched bases are underlined and restriction site linker regions presented in bold print.

[b]Primers K and L were used for RFLP analysis of the del203C mutation. Primers M and N were used for the confirmation of the E41K mutation. Primers O and P were used for P122S confirmation in JB while Q and R were used for MS. Primers L through O were used for the construction of the P122S mutant cDNA.

The presence of a cDNA nucleotide 203 deletion mutation (del203C) in patient BW was confirmed by *Ahd*I restriction digest utilizing mutagenic antisense primer L (Table 2.2). By amplification of BW and control DNA with primers K and L an *Adh*I resriction cut site was introduced in the normal sequence but would remain absent from sequences with the del203C mutation. Following restriction digest, all DNA fragments were analysed by agarose gel electrophoresis. Restriction enzymes were obtained from New England Biolabs (Beverly, MA).

2.2.6 - *Baculovirus Expression*

Recombinant baculovirus containing the glucocerebrosidase cDNA corresponding to the well characterized N370S and L444P mutants were constructed as described previously (Choy et al., 1996). A normal copy of the human glucocerebrosidase cDNA was mutated by overlap extension PCR to introduce the cDNA nucleotide T481C transition that results in P122S. Pfu polymerase (Stratagene, La Jolla, CA) was utilized for the PCR reactions to maximize the fidelity of the amplification reaction. The mutagenesis technique adapted from Ho et al. (1989) and Vallejo et al. (1994) is presented in Figure 2.1. The primers used for the site-directed mutagenesis are presented in Table 2.2. The P122S and normal copies of the cDNA were then cloned into the *EcoR*I sit of the pFastBac1 plasmid (Life Technologies, Bethesda, MD), checked for insert orientation by restriction digest, and the full length cDNA sequenced to ensure the presence of the intended genotype and absence of cloning error. Recombinant bacmids were created and purified using mini Tn7-mediated transposition as described in the Bac-to-Bac^{TM} baculovirus expression manual (Life Technologies, Bethesda, MD).

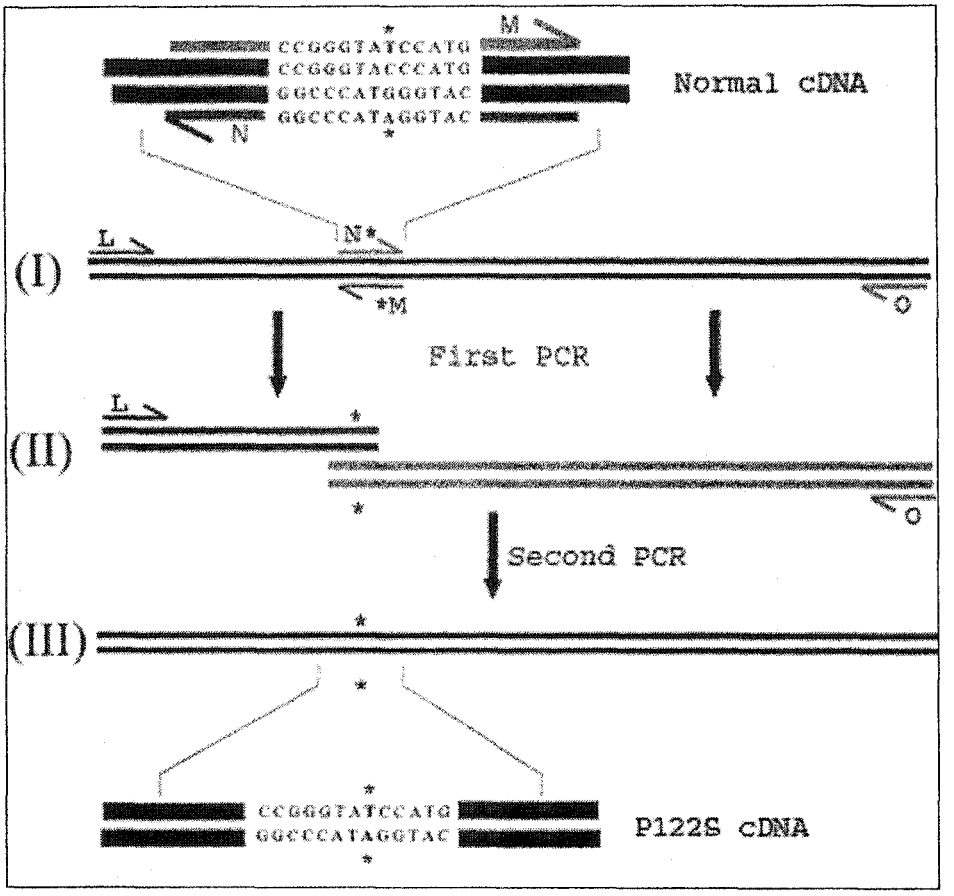

Figure 2.1 - Overlap extension mutagenesis of the GBA gene to create the P122S mutant construct. Primers M and N introduced the CG to TA transition by including the appropriate basepair mismatchs at the base positions denoted by * in the diagram. Initial PCR of the normal GBA cDNA (I) with primers L and M in separate reactions from primers N and O introduced the base changes and amplified the 5' and 3' fragments (II) of the cDNA. A second round of PCR with primers L and O using the two cDNA fragments as template completed the full-length cDNA (III) containing the P122S mutation.

Recombinant bacmids were then transfected into Sf9 cells using CellFectin™ (Life Technologies, Bethesda, MD) lipofection and recombinant baculovirus isolated for expression studies. *Sf*9 cells were cultured in SF900II SFM serum free medium (Life Technologies, Bethesda, MD) at 30°C. Baculovirus titers were determined by neutral red plaque assay as described by O'Reilly and Miller (1994).

2.2.7 - *Western Blotting and Activity Assay*

*Sf*9 cells infected with recombinant baculovirus at a multiplicity of infection of 5 (MOI =5) were harvested at 3-5 days post infection for glucocerebrosidase activity assay. Crude lysates were prepared by repeated freeze/thaw cycles and total protein concentrations determined using Biorad reagent (Biorad Laboratories, Hercules, CA). After standardizing for total protein, 4-methyl umbeliferyl β-D-glucopyranoside (4MUGP) activity assay was performed as described (Choy, 1984). Cell lysates were subjected to tris-glycine SDS-PAGE electrophoresis and electroblotted to Hybond-P PVDF membrane (Amersham Pharmacia Biotech, Inc., Piscataway, NJ) for western blotting (Spector et al., 1998). Human glucocerebrosidase specific mouse monoclonal antibody obtained from Dr. E. Beutler was used in conjunction with HRP conjugated rabbit anti-mouse secondary (Clontech Laboratories, Inc., Palo Alto, CA) for western Blotting as described (Spector et al., 1998). Cross-reactivity was visualized with ECL Plus chemiluminescence (Amersham Pharmacia Biotech, Inc., Piscataway, NJ). Negative control blots were performed using an anti-chicken fibronectin mouse monoclonal antibody generously donated by Dr. R. Burke.

2.3 - **Results**

In the initial RFLP screening of PCR-amplified glucocerebrosidase genomic DNA, patient 1043 was found positive for the L444P mutation (results not shown). Sequence analysis of patient 1043 exon 10 showed base pair substitutions of T>C at cDNA nt 1448, G>C at cDNA nt 1483, and G>C at cDNA nt 1497 suggesting that this L444P mutation was caused by a crossover between the glucocerebrosidase functional gene and pseudogene (data not shown) (Sinclair et al., 1998). Analysis of intron 9 revealed the presence of base pair changes corresponding to those of the pseudogene starting at genomic DNA nucleotide position 5723 and continuing into exon 10 (Figure 2.2). Only functional gene sequences were found upstream at the nearest gene/pseudogene basepair difference at genomic DNA nucleotide position 5689 (Figure 2.2). This suggested that the recombination event took place in the stretch of DNA between these two sites. Previously reported L444P complex alleles all crossed-over upstream of genomic DNA nt 5689 (Eyal et al., 1990; Hong et al., 1990; Latham et al., 1990; Zimran et al., 1990). Using primer I (deleted in the glucocerebrosidase pseudogene) and primer J (present in the metaxin gene but mismatched with the pseudogene), we were able to amplify a DNA fragment of approximately 4.6 Kb in patient 1043, but not in the normal control (Figure 2.2). Sequence analysis of the 5' and 3' ends of this fragment confirmed the presence of a crossover allele that does not cross back by exon 2 of the metaxin functional gene (data not shown). RFLP analysis confirmed these results using a *SnaB*I digest that could not cut such a complex allele. Any false positive PCR products resulting from mispriming of the metaxin functional gene specific 3' primer on the metaxin pseudogene would be digested (data not shown).

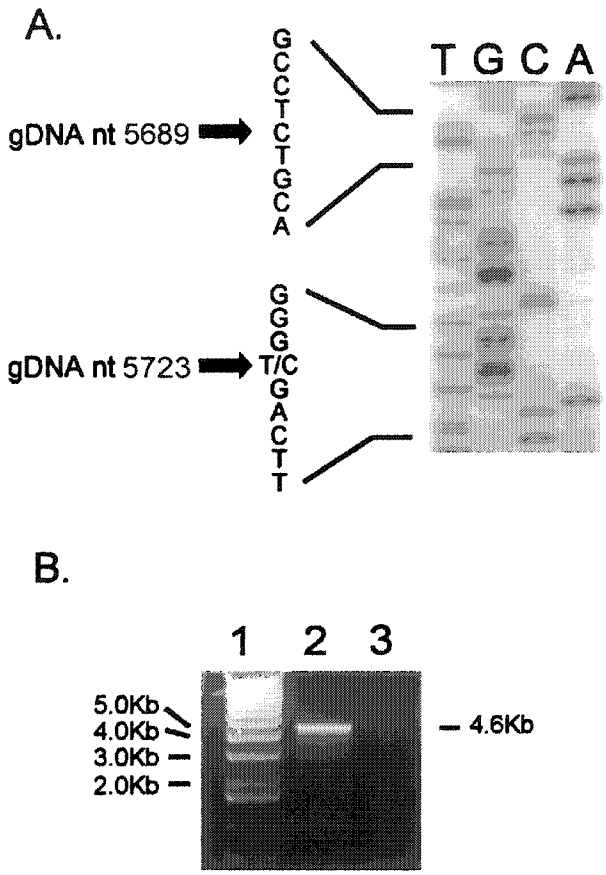

Figure 2.2 - Identification and delineation of the crossover mutant allele in patient 1043. **(A)** Direct sequence analysis of intron 9 from genomic DNA nt 5682 to 5733 showing the site of crossover. Position nt 5689 differs between the glucocerebrosidase gene and pseudogene but only the C corresponding to the functional sequence is seen. The next known sequence difference at nt 5723 shows both a T (function gene) and a C (pseudogene) defining the 5' point of crossover in the region of sequence identity between positions nt 5689 and 5723. **(B)** PCR amplification of genomic DNA from Gaucher patient 1043 and a normal control with primers I and J (see Table 2.1). These primers amplify from the glucocerebrosidase functional gene to the metaxin functional gene, normally a span of 32 Kb, but shortened to only 4.6 Kb in the event of a single crossover within the glucocerebrosidase gene. The 4.6 Kb band is present from patient 1043 (Lane 2) but not in the normal control (Lane 3). The 32 Kb band is not present in either lane as it was beyond the range of the PCR protocol used. A 1 Kb DNA standard ladder (Gibco BRL, Bethesda, MD) was included for reference (Lane 1).

Sequence analysis located the mutation in the second allele of patient 1043. It is a novel point mutation at cDNA nt 238 (genomic DNA nt 1159) (Figure 2.3). This mutation is a G to A transition resulting in the substitution of glutamic acid by lysine at amino acid 41 (E41K). This mutation was confirmed by using the mismatched primer method and *BsrD*I restriction endonuclease analysis (Figure 2.3). Further sequencing of the remaining exons of the glucocerebrosidase gene in patient 1043 revealed no other changes from the expected sequence.

The results of sequence analysis of glucocerebrosidase exon 5 from patient JB are shown in Figure 2.4. The presence of a homozygous C>T transition at cDNA nucleotide 481 that results in a proline to serine amino acid substitution (P122S) was detected (Figure 2.4) (Sinclair et al., 2001). Sequence analysis of the rest of the glucocerebrosidase full length cDNA revealed no further changes. As reported previously (Beutler et al., 1993), a cleavage site for the endonuclease *Kpn*I is abolished by this mutation.

Accordingly, *Kpn*I RFLP analysis of PCR amplified glucocerebrosidase genomic DNA or cDNA was used to confirm the presence of this substitution in the homozygous form (Figure 2.5). This finding also confirms the results of our sequence analysis that both glucocerebrosidase alleles contain the T481C (P122S) mutation. The possibility of a complete deletion of one glucocerebrosidase allele (and heterozygosity for P122S in the other) in JB cannot, however; be ruled out by this analysis alone.

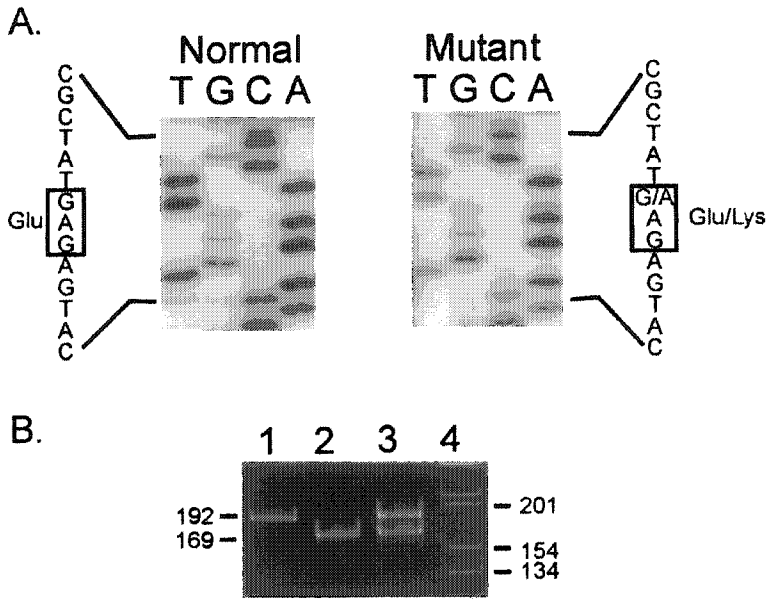

Figure 2.3 - DNA sequence analysis and restriction digest of PCR-amplified genomic DNA of glucocerebrosidase exon 3 from Gaucher patient 1043. **(A)** Direct sequence analysis detected a G to A transition that results in the mutation E41K in patient 1043 as compared to a normal control. **(B)** *BsrD*I restriction digest of a 192 bp exon 3 fragment which is normally digested into 169 bp and 23 bp (not shown) fragments. An undigested control (Lane 1) shows the intact band that is completely digested in a normal control (Lane 2) in the presence of *BsrD*I. The cDNA nt 238 transition abolishes this cut site in one allele of patient 1043 leading to the retention of the undigested 169 bp fragment, along with the digestion products from the second allele (Lane 3). A 1 Kb DNA standard ladder (Gibco BRL, Bethesda, MD) was included for reference (Lane 4).

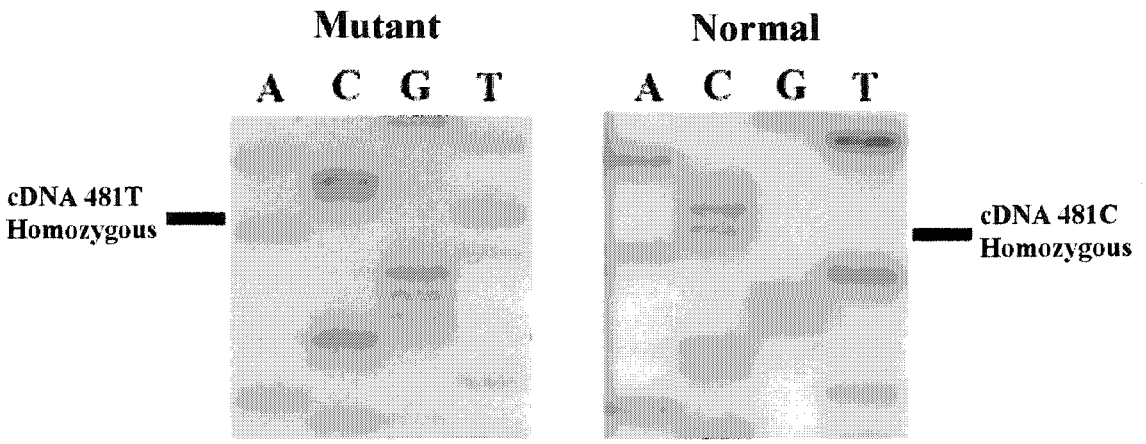

Figure 2.4 - Sequence analysis of PCR-amplified DNA from glucocerebrosidase exon 5 of JB using the chain termination method. The four lanes to the left are the A, C, G, and T lanes of patient JB (Mutant). The four lanes to the right are those from an unaffected individual (Normal). As shown, there is a T>C missense mutation in he homozygous form at cDNA nucleotide position 481(P122S) in patient JB.

The molecular analysis of patients MS was complicated by the fragmented nature of the genomic DNA template extracted from bone marrow aspirate samples and archival tissues in general (Shitaba et al., 1988). The functional gene specific primers utilized for patient JB (see Table 2.2), which amplify a product of 1809bp, were spaced too far apart for efficient amplification of the fragmented template from MS. Successful amplification was only achieved with a smaller PCR product from MS necessitating the use of non-specific primers and the amplification of both functional and pseudogene sequences (Table 2.2). The inclusion of both species in the amplified fragment limited the function of *Kpn*1 RFLP analysis in this case to confirmation of the P122S mutation in only the heterozygous state, as the pseudogene sequences are digested (data not shown). This required the inclusion of a cloning step to isolate functional gene sequences from those of the pseudogene, based on a pseudogene sequence difference 6bp upstream of the P122S mutation site. To rule out cloning and stochastic error, 20 clones were isolated and sequenced, revealing 12 pseudogene sequences normal at cDNA nt 481 (amino acid 122) and 8 functional sequences all with the T481C transition. Based on an equal probability of cloning each of the four possible alleles (two functional and two pseudogene), this confirms with 96% confidence that patient MS was homozygous for the T481C mutation (Sinclair et al., 2001).

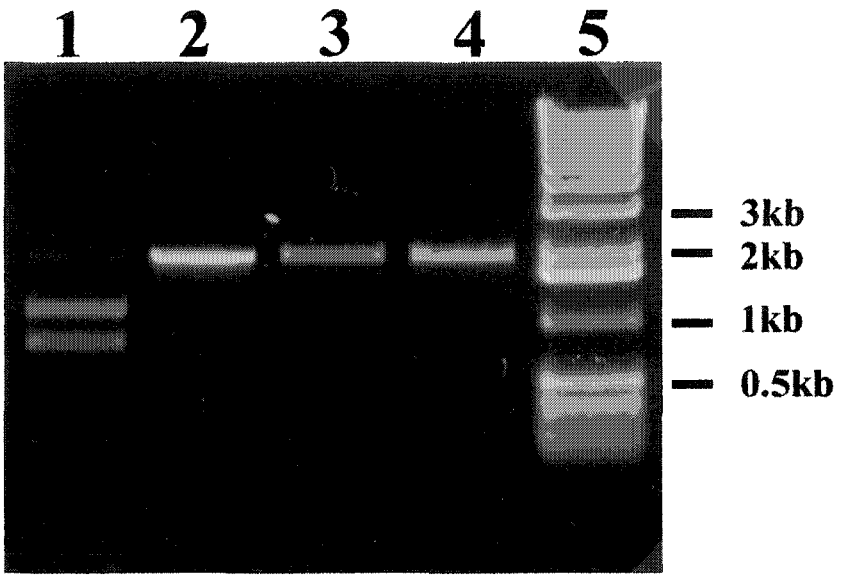

Figure 2.5 - *Kpn*I restriction analysis of mutation P122S. Genomic DNA of glucocerebrosidase exon 6 was amplified by PCR, digested with *Kpn*I endonuclease and analyzed by agarose gel electrophoresis as described in the text. From left to right: Lane 1, normal control negative for mutation P122S digested with *Kpn*I; Lane 2, normal control negative for mutation P122S before digestion; Lane 3, DNA from patient JB in the presence of *Kpn*I endonuclease; Lane 4, DNA from Gaucher patient JB before *Kpn*I addition; Lane 5, DNA standards, 1Kb DNA ladder (GIBO-BRL, Bethesda, MD). The presence of the P122S mutation abolishes the *Kpn*I cut site in patient JB (lane 3).

Mutation screening for P122S was conducted for members of this extended family. Both parents of patient MS were confirmed heterozygotes for the T481C mutation by *Kpn*1 RFLP screening. While the mother of JB was also confirmed as heterozygous, ambiguous paternity in this case hindered analysis of the father. A candidate was screened by RFLP but was homozygous normal at this site, leaving open the possibility of a gene deletion allele in JB. The relatively frequent occurrence of the P122S allele suggests the effects of endogamy within this small isolated community, and argues towards a conclusion of P122S homozygosity for JB. The only previous description of this mutation was in a First Nations P122S homozygote of Cree origin from northern B.C. (Beutler et al., 1993), not knowingly related, also presenting with severe visceral Gaucher symptoms which initially responded well to *Ceredase*^{TM} therapy but progressed to the development of significant neurodegenerative symptoms (Chan et al., 1994).

The P122S mutation was introduced into a normal copy of the glucocerebrosidase cDNA by overlap extension mutagenesis for heterologous expression via a recombinant baculovirus vector. Using a normal glucocerebrosidase clone it was found that an MOI=5 and a 5 day infection of *Sf*9 cells at an initial density of $1x10^5$ cells/ml yielded maximal 4MUGP activity (data not shown). Figure 2.6 shows a western blot of *Sf*9 cell lysates 5 days post infection for the P122S clone along with normal, N370S, L444P, and non-recombinant baculovirus controls.

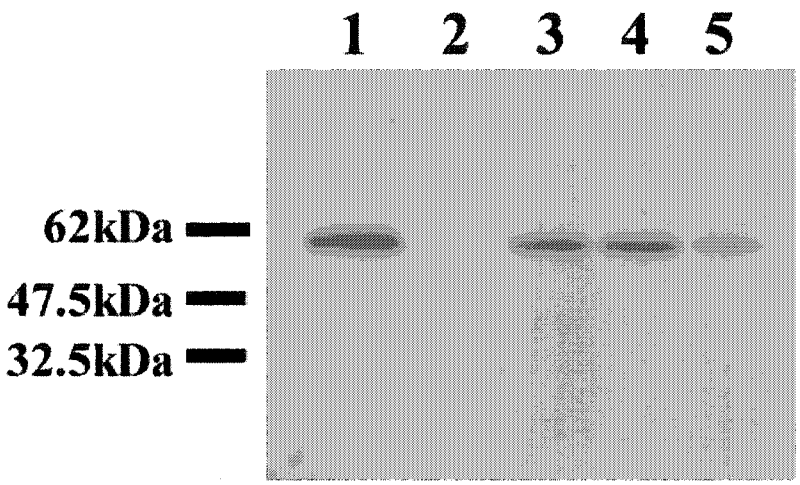

Figure 2.6 - Western Blot of baculovirus infected *Sf*9 cell crude lysates. Each lane was loaded with 5µg of total protein and glucocerebrosidase identified using anti glucocerebrosidase monoclonal antibody AA16B3. Lane 1, Normal glucocerebrosidase construct; Lane 2, Wild-type AcNPV (no glucocerebrosidase insert); Lane 3, P122S mutant glucocerebrosidase; Lane 4, N370S mutant glucocerebrosidase; Lane 5, L444P mutant glucocerebrosidase. The relative positions of prestained broad range protein size markers (New England Biolabs, Beverly, MA) are listed on the left.

The P122S clone produced crossreactive protein at an apparent molecular mass of approximately 63kDa, corresponding to that seen for the normal and N370S clones (Figure 2.6). A reduced amount of crossreactive protein was produced by the L444P mutation clone, previously characterized as unstable (Pasmanik-chor et al., 1997). Activity levels for the P122S mutant clone were 19.2% of normal (Table 2.3) while the severe L444P mutant showed near zero activity and the mild N370S mutant showed 9.4% of normal activity (Sinclair et al., 2001).

PCR analysis of genomic DNA extracted from paraffin-embedded tissues for patient BW was also complicated by the fragmented nature of the template. Direct sequence analysis of DNA amplified with primers K and L revealed the distinct doublet banding pattern diagnostic of a heterozygous deletion mutation (Figure 2.7). The presence of a C deletion in a tract of six cytosine residues ending at cDNA nucleotide 203 (del203C) was identified by determination of the doublet start site and was confirmed by *Ahd*I restriction digest (data not shown). Although the C deletion could have occurred at any point in the 6bp C repeat, the downstream site is given as the point of mutation by convention (Beutler and Gelbart, 1997). Direct sequence analysis of the rest of the glucocerebrosidase coding region and intron splice sites failed to confirm a second deleterious mutation in patient BW. The fragmented nature of the template, however, restricted the PCR analyses required to investigate the presence or absence of a possible complex allele in this individual.

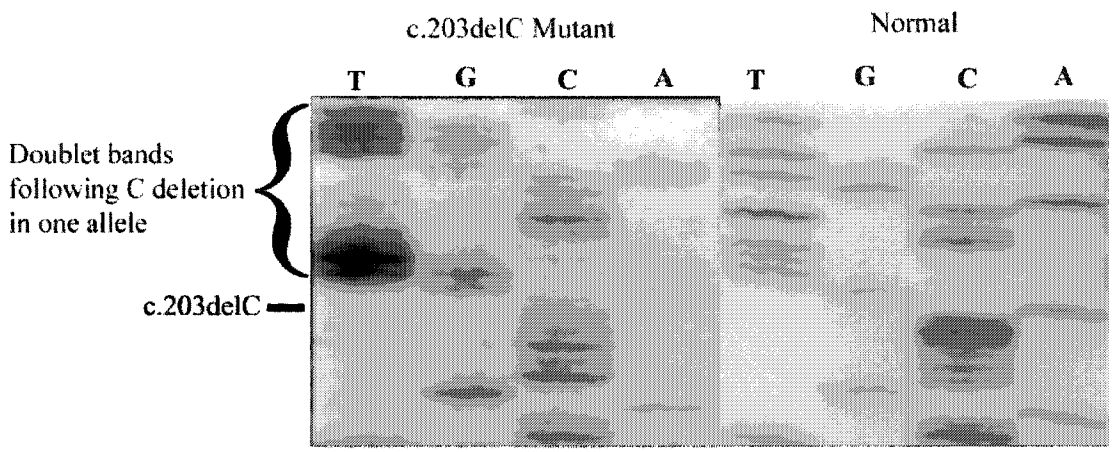

Figure 2.7 – Direct sequence analysis of archival DNA from BW. The four lanes to the left are the T, G, C, and A lanes of patient BW (Mutant). The four lanes to the right are those from an unaffected individual (Normal). As shown, there is a C deletion mutation in one allele in BW resulting in the formation of doublet bands following the deletion site at cDNA nt203.

