The expression of α-*N*-acetylglucosaminidase in two heterologous gene expression systems

by

Joanna Crawford
B.Sc., University of Adelaide, 1991
B.Sc. (Hons), University of Adelaide, 1993
Grad. Dip. Health Counselling, University of South Australia, 1998

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

MASTER OF SCIENCE

in the Department of Biology

The expression of α-*N*-acetylglucosaminidase in two heterologous gene expression systems

by

Joanna Crawford
B.Sc., University of Adelaide, 1991
B.Sc. (Hons), University of Adelaide, 1993
Grad. Dip. Health Counselling, University of South Australia, 1998

Supervisory Committee

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

Dr. Nancy M. Sherwood, Departmental Member
(Department of Biology)

Dr. David B. Levin, Departmental Member
(Department of Biology)

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

Supervisory Committee

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

Dr. Nancy M. Sherwood, Departmental Member
(Department of Biology)

Dr. David B. Levin, Departmental Member
(Department of Biology)

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

Abstract

Mucopolysaccharidosis (MPS) IIIB is an autosomal recessive disorder caused by a defect in *α-N*-acetylglucosaminidase (NAGLU), a lysosomal enzyme involved in the degradation of heparan sulphate. Dysfunctional NAGLU gives rise to a clinical phenotype of severe and progressive mental retardation, often accompanied by hyperactivity and aggressive behaviour. At present, there is no effective treatment for MPS IIIB. However, cloning of the human NAGLU cDNA has made the potential production of human recombinant enzyme for use in enzyme replacement therapy (ERT) a viable option. The work outlined herein focuses on attempts to produce human recombinant NAGLU (rNAGLU) using both yeast and insect cell based expression systems; with the major focus on yeast based expression. Use of a humanized yeast strain, codon optimisation of a portion of the *NAGLU* gene, selection of Mut^{+}, Mut^{S} and multiple integrant strains, and growth at decreased temperature were explored to optimise NAGLU expression in the methylotrophic yeast, *Pichia pastoris.* As none of these measures resulted in abundant NAGLU production, *Sf9* and *Tni* insect cell lines were investigated as an alternate expression system. Additionally, a protein transduction domain (PTD) was fused to NAGLU (NTAT) to circumvent current problems faced in delivering therapeutic enzymes to the brain. NAGLU protein, with and without a fused PTD, were expressed using stable

transfection and baculovirus infection techniques. Small scale experiments utilizing the baculovirus expression vector system (BEVS) have yielded promising results, generating functionally active NAGLU and NTAT protein of the expected approximately 80-85 kDa molecular mass. This preliminary success indicates the BEVS may be an attractive option for the large scale production of rNAGLU and rNTAT.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

4-MU	4-methylumbelliferyl
~	approximately
%	percent
^{o}C	degrees Celsius
Ω	ohm
αα	amino acid
α –MF	α-mating factor
AFU	arbitrary fluorescence unit
AOX	alcohol oxidase gene
BEVS	baculovirus expression vector system
βgal	β-galactosidase
BLAST	basic local alignment search tool
BMGY	buffered glycerol complex medium
BMMY	buffered methanol complex medium
bp	base pair
BSA	bovine serum albumin
CBD	cellulose binding domain
cDNA	complementary DNA
CHO	chinese hamster ovary
CNS	central nervous system
CO	codon optimised
CO.F	codon optimised fragment
ddH_2O	distilled deionized water
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotidetriphosphate
DTT	dithiothreitol
ECM	extracellular matrix

E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
ERT	enzyme replacement therapy
ESF AF medium	Expressions Systems Formula Animal free medium
EtBr	ethidium bromide
F	forward
FGF	fibroblast growth factor
FGFR	fibroblast growth factor receptor
FL	full length
Fx	factor Xa cleavage site
GAG	glucosaminoglycan
Gal	galactose
GBA	glucoceribrosidase
Glc	glucose
GlcNAc	N-acetylglucosamine
H or H_6	hexahistidine tag
HBS	HEPES buffered saline
HEK	human embryonic kidney
His	histidine
His4	histidine dehydrogenase gene
HIV-1	human immunodeficiency virus-1
HS	heparan sulphate
HSPG	heparan sulphate proteoglycans
hr	hour
ie-2 promotor	intermediate-early 2 promotor
IPTG	isopropyl-beta-D-thiogalactopyranoside
kb	kilobase pairs
kDa	kilodalton
LB	luria-bertani medium
LSLB	low salt luria-bertani broth

M6P	mannose 6-phosphate
mA	milliamp
Man	mannose
MAPCs	multipotent adult progenitor cells
MCS	multiple cloning site
MD	minimal dextrose media
MDH	minimal dextrose media with histidine supplement
MeOH	methanol
Met	methionine
mg	milligram
min	minutes
ml	milliliters
mM	millimolar
MM	minimal methanol media
MOI	multiplicity of infection
MPR	mannose 6-phosphate receptor
MPS	mucopolysaccharidosis
mRNA	messenger RNA
Mut^+	wildtype methanol utilization
Mut^S	slow methanol utilization
NAGLU	α-*N*-acetylglucosaminidase gene
NAGLU	α-*N*-acetylglucosaminidase protein
NCBI	national centre for biotechnology information
Ni-NTA	nickel-nitrilotriacetic acid
ng	nanogram
NTAT	NAGLU with TAT moiety
OCH1	α-1,6-mannosyltransferase Och1p
OD_{600}	optical density at 600nm
ORF	open reading frame
pA	poly A signal sequence
PCR	polymerase chain reaction

PMSF	phenylmethylsulphonyl fluoride
P. pastoris	*Pichia pastoris*
PTD	protein transduction domain
PVDF	polyvinylidene difluoride
R	reverse
RER	rough endoplasmic reticulum
RNA	ribonucleic acid
rpm	revolutions per minute
RT	room temperature
RT-PCR	reverse transcription polymerase chain reaction
S. cerevisiae	*Saccharomyces cerevisiae*
SEAP	secreted alkaline phosphatase
sec	seconds
Ser	serine
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
*Sf*9	*Spodoptera frugiperda*
SFM	serum free medium
Sia	sialic acid
T	human transferrin signal
TAT	transcriptional activator of transcription
Thr	threonine
™	trademark
tRNA	transfer RNA
Trp	tryptophan
Tni	*Trichoplusia ni*
U	units
µg	microgram
µl	microliter
UTR	untranslated region
UV	ultraviolet

v/v	volume to volume ratio
w/v	weight to volume ratio
WT	wildtype
WT.F	wildtype fragment
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
x g	times gravity
YPDS	yeast peptone dextrose sorbitol medium

Acknowledgments

I would like to express my thanks to my supervisor, Dr. Francis Choy, without whom I would not have had the opportunity to come to Canada to study. I would also like to thank my committee members, and the members of Dr. Nancy Sherwood and Dr. David Levin's labs for their help, advice and generosity in allowing me the use of various pieces of equipment vital for the completion of my studies. Thank you to Dr. Martin Boulanger and Dr. Tom Pfeifer for their contributions to the baculovirus and stable expression, respectively, of NAGLU and NTAT in insect cells. Special thanks to Dr. Nancy Sherwood and Dr. Martin Boulanger for their support. Thank you to my lab mates April Goebl, Tasha Kulai and Webby Leung. Thanks also to previous Choy lab members Chelsea Patrick, Aggie Zay, Wei Ding and Nuri Nolla, and to numerous undergraduate students whose work in the lab helped support this and other projects. Special and heartfelt thanks go to my wonderful partner Scott Wood, for his never ending support, encouragement and faith in me, and for making it all worthwhile.

1 Introduction

1.1 Mucopolysaccharidosis III (MPS III)

1.1.1 MPS III - an overview of the disease

The mucopolysaccharidoses (MPS) are a group of 11 lysosomal storage disorders which result from deficiencies of specific lysosomal enzymes necessary for the stepwise degradation of glucosaminoglycans (GAGs). These GAGs include dermatan sulphate, heparan sulphate, keratan sulphate, chondroitan sulphate and hyaluronan. Disruption of GAG catabolism leads to partially, or non-degraded GAGs building up in the lysosomes of affected subjects and being excreted at abnormal levels in the urine. Progressive lysosomal accumulation of these various GAG molecules gives rise to cell, tissue and organ dysfunction eventually resulting in clinical onset of disease. The genes encoding all of the enzymes underlying the known MPS have been cloned, and numerous different mutations have been identified as causative for each of the disorders (Neufeld and Muenzer, 2001; Fan *et al.*, 2006).