Table 2.3 - β-glucosidase specific activity measurements for crude extracts of recombinant baculovirus infected *Sf*9 cells with the artificial substrate (4MUGP).

Baculovirus Clone	4MUGP Specific Activity[a] (nmole/hr/mg cell protein)	% Normal Activity
Normal	2681 +/-180 (n=3)	100
N370S	252 +/-4 (n=3)	9.3
L444P	179 +/-10 (n=3)	6.7
P122S	515 +/-36 (n=2)	19.2
Control[b]	213 +/-18 (n=2)	7.5

[a]Mean values plus or minus the standard error of the mean from (n) independent experiments.

[b]AcNPV (non-recombinant baculovirus) without the glucocerebrosidase cDNA construct.

2.4 - **Discussion**

Among the three clinical forms of Gaucher disease, the neuronopathic type 2 and type 3 forms of the disease are the most debilitating and poorly studied. The work reported here identified, using traditional fibroblast and archival DNA samples, one novel point mutation, two rare point mutations, and a novel complex allele in 4 neuronopathic Gaucher patients.

2.4.1 – *Type 2 Cases*

Type 2 Gaucher disease is the most severe form of this disorder and mutations associated with this sub-type have been characterized as severe or null alleles (Beutler and Gelbart, 1997). As patient 1043 died at the early age of 18 months, it appears that the presence of the mutations identified in this individual will lead to a poor prognosis. Heterozygosity for the L444P mutation as either a single point mutation or in the context of a complex mutant allele can lead to a range of clinical subtypes depending on the severity of the second allele (Balicki and Beutler, 1995). Due to the progressive symptoms in patient 1043 and the presence of acute neuronopathy, a speculation of severity can be made for mutation E41K substituting a basic lysine for an acidic glutamic acid in conjunction with a complex L444P allele.

The recent identification and sequencing of several genes contiguous with the glucocerebrosidase gene and pseudogene duplication (Long et al., 1996) has allowed a thorough analysis of the complex allele identified in patient 1043. Possible mechanisms for the creation of complex mutants include gene conversions or single and double

crossover events due to mitotic strand migration or meiotic gene/pseudogene mispairing (Tusie-Luna and White, 1995; Zimran et al., 1990). Gene conversion and double crossover events give rise to short stretches of pseudogene sequence flanked 5' and 3' by the appropriate functional gene sequences with no associated major locus reorganization. The glucocerebrosidase complex alleles RecTL and RecNci fall into this category (Zimran et al., 1990). In contrast, single crossover events can give rise to large deletions in the locus, often truncating or removing intervening genes as a functional/pseudogene fusion is created (Higashi et al., 1988; Zimran et al., 1990). A recent study of *de novo* steroid 21-hydroxylase (*CYP21*) complex alleles revealed that while approximately 25% of the alleles arose from single crossover events, the vast majority were the product of gene conversion with conversion tracts reaching a maximum of 60 basepairs before returning to the normal sequence (Tusie-Luna and White, 1995). Double crossover events were found to make a limited contribution to the creation of new complex alleles.

A diagnostic restriction digest has been used in the past to distinguish gene conversion products from single crossover gene fusion events in L444P complex Gaucher alleles (Zimran et al., 1990). For reference, the locations of these *Ssp*I cut sites are included in Figure 2.8. When digested with *Ssp*I, normal individuals and those with the RecTL and RecNci complex alleles will yield two bands of 17 Kb and 13 Kb in a glucocerebrosidase Southern Blot (Zimran et al., 1990). In contrast, the Crossover or Fusion allele will give rise to a novel 14 Kb band corresponding to the new fusion product as a large portion of the intervening sequences between the glucocerebrosidase gene and pseudogene are lost (Zimran et al., 1990).

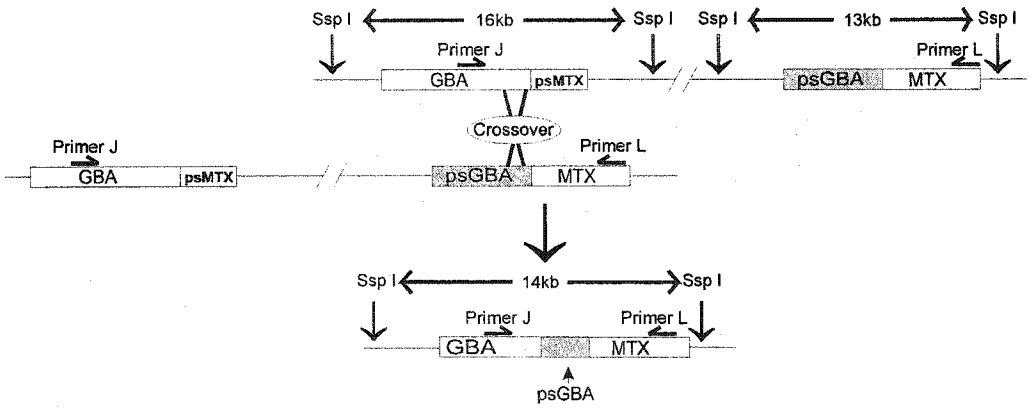

Figure 2.8 - A Mechanism for the creation of a crossover mutant by allelic mispairing and a single unequal crossover event in the glucocerebrosidase gene. *Ssp*I cut sites represent those used in previous Southern blot analyses. GBA=glucocerebrosidase functional gene. psGBA=glucocerebrosidase pseudogene. MTX=metaxin functional gene. psMTX=metaxin pseudogene. Primers I and J correspond to the appropriate primers in Table 2.1 amplifying only the crossover product.

The delineation of the metaxin gene sequence allowed the creation of a primer at it's 5' end and thus the ability to follow crossovers beyond the 3' end of the glucocerebrosidase gene. Due to the inverted orientation of the metaxin gene, primers sense to metaxin will function as antisense 3' primers for glucocerebrosidase amplification. Long PCR analysis confirmed the continuation of the glucocerebrosidase crossover product through into the contiguous metaxin functional gene over a region of 4.6 Kb. The extent of this sequence alteration suggests a single crossover origin for the allele in patient 1043 rather than a gene conversion event. The probable mechanism of such a crossover is presented in Figure 2.8. In fact, primer L (in the metaxin functional gene) is only 400 bp from the terminal *Ssp*I cut site used in previous crossover analyses, enabling this long PCR protocol with primers J and L to be utilized as a simple alternative to the *Ssp*I Southern blot for analysis of glucocerebrosidase complex alleles.

The perinatal lethal phenotype observed in patient BW corresponds to an emerging sub-class of type 2 Gaucher disease associated with homozygosity for null expressing mutant alleles, hydrops fetalis, severe epidermal abnormalities and perinatal death (Sidransky et al., 1992; Strasberg et al., 1994). The identification of a deletion mutation (del203C) early in exon 3 in patient BW correlates with the severe phenotyope seen. This mutation could be considered a null allele as it leads to a frame shift error and introduction of a series of misense mutations and subsequent nonsense stop site 58 basepairs downstream of the deletion site (Germain et al., 1998). A high number of complex alleles have also been identified from infants displaying this perinatal lethal phenotype (Strasberg et al., 1994). Unfortunately, a possible complex allele in patient BW could not be confirmed

due to the fragmented nature of the template and poor consistency in the amplification of the long fragments required. The severe phenotype observed is, however, highly suggestive of the presence of a second null mutation in this individual. Comparisons of the clinical phenotype obseveved in other perinatal lethal Gaucher patients to the available knockout mouse model for Gaucher disease (Tybulewicz et al., 1992) have yielded some insight into the etiology of neuronopathy in Gaucher disease (Sidransky et al., 1992; Strasberg et al., 1994). Similar complications in lipid storage, neurodegeneration, and epidermal barrier formation are seen in the mice and affected infants and these symptoms have been suggested to reflect the null glucocerebrosidase activity present (Sidransky et al., 1992; Strasberg et al., 1994). Importantly, recent work has been done on the level of glucosylphingosine rather than glucocerebroside (glucosylceramide) accumulation in these mice (Orvisky et al., 2000). Glucosylsphingosine is a minor secondary substrate for glucocerebrosidase that is substancially more cytotoxic and its accumulation has been speculated as being involved in the onset of neuronopathy in Gaucher disease (Meivar-Levy et al., 1994). This hypothesis has never reached favour due to experimental inconsistencies and a lack of direct evidence (Meivar-Levy et al., 1994). The Gaucher knockout mouse and a set of perinatal lethal type 2 patients, however, have been shown to accumulate glucosylsphingosine in neural tissue and the level of this accumulation has been shown to correlate with the severity of neuronopathy seen (Orvisky et al., 2000). While this is a preliminary finding, it does suggest a role for glycosylsphingosine storage in the

neuronopathy observed, as such a correlation with glucocerebroside has not been forthcoming (Orvisky et al., 2000).

2.4.2 – *Type 3 Cases*

The P122S mutation identified in JB and MS is unique in that it is present so far only in members of Cree First Nations. Although initially categorized as similar to mutation N370S in the presentation of a non-neuronopathic phenotype (Beutler et al., 1993), the P122S mutation appears now to be associated with a more severe type 3 phenotype, particularly when the more immediate life-threatening complications of the disease have been ameliorated by enzyme replacement therapy. There were hematological, skeletal and visceral symptoms in all three of the affected individuals, leading to the death of the one that was not treated with enzyme replacement. All three individuals were developmentally delayed, and the two survivors show specific neurological features consistent with type 3 disease. Since the hydrophobic imino R group of proline plays a crucial role in determining the secondary structure of proteins (Lehninger et al., 1993), its substitution by the hydrophilic hydroxyamino acid serine could have significant impact on the enzyme biophysical and catalytic properties and function.

Heterologous expression of the P122S protein however, yielded 19.2% of normal glucocerebrosidase activity on the artificial substrate 4MUGP, higher than that seen with the relatively mild N370S mutation (9.4%). Previous studies have measured the 4MUGP residual enzyme activity levels of the N370S protein at 70-621 nmole/h/mg protein, representing a range of 5-20% of normal glucocerebrosidase activity (1500-

8500nmole/hr/mg) (Grace et al., 1990; Grace et al., 1994). This clearly identifies a large variability in the baculovirus expression of glucocerebrosidase alleles, but does allow for the conclusion that the P122S mutant of glucocerebrosidase does produce relatively stable enzyme with a residual activity similar to that seen with a mild type 1 Gaucher disease allele. The presence of neurological features in the two confirmed and third probable P122S homozygote identified to date would suggest the presence of a more severe disruption of enzyme function *in vivo* than that determined *in vitro*. Amino acid 122 is coded in exon 5 of the glucocerebrosidase gene and while this region is outside the predicted active site of the mature protein, little is known of the functional role played by other regions of the polypeptide (Beutler and Grabowski, 2001). The introduction of a proline could cause sufficient alteration in folding of the nascent polypeptide to alter interactions with required chaperones and associated proteins (i.e. *lamp*-1 and *lamp*-2) thus affecting the maturation and, importantly, the lysosomal targeting of the protein (Zimmer et al., 1999). Such alterations in targeting would not be evident using an *in vitro* activity assay provided the glucocerebrosidase had been sufficiently folded and glycosylated to produce an active, but mistargetted enzyme (Zimmer et al., 1999). This lack of a clear correlation between *in vitro* residual enzyme activity and clinical severity has, however, been well documented in Gaucher disease, speaking to the apparent multifactorial nature of disease progression (Sibille et al., 1993; Sidransky et al., 1992; Strasberg et al., 1994; Theophilus et al., 1989; Zimran et al., 1989). As mentioned above, further analysis of *in vitro* and *in vivo* glucosylsphingosine degradation may elucidate the neurodegeneration associated with this and other glucocerebrosidase mutants. As all the P122S affected individuals identified share a similar ethnic (genetic) background,

commonalities in the systemic processing of various sphingolipids in these patients may act in concert with a relatively mild deficiency in glucocerebroside degradation to produce the severity of visceral symptoms and late onset neuronopathy observed.

While these cases suggest a correlation of this Gaucher genotype (homozygous mutation P122S) with a type 3 phenotype, especially when modified by enzyme replacement, a description of more patients is required for confirmation. The *Kpn*I restriction analysis will enable us to screen for the presence of this mutation in either the homozygous or heterozygous form with accuracy and relative ease within this identified population. The added utility of DNA extraction from archival material will increase the availability of samples from this diverse and generally remote population, aiding screening efforts. The use of such tissue stores could facilitate a more complete screening of the population by adding samples from deceased individuals.

2.5 - References

Balicki, D. and Beutler, E. Gaucher disease. *Medicine* 71:305-323, 1995.

Beutler, E. and Gelbart, T. Hematologically important mutations: Gaucher disease. *Blood Cells, Molecules, and Diseases* 23:2-7, 1997.

Beutler, E., Gelbart, T., Kuhl, W., Sorge, J. and West, C. Identification of the second common Jewish Gaucher disease mutation makes possible population-based screening for the heterozygous state. *Proceedings of the National Academy of Science, U.S.A.* 88:10544-10547, 1991.

Beutler, E., Gelbart, T., Kuhl, W., Zimran, A. and West, C. Mutations in Jewish patients with Gaucher disease. *Blood* 79:1662-1666, 1992.

Beutler, E., Gelbart, T. and West, C. The facile detection of the nt 1226 mutation of glucocerebrosidase by 'mismatched' PCR. *Clinica Chimica Acta* 194:161-166, 1990.

Beutler, E., Gelbart, T. and West, C. Identification of six new Gaucher disease mutations. *Genomics* 15:203-205, 1993.

Beutler, E. and Grabowski, G.A. Gaucher disease. In: *The Metabolic and Molecular Bases of Inherited Disease*, edited by Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D. New York: McGraw-Hill, 2001, p. 3645-3668.

Beutler, E., McKusick, V.A., Motulsky, A.G. and Scriver, C.R. Mutation nomenclature: Nicknames, systemic names, and unique identifiers. *Human Mutation* 8:203-206, 1996.

Chan KW, Wong LT, Applegarth D, Davidson AG. Bone marrow transplantation in Gaucher's disease: effect of mixed chimeric state. *Bone Marrow Transplantation.* 14:327-30, 1994.

Choy, F.Y. Gaucher disease: the effects of phosphatidylserine on glucocerebrosidase from normal and Gaucher fibroblasts. *Human Genetics* 67:432-436, 1984.

Choy, F.Y.M., Humphries, M.L. and Shi, H. Identification of two novel and four common missense mutations among Chinese Gaucher disease patients. *American Journal of Medical Genetics* 71:172-178, 1997.

Choy, F.Y.M., Wei, C., Applegarth, D.A. and Yong, S.L. A new missense mutation in glucocerebrosidase exon 9 of a non-Jewish Caucasian type 1 Gaucher disease patient. *Human Molecular Genetics* 3:821-823, 1994.

Choy, F.Y.M., Wei, C. and Levin, D. Gaucher disease: Functional expression of the normal glucocerebrosidase and Gaucher T1366G and G1604A alleles in *Baculovirus*-transfected *Spodoptera frugiperda* cells. *American Journal of Medical Genetics* 65:184-189, 1996.

Coleman, M.B., Steinberg, M.H. and Adams, J.G. Hemoglobin Terre Haute arginine beta 106. A posthumous correction to the original structure of hemoglobin Indianapolis. *Journal of Biological Chemistry* 266:5798-5800, 1991.

Dahl, N., Lagerstrom, M., Erikson, A. and Pettersson, U. Gaucher disease type III (Norrbottnian type) is caused by a single mutation in exon 10 of the glucocerebrosidase gene. *American Journal of Human Genetics* 47:275-278, 1990.

Erikson, A. and Wahlberg, I. Gaucher disease--Norrbottnian type. Ocular abnormalities. *Acta Ophthalmology (Copenhagen)* 63:221-225, 1985.

Eyal, N., Wilder, S. and Horowitz, M. Prevalent and rare mutations among Gaucher patients. *Gene* 96:277-283, 1990.

Germain, D.P., Puech, J.P., Caillaud, C., Kahn, A. and Poenaru, L. Exhaustive screening of the acid beta-glucosidase gene, by fluorescence-assisted mismatch analysis using universal primers: mutation profile and genotype/phenotype correlations in Gaucher disease. *American Journal of Human Genetics* 63:415-427, 1998.

Grace, M.E., Graves, P.N., Smith, F.I. and Grabowski, G.A. Analyses of catalytic activity and inhibitor binding of human acid β-glucosidase by site-directed mutagenesis. *Journal of Biological Chemistry* 265:6827-6835, 1990.

Grace, M.E., Newman, K.M., Scheinker, V., Berg-Fussman, A. and Grabowski, G.A. Analysis of Human acid β-glucosidase by site-directed mutagenesis and heterologous expression. *Journal of Biological Chemistry* 269:2283-2291, 1994.

Higashi, Y., Tanae, A., Inoue, H. and Fujii-Kuriyama, Y. Evidence for frequent gene conversion in the steroid 21-hydroxylase P-450(C21) gene: implications for steroid 21-hydroxylase deficiency. *American Journal of Human Genetics* 42:17-25, 1988.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77:51-59, 1989.

Hong, C.M., Ohashi, T., Yu, X.J., Weiler, S. and Barranger, J.A. Sequence of two alleles responsible for Gaucher disease. *DNA and Cell Biology* 9:233-241, 1990.

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

Koprivica, V., Stone, D.L., Park, J.K., et al. Analysis and classification of 304 mutant alleles in patients with type 1 and type 3 Gaucher disease. *American Journal of Human Genetics* 66:1777-1786, 2000.

Latham, T., Grabowski, G.A., Theophilus, B.D. and Smith, F.I. Complex alleles of the acid beta-glucosidase gene in Gaucher disease. *American Journal of Human Genetics* 47:79-86, 1990.

Lehninger, A.L., Nelson, D.L. and Cox, M.M. *Principles of Biochemistry*, New York:Worth Publishers, 1993. Ed.2nd Edition

Long, G.L., Winfield, S., Adolph, K.W., Ginns, E.I. and Bornstein, P. Structure and organization of the human metaxin gene (MTX) and pseudogene. *Genomics* 33:177-184, 1996.

Meivar-Levy, I., Horowitz, M. and Futerman, A.H. Analysis of glucocerebrosidase activity using N-(1-[14C]hexanoyl)-D-erythroglucosylsphingosine demonstrates a correlation between levels of residual enzyme activity and the type of Gaucher disease. *Biochemical Journal* 303:377-382, 1994.

O'Reilly, D.R. and Miller, L.K. *Baculovirus Expression Vectors: A Laboratory Manual*, New York:W.H. Freeman and Company, 1994.

Orvisky, E., Sidransky, E., McKinney, C.E., et al. Glucosylsphingosine accumulation in mice and patients with type 2 Gaucher disease begins early in gestation. *Pediatric Research* 48:233-237, 2000.

Pasmanik-chor, M., Madar-shapiro, L., Stein, O.E., Aerts, H., Gatt, S. and Horowitz, M. Expression of mutated glucocerebrosidase alleles in human cells. *Human Molecular Genetics* 6:887-895, 1997.

Sanger, F., Nicklen, S. and Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Science, U.S.A.* 74:5463-5467, 1977.

Shitaba, D., Martin, W.J. and Arnheim, N. Analysis of DNA sequences in forty-year-old paraffin-embedded thin-tissue sections: A bridge between molecular biology and classical histology. *Cancer Research* 48:4564-4566, 1988.

Sibille, A., Eng, C.M., Kim, S.J., Pastores, G. and Grabowski, G.A. Phenotype/genotype correlations in Gaucher disease type I: clinical and therapeutic implications. *American Journal of Human Genetics.* 52:1094-1101, 1993.

Sidransky, E., Sherer, D.M. and Ginns, E.I. Gaucher disease in the neonate: A distinct Gaucher phenotype is analogous to a mouse model created by targeted disruption of the glucocerebrosidase gene. *Pediatric Research* 32:494-498, 1992.

Sinclair, G., Choy, F.Y.M., and Humphries, L.M. A novel complex allele and two new point mutations in type 2 (acute neuronopathic) Gaucher disease. *Blood Cells, Molecules, and Diseases* 24:420-427, 1998.

Sinclair, G., Choy, F.Y.M., Ferriera, P. Heterologous expression and characterization of a rare Gaucher disease mutation (c.481C>T) from a Canadian aboriginal population using archival tissue samples. *Molecular Genetics and Metabolism*, accepted for publication, 2001.

Sorge, J.A., West, C., Kuhl, W., Treger, L. and Beutler, E. The human glucocerebrosidase gene has two functional ATG initiator codons. *American Journal of Human Genetics* 41:1016-1024, 1987.

Spector, D.L., Goldman, R.D. and Leinwand, L.A. *Cells, A Laboratory Manual*, Plainview:Cold Spring Harbor Laboratory Press, 1998.

Strasberg, P.M., Skomorowski, M.A., Warren, I.B., Hilson, W.L., Callahan, J.W. and Clarke J.T.R. Homozygous presence of the crossover (fusion gene) mutation identified in a type II Gaucher disease fetus: Is this analogous to the Gaucher knock-out mouse model? *Biochemical Medicine and Metabolic Biology* 53:16-21, 1994.

Theophilus, B., Latham, T., Grabowski, G.A. and Smith, F.I. Gaucher disease: molecular heterogeneity and phenotype-genotype correlations. *American Journal of Human Genetics* 45:212-225, 1989.

Theophilus, B.D., Latham, T., Grabowski, G.A. and Smith, F.I. Comparison of RNase A, a chemical cleavage and GC-clamped denaturing gradient gel electrophoresis for the detection of mutations in exon 9 of the human acid beta-glucosidase gene. *Nucleic Acids Research* 17:7707-7722, 1989.

Tsuji, S., Choudary, P.V., Martin, B.M., et al. A mutation in the human glucocerebrosidase gene in neuronopathic Gaucher's disease. *New England Journal of Medicine* 316:570-575, 1987.

Tusie-Luna, M.T. and White, P.C. Gene conversions and unequal crossovers between CYP21 (steroid 21-hydroxylase gene) and CYP21P involve different mechanisms. *Proceedings of the National Academy of Science, U.S.A.* 92:10796-10800, 1995.

Tybulewicz, V.L., Tremblay, M.L., LaMarca, M.E., et al. Animal model of Gaucher's disease from targeted disruption of the mouse glucocerebrosidase gene. *Nature* 357:407-410, 1992.

Vallejo, A.N., Pogulis, R.J. and Pease, L.R. In vitro synthesis of novel genes: mutagenesis and recombination by PCR. *PCR Methods and Applications* 4:S123-S1301994.

Zimmer, K., le Coutre, P., Aerts, J.M.F.G., et al. Intracellular transport of acid β-glucosidase and lysosome-associated membrame proteins is affected in Gaucher's disease (G202R mutation). *Journal of Pathology* 188:407-414, 1999.

Zimran, A., Sorge, J., Gross, E., Kubitz, M., West, C. and Beutler, E. Prediction of severity of Gaucher's disease by identification of mutations at DNA level. *Lancet* 2:349-352, 1989.

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

Chapter 3 – Heterologous Expression and Purification of Glucocerebrosidase using yeast (*Pichia pastoris*) and insect cell (*Spodoptera frugiperda*) hosts

3.1 - Introduction

As a single gene enzyme deficiency, Gaucher disease has long been a candidate for both enzyme replacement and gene therapy (Beutler, 1993). While the second option has had some limited experimental success, enzyme replacement is now a viable, though expensive option for many affected individuals. Alglucerase, macrophage targeted glucocerebrosidase isolated from human placenta, was approved for clinical use in 1991 and released under the trade name *Ceredase*™ (Barton et al., 1991; Barton et al., 1990; Beutler, 1993). It has been found highly effective in decreasing hepatosplenomegaly and overall sphingolipid storage levels in a large number of type 1 Gaucher cases (Altarescu et al., 2000; Brady and Barton, 1994; Grabowski et al., 1998). Unfortunately, the effort required to isolate and target the enzyme by modification of the protein glycosylation has rendered the enzyme prohibitively expensive. It has been termed the most expensive drug on the planet, costing $150,000 US annually to treat a single individual on an average dose regime (Beutler, 1993; Mistry et al., 1996; Moscicki and Taunton-Rigby, 1993). A recombinant form of the enzyme, imiglucerase, has more recently been developed in a Chinese Hamster Ovary cell (CHO) expression system and marketed for clinical use in the US under the trade name *Cerezyme*™. Although produced at a significantly higher level than the placental enzyme, an overall price decrease for treatment has not accompanied the increased production. The continued high cost of treatment has lead to a heated debate over the appropriate dose regime with many researchers extolling both the financial and therapeutic virtues of an increased frequency of injection but lower overall dose (Altarescu et al., 2000; Grabowski et al., 1998; Moscicki and Taunton-Rigby, 1993).

Regardless of the clinical specifics, enzyme replacement remains the therapy of choice for Gaucher disease and has been highly effective in improving the quality of life of most individuals treated (Grabowski et al., 1998).

It was felt that there existed an opportunity to develop alternative heterologous expression systems for high-level expression of active glucocerebrosidase as both a source of material for structural analysis of the enzyme and as a possible biotherapeutic agent. The alternatives for such an expression system range from well characterized prokaryotic systems, through unicellular eukaryotes, to more recent mammalian cell culture and whole organism recombinant bioreactor systems. While prokaryotic expression systems provide a simple alternative with easily manipulated genomes, inexpensive culture requirements, and high cell density culture, this simplicity limits their usefulness with complex mammalian proteins. Glucocerebrosidase is a highly hydrophobic glycoprotein requiring extensive posttranslational modification to ensure proper folding and activity, a process that cannot be provided by a prokaryotic host cell (Grace and Grabowski, 1990a). Higher eukaryotic systems such as the aforementioned CHO cells, can provide the necessary post-translation modifications and trafficking, but relatively low protein production levels, low achievable cell density, and labour intensive culturing requirements drive up product costs and can limit the availability and corresponding effectiveness of the therapeutic protein (Geisse et al., 1996).