Mucopolysaccharidosis type III (MPS III), also known as Sanfilippo syndrome, is a biochemically diverse, but clinically similar group of disorders caused by a blockage in catabolism of the GAG, heparan sulphate (Lee-Chen *et al.,* 2002). This condition was first described in 1963 in eight children with cognitive impairment and mucopolysaccharidu ria (Sanfilippo *et al.*, 1963). MPS III has a combined incidence estimated at 1/24,000 births in the Netherlands (van de Kamp *et al.*, 1981), 1/66,000 in Australia (Meikle *et al.*, 1999), 1/280,000 in Northern Ireland (Nelson, 1997), and 1/345,000 in British Columbia (Applegarth *et al.*, 2000). It is comprised of 4 subtypes (MPS IIIA, B, C and D), each of which is inherited in an autosomal recessive manner (Neufeld and Muenzer, 1995). The subtypes have been delineated

based on the absence or deficiency of one of four lysosomal enzymes that are sequentially involved in the degradation of heparan sulphate. A lack of function of any one of these enzymes leads to the lysosomal accumulation of both non-degraded and partially degraded heparan sulphate, which eventually gives rise to a clinical phenotype predominantly characterized by severe central nervous system (CNS) degeneration resulting in progressive mental retardation.

Clinical heterogeneity, ranging from attenuated to severe forms of the disease, is seen in MPS III patients, but the phenotypic variation is less notable than that observed in the other MPS disorders. This is due in a large part to the absence of pronounced somatic involvement in MPS III. Unlike most of the MPS, the somatic features of MPS III, which include skeletal pathology, hepatosplenomegaly and joint stiffness, are quite mild. The combination of mild somatic features in conjunction with a high incidence of false negative results in urinary testing for excess heparan sulphate, can frequently lead to a significant delay in diagnosis of MPS III after onset of symptoms. These limitations could also mean that patients with mild clinical phenotypes may not be diagnosed. Disorders with intellectual impairment as their predominant phenotype overlap in much of their clinical presentation, so this, combined with a mild clinical phenotype, makes accurate diagnoses difficult.

The onset of clinical features of MPS III usually occurs between 2 and 6 years of age, but has been observed in patients both earlier and later than this age range. Presenting features can include generalized developmental delay, aggressive behaviour with accompanying hyperactivity, sleep disorders, mild hepatosplenomegaly and coarse or excessive hair. The development of speech may also be delayed, and some patients never learn to speak. Hearing loss and seizures may occur in moderately to severely affected patients. Early onset of puberty is also a common feature in individuals with Sanfilippo syndrome (Neufeld and Muenzer,

2001). The predominant neuropsychiatric anomalies associated with this disease have been grouped into three phases. The first phase occurs between the ages of 1 and 4 years when individuals with Sanfilippo syndrome present with developmental delay alone. Phase II of the disease can start from the ages of 3-4 years in the more severely affected individuals, and is associated with extreme behavioural disturbances such as temper tantrums, hyperactivity, aggression and a rapid narrowing of attention span. Affected individuals are still quite physically strong during this phase of the illness which makes it particularly difficult to manage. The third stage of the illness is characterized by a progressive deterioration in general health and strength. Falls, increased spasticity and seizures are common during this final stage of the illness. Death, usually resulting from respiratory infection, occurs in severely affected individuals in the mid to late teenage years. However in attenuated forms of MPS III, patients may have a later onset of disease symptoms and may still be relatively active at 20-30 years of age (Cleary and Wraith, 1993; Bax and Colville, 1995).

Owing to the clinical heterogeneity in each of the four types of MPS III, it is difficult to classify individuals on the basis of phenotypic presentation. Phenotypic variation for Sanfilippo syndromes A and B has been reported even within kinships (Di Natale, 1991; McDowell *et al.*, 1993). Overall MPS IIIB appears to most often give rise to milder symptoms with occasional reports of patients being functional into the third and even fourth decades (Neufeld and Muenzer, 2001). A specific diagnosis of MPS IIIA, B, C, or D can be based on reduced levels of activity of one of the four enzymes shown to be causative for MPS III. Diagnostic enzymology using fluorogenic or radiolabelled substrates has been established for the four subtypes (Hopwood, 2005). The genes coding for all four enzymes have been cloned and characterized, thereby enabling sequence analysis of the causative gene as the definitive means of diagnosis.

1.2 Mucopolysaccharidosis IIIB (MPS IIIB)

1.2.1 Identification of the causative gene, α-N-acetylglucosaminidase (NAGLU)

The causative gene for MPS IIIB, *NAGLU*, has been fully characterized and resides on chromosome 17q21.1 (Weber *et al.*, 1996; Zhao *et al.*, 1996). It extends approximately 8.2 kb and is comprised of six exons. Northern blot analysis indicates *NAGLU* is transcribed to generate a single, widely expressed approximately 2.7 kb RNA species with high levels of expression in liver, ovary, spleen and peripheral blood leukocytes (Weber *et al.*, 1996). The coding region of the gene consists of 743 amino acids (ααs) with a signal-peptidase cleavage consensus site at position 23 (Weber *et al.*, 1996). The hydrophobic stretch of 23 ααs at the amino terminal of the protein is consistent with a signal peptide (Zhao *et al.*, 1996).

Endogenously expressed NAGLU has been studied in a number of different human tissues including placenta (Weber *et al.*, 1996), liver (Sasaki *et al.*, 1991) and urine (Salvatore *et al.*, 1984) and in cultured human kidney carcinoma cells (Di Natale *et al.*, 1985). The reported molecular masses range from 82-86 kDa for the precursor form of the protein, and 77-80 kDa for the mature protein (Di Natale *et al.*, 1985; Salvatore *et al.*, 1984; Sasaki *et al.*, 1991; Weber *et al.*, 1996).

In silico translation of the 743 αα precursor protein and 720 αα mature protein confirms respective 82.3 kDa and 80.3 kDa molecular masses. The mature 720 αα human *NAGLU* cDNA encodes a protein that has six potential *N*-linked glycosylation sites of the commonly used Asn(N) X Ser(S)/Thr(T) structure (where X is any amino acid except proline) at asparagine residues 261, 272, 435, 503, 526, and 532 (Zhao *et al.*, 1996). While the consensus *N*-linked glycosylation tripeptide is a requirement for glycosylation, it is not always

sufficient to enable the asparagine to be glycosylated. ExPASy analysis of *N*-glycosylation sites in NAGLU identifies all but the asparagines at 526 as potentially glycosylated, though the scores for positions 435 and 503 are quite low (ExPASyNetNGlyc @ http://www.cbs.dtu.dk/services/NetNGlyc/).

N-linked carbohydrates attached at these potential glycosylation sites can play a role in a variety of biological processes including protein folding, stability, and targeting to subcellular locations (Kukuruzinska and Lennon, 1998). At least one of the glycosylation sites must carry a phosphorylated mannose side chain to enable the transport of NAGLU to the lysosomes via the mannose 6-phosphate receptor (MPR) mediated pathway. Weber *et al*. (1996) predict that residue 272 at least carries the mannose 6-phosphate (M6P) moiety necessary for lysosomal targeting, as they found N-terminal sequencing of the peptide was blocked at this position.

1.2.2 Mutations identified in the NAGLU gene

Since the cloning of the *NAGLU* gene, over 100 mutations have been described including missense, nonsense, deletion, insertion and splice site mutations (Coll *et al.*, 2001; Yogalingam and Hopwood, 2001; Emre *et al.*, 2002; Lee-Chen *et al.*, 2002; Tanaka *et al.*, 2002; Beesley *et al.*, 2004; 2005). The mutations have been identified along the length of *NAGLU* at relatively low frequencies. Most of these changes are unique to individual families. Based on this observation, it has been proposed that the variable phenotype seen for MPS IIIB patients may reflect the extensive genetic heterogeneity of mutations in the *NAGLU* gene (Weber *et al.*, 1999).

Most of the mutations characterized in *NAGLU* give rise to a severe clinical presentation in MPS IIIB patients. All nonsense mutations identified to date and the only reported splice site mutation (Tessitore *et al.*, 2000) give rise to a severe phenotype. All

insertion and deletion mutations which cause a frameshift in the resultant protein are also associated with severe disease (Yogalingam and Hopwood, 2001). Several of these deleterious mutations have been expressed in CHO cells and have been shown to give rise to inactive, truncated, rapidly degraded or undetectable levels of protein (Tessitore *et al.*, 2000; Yogalingam *et al.*, 2000). These results are consistent with the clinically severe presentation in the affected individuals. However, the situation is not as clear cut with missense mutations. There have been a number of missense mutations associated with attenuated clinical phenotypes and, as expected, these mutations give rise to expressed proteins with some residual activity (Di Natale, 1991; Yogalingam *et al.*, 2000). However not all missense mutations giving rise to residual enzyme activity result in attenuated disease. Beesley *et al.* (2005) reported residual activity in a compound heterozygote who was severely affected. The low or single incidence of the majority of these missense mutations makes accurate genotype-phenotype correlations difficult.