Based on the cellular characteristics required for the production of active glucocerebrosidase, it was felt that both yeast and insect cell-based expression systems

provided the most promising routes for the high-level heterologous expression of this protein. Yeast expression systems promise the ease of culture and relatively high cell densities of prokaryotic hosts, while providing most of the complex modifications available from higher eukaryotic systems (Gellissen et al., 1992). *S. cerevisiasae* expression has, however, been shown ineffective for the expression of a complex human membrane bound enzyme such as glucocerebrosidase, due to major complications in glycosylation (Gellissen et al., 1992; Skory and Freer, 1996). Conversely, insect cell protein trafficking and post-translational modification pathways have been well characterized and have shown to be effective in the production of glucocerebrosidase at a diagnostic level (Choy et al., 1996; Grace et al., 1990b; Grace et al., 1994). Of the available yeast systems, *Pichia pastoris* presents a potent model for glycoprotein expression as its glycosylation system more closely resembles that of mammals than other unicellular fungi and it contains a well characterized, highly inducible promoter from which to transcribe foreign sequences (Gellissen et al., 1992; Hollenberg and Gellisson, 1997; Sreekrishna et al., 1997). For insect cell expression, the stable genomic integration of the Insect Select™ (Invitrogen, Carlsbad, CA) plasmid-based expression system, rather than transient baculovirus infection, has shown promise for increasing protein production levels in this host cell type (Hegedus et al., 1998; Pfeifer, 1998).

Little is written on the natural history of the methylotrophic budding yeast, *Pichia pastoris*, as much of the research on this yeast has been in a laboratory setting and has been focused on its rare metabolic capacity for methanol (Couderc and Baratti, 1980; Cregg et al., 1989; Sreekrishna et al., 1988). In fact, *P. pastoris* was originally researched

and developed by the Phillips Petroleum Company as a possible mechanism by which to utilize the methanol they produced as a by-product of oil refining (Wegner, 1990). It was envisioned that *P. pastoris,* grown on methanol, could be utilized as a single cell protein source for industrial animal feed (Wegner, 1990). While these far reaching aspirations were not realized, research into methanol utilization in *P. pastoris* revealed that the peculiarities of methanol oxidase induction in this yeast could be co-opted for high-level heterologous protein production. Methanol oxidase catalyses the oxidation of methanol to formaldehyde as the first step in the methanol utilization pathway in *Pichia* (Couderc and Baratti, 1980; Cregg et al., 1989). This enzyme, however, has a low affinity for the oxygen required for this reaction and as a result the transcription of the alcohol oxidase gene (AOX1) is induced over 1,000 fold in the presence of methanol to produce an excess of the AOX protein (Cregg et al., 1989). Under appropriate culture conditions alcohol oxidase can comprise up to 30% of the total cellular protein (Cregg et al., 1989). The fact that this promoter is so tightly regulated is also useful in a heterologous expression setting as it allows the production of cell biomass before the induction of any potentially deleterious heterologous proteins (Tschopp et al., 1987).

The nature of heterologous protein glycosylation in *Pichia* is also an important consideration from a biotherapeutic perspective, as the immunogenicity of recombinant enzymes has been shown to impact therapeutic efficacy in a number of disorders, including *Cerezyme*™ treatment of Gaucher disease (Aviner et al., 1999). Yeasts, in general, process glycans only to the high-mannose stage and are incapable of completing the trimming reactions required for the production of complex glycans (Moremen et al.,

1994; Tanner and Lehle, 1987). Some yeasts, most notably the well characterized *S. cerevisiae* model, tend to hypermannosylate those glycans, creating large cross-linked structures that alter the function and usefulness of the associated heterologous proteins (Jenkins et al., 1996). The glycans on heterologous proteins produced by *Pichia*, however, have been shown to contain an average of 9-18 mannose residues mimmicking the high-mannose structures produced by higher eukaryotes (Miele et al., 1997; Montesino et al., 1998; Richard et al., 1998). *Pichia* glycans have also been shown to be devoid of the α1,3-linked mannose residues responsible for much of the immunogenicity of *S. cerevisiae* produced glycoproteins (Montesino et al., 1998). Accordingly, *Pichia* has been successfully utilized to express functional complex glycoproteins from a number of organisms including human lysosomal α-mannosidase, *Candida* extracellular β-glucosidase, invertase from *S. cerevisiae*, and an HIV-1 envelope protein (Liao et al., 1996; Scorer et al., 1993; Skory and Freer, 1996; Tschopp et al., 1987). As the effective targeting of CHO recombinant and placental glucocerebrosidase for enzyme replacement therapy has required the modification of associated glycans to mannose terminated structures (Barton et al., 1991; Mistry et al., 1996), it appears that the glycosylation patterns of *P. pastoris* are appropriate for the production of this enzyme.

Insect cell expression systems utilizing *Autographa californica* nucleopolyhedrosis virus (*Ac*NPV) baculovirus vectors have been used extensively for the expression of foreign eukaryotic proteins. With much of the research on the baculoviridae family reflecting the role of this group of viruses as potential insect biocontrol agents, the transcriptional regulation and translational processing of proteins in various insect cell lines infected

with baculovirus vectors have been well characterized (Blissard and Rohrmann, 1990; Bonning and Hammock, 1996; Hasnain et al., 1997). Although high levels of transcription have been associated with *Ac*NPV expression systems, the quantity and quality of the heterologous protein produced have been inconsistent due to nature of viral infection (Geisse et al., 1996; Hegedus et al., 1998). As part of the infection process, the baculovirus tends to shut down the production of native insect cell proteins in favour of those encoded by the viral genome (Hegedus et al., 1998). While this leads to increased production levels of simple proteins, the reduced availability of chaperones, glycosidases, and other cellular factors required for the maturation of complex glycoproteins has been shown to affect both the quantity and quality of heterologous proteins, particularly late in infection (Hegedus et al., 1998). Coupled with the transient nature of virus-based protein production, and the potential biohazard of the virus itself, *Ac*NPV expression systems are not optimal for the production of biotherapeutics.

Insect cells have, however, been shown to successfully produce active glucocerebrosidase using *Ac*NPV-based expression systems at a diagnostic level and some investigation of the appropriateness of this system at a large-scale production level was done (Martin et al., 1988). Although a U.S. patent was given for *Ac*NPV-mediated recombinant glucocerebrosidase production in 2000, the product has yet to come to market, most probably due to the constraints presented above. The limitations of a virus-based insect cell expression system can be avoided, however, by the use of a plasmid-mediated expression system allowing the stable genomic integration of the glucocerebrosidase (GBA) transgene. Stable integration of plasmids utilizing a number of baculovirus

immediate-early promoters has been successful for the production of heterologous proteins in a variety of dipteran and lepidopteran cell lines (Hegedus et al., 1998; Pfeifer, 1998). As this avoids the biohazard, low production, and transience problems of baculovirus expression systems, an *Orgyia pseudotsugata* nucleoplyherosis virus (*Op*NPV) promoter-based integrative plasmid was selected for the production of human glucocerebrosidase in *Sf*9 cells.

3.2 - **Materials and Methods**

3.2.1 - *Pichia Strains*

Pichia pastoris strain KM71 was used for the glucocerebrosidase expression studies. KM71 is a strain developed from the basic commercial *Pichia* strain, GS118, to display a constitutive methanol utilization slow (Mut^s) phenotype along with histidine auxotrophy. In the KM71 strain the AOX1 gene has been mutated to a nonfunctional state, requiring the cell to metabolize methanol using the much lower efficiency AOX2 gene, leading to slow growth on methanol as a carbon source. This strain was created commercially in response to preliminary studies suggesting heterologous protein induction from the AOX1 promoter could be maximized only in the absence of the AOX1 gene product (Chiruvolu et al., 1997). The histidine autotrophy allows transformants to be selected using histidase complementation vectors, although such constructs were not utilized in this study.

3.2.2 - *Vector Construction*

All expression constructs utilized the pPICZαA vector from the Invitrogen Corporation (Carlsbad, CA). This vector contains a secretion signal and multiple cloning site (MCS) flanked 5' and 3' by *P. pastoris* AOX1 sequences to direct the genomic integration of the transgene by homologous recombination into the AOX1 locus. The glucocerebrosidase cDNA was *Pfu* PCR amplified by primers A and B to create an open reading frame with its native stop codon but without the native GBA leader sequence, all flanked by *EcoR*I restriction cut sites (Table 3.1). This 1.6Kb fragment was cloned into the pPICZαA vector at the *EcoR*I site in the MCS with the GBA sequence in frame with the α-factor secretion signal to create the vector pαGBA (Figure 3.1). The recombinant pαGBA vector was amplified in *E. coli* TOP10F' cells (Invitrogen, Carlsbad, CA) and linearized at the vector *Bst*XI site (New England Biolabs, Beverley, MA) prior to electroporation into *Pichia pastors* KM71 cells as described (pPICZα Manual, Invitrogen, Carlsbad, CA). The vector was linearized to increase the frequency of genomic integration events and recombinant clones were selected with Zeocin™ (Invitrogen, Carlsbad, CA) resistance in both *E. coli* and *P. pastoris* cells (clone KM1G). Importantly, the pPICZαA vector and subsequent recombinants do not contain a *Pichia* origin of replication so Zeocin™ resistant clones are selected as stable genomic integrants.

3.2.3 - *Dual-cassette Vector*

The pPICZαA vector was designed to allow the entire cassette containing the AOX5', GBA, and AOX3' regions to be tandemly repeated in the vector to increase gene copy number and transcription rates (Clare et al., 1991). The standard protocol for multiple

Table 3.1 – Primers used in the construction of glucocerebrosidae expression plasmids

Primer	Sequence[a] (5'>3')	Location[b]	Orientation
A	TAT**GAATTC**CGCCCGCCCCTGC ATCCCT	GBA nt115-135	Sense
B	GCT**GAATTC**TTTAATGCCCAGG CTG	GBA nt1658-1642	Anti-sense
C	TCATG**AGATCT**AACATCCAAAG ACGAAAGG	pPICZ nt3584-15	Sense
D	AACGAA**GGTCTC**TGATCATCTT CTGTACTC	pPICZ nt1671-1646	Anti-sense
E	ACTC**GAATTC**TTCTTCATCTAAG GACCCTGAGG	GBA nt minus 137-118	Sense
F	TACC**GAATTC**ATGGCTGGCAGC CTCACAGG	GBA nt61-80	Sense
G	CTGCTGCTCTCAACATCCTT	GBA nt380-399	Sense
H	GAAGGGGTATCCACTCAACAG	GBA nt855-835	Anti-Sense

[a]Nucleotides introduced to create restriction enzyme recognition sites are presented in bold print.

[b]Nucleotide numbers identified as GBA represent glucocerebrosidase cDNA sequences with the first base of the upstream initiator ATG as position #1 (Sorge 1987). Numbers for pPICZ represent the circular pPICZαA plasmid with position #1 as the first G in the *Bam*H1 recognition site as presented in the pPICZα Manual (Invitrogen, Carlsbad, CA).

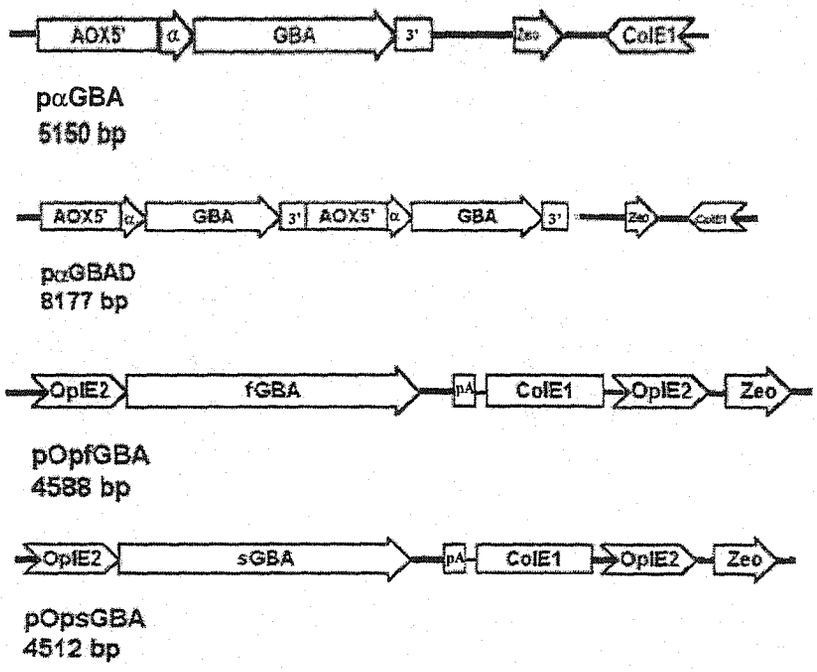

Figure 3.1 – *Pichia pastoris* and *Sf*9 cell expression constructs. pαGBA is the single copy glucocerebrosidase (GBA) vector directing secretion of the protein via the α-factor secretion signal (α). pαGBAD is the double cassette clone with a duplication of the GBA insert and flanking alcohol oxidase promoter (AOX5') and terminal (AOX3' or 3') regions. The Zeocin™ resistance gene (Zeo) and *E. coli* origin of replication (Col E1) are also noted. p*Op*fGBA is the *Sf*9 expression vector directing expression of the GBA cDNA with its full-length leader (fGBA) from the *Op*MNPV immediate-early promoter (OpIE2). p*Ops*GBA contains the GBA cDNA with a shorter, 19 amino acid leader (sGBA). *Op*IE2 polyadenylation signal (pA), *E. coli* origin of replication (Col E1), and Zeocin™ resistance gene (Zeo) are noted.

copy insert formation calls for a double digest of the recombinant plasmid with *Bgl*II and *Bam*HI to release the entire expression cassette (AOX5'-GBA-AOX3') for *in vitro* manipulation (pPICZα Manual, Invitrogen, Carlsbad, CA). Unfortunately GBA contains a *Bam*HI cut site and this enzyme could not be utilized. Multicopy insert formation requires the 5' and 3' restriction sites to yield complementary overhangs that can be ligated together forming a novel sequence unrecognized by either restriction enzyme. A search of *Bgl*ll complements revealed no enzymes with an appropriate unique site in the vector and no cut site in the insert. It was found, however; that the enzyme *Bsa*I did not cut the insert and digested the vector at two sites outside the expression cassette and a third at the 3' end of the AOX3' transcription termination region, 20bp upstream from the original *Bam*HI site. By PCR amplifying the expression cassette out of the vector rather than digesting it, the two external cut sites could be avoided allowing use of this enzyme for multicopy cassette formation. *Bsa*I is a site specific cutter with a recognition site upstream from a fully degenerate cut site that leaves a three base pair overhang (recognition sequence GGTCTCN↓NNN↑N) (New England Biolabs, Beverley, MA). Although the native sequence of the vector following the *Bsa*I recognition site would not result in a *Bgl*ll complementary overhang, an alteration of four bases at the actual cut site would result in the appropriate overhang without altering the recognition sequence.

Accordingly, a primer pair was designed to amplify the target expression cassette out of the vector and in the process mutagenize the 3' end to create the appropriate overhang following *Bsa*I digestion. The 5' primer (Primer C, Table 3.1) was complementary to positions 3584 to 15 of the circular pPICZαA plasmid including the *Bgl*ll cut site at

position 1. The 3' primer (Primer D) was antisense to positions 1671 to 1646 of the pPICZαA vector and included the *Bsa*I recognition sequence and four mismatches at the *Bsa*I cut site to produce the *Bgl*II complementary overhang. Amplification with these two primers liberated a 3.2kb fragment including the AOX5', GBA, and AOX3' sequences flanked by a 5' *Bgl*II cut site and a 3' *Bsa*I cut site. The insert was digested with these two enzymes and ligated into the original pαGBA vector linearized with *Bgl*II to created the double copy clone pαGBAD (Figure 3.1). Zeocin™ resistant clones were picked and the recombinants screened for correct second cassette insert orientation (clones KM26D and KM28D). Attempts to repeat this process and insert a third cassette into the vector were unsuccessful, probably due to the increasing size (12KB) of the potential recombinant plasmid. The double copy plasmid was transformed into KM71 *Pichia pastoris* without linearizing the plasmid as this would fragment the cassettes due to the duplication of the AOX5' region containing the linearization sites.

3.2.4 - *Zeocin™ Selection*

Increased GBA copy number integration was also mediated by Zeocin™ selection. Zeocin™ resistance is dose dependent in yeast and accordingly, multiple plasmid integration events can be selected by plating transformants on growth medium containing increasing concentration of the antibiotic (Higgins and Cregg, 1998). Normally selective at a concentration of 0.1μg/ml, Zeocin™ concentration was increased to 0.5μg/ml and 1.0μg/ml to select for increased plasmid integration, and increased GBA copy number.

3.2.5 - *Pichia Culture and Induction*

Pichia pastoris cells were grown in YPD liqiud medium or agar plates (2% peptone, 1% yeast extract, 2% dextrose, and 15g/l agar) at 30°C for all molecular manipulation. For expression studies small (10ml) pre-cultures in 50ml conical tubes covered with cheese cloth were grown overnight shaking at 250rpm, 30°C in BMGY (0.1M sodium citrate, pH 5.5, 2% peptone, 1% yeast extract, 1.34% yeast nitrogen base, 1% glycerol, $4x10^{-5}$% biotin). Precultures were diluted into a larger volume (500ml-1L) of BMGY in baffled Erlenmeyer flasks covered with cheese cloth and shaken overnight at 250rpm, 30°C to an $O.D._{600}$ of 2.0. Cultures were induced by pelleting the cells at 3,000x*g* for 10 minutes at room temperature and resuspending the cultures in a 1/5 to 1/10 volume of BMMY (0.1M sodium citrate, pH 5.5, 2% peptone, 1% yeast extract, 1.34% yeast nitrogen base, 0.5% methanol, $4x10^{-5}$% biotin) supplemented with 10mM DTT. This yielded a final culture OD_{600} of between 10 and 20. Cultures were shaken at 250rpm, 30°C for up to 144 hours with methanol added to 0.5%, DTT added to 10mM, and pH adjusted to 5.5 every 24 hours. Aliquots were also taken every 12-24 hours, the cells pelletted, and supernatant and cells stored at –70°C for further analysis.

3.2.6- *Sf9 Vector Construction*

All *Sf*9 GBA constructs were made using the vector p2Z*Op*2F, with transcription of the transgene driven from the *Orgyia pseudotsugata* nucleoplyherosis virus (*Op*NPV) immediate early promotor (Hegedus et al. 1998). This vector also contained a Zeocin™

resistance gene for the selection of recombinant plasmids in *E. coli* and selection of stable genomic integrants in transformed *Sf*9 cells. GBA constructs were cloned into the p2Z*Op*2F vector at the *Eco*RI site, downstream of the *Op*NPV immediate-early promotor to produce the p*Op*fGBA and p*Op*sGBA vectors (Figure 3.1). The two constructs were designed to direct transcription of the entire GBA coding region including both the full 39 amino acid (fGBA), or shorter 19 amino acid (sGBA) native leader sequences. The fGBA cDNA was *Pfu* PCR amplified using primers E and B while the sGBA cDNA was amplified using primers F and B (see Table 3.1). The original plasmids and engineering of the final constructs used were generously provided by Dr. T. Pfeifer and Dr.T.Grigliatti.

3.2.7 - *Sf9 Cell Transfection and Stable Cell Line Selection*

Prior to transfection, *Sf*9 cells were grown at 26°C to mid-log phase in ESF921 serum free medium (Expression Systems, Woodland, CA). For transient transfections, 1-2 x 10^6 cells were seeded in a 6 well tissue culture plate with 1 ml of Grace's minimal medium (Life Technologies Canada, Burlington, ON) and allowed to attach. CellFectin™ (Life Technologies Canada, Burlington, ON) lipofection reagent was mixed with 0.1μg of plasmid DNA in 1ml of Grace's medium, incubated for 30 minutes and then added to the seeded *Sf*9 cells. Following a 4 hour incubation, the transfection mix was replaced with 2ml of serum free medium and the cultures incubated at 26°C for 48 hours. For the selection of stable cell lines, Zeocin™ was added to the minimal medium following transfection at a concentration of 1mg/ml. At 48hrs the cells and medium were transferred to T25 tissue culture flasks and allowed to grow to confluence in the presence

of 1mg/ml Zeocin ™. The cells were then passed 2:1 and the selection on Zeocin™ was maintained for a total of three passages to obtain a stable polyclonal cell line.

3.2.8 – *Glucocerbrosidase Activity Assay*

Glucocerebrosidase activity in the crude cell lysates and culture medium was assayed using the artificial substrate 4-methyl-umbelliferyl-β-D-glycopyranoside (4MUGP) (Sigma-Aldrich Canada, Oakville, ON) and the natural substrate N-palmitoyl dihydroglucocerebroside (Sigma-Aldrich Canada, Oakville, ON). *Sf*9 cell lysates were prepared by repeated cycles of freeze/thaw to gently lyse the cells without affecting glucocerebrosidase activity. *Pichia pastoris* cell lysates were prepared by resuspending the cells in breaking buffer (50mM potassium phosphate, pH 6.0; 1mM EDTA; 5% v/v glycerol) supplemented with 1x Minicomplete protease inhibitors (Roche Diagnostics, Indianapolis, IN), and vortexing the suspension with an equal volume of acid washed glass beads (0.1mm diameter) (Sigma-Aldrich Canada, Oakville, ON.). Alternating 30s vortex and 30s ice incubation was performed for 10 minutes followed by a 5 minute centrifugation at 4,000x*g* to pellet the beads and cell debris. The resulting lysate was utilized for activity assay. Culture medium from both yeast and insect cultures was either utilized directly or the soluble proteins were concentrated by ammonium sulfate precipitation (50% w/v) prior to activity assay.

Glucocerebrosidase activity assay with 4MUGP was performed following the protocol of Choy (1984). Briefly, 4MUGP (3.5mM), citrate buffer (0.03M, pH 5.5), sodium taurocholate (0.1%w/v) and the glucocerebrosidase sample were mixed and incubated at

37°C for 30 minutes. The reaction was stopped by adding an excess of 0.2M glycine buffer (pH 10.5) and fluorescence assayed using a Sequoia-Turner Model 450 Fluorometer (Turner Designs, Sunnyvale, CA). Raw fluorescence readings were compared to a 4-methyl-umbelliferone (4MU) standard curve to calculate moles of 4MU released and the relative and/or specific activity of the sample. Natural substrate assays were conducted according to Choy and Davidson (1980) as follows. N-palmitoyl dihydroglucocerebroside was resuspended to 1mg/ml in a buffer containing 0.04M sodium citrate, pH 5.5, 0.8% w/v sodium taurocholate, and 0.1% v/v Triton X-100™. This solution was mixed 10:1 with the glucocerebrosidase sample and incubated at 37°C for 4-6 hours. The reaction was stopped by boiling the sample for 5 minutes followed by a 15 minute centrifugation at 14,000xg to pellet any precipitated components. The glucose released by this reaction was measured by combining the reaction supernatant 1:10 with glucose HK reagent (Sigma-Aldrich Canada, Oakville, ON) and incubation at 37°C for 10 minutes. This reaction couples hexokinase phosphorylation of glucose to NADPH production by glucose-6-phosphate dehydrogenase that was measured on a Spectronic Genesys 5 spectrophotometer (Thermo Spectronic, Rochester, NY) at a wavelength of 340nm. Endogenous glucose present in the culture medium and cell lysate samples was assayed using the identical conditions, in the absence of the glucocerebroside substrate. Net nmoles of glucose released was calculated using a glucose standard curve assayed under identical conditions for the determination of relative and specific glucocerebrosidase activities. All protein concentrations were calculated using the Biorad Reagent (Biorad, Hercules, CA) as adapted from Bradford (Bradford, 1976).

3.2.9 – *Protein Analyses*

Pichia pastoris and *Sf*9 cell lysates were prepared as described above for the analysis of intracellular protein expression. *Sf*9 and *Pichia* induction culture medium was either untreated or concentrated by ammonium sulfate (50% w/v), trichloroacetic acid (15% v/v), or acetone (100%) precipitation prior to tris-glycine, sodium dodecyl-sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) as described (Spector et al., 1998). Proteins were visualized on SDS-PAGE gels using Gel Code Blue stain (Pierce, Rockford, IL) or silver staining following as follows. Gels were microwaved at maximum power for 90 seconds in fixative (50% methanol, 12% acetic acid, 0.1% formaldehyde) followed by a 90s microwave in 50% ethanol. The gels were then pretreated in 0.02% sodium thiosulfate pentahydrate for 90s in the microwave, washed in deionized water for 90s at room temperature, and stained with 2mg/ml silver nitrate in 0.075% formaldehyde by microwaving twice for 40 seconds. Bands were resolved in developer (60mg/ml sodium carbonate, 0.05% formaldehyde, 0.002% sodium thiosulfate pentahydrate) and stopped in 50% methanol following a 90s water wash.

For western blotting, proteins were electroblotted from SDS-PAGE gels onto Hybond-P PVDF membrane (Amersham Pharmacia Biotech, Piscataway, NJ) for one hour at 100V or overnight at 20V in 10% methanol transfer buffer (25mM tris-HCl, 0.2M glycine) using a Mini-protean II Electroblot Apparatus (Biorad, Hercules, CA). PVDF membranes were washed in TTBS (20mM tris-HCl, pH 7.5, 0.05% Tween-20, 500mM sodium chloride) for 5 minutes at room temperature followed by one hour of blocking in

TTBS with 7.5% w/v dry skim milk powder. The membrane was washed twice for 5 minutes in TTBS before a one hour incubation with the glucocerebrosidase specific primary antibody diluted 1:200- 1:400 in blocking solution (7.5% dry milk in TTBS). The antibody used was a mouse monoclonal antibody, AA16B3, raised against the native human placental enzyme and donated by Dr. E. Beutler. Following the primary antibody, membranes were washed 4 times for 5 minutes in TTBS and incubated in a goat anti-mouse horseradish peroxidase (HRP) conjugated secondary antibody (Clontech, Palo Alto, CA) diluted 1:4000-1:6000 in blocking solution for 30 minutes to one hour. Membranes were washed 4 times in TTBS and incubated in ECL Plus chemiluminescent reagent (Amersham Pharmacia Biotech, Piscataway, NJ) for luminescent detection on Biomax autoradiography film (Eastman Kodak, Rochester, NY) or fluorescent detection using a Molecular Dynamics Storm 860 phosphorimager (Molecular Dynamics, Sunnyvale, CA). PNGase F digestion was performed as described in the product literature (Sigma-Aldrich Canada, Oakville, ON)

3.2.10 – *Southern and RNA Dot Blotting*

For Southern blot analysis, genomic DNA was extracted from *Pichia pastoris* cells using Zymolase™ (Seikagaku Kogyo Co., Tokyo) and potassium acetate lysis as described in the *Pichia* Expression Kit Manual (Invitrogen, Carlsbad, CA). DNA concentration was standardized to 20μg for all samples and digested overnight with 10-20 units of *EcoR*I restriction endonuclease (New England Biolabs, Beverley, MA). The digestions were concentrated by ethanol precipitation and loaded onto 0.7% agarose gels for electrophoresis at 2.5V/cm of gel length overnight. The DNA was transferred to

Hybond–N^+ nylon membrane by alkali capillary transfer overnight in 0.4M NaOH. A DNA probe representing 500bp (exons 4-7) of the GBA gene was PCR amplified using primers G and H (Table 3.1) and conjugated with alkaline phosphatase following the AlkPhos Direct Manual (Amershan Pharmacia Biotech, Piscataway, NJ). Membranes were hybridized with 10μg of labeled probe at 55°C overnight and washed as directed. Bands were developed by incubation in ECF chemifluorescent reagent (Amershan Pharmacia Biotech, Piscataway, NJ) as directed for 12-48 hours, and visualized using a Molecular Dynamics Storm 860 phosphorimager (Molecular Dynamics, Sunnyvale, CA).

GBA transcription was investigated by RNA Dot blotting of total RNA extracted from induced *Pichia pastoris* cultures. Total cellular RNA was isloated by SDS/phenol extraction and ethanol precipitation as described in the *Pichia* Expression Manual (Invitrogen, Carlsbad, CA). Total RNA concentrations were standardized and 10μg of each sample was denatured by heating to 65°C in a denaturation buffer of 2.2M formaldehye and 50% formanide in 0.5xMOPS buffer (3-(N-morpholino) propanesulfonic acid and 10 mM EDTA, pH 7.0), dotted onto Hybond-N^+ nylon membrane (Amershan Pharmacia Biotech, Piscataway, NJ) and crosslinked to the membrane by incubation at 120°C for 30 minutes. Probe preparation, hybridization and visualization were all performed as described for Southern Blotting above.

3.2.11 - *FPLC Purification*

Proteins were separated from *Pichia pastoris* induction culture medium and *Sf9* stable cell line medium by hydrophobic interaction chromatograhpy (HIC) with a Biogel TSK

phenyl-sepharose column (150mmx21.5mm) (Biorad, Hercules, CA) using a Pharmacia FPLC system (Ameraham Pharmacia Biotech, Piscataway, NJ). Culture medium was prepared for FPLC by the addition of ammonium sulfate to 1.7M, centrifugation at 14,000x*g* at 4°C for 30 minutes, and filtration/degassing through a 0.45μm membrane (Pall Gelman Laboratories, Ann Arbor, MI). Samples were loaded on to the column previously equilibrated with 1.7M ammonium sulfate at a flow rate of 5ml per minute and held at 100% buffer A (1.7M ammonium sulfate) until the void had passed and a stable baseline monitor reading (λ=230nm) was established. Proteins were eluted from the column with a linear gradient to 100% buffer B (0.1M NaCl) over 8 column volumes at a flow rate of 4ml per minute for *Pichia* samples. To elute possible highly hydrophobic species, as seen with native placental glucocerebrosidase (Choy 1990), the column was then run on a linear gradient to 100% buffer C (5% cholic acid) over 8 column volumes at 4ml per minute.

For *Sf*9 media samples the initial desalting gradient was run over only 4 column volumes and the cholate gradient extended to 12 column volumes due to differences in the elution profiles. Fractions were collected at 2 minute intervals and screened by 4MUGP assay and protein dot blot for the presence of glucocerebrosidase enzyme activity or cross-reactivity. 4MUGP assays were performed as described above with the following alterations. Total volumes were reduced to allow for a 96 well plate format and active fractions were determined by qualificication visually under ultraviolet light rather than quantification by a fluorimeter to speed the screening process. Protein dot blots were performed by prewetting PVDF membranes in methanol and washing in TTBS before

manually dotting 2.5µl aliqouts of each FPLC fraction onto the membrane. The membranes were dried to fix the protein dots and cross-reactive material determined following the immunodetection protocol described in the western blotting section. Volume reports were generated for each dot using the spotfinder function of the Storm 860 phosphorimager (Molecular Dynamics, Sunnyvale, CA) to estimate relative amounts of crossreactive material in each FPLC fraction as compared to the precolumn sample. Using these values an overall purification yield could be calculated.

3.3 – **Results**

3.3.1 – *Pichia Expression*

Initial experiments expressing human GBA in *Pichia pastoris* have shown that intracellular expression of GBA using its native leader sequence does not lead to the production of detectable activity or protein product (Wei, unpublished observation). Accordingly, expression clones were designed to secrete glucocerebrosidase into the culture medium using the α-mating factor secretion signal from *Saccharomyces cerevisiae.* Over 100 *Pichia* clones transformed with plasmids pαGBA and pαGBAD were selected on low, medium, and high zeocin concentrations and screened for the presence of the GBA insert. Figure 3.2 shows a Southern blot of three *Pichia* transformants selected for increasing GBA copy number as compared to a control cell line. Although absolute copy number was not determined, clones KM1G, KM26D and KM28D did show successively intense signals, indicating that the presence of a dual cassette plasmid and increased antibiotic selective pressure was successful in increasing

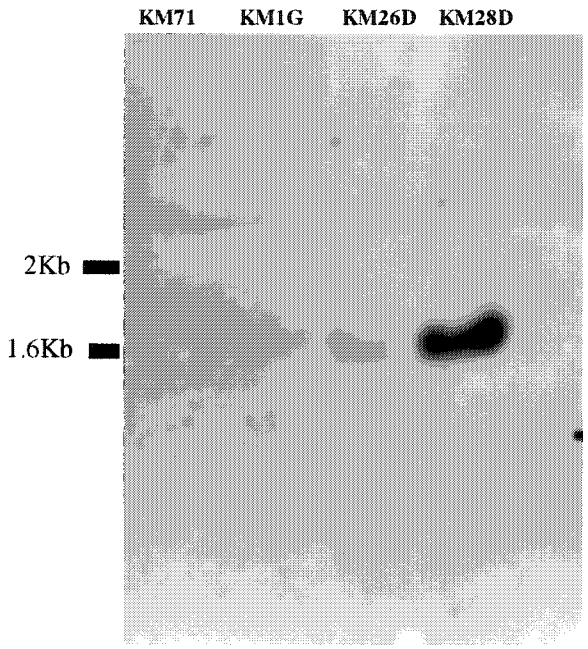

Figure 3.2 – Southern blot of genomic DNA extracted from *Pichia pastoris* strains KM71 (Lane 1), KM1G (Lane 2), KM26D (Lane3), and KM28D (Lane 4). KM71 is a non-recombinant control, while KM1G, KM26D, and KM28D have been selected on sequentially increasing Zeocin™ concentrations, for increased glucocerebrosidase (GBA) transgene integration. The extracted DNA was digested with *Eco*RI and hybridized with a DNA probe representing exons 4-7 of the GBA cDNA.

the copy number of the GBA cDNA in *Pichia pastoris*. These results were further supported by measuring GBA transcription using a total RNA dot blot. When probed for the presence of GBA transcripts, those colonies capable of growth under higher Zeocin™ concentrations (KM26D and KM28D) showed increased signal, roughly correlating gene copy number with GBA transcription rates (Figure 3.3). Interestingly, however, this correlation was not absolute and maximal Zeocin™ resistance did not necessarily correspond to maximal GBA transcription (i.e KM28D, Figure 3.3).

Initial expression screening with the artificial substrate 4MUGP was hampered by the presence of a non-specific, native glucosidase activity present in the induction medium of both transformed and untransformed clones (Table 3.2). Accordingly, clones were screened primarily by RNA dot blot of small scale cultures before scaling up the cultures for activity assay. The induction medium from selected clones was measured for activity using the natural glucocerebroside substrate and total activities ranged from approximately 30-80 nmol glucose/hr/ml of culture meduim as compared to 0-10 nmol glucose/hr/ml of culture meduim for non-transformed *Pichia pastoris* (KM71), and 267 nmol glucose/hr/ml cell extract for crude lysates of the human fibroblast enzyme (Table 3.2). Due to the low measureable quantity of protein in the culture medium, activities are presented as total activity/ml rather than specific activity /mg protein.

Western blotting of induction medium shows a series of crossreactive bands from GBA clone KM1G in comparison to a non-recombinant control strain (Figure 3.4). The putative identity of glucocerebrosidase as the major band at approximately 75kDa in the

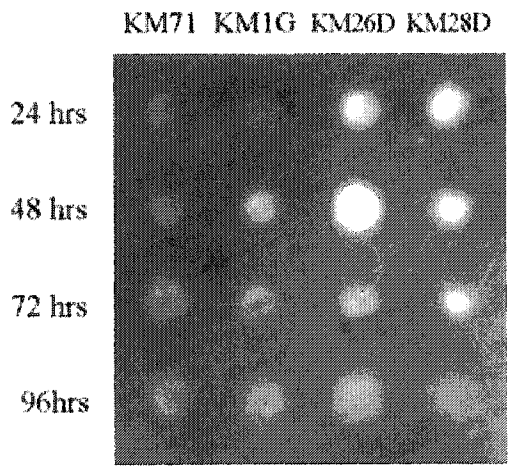

Figure 3.3 – RNA dotblots of total RNA extracted from induced *Pichia pastoris* strains at 24, 48, 72 and 96 hours post-methanol induction. KM71 is a non-transformed negative control; KM1G is a low copy GBA transformant; KM26D was selected on intermediate Zeocin™ concentrations for higher GBA copy number; KM28D was selected on maximal Zeocin™ concentrations. The dots represent 10μg of total RNA hybridized with a DNA probe including exons 4-7 of the GBA cDNA.

Table 3.2 – Glucocerebrosidase activity on the natural and artificial (4MUGP) substrates from *Pichia pastoris* expression clones.

Pichia Clone	Natural Substrate Activity[a] (nmoles/hr/ml culture medium)	4MUGP Activity (nmoles/hr/ml culture medium)
KM71 (-ve control)	10.24 +/- 13.8 (n=6)	3817 +/- 271 (n=3)
KM1G	30.24 +/- 3.47 (n=4)	3391 +/- 481 (n=3)
KM26D	76.04 +/- 11.6 (n=3)	3563 +/- 28.6 (n=3)
KM28D	28.16 +/- 6.68 (n=3)	2852 +/- 10.1 (n=3)
Fibroblast[b]	267.5 +/- 36.9 (n=4)	N/A

[a]Assays represent peak activities for cultures over 96 hours of induction with values presented as averages +/- standard error of the mean from (n) independent experiments.

[b]Fibroblast activities are nmoles/hr/ml cell lysate and have been corrected for relative differences in protein concentration as compared to average culture medium values.

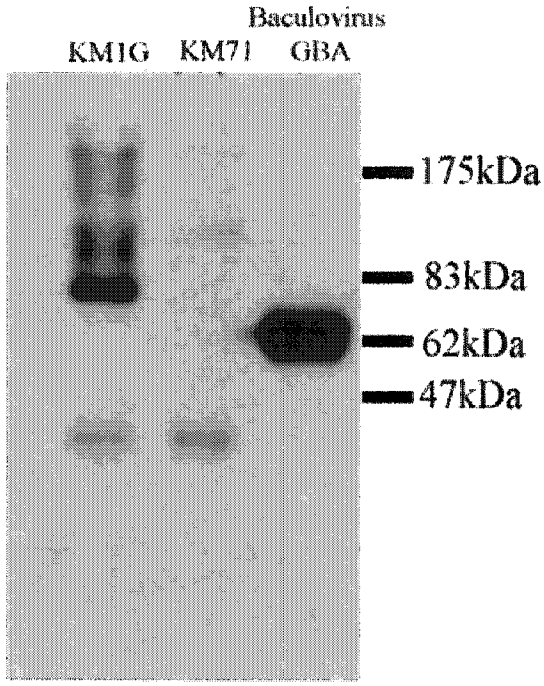

Figure 3.4 – Western blot analysis of glucocerebrosidase secretion from *Pichia pastoris*. KM1G (GBA transformant) and KM71 (non-transformed) *Pichia* clones were induced with methanol and culture medium sampled 48 hours post-induction. 60µl medium samples from equal cell density cultures were run on a 10% tris-glycine SDS-PAGE gel and immunodetected with an anti-glucocerebrosidase monoclonal antibody. A positive control sample of culture medium from human glucocerebrosidase expressed in a baculovirus expression system is presented in the right lane. Relative protein mass markers are noted (Broad Range Protein Standards, New England Biolabs, Beverley, MA).

GBA clone was not the expected size of GBA (62-67kDa) but was consistent in size across analyses of other GBA clones (data not shown). Silver or GelCode Blue stained SDS-PAGE gels, however, revealed a ladder of proteins present in the induction medium for both control and GBA expressing clones with no major band at 75kDa, although a band present in the KM26D GBA clone at ~80kDa was absent from the KM71 control strain (Figure 3.5).

Hydrophobic interaction chromatography was performed to purify the secreted glucocerebrosidase and to clarify the variable background seen in both activity assays and Western Blots. Figure 3.6 shows a comparison of chromatograms and activity assays from GBA high copy clone KM26D and control clone KM71. The major peak of 4MUGP activity that eluted in both strains at T=190 minutes was not associated with glucocerebrosidase induction and most probably reflected the activity of a native glucosidase secreted by *Pichia pastoris*. A second minor activity peak present only in the KM26D clone eluted from the column after the start of the cholate detergent gradient (~0.3% cholate) (Figure 3.6b). The active fractions from this FPLC run were pooled, separated by SDS-PAGE, transferred to PVDF membranes, and visualized by GelCode Blue™ staining (Figures 3.5) and western blotting (Figure 3.7). Both the stained membrane and western blot showed a complex pattern of bands with no major species corresponding to the expected 75kDa protein. Although the most prominent GelCode Blue™ stained band (see Figure 3.5) from the active fractions appeared larger than expected at ~80kDa, it was excised for N-terminal microsequencing. While the microsequencing yielded no discernable product, subsequent amino acid composition

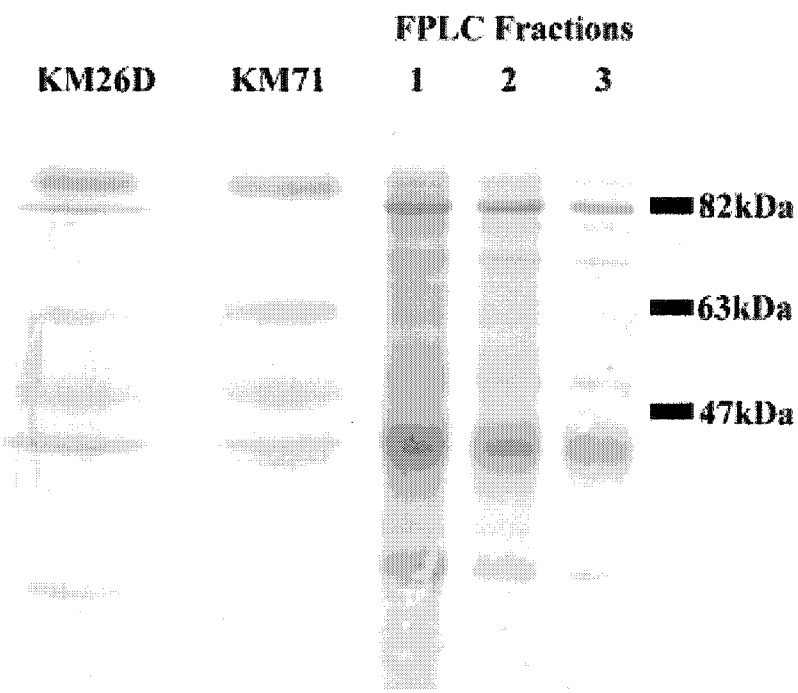

Figure 3.5 – SDS-PAGE and protein staining of culture medium and active FPLC fractions from glucocerebrosidase expression *Pichia* strain KM26D. From left to right, KM26D (GBA) and KM71 (control) culture medium 48 hours post-induction. FPLC pooled active fractions (1, 2,and 3) from hydrophobic interaction FPLC purification of KM26D culture medium 48 hours post-induction. Gels were stained with GelCode Blue™ (Pierce, Rockfor, IL). Protein standards were Broad Range Protein Standards (New England Biolabs, Beverley, MA).

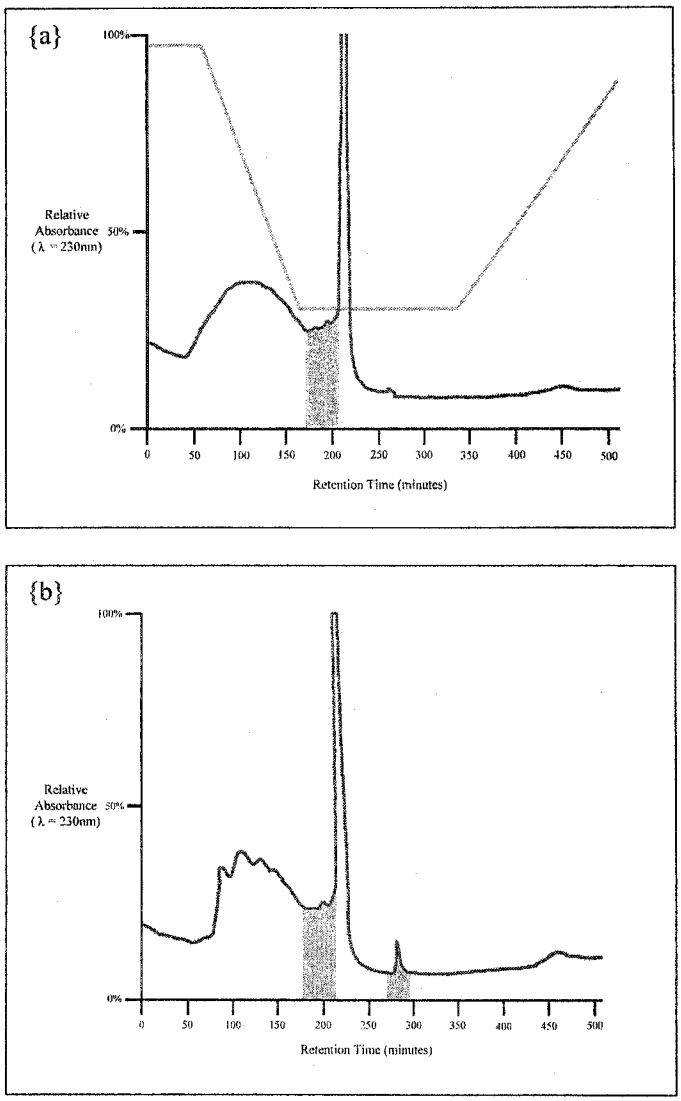

Figure 3.6 – Elution traces from hydrophobic interaction chromatography (HIC- FPLC) purification of 48 hour induction culture meduim from (a) KM71 control, and (b) KM26D glucocerebrosidase expression *Pichia* clones. Fractions displaying activity on the artificial substrate (4MUGP) are noted as shaded regions under the corresponding absorbance trace (λ=230nm). The line representing the desalting and cholate gradients is presented on (a) and was identical for both FPLC runs. Left to right the elution profile runs from 100% buffer A (1.7M ammonium sulfate) to 0% buffer A (0.1M NaCl) and finally to 100% buffer C (5% w/v cholic acid). The activity peak eluted near the end of the desalting gradient was present in both control (a) and GBA expression (b) cultures and does not represent glucocerebroidase activity. The second minor peak at the beginning of the cholate gradient in (b) corresponds to glucocerebrosidase.

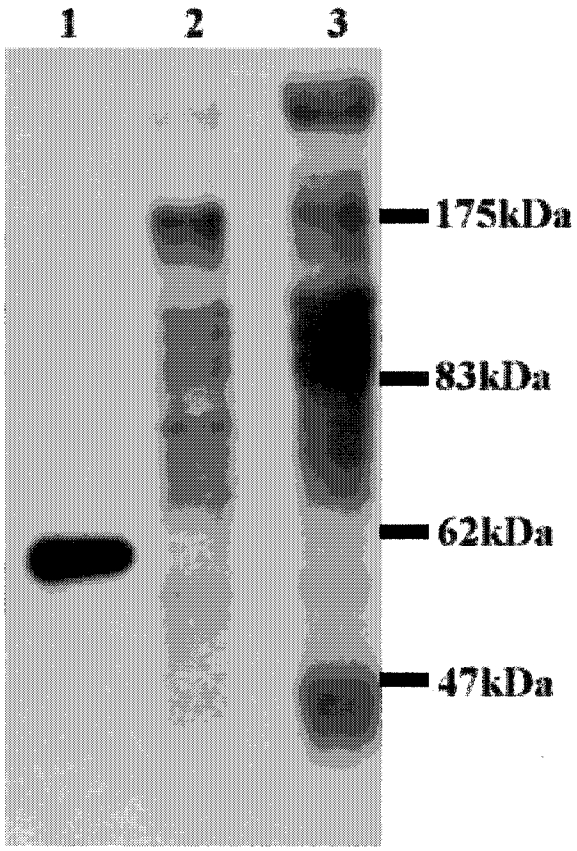

Figure 3.7 – Western Blot of *Pichia* strain KM26D (GBA) culture medium and the pooled active fraction from hydrophobic interaction chromatography. Lane 1= Baculovirus expressed human glucocerebrosidase positive control. Lane 2= Pooled active fractions from HIC-FPLC purification of 48hr induction culture medium from *Pichia* clone KM26D. Lane 3= KM26D culture medium prior to FPLC purification. 60µl samples were loaded onto a 10% tris-glycine SDS-PAGE gel. Immunodetection utilized human glucocerebrosidase specific monoclonal antibody AA16B3. Reletive protein standard positions are from Broad Range Protein Standards (New England Biolabs, Beverley, MA).

analysis of the sample revealed a sufficient quantity of protein was present, suggesting that the sample was blocked at its N-terminus. However, this complicated banding pattern was consistent over multiple culture conditions, strains, and column runs and due to the inconclusive nature of both western blotting and activity assays, attempts to further characterize the protein were unsuccessful.

3.3.2 – *Sf9 Expression*

Sf9 cells tranfected with the p*Op*fGBA and p*Op*sGBA vectors were selected for both transient GBA expression and stable genomic integration of the transgene. Due to the presence of the 19 or 39 amino acid native ER targeting signals on each construct respectively, it was hypothesized that measureable quantities of the nascent glucocerebrosidase would be directed to the insect cell lysosome or remain unlocalized intracellularly as is seen with baculovirus expression (Grabowski et al., 1989; Xu and Grabowski, 1998). Activity assay using the natural or fluorogenic substrates, however, failed to reveal any functional protein in crude cell lysates from transfected Sf9 cells. Figure 3.8 presents a western blot of crude cell lystates with minimal crossreactivity in the GBA transformed *Sf9* strains. A cell lysate showing strong crossreactivity from GBA expression using a baculovirus vector in *Sf9* cells is shown for comparison (Figure 3.8 and see Chapter 2). Analyis of culture meduim from transient expression of fGBA and sGBA in *Sf9* cells by Western Blotting displayed a doublet of cross reactive bands at approximately 60-63kDa (Figure 3.9). Glucocerebrosidase activity was also present in the culture medium when assayed with the natural and fluorogenic substrates (Table 3.3).

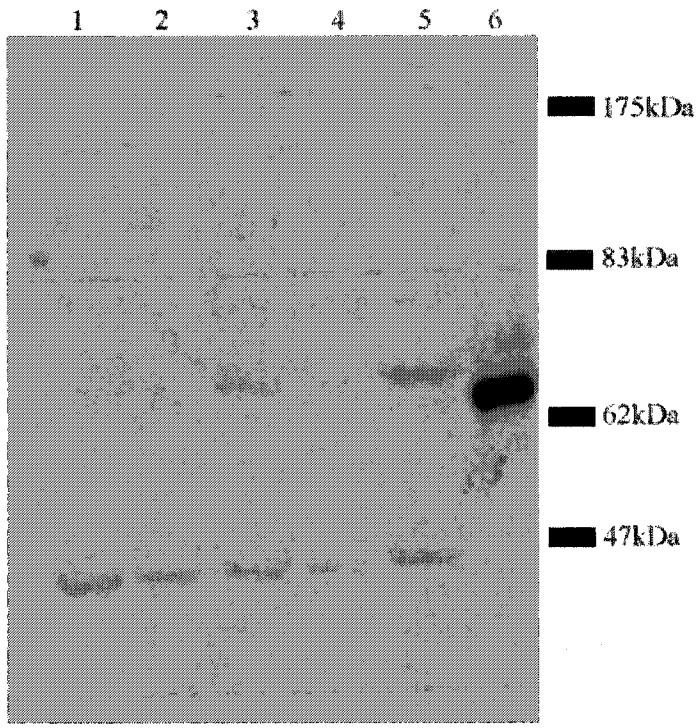

Figure 3.8 – Western blot of transient glucocerebrosidase (GBA) expression Sf9 cell lysates. *Sf*9 cells were harvested 3 days post-transfection and cell lysates prepared for SDS-PAGE and immunodetection using anti-human glucocerebrosidase monoclonal antibody AA16B3. Lane 1, non-transfected Sf9 cells; Lanes 2 and 3, fGBA (GBA with full-length native leader) transient clones; Lanes 4 and 5, sGBA (short native GBA leader) transient clones; Lane 6, cell lysate from baculovirus expressed (*Sf*9 cells) human glucocerebrosidase. Protein standards are Broad Range Protein Standards (New England Biolabs, Beverley, MA).

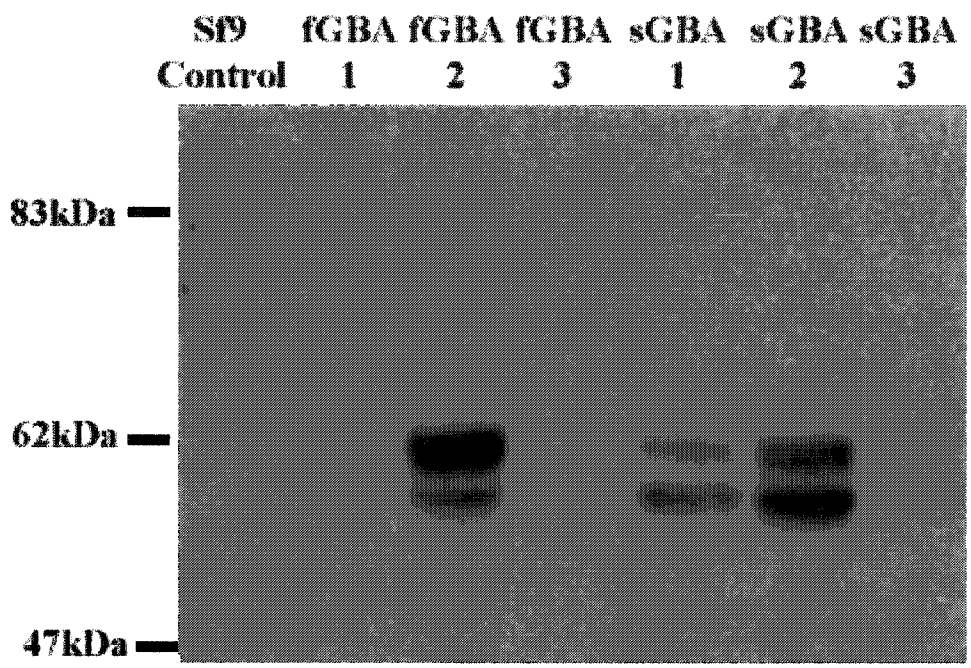

Figure 3.9 – Western blot of culture medium from transient transfected *Sf*9 cells expressing glucocerebrosidase (GBA). Three different clones for GBA constructs with the full-length (fGBA 1, 2, 3) or shorter (sGBA 1, 2, 3) native leader sequence are presented along with a non-transfected control *Sf*9 cell line. 60µl culture medium samples (3 days post-transfection) were loaded for 10% tris-glycine SDS-PAGE separation. Immunodetection was performed using a human glucocerebrosidase specific monoclonal antibody (AA16B3).