There are however a number of recurrent MPS IIIB mutations, the most common of which is the nonsense mutation R297X, which occurs in 11.5% of MPS IIIB patients and has been found in a broad variety of different populations (Yogalingam and Hopwood, 2001). This nonsense mutation occurs at a CpG dinucleotide, a known mutational hotspot in the human genome, which may thereby account for its high frequency and broad distribution range (Cooper and Youssoufian, 1988). Conversely, mutation screening in several countries has uncovered the prevalence of particular mutation in a given population. R626X and H414R are prevalent alleles in the Greek population (Beesley *et al.*, 2004). The latter mutation was found to be in linkage disequilibrium with a polymorphism in all members of that population indicating a possible founder effect. R565P has been reported in seven unrelated families from the Okinawa islands in Japan and was homozygous in five of these families, again suggesting a

founder effect. This Arg565 codon appears quite unstable, with R565W, R565Q and R565P mutations all being reported (Beesley *et al.*, 1998; 2005; Bunge *et al.*, 1999; Weber *et al.*, 1999; Chinen *et al.*, 2005). These changes again occur at a CpG nucleotide, suggesting this codon is a mutational hotspot in the *NAGLU* gene.

In terms of enzyme replacement therapy for MPS IIIB, it may be advantageous to know the genotype and thus the level of protein expressed by an individual, prior to treatment. Patients who express even very low levels of NAGLU may evoke a lesser immune response than those who produce none (Beesley *et al.*, 2005). Thus genotype analysis may be a means to predict, and thereby prevent, an immune reaction.

1.2.3 Heparan sulphate (HS)

MPS IIIB results from a deficiency in the gene encoding NAGLU, an enzyme required for the removal of the *N*-acetylglucosamine residues during degradation of the GAG heparan sulphate (HS) (Lee-Chen *et al.,* 2002). This defect results in the aberrant accumulation of both partially, and non-degraded, HS. The mechanism by which HS storage leads to the CNS degeneration evident in patients with MPS IIIB is not known.

HS is a polysaccharide side chain consisting of a series of repeating disaccharide units which are differentially modified (most notably by the addition of sulphate groups) to give rise to a wide diversity of potential isoforms. Distinct HS isoforms can attach in variable numbers to a variety of core proteins to form HS proteoglycans (HSPGs). These HSPGs are found both as free molecules in the extracellular matrix (ECM) and as bound molecules localized at the cell surface by membrane attachment. It is the large negative charge conferred by the highly sulphated structural motifs of HS that is primarily responsible for the numerous protein binding and regulatory properties of HSPGs (Whitelock and Iozzo, 2005).

HSPGs have been reported to function in many tissues at many stages of development. In light of the cognitive impairment evident for MPS IIIB patients, the involvement of HSPGs in critical events in mammalian neurodevelopment, including neurogenesis and axon guidance is of great interest (Yamaguchi, 2001; Van Vactor *et al.*, 2006).

Neurogenesis, the generation of neurons from neural stem cells and their migration toward genetically determined locations, is regulated by a number of different growth factors including fibroblast growth factor (FGF) 2. There are several lines of evidence that show FGF2 is crucial to this process. Concomitant expression of FGF2 and its receptor during the active phase of neurogenesis is one indicator. Additionally, when FGF2 is added exogenously to neural stem cells in culture, it enables self-renewal of the cells (Yamaguchi, 2001). FGF2 is also pivotal in determining the fate of these neural stem cells. At low FGF2 concentrations stem cells predominantly differentiate into neurons, whereas at higher FGF2 concentrations they differentiate into glial cells (Qian *et al.*, 1997).

HSPGs are highly expressed in the developing brain during embryonic development and have been shown to be expressed on the surface of neural stem cells in culture (Yamaguchi, 2001). The role of HSPGs in assisting and modulating signalling performed by FGF family has been well documented. FGF2 binds HS chains of various HSPGs. Additionally, HS is able to bind FGF receptor (FGFR) molecules. This dual binding ability enables HS to potentiate receptor signalling at low ligand concentrations by enhancing the formation of FGF2-FGFR complexes (Bernfield *et al.*, 1999). Although the exact role played by the various HSPG molecules in modulating FGF2 expression in neurogenesis *in vivo* remains to be determined, it has been established that the FGF2-FGFR-HSPG complex is necessary for the induction of mitosis and optimal biologic responses to FGF2 (Reuss and von Bohlen und Halbach, 2003).

The involvement of HS in regulating axonal growth and guidance was first illustrated by Wang and Denberg (1992) in experiments with whole cultured insect embryos. They showed that exposure of embryos, either to high levels of exogenous HS or to heparinase (an enzyme which cleaves endogenous HS side chains), caused path finding errors in axonal pathways (Wang and Denburg, 1992). Extension of axons towards target cells is not only important in neural development, but also after neural injury. Groves *et al.* (2005) have reported HSPGs to be inhibitory for re-growth of axons. Experimental comparison of the regeneration of peripheral axons with and without enzymatic removal of HS from HSPGs was performed. Treatment of cut peripheral nerves with a combination of heparinase I and III produced an enhancement of axonal regeneration both in number of axons regenerated and in the length of the regenerated axon (Groves *et al.*, 2005). Again, although these results are interesting, the mechanisms by which HS exerts these effects are unknown.

In both MPS I and MPS III, defective enzymes lead to progressive accumulation of heparan sulphate, and in the case of MPS I, the additional accumulation of dermatan sulphate. As HS accumulation appears to be a major feature of these diseases, two groups have employed different approaches to assess the effects of this accumulation on FGF2-FGFR-HS interactions (Li *et al.*, 2002; Pan *et al.*, 2005) and lesions of nerves in the cerebral cortex (Li *et al.*, 2002). Li *et al*. (2002) generated a mouse model for MPS IIIB carrying a disruption of the *NAGLU* gene which provides an *in vivo* system to study the function of HS in the CNS. Pan *et al*. (2005) used *in vitro* conditions to compare multipotent adult progenitor cells from patients with MPS I with those from unaffected individuals.

Autopsies on the brains of MPS IIIB patients have shown both a reduction in the number of neurons and an increased number of astrocytes (Tamagawa *et al.*, 1985). An increase in the relative density of astrocytes in MPS IIIB mouse brain was also noted. To

determine whether these astrocytes were functional, their capacity to react to a lesion was investigated. It was found that the astrocytes in the mutant MPS IIIB mice had a limited capacity to respond to injury unlike their wildtype (WT) litter mates. It could be speculated that this reduced ability to react to injury may reflect the *in vitro* findings of Groves *et al.* (2005) and result from the abnormal accumulation of HS in the tissues of these MPS IIIB mice.

As outlined in a previous paragraph, *in vitro* studies have shown stem cells from the cerebral cortex predominantly differentiate into glial cells (including astrocytes) at higher FGF2 concentrations (Qian *et al.*, 1997). In light of these findings, the observed increased density of astrocytes in the cerebral cortex of NAGLU mouse brain would lead to the expectation that correspondingly high levels of FGF2 would also be found *in vivo*. Whilst this was true for FGF2 levels in the frontal cortex of NAGLU mice at 3 months, levels of FGF2 in the frontal cortex at 6 months and in the caudal cortex at both 3 and 6 months were significantly reduced compared to levels in WT mice (Li *et al.*, 2002). The authors speculated that the reduced levels of FGF2 may be an adaptive response to hugely elevated levels of incorrectly degraded HS. These results indicate that the relationship between FGF2 levels and differentiation pathway of embryonic stem cells *in vivo* are not as straightforward as those reported *in vitro.*

In their study of MPS I, Pan *et al*. (2005) compared multipotent adult progenitor cells (MAPCs) from normal individuals with those from patients with MPS I. They found that the size, structure and composition of proteoglycans from MPS I MAPCs were abnormal. Specifically, the MPS I HS chains were small and abnormally sulphated. A decreased binding of FGF2 to MPS I MAPCs and decreased capability of MPS I HS to facilitate FGF2 binding to FGFR was also reported. They concluded from these observations that the formation of the FGF2-FGFR-HS complex is defective in MPS I MAPCs and speculated that this was probably

due to aberrant HS interfering with the ability for FGF2 to bind its cell surface receptor. Observations of a lack of proliferative reaction of MPS I MAPCs to varying doses of FGF2 (where normal MAPCs show a dose dependent proliferation) indicated abnormal HS may also interfere with mitogenic signalling. The authors hypothesized that the accumulation of structurally and functionally abnormal HS causing perturbations of FGF2-FGFR-HS interactions and defective FGF2 induced proliferation *in vitro,* may reflect the mechanisms by which accumulated HS leads to the progressive neurological dysfunction evident in MPS I patients (Pan *et al.*, 2005). A decreased proliferation of neural progenitor cells in the MPS IIIB mouse model was also reported (Li *et al.*, 2002). This commonality led Pan *et al.* (2005) to speculate that a mechanism of action similar to that seen for the abnormal HS in MPS I may be responsible for the pathogenicity of other storage disorders, including MPS IIIB.