Table 3.3 – Glucocerebrosidase activities (4MUGP) from stable Sf9 transformants

Sf9 Clone	4MUGP Activity[a] (nmoles/hr/mg protein)
Control[b]	249.1 +/- 5.59
fGBA	1942 +/- 15.2
sGBA	1616 +/- 109

[a]Activities represent mean values +/- the standard error of the mean from 3 independent assays 72 hours post-seeding.

[b]Untransformed *Sf*9 cells.

Upon selection of stable expression transformants, there was a 30% difference in protein expression between the fGBA and sGBA clones with 4MUGP specific acitivities of 1942+/-15.2 nmoles/hr/mg protein for the fGBA clone as compared to 1616+/-109 nmoles/hr/mg protein for the sGBA clone. Hydrophobic interaction chromatography was performed to purify the glucocerebrosidase protein present in the culture medium of *Sf*9 cells selected for stable integration of the transgene. As can be seen in Figure 3.10 for the fGBA clone, a single broad peak of 4MUGP activity was eluted from the column at ~2.5% cholate and while no other active fractions were resolved, a second peak of crossreactivity was noted at the beginning of the cholate gradient equivalent to that seen with the *Pichia* expressed protein. A protein dot blot was performed to check all eluted fractions from the FPLC run for the presence of crossreactive material. The major peak in cross-reactive material correlated directly with the peak in 4MUGP activity, confirming the identity of the glucocerebrosidase protein. Western blot and GelCode Blue™ staining of the active fractions separated by SDS-PAGE confirmed the presence of a 63kDa crossreactive protein (Figure 3.11). Again, however, extended electrophoresis of the purified product resolved a doublet of bands at 60-63kDa with both protein staining and immunodetection, suggesting the presence of two distinct glucocerebrosidase species. Identical results were obtained for the purification of glucocerebrosidase from the sGBA clone (data not shown). PNGase F deglycosylation was performed on the purified glucocerebrosidase product and silver staining of the protein resolved the major 63kDa band to a single species of 58kDa, confirming the

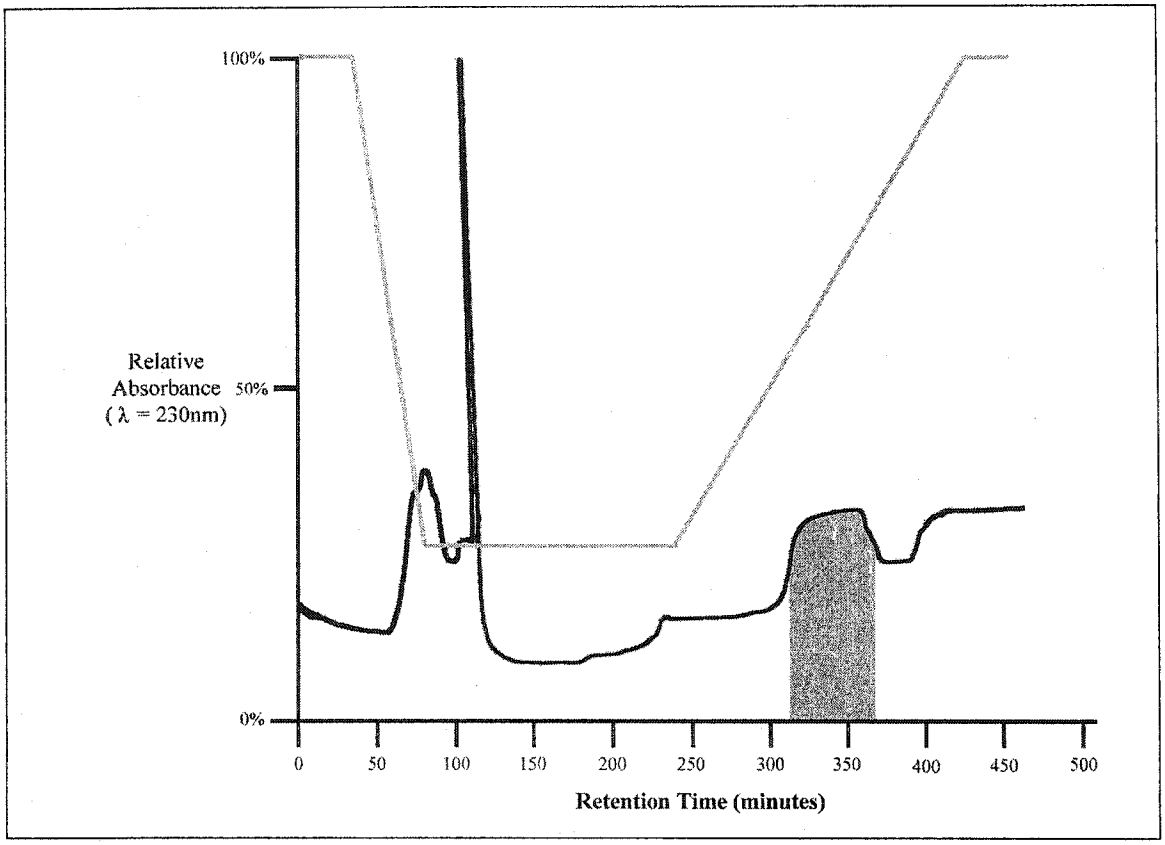

Figure 3.10 – Elution profile from hydrophobic interaction chromatography (HIC-FPLC) of stable transfected *Sf*9 cell medium expressing glucocerebrosidase construct fGBA (full-length native leader). The grey trace in the upper portion of the figure represents, from left to right, the desalting gradient from 100% buffer A (1.7M ammonium sulfate) to 0% buffer A (100% B, 0.1M NaCl) and the final cholate gradient to 100% buffer C (5%w/v cholic acid). Those fractions displaying activity on the artificial 4MUGP substrate are highlighted below the absorbance trace (λ=230nm). A single broad peak of active and crossreactive protein eluted at ~2-2.5% cholate.

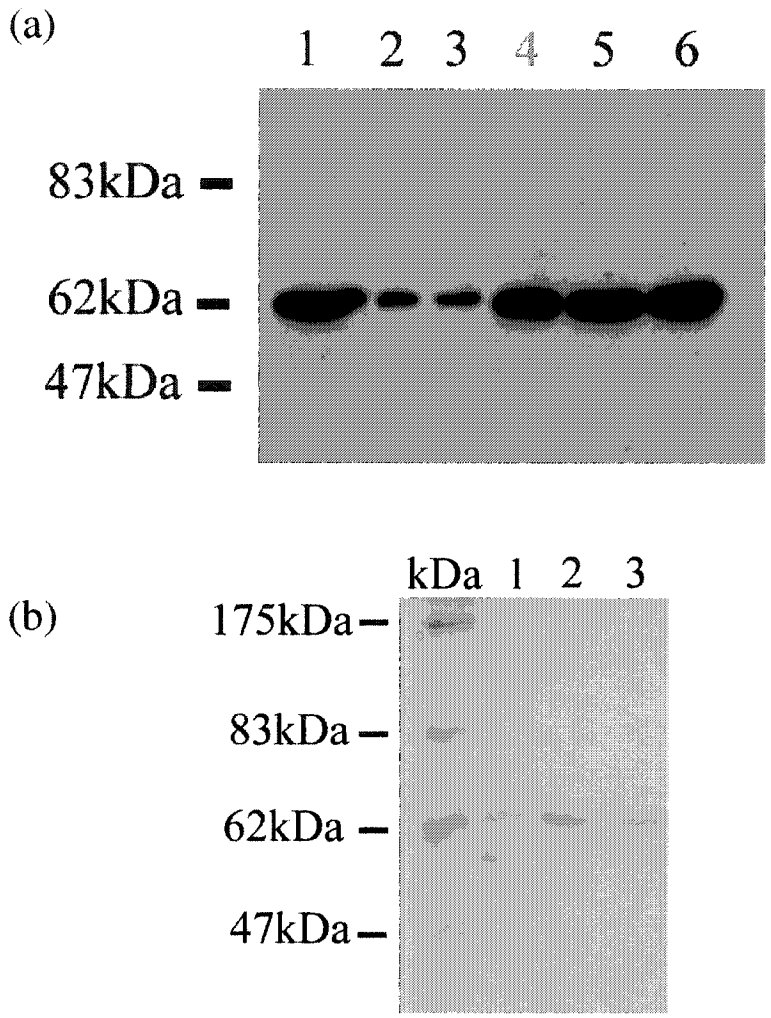

Figure 3.11 – Glucocerebrosidase purified from the culture medium of a stable Sf9 cell line expressing the glucocerebrosidase construct containing the full-length native leader sequence (fGBA). (a) Western blotting analysis of pre-colum media (Lane 1), and the minor (Lanes 2 and 3) and major (Lanes 4, 5, and 6) cross-reactive fractions eluted from the hydrophobic interaction colunm during the cholate gradient using human glucocerebrosidase specific monoclonal antibody AA16B3. (b) Gelcode Blue™ protein staining of the highly cross-reactive fractions (Lanes 1, 2 and 3) corresponding to Lanes 4, 5, and 6 of the western blot. Broad Range Protein Standards (New England Biolabs, Beverley, MA) were used for both (a) and (b).

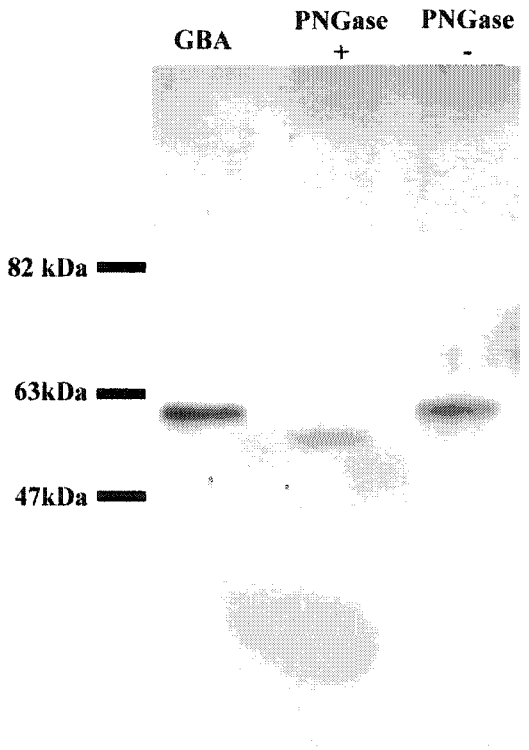

Figure 3.12 – PNGase F enzymatic deglycosylation of partially purified glucocerebrosidase produced in stable transfected *Sf*9 cells. The original glucocerebrosidase sample (GBA), and GBA incubated in digestion buffer in the presence (PNGase +) and absence (PNGase -) of the N-glycanase (PNGase F, Sigma-Aldrich Canada, Oakville, ON) are presented. Protein size markers correspond to the Broad Range Protein Standards (New England Biolabs, Beverley, MA). Proteins were separated on 10% tris-glycine gels and silver stained.

presence of N-linked glycans and suggesting that the lower band of the high resolution doublet represented a hypoglycosylated form of the secreted protein (Figure 3.12). The purified glucocerebrosidase was assayed using the artificial 4MUGP substrate under varying pH conditions to confirm its identity as an acid β-glucosidase. Figure 3.13 depicts the activity curve of the purified product with the activity peaking at a pH of 5.5.

Scanning volume reports were generated on protein dotblots from the FPLC using the Storm 860 phosphorimager (Molecular Dynamics, Sunnyvale, CA) to estimate the yield of the purification. As all the material eluted from the column was collected and sampled (including the flowthrough), volume reports from all collected fractions were summed and compared to dots of the sample prior to loading (data not shown). From this comparison it was calculated that 67% of the glucocerebrosidase crossreactivity applied to the column was contained within the pooled fractions representing the single activity peak. Although this analysis was an estimate of the glucocerebrosidase yield due to variations in crossreactivity outside the linear range of the instrument, a similar densitometric analysis of a western blot including pooled crossreactive fractions and the precolumn sample calculated a yield of 44% (data not shown). Based on a protein concentraion of 6.6μg/ml in the pooled cross-reactive glucocerebrosidase fractions and a purification yeild of 44-67% this suggests an initial protein production of 1.0-1.3mg/L of culture medium.

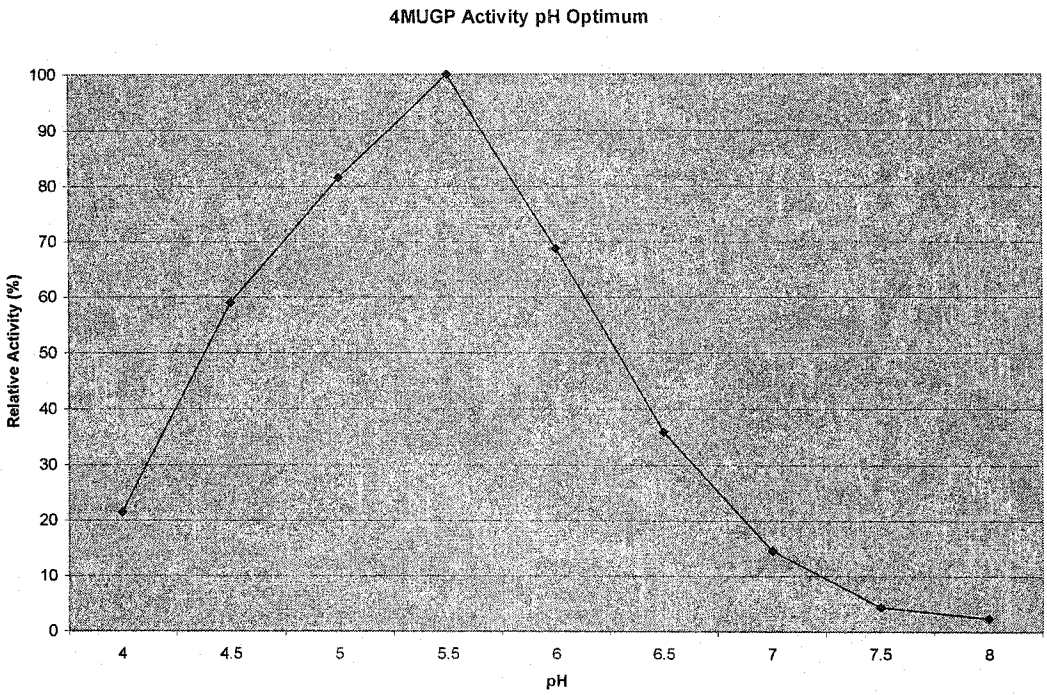

Figure 3.13 – pH optimum on the artificial substrate (4MUGP) activity of partially purified glucocerebrosidases produced in stable transfected *Sf*9 cells.

3.4 – **Discussion**

3.4.1 – *Pichia Expression*

The expression host *Pichia pastoris* has been shown to integrate multiple copies of a glucocerebrosidase transgene and produce significant quantities of the associated transcript. Based on both glucocerebrosidase activity assay and immunoblotting however, it appears that minimal amounts of active or crossreactive protein were being produced. While activity assay using the natural glucocerebroside substrate did suggest that high gene copy glucocerebrosidase transformants could secrete functional enzyme into the culture medium, high endogenous background activity was observed with the artificial (4MUGP) substrate disallowing an independent confirmation of these results. Hydrophobic interaction chromatography was successful in separating a minor peak of 4MUGP activity from the native *Pichia* non-specific activity, but subsequent immunoblotting of this active fraction failed to resolve consistently measurable quantities of crossreactive protein. These results suggest that although it is possible to secrete functional human glucocerebrosidase from the yeast *Pichia pastoris,* levels of protein production and/or stability are such that it is difficult to characterize the product. Whether this low production level is due to inefficient translational and/or post-translational processing of the enzyme, or low stability of the protein product in the yeast culture medium remains to be elucidated.

3.4.2 – *Sf9 Expression*

This low level of active protein production in *Pichia pastoris* is contrasted with the significant levels of active and crossreactive glucocerebrosidase produced by stable

transfected *Sf*9 cell lines. Interestingly, the protein from GBA constructs with either the full-length or truncated leader was secreted into the culture medium with only minimal levels of cross-reactive protein evident in the cells. The secretion of recombinant glucocerebrosidase appears to be a host-cell dependent phenomenon with some cell types secreting significant amounts of protein and others releasing no measurable glucocerebrosidase to the culture medium. Xu and Grabowski (1998) found that Chinese hamster ovary (CHO) and mouse myoblast (C2C12) cell lines expressing glucocerebrosidase via a retroviral vector secreted 50-75% of the recombinant glucocerebrosidase into the culture medium while transformed fibroblasts secreted no enzyme. Glucocerebrosidase secretion was also seen with *Sf*9 cells expressing GBA via a baculovirus vector in their study, with 50-75% of the enzyme activity present in the meduim (Xu and Grabowski, 1998). In contrast to these findings, Berg-Fussman et al. (Berg-Fussman et al., 1993) reported that little enzyme was secreted from *Sf*9 or COS-1 cells expressing glucocerebrosidase. Unfortunately, most other studies expressing glucocerebrosidase with baculovirus infected *Sf*9 cells focused on cellular activities and include no investigation of glucocerebrosidase secretion to the culture medium (Grace and Grabowski, 1990a; Grace et al., 1994). Observations from the baculovirus expressed glucocerebrosidase presented in Chapter 2 suggest that significant amounts of protein were also released in a baculovirus-based system, although formal comparative assays were not completed and enzyme release could simply reflect the lysis of cells in the late stages of viral infection (data not shown). Despite the conflicting data in the literature, it appears that plasmid-based *Sf*9 expression of glucocerebrosidase is capable of secreting the majority of the protein into the culture medium.

Although the native targeting sequence from glucocerebrosidase is required for localization to the endoplasmic reticulum (ER), the mechanism of lysosomal localization in human cells is independent of this sequence and the lysosomal targeting signal for glucocerebrosidase remains to be elucidated (Glickman and Kornfeld, 1993; Zimmer et al., 1999). The required signal may not be recognized or present in Sf9 cells (and other expression hosts) leading to the default secretion of the recombinant product seen. The near absolute secretion of glucocerebrosidase by stable transformed *Sf*9 cells as compared to the partial (or absent) secretion of baculovirus-based systems could result from the maintenance of regular cellular metabolism in plasmid transformed cells. Baculovirus infection of *Sf*9 cells inhibits the transcriptional activity of host genes in preference to those transcribed from the viral genome (Ailor and Betenbaugh, 1999; Hegedus et al., 1998). This leads to a saturation and degradation of cellular enzymes and cofactors involved in protein processing and can alter the maturation and trafficking of overexpressed proteins (Ailor and Betenbaugh, 1999; Hegedus et al., 1998).

The GBA construct containing both native initiator ATG's (fGBA) produced approximately 30% more enzymatic activity than the construct with only the second ATG and a shorter 19 amino acid leader (sGBA). The relative GBA gene copy numbers of these clones was not investigated and this alteration may reflect simple variations in gene dosage. This result is, however, in agreement with the overexpression of glucocerebrosidase in retrovirus transformed human cells (HeLa) where transcription from the upstream initiator ATG alone, or in concert with the second ATG, produced 25-

50% more protein activity than seen with only the second ATG (Pasmanik-chor et al., 1996). Although the fGBA construct could initiate translation from either in-frame ATG, these results suggest that the upstream initiator could be preferred in *Sf9* cells as is seen with both *in vitro* and *in vivo* mammalian translation of human glucocerebrosidase (Pasmanik-chor et al., 1996; Sorge et al., 1987). Western blots suggest that alteration in leader sequence length had little impact on glucocerebrosidase secretion in Sf9 cells as both constructs yielded minimal intracellular expression (Figure 3.8). Heterologous expression in murine cells (NIH/3T3) supports this finding, as no alterations in intracellular trafficking were observed between GBA constructs translated from either initiation codon (Sorge et al., 1987).

The presence of a 63-Kda species in the culture medium and partially purified active FPLC pools corresponds to the published size of *Sf9* expressed glucocerebrosidase (Grabowski et al., 1989; Grace and Grabowski, 1990a; Grace et al., 1990b; Grace et al., 1994). The presence of a second cross-reactive species of 60kDa suggests that not all of the secreted protein is fully processed or glycosylated. Enzymatic deglycosylation of the partially purifed product resolved a single 58kDa species as presented in Figure 3.12. A reduction of the 63Kda glucocerebrosidase species by approximately 2 kDa was observed by Grace and Grabowski (1990a) following PNGase digestion correlating with the removal of a single glycan from the mature protein. This would suggest that the minor 60kDa band observed represents a hypoglycosylated form, although altered size due to incorrect processing or proteolytic degradation cannot be ruled out.

The partially purified glucocerebrosidase shows activity towards both the artificial substrate (4MUGP) as stated above, and natural substrate at 330–668 nmoles/hr/mg. The purified human placental enzyme, partially purified with a similar hydrophobic column had a specific activity ranging from $2.8x10^4$-$1.5x10^5$ nmoles/hr/mg suggesting that the partially purified *Sf*9 enzyme has reduced specific activity as compared to the native enzyme (Choy, 1986; Furbish et al., 1977). Clearly, the insect cell culture conditions can be further optimized to maximize the stability, activity, and expression of the recombinant product. As the current production of catalytically active human glucocerebrosidase was estimated in mg/L quantities using a non-optimized construct, it appears that a stable transfected *Sf*9 cell expression system could be suitable for larger-scale expression of this valuable protein for basic research and potential biotherapeutics following the appropriate optimization and scale-up. Much work remains, however, to further analyze the biochemical properties of this recombinant glucocerebrosidase, with particular attention to the extent and pattern of N-glycosylation. Although *Sf*9 cells tend to only produce simple tri-and penta-mannosyl core N-glycan structures, a significant amount of research has been conducted recently into the engineering of the insect cell N-glycosylation pathway (Ailor and Betenbaugh, 1999). As the addition of the appropriate glycosyl-transferase genes to the host insect cell can lead to the production of mammalian-type complex N-glycans, it appears that any current limitations in recombinant glucocerebrosidase glycosylation in insect cells could be addressed (Jarvis et al., 1998; Wolff et al., 1999).

3.5 - **References**

Ailor, E. and Betenbaugh, M.J. Modifying secretion and post-translational processing in insect cells. *Current Opinion in Biotechnology* 10:142-145, 1999.

Altarescu, G., Schiffmann, R., Parker, C.C., et al. Comparative efficacy of dose regimes in enzyme replacement therapy of type I Gaucher disease. *Blood Cells, Molecules, and Diseases* 26:285-290, 2000.

Aviner, S., Levy, Y., Yaniv, L. and Cohen, I.J. Anaphylactoid reaction to Imiglucerase but not to alglucerase, in a type I Gaucher patient. *Blood Cells, Molecules, and Diseases* 25:90-92, 1999.

Barton, N.W., Brady, R.O., Dambrosia, J.M., et al. Replacement therapy for inherited enzyme deficiency - Macrophage targetted glucocerebrosidase for Gaucher's Disease. *New England Journal of Medicine* 324:1464-1470, 1991.

Barton, N.W., Furbish, F.S., Murray, G.J., Garfield, M. and Brady, R.O. Therapeutic response to intravenous infusions of glucocerebroisdase in a patient with Gaucher disease. *Proceedings of the National Academy of Sciences, U.S.A* 87:1913-1916, 1990.

Berg-Fussman, A., Grace, M.E., Ioannou, Y. and Grabowski, G.A. Human acid β-glucosidase. *Journal of Biological Chemistry* 268:14861-14866, 1993.

Beutler, E. Gaucher disease as a paradigm of current issues regarding single gene mutations of humans. *Proceedings of the National Academy of Sciences, U.S.A* 90:5384-5390, 1993.

Blissard, G.W. and Rohrmann, G.F. Baculovirus diversity and molecular biology. *Annual Review of Entomology* 35:127-155, 1990.

Bonning, B.C. and Hammock, B.D. Development of recombinant baculoviruses for insect control. *Annual Review of Entomology* 41:191-210, 1996.

Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anals of Biochemistry* 72:248-54.:248-254, 1976.

Brady, R.O. and Barton, N.W. Enzyme replacement therapy for Gaucher disease: Critical investigations beyond demonstration of clinical efficacy. *Biochemical Medicine and Metabolic Biology* 52:1-9, 1994.

Chiruvolu, V., Cregg, J.M. and Meagher, M.M. Recombinant protein production in an alcohol oxidase-defective strain of *Pichia pastoris* in fedbatch fermentations. *Enzyme and Microbial Technology* 21:277-283, 1997.

Choy, F.Y. Gaucher disease: the effects of phosphatidylserine on glucocerebrosidase from normal and Gaucher fibroblasts. *Human Genetics* 67:432-436, 1984.

Choy, F.Y.M. Purification of human placental glucoceerebrosidase using a two-step high-performance hydrophobic and gel permeation column chromatography method. *Analytical Biochemistry* 156:515-520, 1986.

Choy, F.Y.M. and Davidson, R.G. Gaucher disease III: Substrate specificity of glucocerebrosidase and the use of nonlabeled natural substrates for the investigation of patients. *American Journal of Human Genetics* 32:670-680, 1980.

Choy, F.Y.M., Wei, C. and Levin, D. Gaucher disease: Functional expression of the normal glucocerebrosidase and Gaucher T1366G and G1604A alleles in *Baculovirus*-transfected *Spodoptera frugiperda* cells. *American Journal of Medical Genetics* 65:184-189, 1996.

Clare, J.J., Romanos, M.A., Rayment, F.B., 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.

Couderc, R. and Baratti, J. Oxidation of methanol by the yest, Pichia pastoris. Purification and properties of the alcohol oxidase. *Agricultural and Biological Chemistry* 44:2279-2289, 1980.