1.2.4 Treatment for MPS IIIB

At present, there is no effective treatment for MPS IIIB. Care is limited to the clinical management of the variety of complications arising from this disease, or behaviour modification for aggression or hyperactivity issues (Neufeld and Muenzer, 2001). However the cloning of human *NAGLU* cDNA (Weber *et al.*, 1996; Zhao *et al.*, 1996) has made the potential production of human recombinant enzyme for use in enzyme replacement therapy (ERT) a viable option.

ERT is a treatment that has proven effective for other lysosomal storage diseases, such as Gaucher and Fabry disease, and more recently for MPS types I and VI (Kakkis *et al.*, 1996; Schiffmann and Brady, 2002; Wraith, 2006). Lysosomal storage diseases lend themselves readily to ERT because proper cell phenotype can be restored with only a small amount of normal activity of the enzyme in question (often less than 10%). However, most lysosomal

enzymes need to undergo complex posttranslational modification to ensure their correct folding, stability, bioactivity and localization. Production of *N*-linked glycosylated enzyme with an intact M6P targeting signal is necessary for the lysosomal uptake of the recombinant enzyme via the endogenous MPR mediated pathway (Dahms *et al.*, 1989).

This lysosomal targeting is a multi step process. Initially, whilst in the rough endoplasmic reticulum (RER), the lysosomal enzymes are co-translationally glycosylated at select asparagine residues. Following signal sequence cleavage and preliminary processing of the *N*-linked oligosaccharides, the proteins are transported to the Golgi apparatus where they undergo phosphomannosylation. The M6P moieties on these lysosomal enzymes serve as high affinity ligands for binding to MPRs in the trans-Golgi network. The ligand/receptor complex then exits the Golgi via a coated vesicle and is delivered to a pre-lysosomal compartment where the ligand dissociates and the released lysosomal enzyme is finally packaged into a lysosome (Dahms *et al.*, 1989). It is this mannose 6-phosphate receptor mediated uptake of secreted recombinant NAGLU that has been one of the limiting steps in uptake assays to date.

Recombinant NAGLU has been expressed in Chinese hamster ovary (CHO) cells (Zhao and Neufeld, 2000; Weber *et al.*, 2001), human embryonic kidney cells (HEK 293) (Zhao and Neufeld, 2000), HeLa cells, human skin fibroblasts (Yogalingam *et al.*, 2000) and *Spodoptera frugiperda (Sf9)* cells (Bandsmer *et al.*, 2006). Although produced at very low levels (μg/L) in *Sf9* cells, NAGLU was produced at promising levels in CHO cells (mg/L). However the MPR-mediated uptake of recombinant NAGLU into cultured cells was negligible (Weber *et al.*, 2001). Weak phosphorylation and subsequent poor uptake of secreted recombinant NAGLU expressed in HeLa cells was also reported (Yogalingam *et al.*, 2000). The causative factors leading to poor mannose 6-phosphorylation of secreted WT NAGLU are unknown.

Irrespective of cause, the use of secreted enzyme in enzyme replacement studies is currently severely limited by its poor mannose 6-phosphorylation.

Another limitation of ERT is its inability to address neurodegeneration due to restricted access to the brain. As MPS IIIB is a disease that predominantly affects the CNS, research needs to be focused on discovering a mechanism which allows the effective delivery of recombinant enzyme to the most affected tissue. New treatment approaches are needed to restore NAGLU activity in the CNS and thereby prevent the neurological degeneration in MPS IIIB patients.

In 1988, two groups independently reported that the transcriptional activator of transcription (TAT) protein from human immunodeficiency virus-1 (HIV-1) had the unique potential to enter cells in culture when added exogenously (Frankel and Pabo, 1988; Green and Loewenstein, 1988). The domain responsible for this transduction has now been ascribed to a short arginine- and lysine-rich region encompassing residues 47-57 (YGRKKRRQRRR) of the TAT protein (Vives *et al.*, 1997; Ezhevsky *et al.*, 1997; Becker-Hapak *et al.,* 2001). This stretch of 11 highly basic residues has been termed the TAT protein transduction domain (PTD). Studies have shown that TAT-PTD peptides conjugated to heterologous proteins can deliver these proteins both cytoplasmically and in the nucleus and that this delivery is independent of cell type and has very low toxicity in cell culture (Mann and Frankel, 1991; Fawell *et al.*, 1994; Vives *et al.,* 1997).

Initial *in vitro* experiments with these peptides focused on establishing whether the TAT-PTD domain affects the function of its fused protein partner (Xia *et al.*, 2001; Lee *et al.*, 2005). Lee *et al*. (2005) demonstrated the uptake and expression of functional glucocerebrosidase in NIH/3T3 cells and Gaucher fibroblasts from transduction with a variety of TAT-modified chimeric constructs. Xia *et al*. (2001), also working with NIH/3T3 cells, reported that

TAT modifications to the C terminus of the β-glucuronidase, the protein deficient in MPS VII, again did not inhibit enzyme activity.

Perhaps the most promising findings have been those demonstrating *in vivo* distribution of TAT-PTD fusion proteins to a broad range of tissues in mice (Fawell *et al*., 1994; Schwarze *et al.*, 1999). Fawell *et al*. (1994) injected mice intravenously with TAT-PTD fused to the 120 kDa β-galactosidase (βgal) protein. They reported delivery to several tissues with high levels seen in heart, liver and spleen and lower levels in lung and skeletal muscle, but they did not detect activity in either kidney or brain. However, when Schwarze *et al.* (1999) conducted similar experiments with a TAT-PTD βgal fusion protein, they reported a much wider tissue distribution and overall higher levels of expression. They demonstrated that intraperitoneal injection of the TAT chimera resulted in delivery of biologically active fusion protein to blood cells, spleen, liver, heart, lung, kidney and importantly, brain tissue. It was also noted in this study that the blood-brain barrier remained intact in the TAT-PTD (βgal) treated mice. This illustrated that even large proteins, when fused to TAT-PTDs, are able to harmlessly cross the blood brain barrier. These initially promising studies have been followed by those of other groups who have fused endogenously expressed proteins to the TAT-PTD (Xia *et al.*, 2001; Cao *et al.*, 2002; Elliger *et al.*, 2002).

The β-glucuronidase deficient mouse is an animal model for MPS VII, a lysosomal storage disease with CNS involvement. Xia *et al.* (2001) injected vectors expressing β-glucuronidase and β-glucuronidase-TAT-PTD into the right hemispheres of the mice brains. They found the TAT modification both enhanced the penetration of β-glucuronidase into the brain, and showed a significant reduction in the storage material in those areas above that observed for the native β-glucuronidase protein. Elliger *et al.* (2002) injected TAT-PTD fused

with β-glucuronidase into the spinal cord of MPS VII newborn mice and showed distribution of functional enzyme to a variety of tissues including the brain. The functionality of the enzyme was illustrated by a reversal of excessive GAG storage in the brains of treated mice to levels indistinguishable from those in wildtype mouse brain tissue.

Cao *et al.* (2002) created a TAT-PTD Bcl-xL fusion protein (designated PTD-HA-Bcl-xL) and demonstrated this fusion protein was efficiently transduced into primary neural cells where it was shown to be functionally active. Specifically, upon intraperitoneal injection of PTD-HA-Bcl-xL into mice, robust protein transduction was noted in neurons in various brain regions, including the cortex, hippocampus, cerebellum and spinal cord. In contrast, injection of the HA-Bcl-xL non-PTD containing control construct did not achieve detectable protein transduction levels. Furthermore, the biological activity of PTD-HA-Bcl-xL was illustrated by its ability to decrease cerebral infarction (stroke) and attenuate activity of the apoptotic marker caspase-3 (Cao *et al.,* 2002).

Despite the distinct potential of the TAT and other modified PTDs, the mechanism involved in the cellular uptake of PTD fused proteins remains controversial. Early reports indicated that PTD-mediated internalization of proteins occurred in an energy and receptor independent manner as it occurred at 4 °C, a temperature which abolishes active endocytic transport mechanisms (Vives *et al.,* 1997). However, it has been since recognized that experimental artifacts resulting both from inadequate removal of cell surface bound proteins and increased cell permeability due to cell fixation prior to microscopic observation, may have led to the erroneous assumption of energy and receptor independence of PTD-mediated internalization. More recently it has been demonstrated that internalization is almost completely suppressed at 4 °C in unfixed conditions (Liu *et al.*, 2000; Richard *et al.,* 2003),

consistent with the involvement of endocytosis in the cellular internalization of cell penetrating peptides.

Cellular uptake of TAT-PTD fused molecules has been shown to be inhibited by addition of HS to the media (Orii *et al.,* 2005) and is competitively inhibited by enzymatic removal of membrane-associated HS (Tyagi *et al.*, 2001; Ziegler *et al.,* 2005). Tyagi *et al*. (2001) further demonstrated that cells genetically defective in the biosynthesis of fully sulphated HS are selectively impaired in the internalization of recombinant TAT-PTD fused macromolecules. These findings suggest that membrane-bound HS plays an important role in mediating TAT-PTD cell surface binding and its endocytosis into the cell. Further support comes from the observation that TAT-PTD can be taken up by a variety of different cell types, which suggest that a conserved, widely expressed cell membrane receptor such as an HSPG is responsible for internalization.

A recent study indicating that there is no single mechanism for the cellular uptake of cell penetrating peptides has helped to clarify the situation. At low micromolar concentrations, uptake was shown to occur simultaneously via three endocytic pathways, whereas at higher concentrations, uptake was endocytic pathway independent and was mediated by the presence of HS (Duchardt *et al.,* 2007).

The use of TAT-PTDs for treatment of lysosomal storage disorders is an attractive prospect. The ability to transduce proteins into a variety of cells, including those of the brain, not only bypasses the need for MPR-mediated uptake of lysosomal enzymes, but also allows access of recombinant proteins to previously inaccessible areas. This is invaluable for the treatment of lysosomal storage diseases such as MPS IIIB where CNS involvement is the major disease phenotype.

1.3 Heterologous Gene Expression Systems

1.3.1 Heterologous gene expression in Pichia pastoris

An expansion in the field of gene technology coupled with the completion of the human genome sequencing project has necessitated the discovery of suitable expression systems in which to produce these genes. The methylotrophic yeast *Pichia pastoris* is a single-celled eukaryote that has been successfully utilized for the expression of more than 550 recombinant proteins (http://faculty.kgi.edu/cregg/index.htm). A number of factors have led to the increasing popularity of this system. *P. pastoris* is easier to genetically manipulate and culture than mammalian cells, and is commercially available. It can be grown to high cell densities to produce foreign proteins at high levels and these recombinant proteins can be expressed either extra- or intra-cellularly. Unlike the widely used *Escherichia coli* bacterial expression systems, the eukaryotic nature of *P. pastoris* provides the potential for producing soluble, correctly folded recombinant proteins that have undergone some or all of the posttranslational modifications required for functionality.

A unique feature of methylotrophic yeasts such as *Pichia* is their ability to utilize methanol as the sole carbon and energy source. They achieve this through induction of the alcohol oxidase (*AOX*) gene which encodes the enzyme responsible for the first step in the methanol utilization pathway (Cregg *et al.*, 2000; Macauley-Patrick *et al.,* 2005). The presence of methanol is essential to induce high levels of transcription from this promoter. In methanol grown cultures, alcohol oxidase can constitute up to 30% of the total cellular protein (Gellissen, 2000). This strong *AOX1* promoter has therefore been utilized to drive the expression of recombinant proteins to high levels. Another advantage of this promoter is that it will be switched off by growth on most other carbon sources. Alcohol oxidase levels are undetectable

in cells grown on glucose, ethanol or glycerol (Wegner and Harder, 1986). The presence of a highly inducible and stringently regulated promoter is ideal for the expression of recombinant proteins and even allows production of proteins that may be toxic to the *P. pastoris* cells (Daly and Hearn, 2005).

P. pastoris contains two genes that encode alcohol oxidase, *AOX1* and *AOX2*. The protein coding regions of the two genes are highly conserved showing greater than 95% αα sequence similarity, however no homology is observed in their regulatory regions. When induced with methanol, the *AOX1* promoter is responsible for the vast majority of alcohol oxidase activity in the cell (Cregg *et al.,* 1989). Disruption of the *AOX1* gene or its promoter leads to a slow methanol utilization (Mut^S) phenotype. As the cells must rely on the weaker *AOX2* for methanol metabolism, and this gene yields 10-20 times less alcohol oxidase activity than the *AOX1* gene, a slower growing and slower methanol utilizing strain is produced. The Mut^S phenotype, because of its slower growth, may be desirable when a gene product is difficult to synthesize, slow to fold, or must undergo other complex posttranslational modifications (Daly and Hearn, 2005).

A variety of proteins that could not be expressed in *E. coli* due to a lack of correct posttranslational maturation, have subsequently been produced in *P. pastoris* (Daly and Hearn, 2005). This is presumably due to the ability of *P. pastoris* to perform posttranslational modifications such as proteolytic processing of signal sequences, and *O*- and *N*-linked glycosylation, which confers stability on the heterologous protein.

P. pastoris can produce foreign proteins which are expressed either intracellularly or extracellularly. As yeast secrete only low levels of native protein, extracellular production of recombinant protein is most desirable as the secreted heterologous protein will constitute the vast majority of the protein in the medium. The secretion signal sequence used with most

success in *P. pastoris* is the *Saccharomyces cerevisiae* α-factor prepro peptide (Macauley-Patrick *et al.,* 2005).

It is thought that since many native proteins are glycosylated, it must be necessary to have the correct glycosylation patterns on recombinant proteins to ensure their biological activity. Yeasts are capable of adding both *O*- and *N*-linked carbohydrates to secreted proteins. However, important differences exist between yeast and mammalian glycosylation abilities. Although *P. pastoris* and higher eukaryotes both add *O*-oligosaccharides to the oxygen molecule of hydroxyl groups of serine and threonine residues of secreted proteins, these oligosaccharides are composed solely of mannose residues in *Pichia* whereas in mammals, *O*-oligosaccharides are composed of a variety of sugars including *N*-acetylgalactosamine, galactose and sialic acid. Additionally, it is possible that *P. pastoris* may not glycosylate heterologous proteins on the same serine and threonine residues as the native host, if at all (Cereghino and Cregg, 2000; Macauley-Patrick *et al.* 2005). No consensus primary amino acid sequence for *O*-glycosylation (in either mammals or lower eukaryotes) appears to exist, and unlike *N*-glycosylation, which has been shown to be crucial for protein function, relatively little is known about *O*-glycosylation and its biological role (Wildt and Gerngross, 2005).

There are also both similarities and differences in *N*-linked glycosylation in the two species (Figure 1.1). *N*-glycosylation in all eukaryotes begins in the endoplasmic reticulum (ER) with the transfer of a pre-formed lipid-linked oligosaccharide unit $Glc_3Man_9(GlcNAc)_2$ (Glc = glucose; Man = Mannose; GlcNAc = *N*-acetylglucosamine) to the amide nitrogen of an asparagine residue at a specific recognition site in the protein. The structure is then trimmed to $Man_8(GlcNAc)_2$, but at this point glycosylation patterns become organism specific. In the mammalian system, further trimming of the mannose residues occurs in the Golgi apparatus to generate $Man_5(GlcNAc)_2$ and subsequent trimming and addition of a variety of different sugars