Cregg, J.M., Madden, K.R., Barringer, K.J., Thill, G.P. and Stillman, C.A. Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. *Molecular and Cellular Biology* 9:1316-1323, 1989.

Furbish, F.S., Blair, H.E., Shiloach, J., Pentchev, P.G. and Brady, R.O. Enzyme replacement therapy in Gaucher's disease: large-scale purification of glucocerebrosidase suitable for human administration. *Proceedings of the National .Academy of Science, U.S.A.* 74:3560-3563, 1977.

Geisse, S., Gram, H., Kleuser, B. and Kocher, H.P. Eukaryotic expression systems: A comparison. *Protein Expression and Purification* 8:271-282, 1996.

Gellissen, G., Melber, K., Janowicz, Z.A., et al. Heterologous protein production in yeast. *Antonnie von Leeuwenhoek* 62:79-93, 1992.

Glickman, J.N. and Kornfeld, S. Mannose 6-phosphate-independent targeting of lysosomal enzymes in I-cell disease B lymphoblasts. *Journal of Cell Biology* 123:99-108, 1993.

Grabowski, G.A., Leslie, N. and Wenstrup, R. Enzyme therapy for Gaucher disease: the first 5 years. *Blood Reviews* 12:115-133, 1998.

Grabowski, G.A., White, W.R. and Grace, M.E. Expression of functional human acid β-glucosidase in COS-1 and *Spodoptera frugiperda* cells. *Enzyme* 41:131-142, 1989.

Grace, M.E. and Grabowski, G.A. Human acid β-glucosidase: glycosylation is required for catalytic activity. *Biochemical and Biophysical Research Communications* 168:771-777, 1990a.

Grace, M.E., Graves, P.N., Smith, F.I. and Grabowski, G.A. Analyses of catalytic activity and inhibitor binding of human acid β-glucosidase by site-directed mutagenesis. *Journal of Biological Chemistry* 265:6827-6835, 1990b.

Grace, M.E., Newman, K.M., Scheinker, V., Berg-Fussman, A. and Grabowski, G.A. Analysis of Human acid β-glucosidase by site-directed mutagenesis and heterologous expression. *Journal of Biological Chemistry* 269:2283-2291, 1994.

Hasnain, S.E., Jain, A., Habib, S., et al. Involvement of host factors in transcription from baculovirus very late promotors - a review. *Gene* 190:113-118, 1997.

Hegedus, D.D., Pfeifer, T.A., Hendry, J., Theilmann, D.A. and Grigliatti, T.A. A series of broad host range shuttle vectors for constitutive inducible expression of heterologous proteins in insect cell lines. *Gene* 207:241-249, 1998.

Higgins, D.R. and Cregg, J. *Pichia Protocols: Methods in Molecular Biology*, Totowa, NJ:The Humana Press, 1998.

Hollenberg, C.P. and Gellisson, G. Production of recombinant proteins by methylotrophic yeasts. *Current Opinion in Biotechnology* 8:554-560, 1997.

Jarvis, D.L., Kawar, Z.S. and Hollister, J.R. Engineering N-glycosylation pathways in the baculovirus-insect cell system. *Current Opinion in Biotechnology* 9:528-533, 1998.

Jenkins, N., Parekh, R.B. and James, D.C. Getting the glycosylation right: Implications for the biotechnology industry. *Nature Biotechnology* 14:975-981, 1996.

Liao, Y., Lal, A. and Moremen, K.W. Cloning, expression, purification, and characterization of the human broad specificity lysomal acid α-mannosidase. *Journal of Biological Chemistry* 271:28348-28358, 1996.

Martin, B.M., Tsuji, S., LaMarca, M.E., Maysak, K., Eliason, W. and Ginns, E.I. Glycosylation and processing of high levels of active human glucocerebrosidase in invertebrate cells using a baculovirus expression vector. *DNA* 7:99-106, 1988.

Miele, R.G., Nilsen, S.L., Brito, T., Bretthauer, R.K. and Castellino, F.J. Glycosylation properties of the Pichia Pastoris -expressed recombinant kringle 2 domain of tissue-type plasminogen activator. *Biotechnology and Applied Biochemistry* 25:151-157, 1997.

Mistry, P.K., Wraight, E.P. and Cox, T.M. Therapeutic delivery of proteins to macrophages: implications for treatment of Gaucher's disease. *Lancet* 348:1555-1559, 1996.

Montesino, R., Garcia, R., Quintero, O. and Cremata, J.A. Variation in N-linked oligosaccharide structures on heterologous proteins secreted by the methylotrophic yeast *Pichia pastoris*. *Protein Expression and Purification* 14:197-207, 1998.

Moremen, K.W., Trimble, R.B. and Herscovies, A. Glycosidases of the asparigine-linked oligosaccharide processing pathway. *Glycobiology* 4:113-125, 1994.

Moscicki, R.A. and Taunton-Rigby, A. Treatment of Gaucher's Disease. *New England Journal of Medicine* 328:1564-1568, 1993.

Pasmanik-chor, M., Elroy-Stein, O., Aerts, H., Agmon, V., Gatt, S. and Horowitz, M. Overexpression of human glucocerebrosidase containing different-sized leaders. *Biochemical Journal* 317:81-88, 1996.

Pfeifer, T.A. Expression of heterologous proteins in stable insect cell culture. *Current Opinion in Biotechnology* 9:518-521, 1998.

Richard, F., Robert, P., Remy, J., et al. High-level secretion of biologically active recombinant porcine follicle-stimulating hormone by the methylotrophic yeast *Pichia pastoris*. *Biochemical and Biophysical Research Communications* 245:847-852, 1998.

Scorer, C.A., Buckholz, R.G., Clare, J.J. and Romanos, M.A. The intracellular production and secretion of HIV-1 envelope protein in the methylotophic yeast *Pichia pastoris*. *Gene* 136:111-119, 1993.

Skory, C.D. and Freer, S.N. Expression and secretion of the Candida wickerhamii extracellular β-glucosidase gene, *bglB*, in *Saccharomyces cerevisiae*. *Current Genetics* 30:417-422, 1996.

Sorge, J.A., West, C., Kuhl, W., Treger, L. and Beutler, E. The human glucocerebrosidase gene has two functional ATG initiator codons. *American Journal of Human Genetics* 41:1016-1024, 1987.

Spector, D.L., Goldman, R.D. and Leinwand, L.A. *Cells, A Laboratory Manual*, Plainview:Cold Spring Harbor Laboratory Press, 1998.

Sreekrishna, K., Brankamp, R.G., Kropp, K.E., et al. Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*. *Gene* 55:1997.

Sreekrishna, K., Rica, H.B., Potenz, R., et al. High level expression of heterologous proteins in methylotrophic yeast *Pichia pastoris*. *Basic Microbiology* 28:265-278, 1988.

Tanner, W. and Lehle, L. Protein glycosylation in yeast. *Biochimica et Biophysica Acta* 906:81-99, 1987.

Tschopp, J.F., Sverlow, G., Kosson, R., Craig, W. and Grinna, L. High-level secretion of glycosylated invertase in the methylotrophic yeast, *Pichia pastoris*. *Biotechnology* 5:1305-1308, 1987.

Wegner, G.H. Emerging applications of the methylotrophic yeasts. *FEMS Microbiology Reviews* 7:279-283, 1990.

Wolff, M.W., Murhammer, D.W., Jarvis, D.L. and Linhardt, R.J. Electrophoretic analysis of glycoprotein glycans produced by lepidopteran insect cells infected with an immediate early recombinant baculovirus encoding mammalian beta1,4-galactosyltransferase. *Glycoconjugate Journal* 16:753-756, 1999.

Xu, Y.H. and Grabowski, G.A. Translational inefficiency of acid beta-glucosidase mRNA in transgenic mammalian cells. *Molecular Genetics and Metabolism* 64:87-98, 1998.

Zimmer, K., le Coutre, P., Aerts, J.M.F.G., et al. Intracellular transport of acid β-glucosidase and lysosome-associated membrame proteins is affected in Gaucher's disease (G202R mutation). *Journal of Pathology* 188:407-414, 1999.

Chapter 4 – Synonymous Codon Usage Bias and Translational Inefficiency in *Pichia pastoris*

4.1 – Introduction

The heterologous production of human glucocerebrosidase has been attempted in a number of expression systems including many mammalian cells, insect cells, transgenic tobacco plants, and now yeast cells with the range of protein production levels equalling the variety in host cells used (Martin et al., 1988; Liu et al., 1998; Cramer et al., 1996). While all the systems utilize highly efficient transcriptional promoters and/or high gene copy strategies to maximize transcription, this did not universally result in high protein production levels. Even the commercial product *Cerezyme*™ is faced with high production costs due to current limitations in enzyme yield (Mistry et al., 1996). In mammalian cells this translational inefficiency has been shown to result from the interaction of the glucocerebrosidase transcript with an RNA binding protein (TCP80/NF90) that subsequently inhibits translation (Xu and Grabowski, 1998; Xu and Grabowski, 1999). Although this inhibitory effect can be reconstituted by co-expression of GBA and TCP80/NF90 in insect cells, the protein is not naturally expressed in non-mammalian cells and was not implicated in translational inefficiency with other eukaryotic cell types (Xu and Grabowski, 2000). In fact, *Sf*9 cells have consistently produced relatively high levels of glucocerebrosidase as presented in Chapter 3 and previous published studies (Martin et al., 1988; Grace et al., 1994; Choy et al., 1996). In contrast, the consistently low levels of glucocerebrosidase expression in *Pichia pastoris* cells presented in Chapter 3 exist despite significant levels of transcript production, suggest that some level of translational inefficiency occurs in this yeast host. The evolutionary distance between the human target gene and yeast expression host brings

forth the possibility that differences in synonymous codon usage bias between the two species could impact translational efficiency.

The concept that organisms display a non-random pattern of synonymous codon usage was first established by Grantham et al. (1980) and has been well substantiated by the explosion of sequence data available as a product of recent genome sequencing projects (Nakamura et al., 1999). Although the neutral theory of molecular evolution suggests that mutations at degenerate coding positions should be selectively neutral and thus synonymous codon choice random, all organisms investigated to date show some general bias towards a subset of the 61 possible codons (Li and Graur, 1991; Nakamura et al., 1999). While a number of hypotheses have been but forward to explain the origin of this bias, a model involving selection for translational efficiency has been well supported for prokaryotes, unicellular eukaryotes, and to some extent, for insects (Powell and Moriyama, 1997; Bulmer, 1987; Bagnoli and Lio, 1995).

This so called "Major Codon Preference" model of codon bias stems from the observation that preferred codons are consistent across all genes within a given genome and correlate with the most abundant aminoacyl-tRNA's present in the cell (Moriyama and Powell, 1997). As individual codons have been found to be translated at different rates and the limiting factor in translational elongation appears to be the wait time for the appropriate aminoacyl-tRNA, it follows that to maximize protein translation rate, coding sequences should utilize codons matching the most common tRNA's (Bagnoli and Lio, 1995). Large-scale analysis of prokaryotic coding sequences has found that codon usage

bias levels positively correlate with the level of expression of the gene further supporting this hypothesis (Karlin et al., 1998; Nakamura and Tabata, 1997; Nakamura and Tabata, 1997). It is argued that the benefit of rapid translation of proteins required in high molar concentrations would select for high bias in these genes while selective pressures would be minimal on less abundant species and bias would lost to the effects of random mutation (Bagnoli and Lio, 1995).

Significant experimental work has been done to examine the effects of varying codon bias in heterologous expression systems with the assumption that disparate patterns of codon bias between the transgene and the expression host will have significant impact on the levels of recombinant protein produced. The complete optimization of coding regions towards the codon bias of the host cell has lead to an average 10 to 50 fold increase in heterologous protein production in a variety of host cells expressing an equally diverse range of foreign proteins (Nagata et al., 1999; Brocca et al., 1998; Apeler et al., 1997; Andre et al., 1998; Zolotukhin et al., 1996; Uchijima et al., 1998). Even the regional optimization of the 5' end of the coding region or the removal of only particularly rare codons throughout the cDNA has shown to have a significant impact on heterologous protein production (Kane, 1995; Forman et al., 1998; Zhou et al., 1999). Interestingly, however, the causal role of increased translational efficiency on many of these results is clouded by alterations in G+C content that are a by-product of changing codon bias. It appears that codon bias in mammals tends towards a high G+C content and that optimization towards the A+T biased codon choice of lower eukaryote hosts may affect mRNA secondary structure and increase protein production irrespective of the actual

codon choice (Johansson et al., 1999; Kim et al., 1997). Regardless of the mechanisms involved it was suggested that codon optimization of the glucocerebrosidase cDNA towards the bias of *Pichia pastoris* could have a positive impact on expression levels.

The success of such an endeavour is dependent, however, on the establishment of a clear set of preferred codons in the host cell, a table of which has yet to be determined for *Pichia pastoris.* Although the complete genome sequencing of *Saccharomyces cerevisiae* has allowed the determination of a definitive set of preferred codons for this yeast, the codon bias of *Pichia* may not necessarily reflect their evolutionary relatedness. It is possible to confirm the preferred codon choice of *Pichia pastoris* through a statistical analysis of bias in highly expressed genes from this organism using the program CodonW. Based on this table of preferred codons, it would also be possible to quantitate the level of codon bias disparity between this expression host and the GBA gene to confirm the speculation that differences in this codon choice are significant (Comeron and Aguade, 1998).

Due to the excessive cost of recoding the glucocerebrosidase cDNA (approximately $10,000 USD) it was felt pertinent to first establish a theoretical pool of preferred codons from the available *Pichia pastoris* sequence data and to test this model through the optimization of a short fragments of the GBA cDNA. The creation of multiple constructs both optimizing the codon region and altering the G+C content could tease apart the relative effects of these two related factors. By fusing these GBA fragments to the 5' end of a quantitative reporter gene such as luciferase, the efficiency of translational read-

through for the altered sequences relative to the native human clone could be determined. Luciferase is a well characterized reporter gene that has been successfully utilized to measure the effect of codon optimization on the heterologous expression of a *L. monocytogenes* antigen as a DNA vaccine in mice (Nagata et al., 1999). Accordingly, this study was undertaken to; (1) determine the preferred set of codons for *Pichia pastoris*; (2) measure the synonymous codon bias differences between this host and the glucocerebrosidase cDNA; (3) quantitate the impact of codon optimization on the read-through of a GBA-luciferase fusion construct; and (4) investigate the effect of codon optimization versus the alteration of G+C content without changing the codon bias of GBA.

4.2 – **Materials and Methods**

4.2.1- *Measures of Synonymous Codon Usage Bias*

The quantitation of synonymous codon usage bias in an organism or gene of interest can be accomplished by a variety of measures addressing both the magnitude and pattern of the observed bias. The measures used in this study are based on the relative synonymous codon usage (RSCU) of the data set measuring the observed frequency of a codon divided by the frequency expect under random usage of all possible synonymous codons for that amino acid (Sharp and Li, 1986). Once the RSCU is determined for each amino acid, a quantitative measure for the codon bias can be calculated as the Effective Number of Codons (ENC) utilized by the gene of interest (Wright, 1990). This measure is equivalent to the determination of the level of homozygosity at a locus, yielding values ranging from 20 for a gene with extreme bias using only one codon per amino acid, to 61

for a gene with no bias utilizing synonymous codons randomly (Wright, 1990). While this measure allows for the analysis of the magnitude of codon bias present in a gene, it does not address the pattern of that bias. As different organisms will show a bias towards different sets of preferred codons, an analysis of codon bias pattern is crucial to investigating the translational efficiency of heterologous coding regions.

The Codon Adaptation Index (CAI) allows such a measure of codon bias pattern by comparing those codons preferred by the host organism to the pattern of codon usage observed in the gene of interest (Sharp and Li, 1987). This measure is calculated by determining the relative adaptiveness for each codon in a gene in comparison with a preferred codon set. By dividing the mean RSCU for all codons in a gene by the mean RSCU of the corresponding synonymous codon preferred by the host, the CAI for a gene can be determined (Sharp and Li, 1987). Highly adapted genes with codon choice identical to that preferred by the organism in question will yield a CAI value of 1, while genes with low overall magnitudes of bias, or bias towards a different set of preferred codons with yield values approaching zero.

4.2.2 – *Correspondence Analysis of Codon Usage*

The determination of those synonymous codons preferred by an organism is not a trivial task and is tabulated based on a comparison of codon choice between those genes expressed in high quantities and those genes displaying a low level of expression. This follows from the assumption that selection for optimal codon choice will be greater in highly expressed genes and bias will be lost due to mutational drift and low selective

pressures in genes expressed in low levels (Sharp and Li, 1987). To be considered a preferred codon the RSCU of that codon in highly expressed genes must be significantly different from its RSCU value in low expressed genes based on a 2-way Chi-squared contingency test (p=0.001) (Sharp and Cowe, 1991). Although limited experimental data is available on the expression levels of the *Pichia pastoris* genes present in the public sequence databases, multivariate analysis of synonymous codon choice in these genes can be done to determine the major trends in codon usage and differentiate potentially highly expressed genes from those with potentially low expression levels. From this correspondence analysis (COA), a pool of preferred codons can be determined for *Pichia pastoris* and the appropriate measures of codon bias calculated.

CodonW (John Peden, Oxford University), an integrated codon bias and correspondence analysis program, was utilized to perform a multivariate analysis of codon usage in *Pichia pastoris.* A pool of coding sequences submitted from this organism were analyzed for RSCU and the genes ordered along 58 axes representing RSCU values for each of the 59 synonymous sense codons. Major trends within this pool of data were determined using measures of relative inertia (*eigenvalue*) and genes ordered according to their positions along these axes of major inertia. Assuming that the primary trend in the data correlated with gene expression level, a set of preferred codons for the analyzed genes was determined from the RSCU of genes at the extremities of the primary axis. Independent measures of codon bias (ENC) and G+C content were then used to determine if the ordination of genes on the major axis was correlated with expression level (inferred from ENC values), or another bias in nucleotide choice such as overall

G+C content, using Pearson's correlation coefficient in Excel (Microsoft, Redmond, WA). CodonW was then used to calculate the ENC, CAI (based on the calculated preferred codon set for *Pichia pastoris)*, and overall G+C content of the human glucocerebrosidase cDNA.

The pool of sequence data utilized for the determination of *Pichia pastoris* preferred codons was extracted from the Genbank database (June 7, 2001). Of a possible 73 *Pichia pastoris* coding sequences present in the database, 52 sequences representing complete coding regions, including annotated initiation and termination codons, were selected. The sequences were edited to remove 5' untranslated sequences so that position #1 of each entry corresponded to the A of the initiator ATG. Although both genomic and cDNA sequences were included in the data set, introns were manually edited out of the genomic DNA species to ensure appropriate *in silico* translation.

4.2.3 – *Luciferase Fusion and Plasmid Construction*

To determine if GBA coding sequences could inhibit the level of production of active luciferase in *Pichia pastoris*, in frame genetic fusions of varying length fragments of the GBA cDNA to the luciferase cDNA were constructed. The GBA cDNA was obtained as described previously (Choy et al., 1997) and the luciferase cDNA was excised from the pGL2 vector (Promega, Madison, WI). The appropriate GBA cDNA fragments were PCR amplified using *Pfu* polymerase (Stratagene, La Jolla, CA) and the primers used are presented in Table 4.1. All GBA fragments were amplified using A (Table 4.1) as the

Table 4.1 – Primers used in the construction of GBA–luciferase fusion vectors

Primer	Sequence[a]	Location[b]	Orientation
A	TATGAATTCCGCCCGCCCCTGCAT CCCT	GBA nt115-135	Sense
B	TCAGTGGCCTACTTGGCCCCGTGT GATTAG	GBA nt 303-290	Anti-sense
C	CTGTTGGGCCGGGTAGGCCCTAG CTCACGGGCAAT	GBA nt 911-895	Anti-sense
D	CTGTTGGGCCGGGTAGGCCCTTTA ATGCCCAGGCTG	GBA nt1658-1642	Anti-sense
E	GGAGGTACCATGGAAGACGCCAA AAACAT	pGL2 nt76-96	Sense
F	CCGTGGTACCCGCTGAATACAGT TACATTT	pGL2 nt1747-1726	Anti-Sense
G	GACTGGTTCCAATTGACAAGC	pPICZ nt855-875	Sense

[a]Restriction enzyme recognition sites introduced for cloning steps are presented in bold type.

[b]Nucleotide numbers identified as GBA represent glucocerebrosidase cDNA sequences with the first base of the upstream initiator ATG as position #1 [Sorge 1987]. pGL2 numbers represent the pGL2 (luciferase) plasmid (Promega, Madison, WI) with position #1 as the first C in the unique *Sma*I recognition site. Numbers for pPICZ represent the circular pPICZαA plasmid with position #1 as the first G in the *BamH*1 recognition site as presented in the pPICZα Manual (Invitrogen, Carlsbad, CA).

forward primer to introduce a flanking *EcoR*I site and allow for the cloning the fragments in frame with the α-factor secretion signal of the pPICZαA *Pichia pastoris* expression vector. This primer also effectively removed the native GBA leader sequence, initiating the cDNA at the first codon of exon 3 of the GBA gene. Primers B, C, and D were used as the reverse primers to amplify the 200bp (GB200), 900bp (GB900), and full-length GBA cDNA (GBA) fragments respectively (see Table 4.1). These primers incorporated a 3' flanking *Sfi*I restriction cut site for cloning into the corresponding site in the pPICZαA vector.

The *Photinus pyralis* luciferase cDNA was amplified out of the pGL2 vector using primers E and F, which incorporated flanking *Kpn*I restriction sites for cloning into the expression vector downstream of the GBA fragments (Table 4.1). These primers were designed to introduce the luciferase cDNA in frame with the GBA constructs or the α-factor secretion signal, including a 9 amino acid linker (amino acid sequence PSRPSRIGN) prior to the first luciferase codon necessitated by the cloning strategy and intervening vector sequences. The completed vector constructs pαLuc pαGB200Luc, pαGB900Luc and pαGBALuc are presented in Figure 4.1.

4.2.4 – *Construction of Codon Optimized and G+C Altered GBA Fragments*

The 200bp GBA fragment (GB200) encoding the first 60 amino acids of the mature polypeptide (excluding the native leader sequence) was utilized for codon optimization and G+C content alteration studies. Using the table of preferred *Pichia pastoris* codons as a guide (Table 4.2), the codon choice of the native sequence was altered to create a

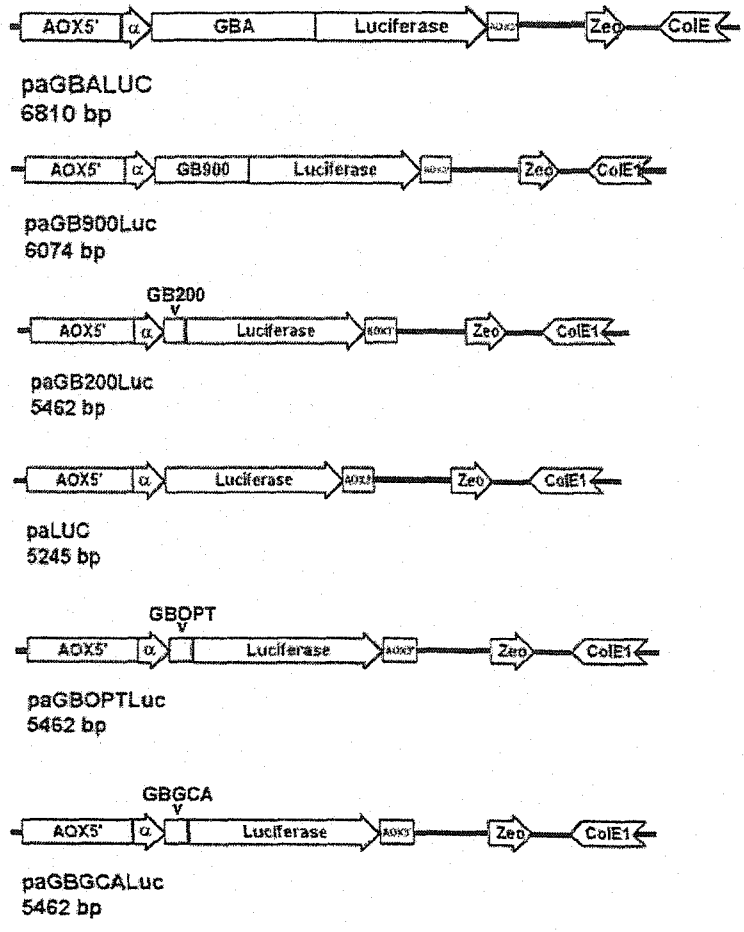

Figure 4.1 – Glucocerebrosidase and luciferase fusion vectors constructed for the expression analysis of codon usage bias. Expression for all constructs was initiated from the inducible alcohol oxidase promoter (AOX5') for secretion using the α-factor secretion signal present in the pPICZαA vector backbone (Invitrogen, Carlsbad, CA). Regions of AOX1 homology (AOX5'and AOX3') directed integration of the expression constructs into the AOX1 locus in the *Pichia pastoris* KM71 host. The *Photinus pyralis* luciferase cDNA insert (Luciferase), *E. coli* origin of replication (ColE1) and Zeocin™ selectable marker (Zeo) are noted. GBA= the full-length glucocerebrosidase cDNA; GB900= 900bp from the 5' end of the GBA cDNA; GB200= 200bp from the 5' end of the GBA cDNA; GBOPT= Codon optimized GB200; GBGCA= G+C content altered GB200. All GBA constructs exclude the sequences corresponding to the native glucocerebrosidase leader with a 5' terminus at cDNA nt115 (Sorge et al., 1987).

Table 4.2 – Comparison of preferred codons from *P. pastoris* and *S. cerevisiae*

Amino Acid	*P. pastoris* Codon[a]	*S. cerevisiae* Codon[b]	Amino Acid	*P. pastoris* Codon	*S. cerevisiae* Codon
Phe	TTC	TTC	Glu	GAG	GAA
Leu	TTG	TTG	Ser	TCT/TCC	TCT
Ile	ATC	ATC	Pro	CCA	CCA
Val	GTT/GTC	GTT	Thr	ACC	ACC
Tyr	TAC	TAC	Ala	GCT/GCC	GCT
His	CAC	CAC	Arg	AGA	AGA
Asn	AAC	AAC	Gly	GGT	GGT
Lys	AAG	AAG	Gln	CAA/CAG	CAA
Asp	GAC	GAC	Cys	TGT/TGC	TGT

[a]*Pichia pastoris* preferred codons were determined from a correspondence analysis of 52 full-length coding regions. For those amino acids with two codons listed, the difference in usage between the two was not statistically significant based on a 2-way Chi-squared contingency test (p=0.001).