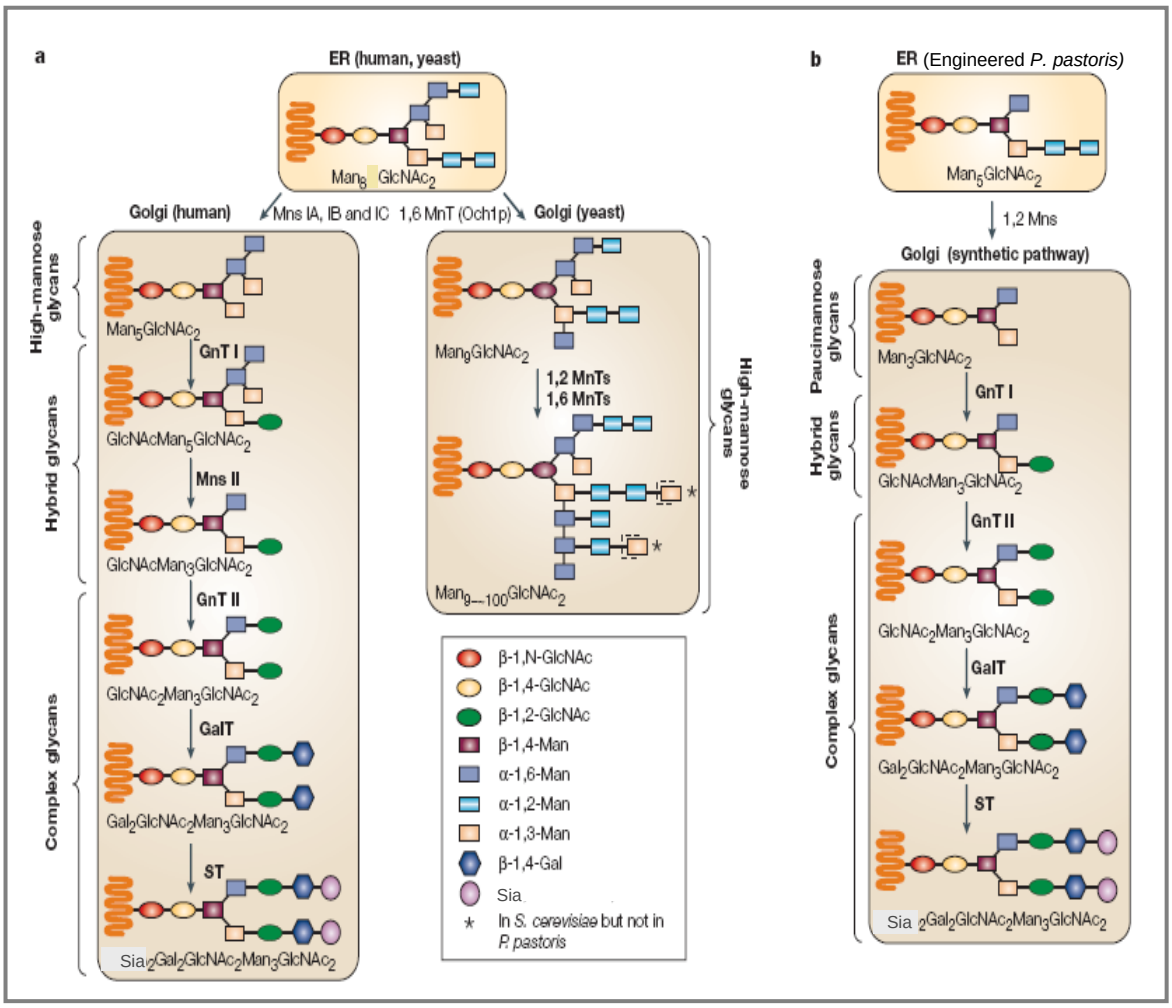


Figure 1.1 Major *N*-glycosylation pathways in humans and yeast. **a)** Representative pathway of N-glycosylation pathways in humans (left) provides a template for humanizing *N*-glycosylation pathways in yeast (right). **b)** Glycoengineering can be implemented to create synthetic glycosylation pathways that lead to complex *N*-glycosylation in yeast. ER: endoplasmic reticulum; GalT: galactosyltransferase; GlcNAc: *N*-acetylglucosamine; GnT1: *N*-acetylglucosaminyl transferase I; GnTII: *N*-acetylglucosaminyltransferaseII; Man: mannose; MnsI: α-1,2-mannosidase; MnsII: mannosidase II; MnTs: mannosyltransferase; Sia: sialic acid; ST: α-2,6-sialytransferase.
[Adapted from Wildt and Gerngross, 2005, pg 20]

generates high mannose, hybrid or complex type oligosaccharides (Cereghino and Cregg, 2000). In contrast, $Man_8(GlcNAc)_2$ glycans in yeast are not trimmed and can in fact be further elongated, sometimes forming hypermannosyl glycans (Vervecken *et al.,* 2004). Methylotrophs have a less pronounced tendency towards over glycosylation than *S. cerevisiae.* A typical outer chain on a *P. pastoris* secreted protein is $Man_8(GlcNAc)_2$ or $Man_9(GlcNAc)_2$ (Montesino *et al.,* 1998). However, examples of over glycosylation in *P. pastoris* have been reported (Scorer *et al.,* 1993). Hypermannosylation of proteins represents a significant problem, as the recombinant proteins can be extremely antigenic when introduced to patients through intravenous injection into the blood stream and will be rapidly cleared from the blood by the liver. Importantly, and unlike the situation in *S. cerevisiae*, no hyper-immunogenic terminal α-1,3-linked mannose residues are incorporated in the *N*-glycans on glycoproteins produced by *P. pastoris*. However the extent and positioning of the long outer mannose chains may affect the activity of the protein by interfering with its ability to fold correctly (Macauley-Patrick *et al.,* 2005). To overcome potential posttranslational processing problems, research has been focused on producing more "humanized" strains of *Pichia* which have been engineered to perform glycosylation in a manner more similar to that in mammalian systems.

1.3.2 Glycoengineered strains of P. pastoris

Generating a properly folded active protein is a necessary requirement of a recombinant protein expression host. The production of heterologous proteins that are *N*-glycosylated in their native state has historically required a mammalian expression system which has the ability to mimic human glycosylation, as aglycosylated forms of glycoproteins tend to be misfolded, biologically inactive or rapidly cleared from circulation (Gerngross, 2004). CHO cells have frequently been used to express glycoproteins with human-like glycosylation patterns.

However the produced glycoproteins are an inherently heterogeneous mixture of glycoforms, which invariably differ in composition to their human counterparts (Wildt and Gerngross, 2005). Further disadvantages of mammalian cell culture are that it is an expensive and lengthy process and the cells are very sensitive to environmental changes during culturing. Process time, pH, glutamine levels and the availability of nucleotide sugars are all factors that have been associated with a corresponding increase in heterogeneity in *N*-glycosylation (Gerngross, 2004). Thus production of recombinant proteins using an expression system that enables modulation and control of glycosylation is advantageous.

Several groups have investigated the possibility of humanizing glycosylation pathways in simple eukaryotes to produce human-like glycoproteins. As mentioned previously, the *N*-glycosylation pathways in yeast and mammals are identical up to the formation of the $Man_8(GlcNAc)_2$ intermediate in the ER (Figure 1.1). However, following transport of the protein to the Golgi apparatus, their *N*-glycan processing diverges significantly. In yeasts, including the methylotroph *P. pastoris*, processing is limited to the addition of mannose and mannosylphosphate sugars. This produces *N*-glycan structures which are mannosylated and hypermannosylated to various extents. This contrasts with the sequence of events in humans which involves the removal of mannose by mannosidases II, followed by the stepwise addition of *N*-acetylglucosamine, galactose and sialic acid (Figure 1.1). Engineering human-like *N*-glycan processing in *P. pastoris* has involved eliminating genes responsible for the non human *N*-glycosylation reactions and introducing genes that mimic human *N*-glycosylation reactions.

The first successfully engineered *P. pastoris* strain that produced high levels of homogenously humanized glycoprotein was published in 2001 (Callewaert *et al.*, 2001). This strain was an *och1* mutant deficient in initiating yeast specific hyperglycosylation and also contained an ER-tagged α-1,2-mannosidase to further trim mannose residues to be more like

the human counterpart. Analysis of the *N*-glycan structure of test proteins showed a >85% decrease in the number of α-1,2-linked mannose residues and that the predominant glycan chain was the desired $Man_5(GlcNAc)_2$. The next step in the mammalian pathway is the addition of *N*-acetylglucosamine (GlcNAc) to the oligosaccharide chain. Two groups reported the production of *P. pastoris* strains which were able to produce almost complete conversion of $Man_5(GlcNAc)_2$ to the hybrid glycan intermediate $GlcNAcMan_5(GlcNAc)_2$ (Choi *et al.*, 2003; Vervecken *et al.,* 2004). In mammals, this hybrid glycan is the substrate for mannosidase II. In 2003, Hamilton *et al.* published reports of their strain of *P. pastoris* which built upon that of their predecessors Choi *et al*. (2003). This strain not only incorporated the mannosidase function needed next in the step-wise progress of human-like glycosylation, but also included the addition of GlcNAc transferase II. This was the first report of a strain that was able to secrete homogenous complex glycans, in this case $(GlcNAc)_2Man_3(GlcNAc)_2$ (Hamilton *et al.,* 2003). A further strain of *P. pastoris* conferring the transfer of galactose onto both terminal GlcNAc residues of the previous glycan form to give $Gal_2(GlcNAc)_2Man_3(GlcNAc)_2$, achieved the penultimate step in creation of fully complex glycans in yeast (Bobrowicz *et al.,* 2004).

In 2006, Hamilton *et al*. reported the production of an engineered yeast with the ability to secrete human glycoproteins with fully complex terminal sialylated *N*-glycans. This strain included the knock-out of four endogenous genes to eliminate yeast specific glycosylation, combined with the introduction of 14 heterologous genes. These engineered changes culminated in the generation of a strain of *P. pastoris* which mimicked the sequential pattern of human glycosylation, giving rise to complex glycoproteins with over 90% $Sia_2Gal_2(GlcNAc)_2$ $Man_3(GlcNAc)_2$ (Hamilton *et al.*, 2006) (Figure 1.1).

The availability of this cell line and its precursors gives researchers the option of producing therapeutic glycoproteins in a non-mammalian host. Yeast strains engineered to

secrete glycoproteins with a high level of glycan uniformity should improve clinically relevant characteristics of the recombinant protein such as tissue distribution, activity and half life. The variety of "humanized" cell lines allows researchers to express proteins with different *N*-glycan sidechains to establish which is the most biologically relevant for each specific glycosylated protein. Li *et al*. (2006) have already taken advantage of this opportunity. They compared the function of a human antibody expressed in the glycoengineered lines of Choi *et al.* (2003), Hamilton *et al.* (2003) and Bobrowicz *et al*. (2004) to find which best optimised the receptor binding affinity of their antibody (Li *et al.*, 2006).

1.3.3 Codon optimisation to improve heterologous expression in P. pastoris

The methylotrophic yeast *P. pastoris* has been developed and extensively used as a host organism for recombinant protein production, however the levels of expression achieved for different heterologous proteins have been widely varied (Cregg *et al.,* 2000). The expression of foreign proteins may exert severe stress on the host cell at a variety of different levels and hence, depending on the specific features of the protein product, different steps may be rate limiting. If it has been established that a gene is both correctly inserted without mutation and efficiently transcribed, but there is a low or undetectable level of protein, it can be speculated that the potential bottleneck is due to a problem at the translational or posttranslational level. Codon optimisation is an approach described in the literature to overcome translational difficulties.

The concept of non random synonymous codon selection was first established by Grantham and colleagues in 1980. They recognized that the DNA sequence used to encode a protein in one organism is frequently substantially different from that which would be used to code for that same protein in a different species (Grantham *et al.*, 1980). This arises due to the

fact that there is more than one codon coding for each αα (with the exception of Met and Trp that are each coded for by a single codon). Not only can the frequencies with which the various codons are used differ notably between different organisms, they can even vary between highly or modestly expressed proteins within the same organism, with highly expressed genes tending to use only a limited number of codons (Gouy and Gautier, 1982). Because translation is the most energetically expensive process occurring in exponentially growing cells, its efficiency is under considerable selective pressure (Rocha, 2004). Natural selection for increased translational efficiency has been proposed as the major cause of inter- and intra-genome differences in codon usage (Bulmer, 1991; Sharp *et al.,* 1993). Several lines of evidence support this hypothesis. Firstly, a strong positive correlation between the frequency of codon usage and the number of available anti-sense tRNA molecules in a given organism has been reported (Gouy and Gautier, 1982; Ikemura, 1985). Secondly, the degree of this usage bias has been found to be related to the level of gene expression, with more highly expressed genes exhibiting greater codon bias than poorly expressed genes (Sharp *et al.*, 1986; Dong *et al.,* 1996). Thirdly, mRNA consisting of preferred codons is translated faster than mRNA molecules which have been engineered to contain rare codons (Sorensen *et al.,* 1989), and the rate limiting step in the elongation of the various nascent polypeptide chains has been shown to be the availability of the given tRNA molecules (Varenne *et al.,* 1984).

Preferred codon bias is not limited to occurring only between genomes or amongst genes within the same genome. Differences in codon usage in different regions of the same gene have also been noted. This is hypothesized to result from selection for increased accuracy of translation, as some codons are more prone to allowing mistakes in translation or a premature stop in elongation of the nascent polypeptide chain (Rocha, 2004). This concept was illustrated in *Drosophila melanogaster*, where preferred codons coded for ααs in highly

conserved positions more frequently than they coded for ααs in non conserved positions within the same gene (Akashi, 1994).

Regardless of the relative translational impact of each of the observations above, it has become increasingly clear that codon biases can have a profound impact on the expression of heterologous proteins. Logically it follows that the more codons a gene contains that are rarely utilized in the host cell, the lower the levels of resultant heterologous gene expression. Altering these rare codons in the target gene (without modifying the amino acid sequence of the protein) so they more accurately reflect both the G+C content and the codon usage of the host, is now a common strategy to try and improve heterologous gene expression (Gustafsson *et al.,* 2004).

Several studies have utilized codon optimisation of just a portion their respective genes as a means to increase levels of heterologous proteins expression in *P. pastoris*. In 2002 and 2003 respectively, two independent groups optimised codons of various gene fragments and found increased levels of expression when compared to WT sequences (Sinclair and Choy, 2002; Yadava and Ockenhouse, 2003). The reverse approach was taken in *S. cerevisiae,* where preferred codons at the 5' end of the coding sequence of a gene were incrementally replaced with an increasing number of synonymous non-preferred ones, leading to a dramatic decline in the level of protein expression (Hoekema *et al.*, 1987). Thus it appears that codon optimisation of even just the 5' region of a heterologous gene can dramatically affect its protein expression levels in yeast.

Several groups have found a combination of abolishing AT-rich regions in conjunction with codon optimising the DNA to contain *P. pastoris* preferred codons, has increased the expression of their recombinant gene in this yeast (Outchkourov *et al.*, 2002; Woo *et al.*, 2002; Xiong *et al.,* 2006). However, Woo *et al.* (2002) attributed their increased expression levels to an increase in mRNA stability and transcription efficiency, rather than an increase in

translation. Outchkourov *et al.* (2002) reported that full gene optimisation approximately doubled the level of both mRNA and protein expression of their gene. Interestingly, both Outchkourov *et al*. (2002) and Xiong *et al*. (2006) noted that modifications of vector sequences upstream of their recombinant gene had an even greater effect on protein expression levels. The removal of a short stretch of 12 additional nucleotides from the multiple cloning site in the *Pichia* expression vector pPIC9 doubled the mRNA expression for Outchkourov *et al*. (2002) and led to a 4 to10-fold increase at the protein level. Modification of the *S. cerevisiae* mating factor α-prepro leader sequence to a codon optimised synthetic signal peptide increased protein expression to over three times that obtained by codon optimising the gene sequence alone (Xiong *et al.,* 2006).

Codon optimisation for expression of human genes in *P. pastoris* generally results in an accompanying decrease in the overall G+C content of the gene. The relatively low G+C bias of *P. pastoris* preferred codons reflects the tendency of *Pichia* towards A/T ended codons (while mammals prefer G/C in the degenerate positions). Hu *et al*. (2006) concurrently reduced the G+C content and selected *P. pastoris* preferred codons for their human ERB2 protein and noted a 3 to 5-fold increase of expression when compared to the wildtype human sequence. Northern blots showing consistent RNA levels, confirmed that the increase in expression was mainly due to enhanced translational efficiency (Hu *et al.*, 2006). Sinclair and Choy (2002) also found reduction of G+C content and codon optimisation increased levels of protein expression. When G+C content and codon usage were independently optimised, similar transcription rates of all constructs but increased levels of protein expression were observed. This led the authors to conclude that the increased expression was due to posttranslational effects.

Two groups have reported initially undetectable levels of protein expression when the native form of their genes was cloned into the *P. pastoris* expression system (Qian *et al.*, 2003;

Delroisse *et al.,* 2005). Although RNA levels were not tested by Qian *et al*. (2003), Delroisse *et al.* (2005) detected full length RNA of the expected size of their native non-optimised gene by Northern blot analysis. After codon optimisation of their respective genes, in conjunction with modification of the poly (A) signal and secondary structure of the 5' end of the cDNA in the case of Qian *et al.* (2003), both groups reported detectable levels of expression of their modified proteins. This indicates, at least in the case of Delroisse and colleagues, that the increased level of protein expression seen after codon optimisation was due to an increased ability of the recombinant clone to be expressed at the translational or posttranslational level.

1.3.4 Heterologous expression in insect cells

The lepidopteran insect cell baculovirus expression vector system (BEVS) was first utilized as a means of recombinant protein expression in 1983 (Smith *et al.*, 1983). Since its introduction over 20 years ago, the system has been modified and improved and is now one of the most popular systems for recombinant protein expression (Ho *et al.*, 2004; Kost *et al.*, 2005). One of the early disadvantages of the system was the difficulty in generating recombinant virus. The standard method of producing virus expressing recombinant proteins is to co-transfect insect cells with viral DNA and DNA of a transfer vector modified to incorporate the foreign gene. This produces recombinant virus, but at a very low frequency of approximately 0.1%. In 1993, the efficiency of this system was vastly improved when baculovirus DNA was engineered to contain a deletion that functionally inactivated an essential gene. This precluded replication of parental virus and increased the frequency of recombinant virus to over 99% (Kitts and Possee, 1993).

The BEVS has several advantages over other recombinant protein expression systems. One of the most pronounced is the high recombinant protein yields enabled by the strong very

late polyhedrin promoter. In the native baculovirus, the polyhedrin protein accounts for 30-50% of the total insect protein at the late stage of the baculovirus infection cycle. Up to 1 mg of polyhedrin protein may be synthesized per 1-2 x 10^6 infected cells. As the polyhedrin promoter is not essential for the baculovirus life cycle in tissue culture, it can be replaced by a recombinant gene of interest, thereby enabling very high levels of expression of recombinant protein. Recombinant protein levels may reach between 0.1% and 50% of total insect cellular protein (Ong *et al.*, 1998). Expression levels of recombinant protein under the control of the viral polyhedrin promoter have commonly been found to surpass the productivity obtained with mammalian *in vitro* systems (Altmann *et al.*, 1999). The three most frequently used insect cell lines are *Sf9* and *Sf21* from the fall armyworm, *Spodoptera frugiperda,* and TN5B1-4 (*Tni*) from the cabbage looper, *Trichoplusia ni* (Patterson *et al.*, 1995). Comparison of expression levels of a variety of different recombinant proteins has indicated *Tni* cells typically exhibit greater recombinant protein synthesis than the *Sf9* or *Sf21* cell lines (Davis *et al.*, 1993; Patterson *et al.*, 1995; Mukherjee *et al.*, 1995; Zhang *et al.*, 2002; Andersen, 2004).