[b]*Saccharomyces cerevisiae* prefered codons are from published empirical data (Sharp and Li, 1987).

fully codon optimized fragment (GBOPT) with a *P. pastoris* CAI value of 1.0. This optimization procedure resulted in a significant decrease in the overall G+C content of the fragment due to the relatively lower G+C bias of *P. pastoris* preferred codons. Accordingly, a second construct (GBGCA) was designed to maintain the *P. pastoris* CAI of the native GBA sequence while decreasing the G+C content overall to match that of the GBOPT fragment. This was achieved by changing all available degenerate positions to A or T nucleotides without introducing a preferred codon or removing a codon considered rare in *Pichia pastoris*. The sequences and alterations introduced are presented in Figure 4.2, aligned with the native GBA sequence.

To construct these altered GBA fragments, a series of overlapping 48bp primers were synthesized for each construct (Life Technologies Canada, Burlington, ON) to span the entire 200bp sequence and introduce the required basepair substitutions. These primers, presented in Table 4.3, were ordered as alternating forward and reverse (complimentary) sequences with each primer overlapping the previous one by 15bp. *EcoR*I and *Sfi*I cut sites were included at the 5' and 3' termini, respectively, for cloning into the expression vector.

The GBOPT construct was assembled in a stepwise fashion using four PCR reactions. Three separate reactions were set-up with primer pairs OPT1/OPT2, OPT3/OPT4, and OPT5/OPT6 respectively (0.5μM final concentration of each primer) in a *Pfu* PCR mixture (1X Native *Pfu* buffer, 0.2mM dNTP, 1U *Pfu* polymerase) (Stratagene, La Jolla, CA). Amplification proceeded with 30 cycles of 30s at 94°C, 30s at 63°C, and 1 min. at

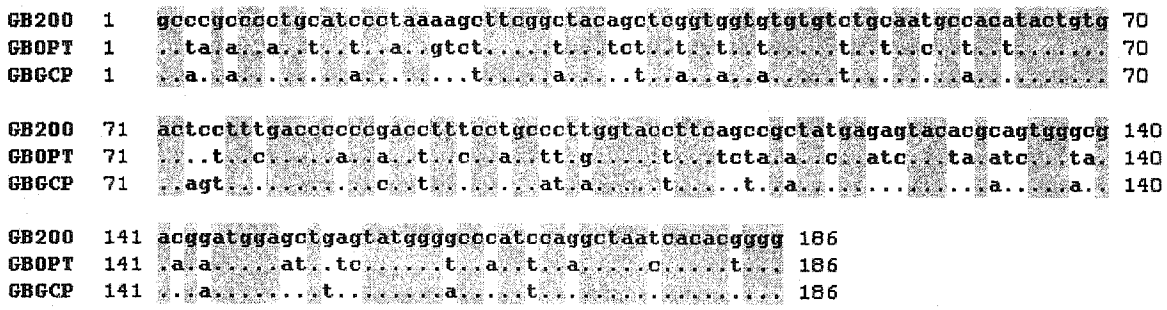

Figure 4.2 – Sequence alignment of the first 200bp of the native glucocerebrosidase cDNA (excluding the native leader sequence) with the *Pichia pastoris* codon optimized (GBOPT) and G+C altered (GBGCA) constructs. All regions of sequence identity are highlighted and noted as a dot.

Table 4.3 – Primers used for the construction of codon altered GBA fragments.

Primer	Sequence (5'>3')[a]	Orientation
OPT1	TAT**GAATTC**GCTAGACCATGTATTCCAAAGTCTTTCGGT TACTCTTCT	Sense
OPT2	GTCACAGTAAGTAGCGTTACAAACACAAACAACAGAAG AGTAACCGAA	Anti-sense
OPT3	GCTACTTACTGTGACTCTTTCGACCCACCAACTTTCCCA GCTTTGGGT	Sense
OPT4	ACCAGATCTAGTAGATTCGTATCTAGAGAAAGTACCCA AAGCTGGGAA	Anti-sense
OPT5	TCTACTAGATCTGGTAGAAGAATGGAATTGTCTATGGG TCCAATTCAA	Sense
OPT6	TCAGT**GGCCGGCTGGGCC**CAGTGTGGTTAGCTTGAAT TGGACCCAT	Anti-sense
GCA1	TAT**GAATTC**GCACGACCCTGCATACCTAAAAGTTTCGG ATACAGTTCA	Sense
GCA2	GTCACAGTATGTTGCATTGCAAACACATACTACTGAACT GTATCCGAA	Anti-sense
GCA3	GCAACATACTGTGACAGTTTTGACCCCCCCACTTTTCCT GCATTAGGT	Sense
GCA4	TCCACTTCGTGTACTCTCATATCGACTGAAAGTACCTAA TGCAGGAAA	Anti-sense
GCA5	AGTACACGAAGTGGACGACGAATGGAGCTTAGTATGGG ACCCATTCAG	Sense
GCA6	TCAGT**GGCCGGCTGGGC**CCCGTGTGATTAGCCTGAAT GGGTCCCAT	Anti-sense

[a]Sequences added to introduce restriction endonuclease recognition sites are presented in bold type.

72°C in a Genamp thermocycler (Perkin Elmer, Boston, MA). Following this initial PCR, 20µl aliquots from each reaction were pooled, 1ul of *Pfu* polymerase and 5µl of 2mM dNTPs added, and four more thermocycles of the previous program repeated. At the start of the fifth cycle, primers OPT1 and OPT6 were added to 0.5µM and the thermocycles continued for a total of 30 cycles. This procedure allowed for the assembly of the full-length 200bp product with only four total PCR reactions. An identical protocol was used for the construction of the GBGCA fragment using the appropriate GCA1-GCA6 primers. Attempts to construct these fragments using all six corresponding primers in a single PCR reaction failed to produce sufficient quantities of full-length product.

These constructs, GBOPT and GBGCA, were then ligated into the pPICZαA vector using the terminal *EcoR*I and *Sfi*I cut sites and sequenced to ensure the fidelity of the PCR and cloning steps. Clones with the correct sequence were then utilized for the introduction of the luciferase cDNA via the downstream *Kpn*I site in the vector as described above. The completed vectors pαGBOPTLuc and pαGBGCALuc are presented in Figure 4.1.

4.2.5 - *Pichia pastoris Transformation*

The various GBA-Luciferase fusion constructs (Figure 4.1) were isolated, the purified vectors linearized at the *Bst*XI cut site in the 5' alcohol oxidase (AOX1) sequence to promote homologous recombination, and electroporated into KM71 *Pichia pastoris* cells

following the *Pichia* Expression Manual protocol (Invitrogen, Carlsbad, CA) using a Gene Pulser electroporator (BioRad). Transformants were selected by resistance to Zeocin™ (0.1mg/ml) on YPDS plates according to the pPICZα Manual (Invitrogen, Carlsbad, CA) and the presence of genomic integration confirmed by direct yeast PCR using primers G and F as follows (Table 4.1). Sterile toothpicks were used to touch the surface of isolated transformant colonies and the small amount of cells resuspended in 10μl of sterile water. Zymolase™ was added to 1mg/ml final concentration and the samples incubated at 30°C for one hour. The tubes were flash frozen in liquid nitrogen for 1 minute and the resulting lysate used as template for a standard PCR reaction as described above for the initial construction of the clones. Primer G was complementary to bases 855-875 of the 5' AOX1 fragment in the pPICZαA vector and primer F was a reverse complement to the 3' end of the luciferase insert (Table 4.1). Accordingly, this PCR confirmed both the presence of the genomic integration and correct alignment of the overall construct.

4.2.6 – *Pichia Expression and Luciferase Assay*

Transformed *Pichia* clones were cultured and induced as describe previously (Chapter 3) and aliquots of cells and culture media taken every 24 hours and frozen at –20°C. Luciferase activity assay was performed on culture medium using the Promega Luciferase Assay Kit (Promega, Madison, WI) and quantitated using a Wallace 1410 Liquid Scintillation counter (Amersham Pharmacia Biotech, Piscataway, NJ). Briefly, 10μl of culture medium was mixed with 50μl of luciferase assay reagent in a 0.5ml microfuge tube. This tube was placed in a capped scintillation vial and the light emission

summed over 1 minute with the coincidence circuit off. There was a 30 second delay between the initiation of the reaction and the start of signal summation to allow the reaction to reach maximal light output as described in the Luciferase Assay Kit Manual (Promega, Madison, WI).

4.2.7 – *Northern Blotting*

Total RNA was extracted from the induction clones following the SDS/phenol extraction and ethanol precipitation protocol described in the *Pichia* Expression Manual (Invitrogen, Carlsbad, CA). Total RNA concentrations were standardized and 15µg of each sample was denatured by heating to 65°C in denaturation buffer of 2.2M formaldehye and 50% formamide in 0.5xMOPS buffer (0.2M 3-(N-morpholino propanesulfonic acid and 10 mM EDTA, pH 7.0) and prepared for electrophoresis by the addition of 0.5% w/v bromophenol blue in 50% glycerol. The samples were electrophoresed in a 1% agarose MOPS/formaldehyde (1xMOPS buffer, 6% formaldehyde) gel at 70V for 1.5 hours in 1xMOPS running buffer. The RNA was transferred to Hybond-N^+ nylon membrane (Amersham Pharmacia Biotech, Piscataway, NJ) by capillary transfer overnight in 20xSSC (3M NaCl; 0.3 M Na-citrate) and crosslinked to the membrane by incubation at 80°C for 4 hours. A DNA probe representing the full-length the luciferase cDNA was PCR amplified using primers E and F (Table 4.1) and conjugated with alkaline phosphatase following the AlkPhos Direct Manual (Amersham Pharmacia Biotech, Piscataway, NJ). Membranes were hybridized with 10µg of labeled probe at 55°C overnight and washed as directed. Bands were developed by incubation in ECF chemifluorescent reagent (Amersham Pharmacia Biotech, Piscataway, NJ) as directed for

12-48 hours, and visualized using a Molecular Dynamics Storm 860 phosphorimager (Molecular Dynamics, Sunnyvale, CA).

4.3 – Results

4.3.1 – *Correspondence Analysis of Codon Usage*

Correspondence analysis of RSCU for all codons in the 52 *Pichia pastoris* genes analyzed identified a single major trend in the RCSU that explained 34.9% of the inertia in the data set. The second axis of inertia explained only 7% of the total inertia confirming that the primary axis was the main factor explaining RSCU in this data set. A scatter plot of the ordination of genes along the primary axis as compared to the calculated effective number of codons (ENC) is presented in Figure 4.3a. There is a strong correlation between the two factors (r=0.932) further supporting the assumption that genes on the primary axis are ordered based on expression level rather than unrelated nucleotide bias (Sharp and Cowe, 1991). The primary axis also shows some correlation with G+C content (r=-0.713) suggesting that higher expressed gene do have a higher G+C content, but with variation only between 40-50% G+C it appears *Pichia pastoris* genes as a whole have a slight A+T bias (Figure 4.3b). An examination of genes at either end of the primary axis further supported the hypothesis that the primary axis was ordered according to gene expression levels. Those genes at the left extremity, correlated with low ENC values should be highly expressed genes. The first five genes along the axis include the alcohol oxidase (AOX1) and glyceraldehyde-3-phosphate dehydrogenase (GAP) genes, both of which have been well characterized as highly expressed genes and have been utilized for heterologous expression studies due to the strength of their

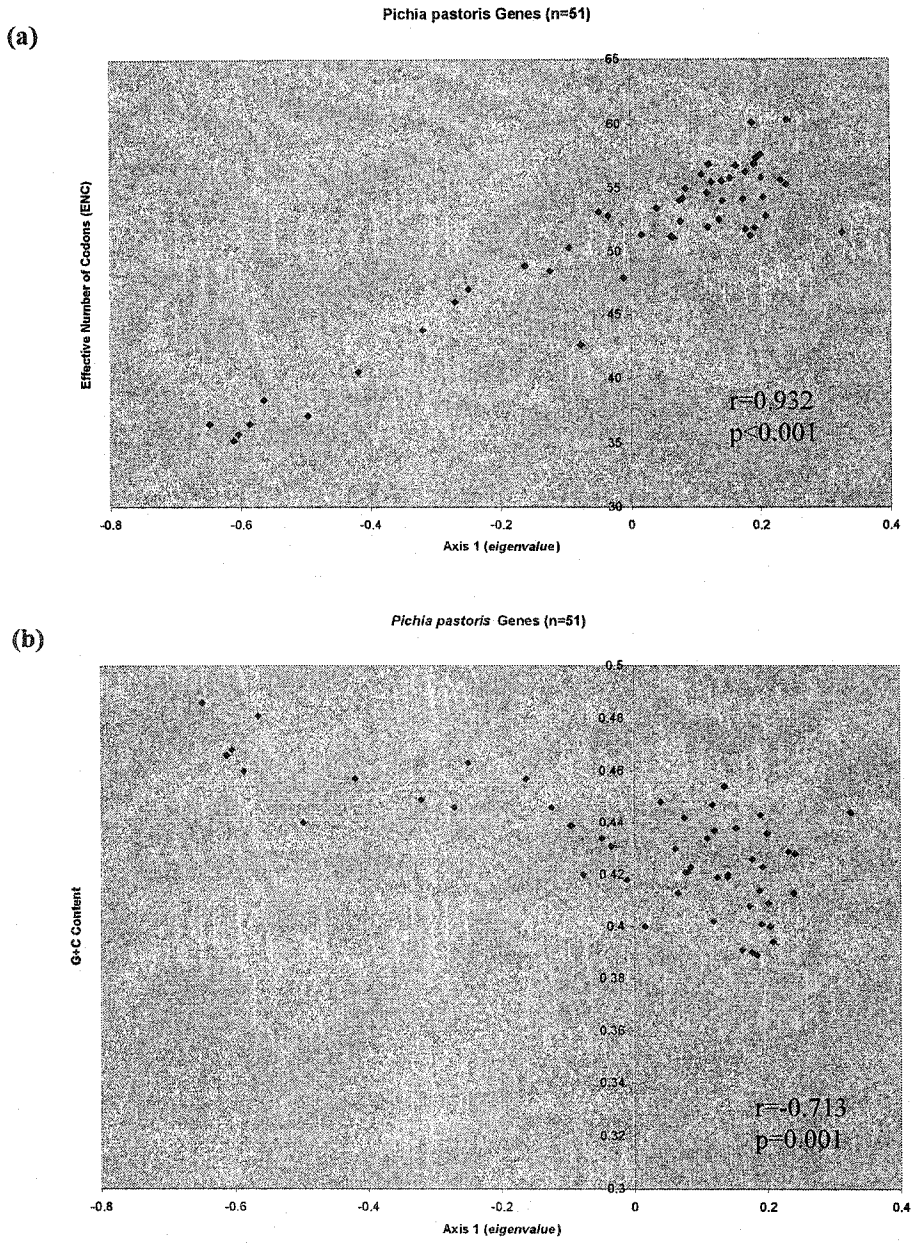

Figure 4.3 – (a) Scatter plot of the effective number of codons (ENC) used in a coding sequence and the position of that coding region on the correspondence analysis of codon usage (COA) primary axis (*eigenvalue*). Highly expressed genes will show a lower ENC value. (b) Scatter plot correlating G+C content of the coding region with the COA primary axis. The coding regions are 52 full-length cDNA sequences from *Pichia pastoris*. Pearson's correlation coefficient (r) for each scatter plot is presented.

promoters (Waterham et al., 1997; Cregg et al., 1989). The other top genes include the O-acetylhomoserine sulfhydrylase (Met17) gene involved in sulfur utilization, and the lysyl oxidase (AOC1) gene involved in proteolysis, both of which have not been well characterized for expression level (Kucha and Dooley, 2000; Spiropoulos and Bisson, 2000). Conversely, the 5 genes correlated with the highest ENC (lowest codon usage bias) are the SAR1, PEX19, PEX17, PEX7, and PEX2 genes all involved in signal transduction for peroxisome biogenesis (PEX genes) or ER to Golgi trafficking (SAR1). Due to the tight regulation required for effective signal transduction it can be speculated that these would be low expressed genes, although no data on their relative expression levels are available (Payne et al., 2000; Johnson et al., 1999).

Based on the conclusion that the primary axis correlated with gene expression levels, a pool of preferred codons was calculated using the RSCU of the top five and bottom five genes as is presented in Table 4.3, with the preferred codon choice of *Saccharomyces cerevisiae* as a reference (Sharp and Li, 1987). The overall similarity of these two codon sets further supports the correspondence analysis data. This table of preferred *Pichia* codons was then used to measure the codon adaptation index (CAI) of the human GBA gene, in conjunction with measures of G+C content and ENC (Table 4.4). The CAI of GBA was 0.302 (maximum possible=1.0), as compared to a range of CAI values for the *Pichia* genes analyzed from a low CAI=0.304 to a maximum CAI=0.718. The codon and nucleotide statistics of the two other GBA variant fragment length constructs (GB200 and GB900) are also presented in Table 4.4.

Table 4.4 – Measures of codon and nucleotide bias from native and codon altered glucocerebrosidase cDNA fragments.

Construct	Effective Number of Codons (ENC)	Codon Adaptation Index (CAI)[a]	G+C Content
Native GBA cDNA	51.15	0.302	0.558
Native GB900	50.72	0.307	0.555
Native GB200	43.79	0.366	0.593
Optimized GBOPT Fragment	20	1.0	0.418
Altered G+C GBGCA Fragment	34.77	0.395	0.468

[a]Codon Adaptation Index calculated using the set of optimal codons determined for *Pichia pastoris* using correspondence analysis (CodonW).

4.3.2 – *Pichia pastoris Expression of Luciferase Fusions*

The luciferase fusion constructs made using the full length and shorter GBA fragments were expressed and luciferase activities compared to a luciferase construct without a 5' GBA fusion (Figure 4.4). Culture medium from both the GB900 and full-length GBA constructs yielded baseline levels of luciferase activity similar to an untransformed KM71 control suggesting a significant impairment of luciferase transcription or translation with these clones. The GB200 fusion construct did, however, produce measurable quantities of active luciferase, although the overall expression level was greatly reduced in comparison to the non-fusion luciferase clone (Figure 4.4). As the GB200 clone represented the best candidate for observing the impact of codon optimization on luciferase expression, this fragment was utilized for subsequent sequence alteration.

4.3.3 - *Analysis and Expression of Codon Altered Luciferase Fusions*

The codon and nucleotide statistics of the codon optimized (GBOPT) and G+C altered (GBGCA) 200bp GBA constructs are presented in Table 4.4 and an alignment of the native (GBA), GBOPT, and GBGCA sequences are presented in Figure 4.2. As the GBOPT sequence was altered to reach a CAI=1.0, the G+C content of the fragment decreased from that of the native GBA sequence due to the initial high G+C content of the 5' end of the GBA gene. The construct GBGCA was developed to hold the CAI value near to that of the native sequence (CAI=0.395) while decreasing the G+C in proportion with that of the GBOPT fragment (G+C=0.468) to help differentiate between the effects of codon optimization and altered G+C content on the translational efficiency

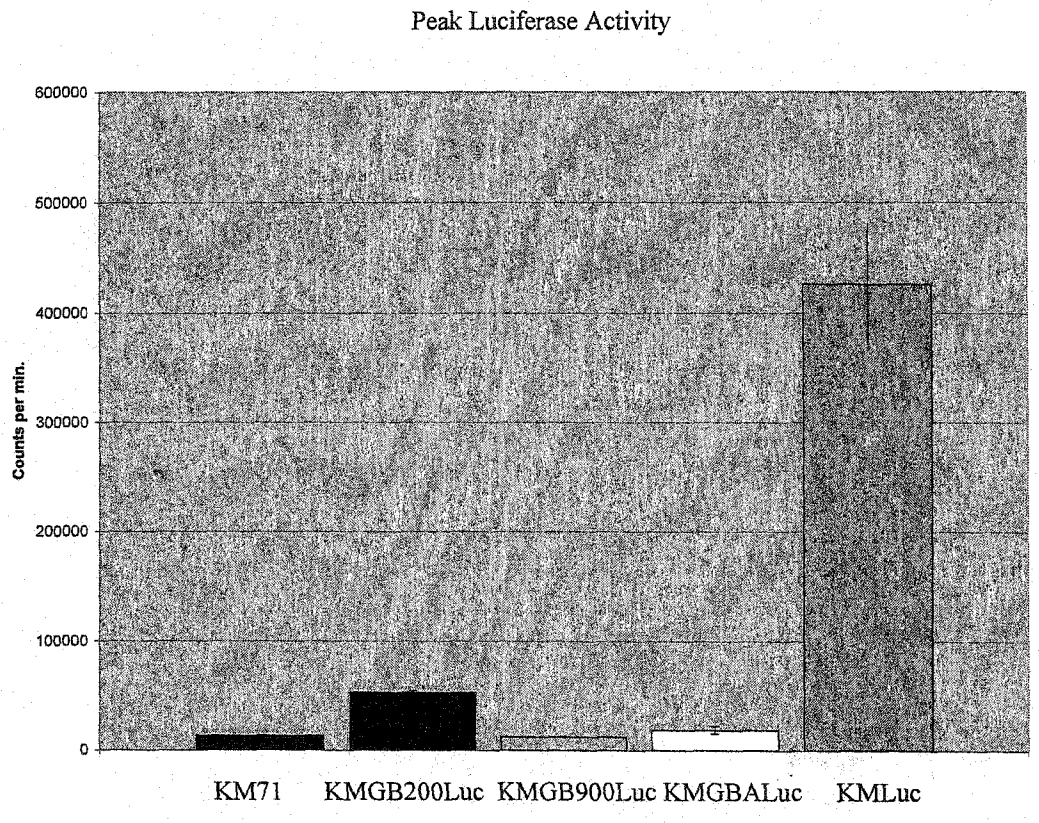

Figure 4.4 – Peak luciferase activities for variable length glucocerebrosidase (GBA) fragments fused to the luciferase cDNA as a translational read through reporter. Activity was measured in the induction culture medium of transformed *Pichia pastoris* clones (KM71 base strain) over 48 hours of induction and the peak value for each clone determined. Error bars represent the standard error of the mean over two independent experiments. KMLuc= luciferase with no GBA fusion; KMGBALuc=a full-length GBA-luciferase fusion; KMGB900Luc=the 5'terminal 900bp of GBA fused to the luciferase gene; KMGB200Luc= the 5'terminal 200bp of GBA fused to the luciferase gene. All GBA constructs exclude the native leader sequence, with secretion directed by the α-factor secretion signal.

of the fragments (Table 4.4). Importantly, this was done without changing any codons from the native sequence if they were considered rare in yeasts (*S. cerevisiae*) as the presence of rare codons alone can have a significant impact on translational efficiency (Zhang et al., 1991). A comparison of the peak luciferase activity obtained with each GBA-luciferase fusion construct induced in *Pichia pastoris* is presented in Figure 4.5. Both the OPT construct and the GCP construct displayed consistently higher luciferase activity than the native GBA construct with the OPT sequence leading to a more than 10 fold increase in luciferase activity, while the GCA construct was 7.5 fold higher than the native GBA fusion (Figure 4.5). The OPT fusion still showed lower activity than the luciferase construct with no 5' fusion (Figure 4.5) and while this may have reflected differences in translational efficiencies, it appeared from Northern blotting that there were some differences in transcription between the fusion and non-fusion constructs (Figure 4.6). While the GBA, OPT, and GCA fusions all displayed similar levels of GBA-luciferase fusion transcription, the luciferase construct by itself produced noticeably more transcript (Figure 4.6).

4.4 – **Discussion**

With the original identification of genome wide trends in synonymous codon usage (SCU), much speculation arose to explain the tendency of an organism to prefer the use of one synonymous codon over another, irrespective of overall trends in amino acid usage (Grantham et al., 1980). While issues of G+C content, dinucleotide relative abundance and mRNA stability have all been evoked to explain codon bias, in unicellular organisms (at least) the major deciding factor appears to be selection for translational efficiency

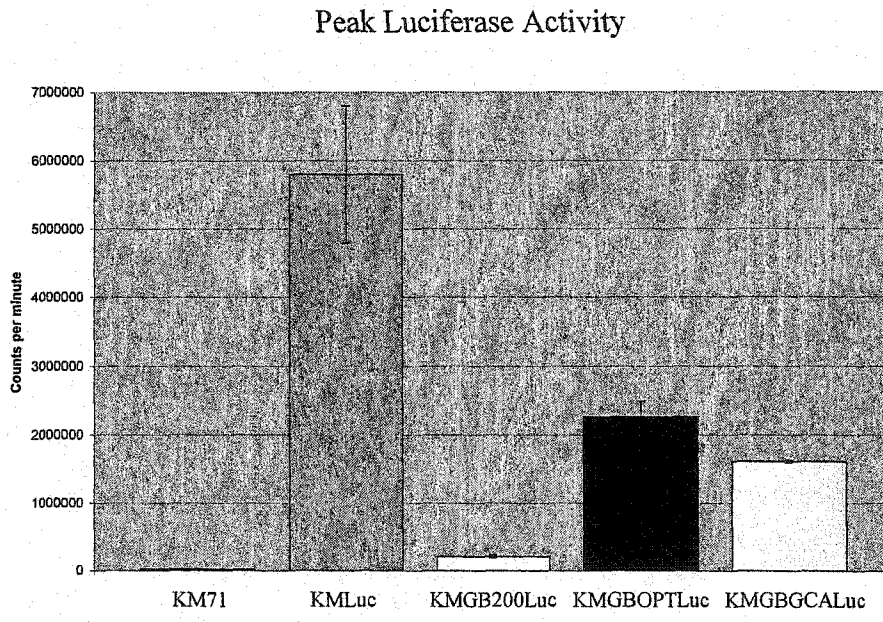

Figure 4.5 - Peak luciferase activities for codon altered glucocerebrosidase (GBA) fragments fused to the luciferase cDNA as a translational read thorough reporter. Activity was measured in the induction culture medium of transformed *Pichia pastoris* clones (KM71 base strain) over 96 hours of induction and the peak value for each clone determined. Error bars represent the standard error of the mean over three independent experiments. KMLuc= luciferase with no GBA fusion; KMGB200Luc=The 5'terminal 200bp of GBA fused to the luciferase gene; KMGBOPTLuc= A Codon optimized version of GB200 (CAI=1.0 in P*ichia pastoris*); KMGBGCALuc=A G+C content altered but non-codon optimized version of GB200. All GBA constructs exclude the native leader sequence, with secretion directed by the α-factor secretion signal.

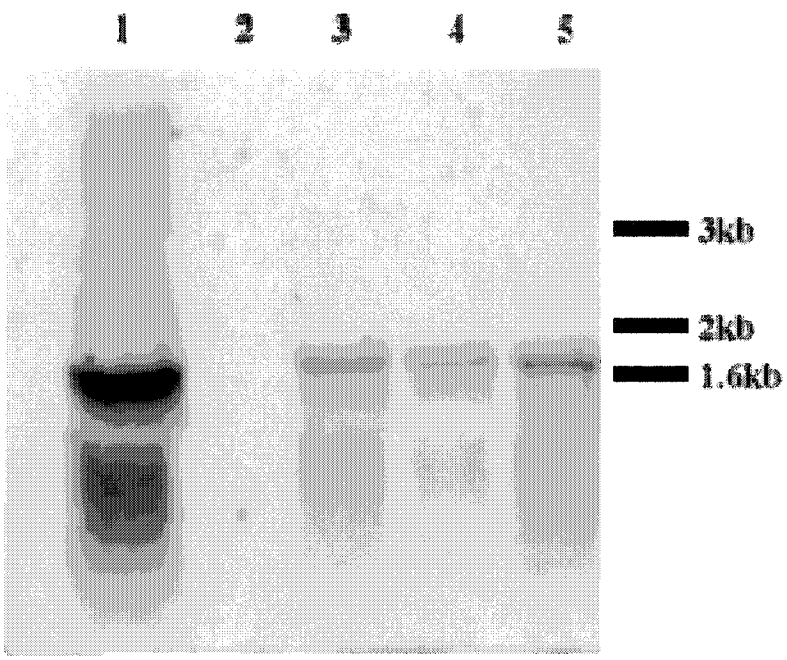

Figure 4.6 – Northern blot of total RNA extracted from GBA-luciferase fusion transformed *Pichia pastoris* clones following 48 hours of methanol induction. A DNA probe representing the full-length luciferase cDNA was used for hybridization. Lane1= KMLuc strain expressing luciferase with no GBA fursion. Lane 2= A non-recombinant KM71 negative control. Lane 3= KMGB200Luc non-codon optimized GBA-luciferase fusion strain. Lane 4= KMGBOPTLuc codon optimized GB200 strain. Lane 5= KMGBGCALuc G+C content altered GB200 strain.

(Sharp et al., 1993; Xia, 1998; Xia, 1996; Bulmer, 1987; Karlin and Mrazek, 1996). Although this correlation has been put to use in bioinformatics to help identify putative open reading frames in large stretches of sequence data (Termier and Kalogeropoulos, 1996), it also has significant implications in the field of heterologous protein production (Smith, 1996). Numerous reports have been submitted delineating significant increases in heterologous protein production upon codon optimization, but these studies have been limited to those expression hosts for which a preferred codon set has been determined, or for which a preferred set could be inferred from a related but well characterized host (Kim et al., 1997; Nagata et al., 1999; Andre et al., 1998; Uchijima et al., 1998; Forman et al., 1998). Although the codon preference of *Saccharomyces cerevisiae* has been well characterized (Xia, 1998; Moriyama and Powell, 1998) and could be used to infer *Pichia pastoris* codon bias, the increasing interest in *Pichia* as an expression host argued towards a thorough analysis of synonymous codon usage in this methylotrophic yeast. Correspondence analysis of available sequences from *Pichia pastoris* was able to identify a single major trend in relative synonymous codon usage (RSCU) correlating with an independent measure of codon bias (ENC). While this trend did show a weaker correlation with G+C content, the arrangement of a number of experimentally determined highly expressed metabolic genes at the high codon bias extremity of the primary axis supported the correct ordination of genes based on expression level (Nakamura and Tabata, 1997). The resultant pool of preferred codons also corresponded to those determined for *Saccharomyces cerevisiae* with the single exception of GAG for glutamic acid in *Pichia pastoris* as compared to GAA in *S. cerevisiae* (Sharp and Li, 1987). This clearly supports the correct determination of the preferred codons for *Pichia pastoris* and

the appropriateness of previous analyses assuming *S. cerevisiae* bias patterns for *Pichia* expression optimization (Withers-Martinez et al., 1999; Brocca et al., 1998). Using the *Pichia* preferred codon set, the CAI value for human glucocerebrosidase was determined as 0.302. Optimization studies expressing a *L. monocytogenes* antigen in mouse (COS7) cells showed that optimization of coding regions to a CAI=1 from an initial CAI=0.442 led to increases in expression level of at least an order of magnitude, although there was not a direct proportionality between increasing CAI and expression level (Nagata et al., 1999). This suggested that poor codon bias match between the human glucocerebrosidase gene and the *Pichia* expression host could be playing a major role in the disappointing levels of protein expression discussed in Chapter 3.

The possible relationship between translational inefficiency and glucocerebrosidase expression in *Pichia pastoris* was investigated using GBA fragments fused to the luciferase cDNA as a translational read-through reporter. Initial expression studies using the full-length GBA cDNA and two smaller 5' GBA fragments fused to the luciferase cDNA established that even a short 200bp fragment was sufficient to significantly impact luciferase expression. This 200bp fragment was then altered to increase the codon optimization from an initial CAI=0.366 to a CAI=1.0 in *Pichia pastoris.* This alteration lead to a 10.6 fold increase in luciferase expression which is consistent with the order of magnitude increases published for other codon optimization studies (Nagata et al., 1999; Kim et al., 1997; Uchijima et al., 1998). Although highly biased *Pichia* genes appear to have a minor bias towards increased G+C content, this nucleotide bias is still lower than that seen for the native human GBA coding region. Accordingly, codon optimization of

the 5' fragment of GBA decreased the G+C content from 59.8% to 41.3%. As this significant decrease in G+C content alone could have an impact on GBA transcription or translation, as second construct that maintained the CAI value of the native sequence while simultaneously decreasing the overall G+C content was investigated. Interestingly, this construct also displayed increased luciferase expression over that of the native coding sequence with an increase in activity of 7.5 fold.

Northern blotting suggested that transcription rates were similar between all GBA-luciferase fusion constructs, leading to the conclusion that the increased expression effects seen were post-transcriptional in nature. Although the alteration in G+C content of the GCA clone did increase the CAI value of the construct slightly, from a CAI=0.336 to a CAI=0.395, this should not be a sufficient alteration in codon bias to affect the translational efficiency of the construct. The nucleotide composition of the messenger RNA has, however, been implicated in the relative efficiency of protein expression through effects on mRNA secondary structure and stability. Analyses of codon bias and mRNA stability have suggested that both increases and decreased in the stability of secondary structures can have a significant impact on mRNA degradation and the efficiency of translational initiation depending on the structures formed or lost (Seffens and Digby, 1999; Oliveira et al., 1993; Tranque et al., 1998). As well as mRNA secondary structure, nucleotide interactions between the mRNA and rRNA have been implicated in controlling translational initiation (Tranque et al., 1998). As mentioned above, recent analysis of translational inefficiency in mammalian cell expression of human glucocerebrosidase revealed the presence of an mRNA binding protein which

appears to regulate GBA translation in certain cell types (Xu and Grabowski, 1998; Xu and Grabowski, 1999). Although this protein was found to be absent from non-mammalian cells, the possibility of a similar cytoplasmic protein binding GBA messenger in a sequence specific manner in yeasts still exists (Xu and Grabowski, 2000). The exact location of interaction between the mRNA and protein has yet to be established, although it appears to be in a coding rather than 5' or 3' untranslated region of the messenger (Xu and Grabowski, 2000). A further elucidation of this interaction could have implications on the codon alteration of GBA constructs in a variety of expression hosts.

Although the mechanism of action is not evident for the increased luciferase activity observed with decreased G+C content of the GBA gene, it appears to work in a complementary fashion with changes in codon bias. As the effects of altered G+C content appear to be as significant to increased expression levels as codon optimization in this case, it will be important to investigate this phenomenon to determine if it is consistent across a variety of coding regions or simply specific to this gene and expression host. Regardless of the mechanism behind the increased expression levels observed, it is predicted that codon optimization of the glucocerebrosidase gene will significantly impact its heterologous expression in *Pichia pastoris*. Based on these results it appears that the financial investment required to recode the entire GBA cDNA would be warranted if *Pichia pastoris* was to be further investigated as a potential system for the high-level heterologous production of human glucocerebrosidase.

4.5 - **References**

Andre, S., Seed, B., Eberle, J., Schraut, W., Bultmann, A. and Haas, J. Increased immune response elicited by DNA vaccination with a synthetic gp120 sequence with optimized codon usage. *Journal of Virology* 72:1497-1503, 1998.

Apeler, H., Gottschalk, U., Guntermann, D., et al. Expression of natural and synthetic genes encoding herpes simplex virus 1 protease in *Escherichia coli* and purification of the protein. *European Journal of Biochemistry* 247:1997.

Bagnoli, F. and Lio, P. Selection, mutations and codon usage in a bacterial model. *Journal of Theoretical Biology* 173:271-281, 1995.

Brocca, S., Schmidt-Dannert, C., Lotti, M., Alberghina, L. and Schmid, R.D. Design, total synthesis, and functional overexpression of the *Candida rugosa* lip1 gene coding for a major industrial lipase. *Protein Science* 7:1415-1422, 1998.

Bulmer, M. Coevolution of codon usage and transfer RNA abundance. *Nature* 325:728-730, 1987.

Choy, F.Y.M., Humphries, M.L. and Shi, H. Identification of two novel and four common missense mutations among Chinese Gaucher disease patients. *American Journal of Medical Genetics* 71:172-178, 1997.

Choy, F.Y.M., Wei, C. and Levin, D. Gaucher Disease: Functional expression of the normal glucocerebrosidase and Gaucher T1366G and G1604A alleles in *Baculovirus*-transfected *Spodoptera frugiperda* cells. *American Journal of Medical Genetics* 65:184-189, 1996.

Comeron, J.M. and Aguade, M. An evaluation of measures of codon usage bias. *Journal of Molecular Evolution* 47:268-274, 1998.

Cramer, C.L., Weissenborn, D.L., Oishi, K.K., et al. Bioproduction of human enzymes in transgenic tobacco. *Annals of the New York Academy of Science* 792:62-71.:62-71, 1996.

Cregg, J.M., Madden, K.R., Barringer, K.J., Thill, G.P. and Stillman, C.A. Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. *Molecular and Cellular Biology* 9:1316-1323, 1989.

Forman, M.D., Stack, R.F., Masters, P.S., Hauer, C.R. and Baxter, S.M. High level, context dependant misincorporation of lysine for arginine in *Saccharomyces cerevisiae* a1 homeodomain expressed in *Escherichia coli*. *Protein Science* 7:500-503, 1998.

Grace, M.E., Newman, K.M., Scheinker, V., Berg-Fussman, A. and Grabowski, G.A. Analysis of Human acid β-glucosidase by site-directed mutagenesis and heterologous expression. *Journal of Biological Chemistry* 269:2283-2291, 1994.

Grantham, R., Gautier, C. and Gouy, M. Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucleic Acids Research* 8:1893-1912, 1980.

Johansson, A., Bolton-Grob, R. and Mannervik, B. Use of silent mutations in cDNA encoding human glutathione transferase M2-2 for optimized expression in *Escherichia coli*. *Protein Expression and Purification* 17:105-112, 1999.

Johnson, M.A., Waterham, H.R., Ksheminska, G.P., et al. Positive selection of novel peroxisome biogenesis-defective mutants of the yeast Pichia pastoris. *Genetics* 151:1379-1391, 1999.

Kane, J.F. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology* 6:494-500, 1995.

Karlin, S. and Mrazek, J. What drives codon choices in human genes. *Journal of Molecular Biology* 262:459-472, 1996.

Karlin, S., Mrazek, J. and Campbell, A.M. Codon usages in different gene classes of the *Escherichia coli* genome. *Molecular Microbiology* 29:1341-1355, 1998.

Kim, C.H., Younghoon, O. and Lee, T.H. Codon optimization for high-level expression of human erythropoietin (EPO) in mammalian cells. *Gene* 199:293-301, 1997.

Kucha, J.A. and Dooley, D.M. Cloning, sequence analysis, and characterization of the 'lysyl oxidase' from Pichia pastoris. *Journal of Inorganic Biochemistry* 83:193-204, 2000.

Li, W. and Graur, D. *Fundamentals of Molecular Evolution*, Sunderland, Massachusetts:Sinauer Associates, Inc., 1991. pp. 91

Liu, C., Bahnson, A.B., Dunigan, J.T., Watkins, S.C. and Barranger, J.A. Long-term expression and secretion of human glucocerebrosidase by primary murine and human myoblasts and differentiated myotubes. *Journal of Molecular Medicine* 76:773-781, 1998.

Martin, B.M., Tsuji, S., LaMarca, M.E., Maysak, K., Eliason, W. and Ginns, E.I. Glycosylation and processing of high levels of active human glucocerebrosidase in invertebrate cells using a baculovirus expression vector. *DNA* 7:99-106, 1988.

Mistry, P.K., Wraight, E.P. and Cox, T.M. Therapeutic delivery of proteins to macrophages: implications for treatment of Gaucher's disease. *Lancet* 348:1555-1559, 1996.

Moriyama, E.N. and Powell, J.R. Codon usage bias and tRNA abundance in *Drosophila*. *Journal of Molecular Evolution* 45:5231997.

Moriyama, E.N. and Powell, J.R. Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Esherichia coli*. *Nucleic Acids Research* 26:3188-3193, 1998.

Nagata, T., Uchijima, M., Yoshida, A., Kawashima, M. and Koide, Y. Codon optimization effect on translational efficiency of DNA vaccine in mammalian cells: Analysis of plasmid DNA encoding a CTL epitope derived form microorganisms. *Biochemical and Biophysical Research Communications* 261:445-451, 1999.

Nakamura, Y., Gojobori, T. and Ikemura, T. Codon usage tabulated from the international DNA sequence database; its status 1999. *Nucleic Acids Research* 27:292-292, 1999.

Nakamura, Y. and Tabata, S. Codon-anticodon assignment and detection of codon usage trend in seven microbial genomes. *Microbial and Comparative Genomics* 2:299-312, 1997.

Oliveira, C.C., van den Heuvel, J.J. and McCarthy, J.E.G. Inhibition of translational initiation in Saccharomyces cerevisiae by secondary structure: the roles of the stability and position of stem-loops in the mRNA leader. *Molecular Microbiology* 9:521-532, 1993.

Payne, W.E., Kaiser, C.A., Bevis, B.J., et al. Isolation of Pichia pastoris genes involved in ER-to-Golgi transport. *Yeast* 16:979-993, 2000.

Powell, J.R. and Moriyama, E.N. Evolution of codon usage bias in *Drosophila*. *Procedings of the National Academy of Sciences, U.S.A* 94:7784-7790, 1997.

Seffens, W. and Digby, D. mRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences. *Nucleic Acids Research* 27:1578-1584, 1999.

Sharp, P.M. and Cowe, E. Synonymous codon usage in Saccharomyces cerevisiae. *Yeast* 7:657-678, 1991.

Sharp, P.M. and Li, W. An evolutionary perspective on synonymous codon usage in unicellular organisms. *Journal of Molecular Evolution* 24:28-38, 1986.

Sharp, P.M. and Li, W. The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research* 15:1281-1295, 1987.

Sharp, P.M., Stenico, M., Peden, J.F. and Lloyd, A.T. Codon usage: mutational bias, translational selection, or both? *Biochemical Society Transactions* 21:835-846, 1993.

Smith, D.W.E. Problems of translating heterologous genes in expression systems: The role of tRNA. *Biotechnology Progress* 12:417-422, 1996.

Spiropoulos, A. and Bisson, L.F. MET17 and hydrogen sulfide formation in Saccharomyces cerevisiae. *Applied Environmental Microbiology* 66:4421-4426, 2000.

Termier, M. and Kalogeropoulos, A. Discrimination between fortuitous and biologically constrained open reading frames in DNA sequences of Saccharomyces cerevisiae. *Yeast* 12:369-384, 1996.

Tranque, P., Hu, M.C.Y., Edelman, G.M. and Mauro, V.P. rRNA complementarity within mRNAs: A possible basis for mRNA-ribosome interactions and translational control. *Procedings of the National Academy of Sciences, U.S.A* 95:12238-12243, 1998.

Uchijima, M., Yoshida, A., Nagata, T. and Koide, Y. Optimization of codon usage of plasmid DNA vaccine is required fo the effective MHC class I-restricted T cell responses against an intracellular bacterium. *Journal of Immunology* 161:5594-5599, 1998.

Waterham, H.R., Digan, M.E., Koutz, P.J., Lair, S.V. and Cregg, J.M. Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene* 186:37-44, 1997.

Withers-Martinez, C., Carpenter, E.P., Hackett, F., et al. PCR-based gene synthesis as an efficient approach for expression of the A+T-rich malaria genome. *Protein Engineering* 12:1113-1120, 1999.

Wright, F. The effective number of codons' used in a gene. *Gene* 87:23-29, 1990.

Xia, X. Maximizing transcription efficiency causes codon usage bias. *Genetics* 144:1309-1320, 1996.

Xia, X. How optimized is the translational machinery in *Esherichia coli, Salmonenna typhurium* and *Saccharomycese cerevisiae*. *Genetics* 149:37-44, 1998.

Xu, Y.H. and Grabowski, G.A. Translational inefficiency of acid beta-glucosidase mRNA in transgenic mammalian cells. *Molecular Genetics and Metabolism* 64:87-98, 1998.

Xu, Y.H. and Grabowski, G.A. Molecular cloning and characterization of a translational inhibitory protein
that binds to coding sequences of human acid beta-glucosidase and other mRNAs. *Molecular Genetics and Metabolism* 68:441-454, 1999.

Xu, Y.H. and Grabowski, G.A. Reconstitution of TCP80/NF90 translation inhibition activity in insect cells. *Molecular Genetics and Metabolism* 70:106-115, 2000.

Zhang, S., Zubay, G. and Goldman, E. Low-usage codons in *Escherichia coli*, yeast, fruit fly and primates. *Gene* 105:61-72, 1991.

Zhou, J., Liu, W.J., Peng, S.W., Sun, X.Y. and Frazer, I. *Papillomavirus* capsid protein expression level depends on the match between codon usage and tRNA availability. *Journal of Virology* 73:4972-4982, 1999.

Zolotukhin, S., Potter, M., Hausworth, W.W., Guy, J. and Muzyczka, N. A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells. *Journal of Virology* 70:4646-4654, 1996.

Chapter 5 – **Summary of Glucoerebrosidase Analysis and Expression**

Gaucher Disease has emerged as a model system for investigating the genetics, biochemistry, and therapeutic options for single gene disorders (Beutler, 1993a). The existence of a high incidence Ashkenazi Jewish population has allowed for a focused investigation of the most common (Type 1) form of Gaucher Disease. Even within this identifiable population however, there exists a great deal of clinical heterogeneity. Beyond the mild N370S allele, few clear correlations can be made between individual genotypes and *in vitro* residual enzyme activity or clinical severity (Beutler and Grabowski, 2001). The pool of data from neuronopathic (Types 2 and 3) Gaucher Disease is even less cohesive with most mutations represented in only single individuals or small focal populations. The direct biochemical pathway that leads to neuronopathy in Gaucher Diseaese has not, in fact, been clearly delineated (Orvisky et al., 2000).

The work presented in Chapter 2 attempted to address this deficiency in understanding of neuronopathic Gaucher disease through the analysis of a number of individuals with type 2 and type 3 disease, including two cases from an isolated community in northern Alberta. Importantly, this investigation demonstrated the effectiveness of archival tissue analysis for the molecular characterization of unique cases for which traditional DNA sources are unavailable. Modern molecular techniques can be used to amplify sufficient quantities of DNA, although highly fragmented, for subsequent mutational characterization. The identification of two novel, and two rare mutations associated with neuronopathy will add to the total pool of data being compiled for neurodegenerative

Gaucher disease. From this global collection of data, individual genotype / phenotype correlations similar to that seen with the N370S mutation may emerge.

The identification and characterization of a rare Gaucher allele in an isolated aboriginal population has implications for both disease diagnosis and possible therapeutic options within this group. Originally identified as associated with type 1 disease in an individual of Cree descent (Beutler et al., 1993b), the P122S allele can now be categorized as a more severe type 3-associated mutation. This change in categorization results from an updated assessment of the original patient and two others genotyped in this study. Importantly, this mutation appears to be unique to this isolated population and is consistently associated with type 3 Gaucher disease. The possibility of increased allele frequencies due to the genetic dynamics of small populations argues for a more broad based population screening within this community. The early identification of affected individuals and carriers could significantly impact the physical and genetic health of the community as a whole.

The heterologous expression of the P122S allele found only a minor impairment in residual enzyme activity emphasizing the complexity of disease progression in Gaucher disease. Although the basic biochemical defect is clear, the multiple factors leading into and out of this step of the sphingolipid degradation pathway must be playing a significant role in the onset of symptoms. Issues of *in vitro* versus *in vivo* activity, plus possible roles for the primary glucocerebroside and secondary glucosylsphingosine substrates in

neuronal cytotoxicity and signal transduction must be investigated to further clarify the pathology observed in this disease (Kolter et al., 1999; Orvisky et al., 2000).

While the details of disease progression and neuronopathy in Gaucher Disease remain to be clarified, obvious targets for treatment do exist. The development of alglucerase and imiglucerase for the enzyme replacement therapy (ERT) of Gaucher disease have been invaluable to increased longevity and quality of life for those challenged by this debilitating disorder (Grabowski et al., 1998). Unfortunately, the restrictive cost of therapy remains an issue for many patients, particularly those with severe neuronopathic disease (Elstein et al., 1998). The development of alternative high-level expression systems for enzyme production could improve the availability and cost of ERT. Two potential heterologous expression systems were investigated for the production of recombinant human glucocerebrosidase. While the methylotrophic yeast *Pichia pastoris* has been shown to produce g/L quantities of simple soluble proteins (Sreekrishna et al., 1997), the levels of glucocerebrosidase expression in this host were disappointing. Detectable levels of active enzyme could be secreted into the culture medium, but at levels low enough to make further purification and characterization extremely difficult.

The application of stable transfected *Sf*9 cells, however, proved more consistent, with mg/L quantities of active enzyme secreted into the culture medium. The expression conditions and constructs were not optimized in this study, and there is clear potential for further improvements to the high-level production of human glucocerebrosidase using the vector system described. Importantly, the *Op*MNPV promotor present in the plasmid

makes the system portable between a variety of dipteran and lepidopteran cell lines allowing the investigation of varied glycosylation between these hosts (Hegedus et al., 1998). The enzyme produced by stable transfected *Sf*9 cells is similar to that observed with baculovirus expression of glucocerebrosidase, although it appears that a higher proportion of the protein is secreted from the cells with a plasmid based system. This could reflect the increased metabolic health of the host cell as compared to a virally infected cell, leading to a potential increase in the quantity and qualtity of recombinant protein produced (Hegedus et al., 1998). Further optimization and scale-up of this expression system will allow for a more thorough characterization of the glucocerebrosidase produced, and the potential of this system for large-scale biotherapeutic protein production.

The sharp contrast in the success of expression attempts between *Pichia pastoris* and insect cells lead to an analysis of post-transcriptional factors that could impact protein expression levels. Synonymous codon usage bias was identified as a potential barrier to successful production of glucocerebrosidase in *Pichia* cultures. An analysis of codon usage bias in this yeast identified the preferred codon set utilized by highly expressed genes and a subsequent comparison of this preferred codon choice to those codons present in the human GBA gene revealed significant differences. Based on the hypothesis that synonymous codon choice is driven by selection for translational efficiency in unicellular organisms, it was predicted that this discrepancy between the recombinant gene and expression host could limit protein expression levels (Nakamura and Tabata, 1997).

To investigate this possibility a translational read through assay utilizing a luciferase reporter gene was employed. Initial investigation revealed that even a 200bp fragment from the 5' end of the GBA gene could inhibit luciferase production. Codon optimization and G+C content alteration of this fragment revealed that both optimizing the codon choice and decreasing the overall G+C content improved translational read through up to 10 fold. This finding suggests that recoding of the full-length GBA gene to fit the codon bias of the expression host could significantly impact overall protein production levels, particularly where codon choice and/or G+C content differ significantly from highly expressed host genes. As issues of G+C content and codon choice are invariably codependent, it has been difficult in past to tease apart the impact of each factor in isolation. The fact that G+C content alteration alone can increase protein production levels in *Pichia pastoris* clearly indicate that this factor should be considered when designing optimal expression constructs for heterologous systems.

This dissertation has taken a broad view of the molecular analysis of Gaucher disease, ranging from fine scale characterization of individual mutations to fundamental questions regarding the expression of human proteins in heterologous hosts. Although the results of this work made direct contributions to the understanding of the complex nature of the glucocerebrosidase protein and the disease that results from its dysfunction, several promising lines of future research have also been revealed. Further characterization of a possible high-incidence aboriginal population will be of both clinical and basic science interest, while the further development of heterologous expression systems for

biotherapeutic glucocerebrosidase production could have a significant impact on the quality of life for those suffering from Gaucher disease.

5.1 - **References**

Beutler, E. Gaucher Disease as a paradigm of current issues regarding single gene mutations of humans. *Proceedings of the National Academy of Sciences, U.S.A* 90:5384-5390, 1993a.

Beutler, E., Gelbart, T. and West, C. Identification of six new Gaucher disease mutations. *Genomics* 15:203-205, 1993b.

Beutler, E. and Grabowski, G.A. Gaucher Disease. In: *The Metabolic and Molecular Bases of Inherited Disease*, edited by Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D. New York: McGraw-Hill, 2001, p. 3645-3668.

Elstein, D., Abrahamov, A. and Zimran, A. Ethical considerations for enzyme replacement therapy in neuronopathic Gaucher disease. *Clinical Genetics* 54:179-184, 1998.

Grabowski, G.A., Leslie, N. and Wenstrup, R. Enzyme therapy for Gaucher disease: the first 5 years. *Blood Reviews* 12:115-133, 1998.

Hegedus, D.D., Pfeifer, T.A., Hendry, J., Theilmann, D.A. and Grigliatti, T.A. A series of broad host range shuttle vectors for constitutive inducible expression of heterologous proteins in insect cell lines. *Gene* 207:241-249, 1998.

Kolter, T., Doering, T., Wilkening, G., Werth, N. and Sandhoff, K. recent advances in the biochemistry ofglycosphingolipid metabolism. *Biochemical Society Transactions* 27:409-415, 1999.

Nakamura, Y. and Tabata, S. Codon-anticodon assignment and detection of codon usage trend in seven microbial genomes. *Microbial and Comparative Genomics* 2:299-312, 1997.

Orvisky, E., Sidransky, E., McKinney, C.E., et al. Glucosylsphingosine accumulation in mice and patients with type 2 Gaucher disease begins early in gestation. *Pediatric Research* 48:233-237, 2000.

Sreekrishna, K., Brankamp, R.G., Kropp, K.E., et al. Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*. *Gene* 55:1997.