One disadvantage of a viral-based system is the lytic nature of viral infection. The advent of baculovirus transfer vectors designed to secrete the recombinant protein into the culture media has overcome many of the problems with lytic infection because it allows harvesting of recombinant proteins from the supernatant prior to cell lysis. However, as recombinant protein (under transcriptional regulation of the very late polyhedrin promoter) is being produced late in the baculovirus infection cycle, some disruption of the cellular machinery may occur. For proteins requiring posttranslational processing such as those destined for lysosomal uptake, this could result in low level expression of a heterogeneous recombinant protein product (Ailor and Betenbaugh, 1999).

This possible limitation of the baculovirus based insect cell expression systems can be avoided by the use of stably transformed insect cell lines. These are plasmid mediated expression systems whereby the gene of interest is cloned into an expression vector and, upon introduction to the cell, is stably integrated into the insect chromosomal DNA. Promoters utilized to drive heterologous gene expression in these systems are derived from baculovirus genes active early in the infection process, which presumably allows more accurate and complete posttranslational modification of recombinant proteins. Stably transformed insect cell lines with constitutive expression mediated by the *Orgyia pseudotsugata* nucleopolyhedrosis virus (*Op*MNPV) immediate-early 2 (ie2) promoter used in conjunction with a human transferrin secretion signal have successfully produced a number of secreted recombinant proteins (Theilmann and Stewart, 1992; Pfeifer *et al.*, 1997, 2001; Hegedus *et al.*, 1998). As the processing machinery of the cell is not compromised by viral infection, it is proposed that secreted proteins of a more homogeneously posttranslationally modified nature can be purified from the cell media.

Historically one of the appealing features of insect expression systems, along with the ease of growth, handling and scale-up for large scale culture, is their eukaryotic protein processing capabilities such as *N*- and *O*-linked glycosylation which enable correct protein folding. Insect cells are reported to be second only to mammalian cell lines with regard to posttranslational processing (Pfeifer, 1998; Altmann *et al.*, 1999). Consequently there is a greater probability that a protein expressed in insect cells will have normal biological activity than a protein produced in the frequently used *E. coli* expression system.

As with *P. pastoris*, there is relatively little known about the details of *O*-linked glycosylation in insect cells. Although it appears that for most human proteins *O*-linked glycosylation in insect cells occurs on the same residues as it does in the naturally occurring

protein, the extent of glycosylation and type of *O*-linked glycans added appear quite frequently to differ (Altmann *et al.,* 1999).

The mechanism of *N*-linked glycosylation, however, has been more extensively studied. The complexity of insect cell *N*-glycosylation lies between that of *P. pastoris* and mammalian cells. In all three systems, *N*-glycosylation begins in the ER with the transfer of the oligosaccharide unit, $Glc_3Man_9(GlcNAc)_2$, to an asparagine residue at a specific recognition site in the protein. The structure is then trimmed to the common intermediate, $Man_8(GlcNAc)_2$. At this point the glycosylation pathway of native *P. pastoris* diverges to generate high mannose and occasionally hypermannosylated *N*-glycans. In insect and mammalian cells, sequential trimming and elongation reactions occur to yield $GlcNAcMan_3(GlcNAc)_2$. However subsequent elongation reactions to produce the complex *N*-glycans seen in mammalian cells generally do not occur in insect cell lines. In fact, further trimming occurs. The major processed *N*-glycan produced in lepidopteran cell lines is the paucimannosidic structure, $Man_3(GlcNAc)_2$. Insect cell lines appears to contain all the enzymes involved in *N*-glycan trimming, but few of those that give rise to elongation of mammalian glycans (Altmann *et al.,* 1999). Recent genomic studies indicate that insects themselves may have the capacity to produce complex glycosylation patterns which more closely reflect those produced in mammals, but that this capability has been lost in most of the cell lines derived for cell culture (Tomiya *et al.,* 2004).

As with *P. pastoris*, a number of groups have been working to genetically engineer insect *N*-glycosylation patterns so they more closely resemble those of mammalian cells. Transgenic *Sf9* cells have been successfully manipulated to express mammalian genes encoding functions that are missing or limited relative to mammalian cells (Jarvis, 2003). SfSWT-3, the insect cell line containing the most efficient and extensive *N*-glycan processing pathway

produced to date, has been engineered to contain a galactosyltransferase (for addition of Gal residues), five *N*-acetylglucosaminyltransferases (for addition of GlcNAc residues), a sialic acid synthase and a CMP-sialic acid synthetase (for addition of Sia residues). This transgenic cell line has been shown to produce extensively processed biantennary, terminally disialylated *N*-glycans, thus mimicking the human complex glycoprotein pattern of $Sia_2Gal_2(GlcNAc)_2$ $Man_3(GlcNAc)_2$ (Aumiller *et al.,* 2003). The use of a glycoengineered *Sf9* cell line such as SfSWT-3 for the production of lysosomal enzymes certainly warrants further investigation. Nonetheless, previous work in our laboratory has shown that the wildtype *Sf9* system is able to produce active recombinant NAGLU estimated to be approximately 83 kDa by immunoblot (Bandsmer *et al.,* 2006). As the 720 αα mature NAGLU protein is predicted to have a molecular mass of 80.3 kDa, the presence of an 83 kDa band generated by *Sf9* cells may indicate some posttranslational glycosylation has occurred.

1.4 Project Overview and Objectives

This work focuses on attempts to express human NAGLU for ultimate use in enzyme replacement therapy. To date, recombinant NAGLU has been expressed in CHO cells (Zhao and Neufeld, 2000; Weber *et al.,* 2001), HEK 293 cells (Zhao and Neufeld, 2000), HeLa cells, human skin fibroblasts (Yogalingam *et al.,* 2000) and *Sf9* cells (Bandsmer *et al.,* 2006). However inefficient MPR-mediated protein uptake and low levels of protein expression have limited the usefulness of the recombinant protein produced thus far. Finding an inexpensive and reliable production process which generates functionally active protein is vital for proteins destined for use in ERT.

With the aim of achieving this goal, two different approaches were utilized. In the first approach, which is the major focus of this thesis, the use of the single-celled eukaryote *P.*

pastoris for the secreted expression of recombinant human NAGLU was investigated. In the second, insect cells both stably transformed and infected with baculovirus, were used in attempts to produce NAGLU fusion proteins with and without the addition of protein transduction domains.

The aim in expressing recombinant NAGLU in *P. pastoris* was to produce stable, functionally active protein that is able to correctly localize in cells. To this end, the "humanized" *P. pastoris* strain pGlycoSwitchM5 was chosen. This strain has the ability to produce a homogeneous $Man_5(GlnAc)_2$ high mannose form of recombinant NAGLU. The advantages of this are twofold. Firstly, the mammalian-like glycosylation pattern endowed by this strain on NAGLU will presumably prevent the degradation of the recombinant protein and its clearance from the blood. Secondly, the high mannose form of *N*-glycans produced may enable more efficient mannose 6-phosphorylation of these residues, circumventing the uptake problems experienced by researchers expressing heterogeneously glycosylated forms of recombinant NAGLU from mammalian cells. The ability to produce a homogeneously glycosylated protein could also result in an improvement of therapeutic function and concomitantly enable possible elimination of undesirable functions.

Previous research conducted in our laboratory using the *P. pastoris* expression system generated easily detectable levels of recombinant mRNA, but low to undetectable levels of NAGLU protein (Patrick, 2006). To address the problem of low level expression of the NAGLU protein, the use of codon optimisation - an approach described in the literature as a means of successfully overcoming translational difficulties, was investigated. The premise of codon optimisation suggests that if codons in the recombinant gene are modified to more closely reflect the codon usage of the host cell, then the resultant levels of heterologous gene expression may be increased. To evaluate the effectiveness of this approach, codons of a 5'

portion of the human *NAGLU* gene were initially altered to more accurately reflect both the G+C content and the codon bias of *P. pastoris*. Had sufficiently increased expression levels been attained, complete codon optimisation and further protein modifications such as tags to allow either transduction or increased ease of protein purification would have been incorporated.

In the second approach, work was initially focused on the production of *NAGLU* fusion constructs with and without TAT-PTDs, in stably transformed *Sf9* and *Tni* insect cell lines. A limitation of ERT has been its inability to address neurodegeneration due to restricted access to the brain. The postulate of this work is that the NAGLU-TAT fusion proteins produced in insect cells may have enhanced trans-membrane capabilities compared to that of the untagged protein. Previous studies in this laboratory using *Sf9* cells stably transformed with *NAGLU* constructs fused with a cellulose binding domain (CBD) for purification, showed low levels of CBD-tagged NAGLU expression (Bandsmer, 2004). With the aim of obtaining increased expression levels in this current project, constructs encoding *NAGLU* and *NAGLU*-TAT were created such that the large CBD domain was replaced with a small cleavable hexahistidine tag to aid purification. These constructs were sent to our collaborator Dr. T. Pfeifer (University of British Columbia, BC) who has extensive experience with expression of heterologous proteins in stable insect cell culture (Pfeifer, 1998; Pfeifer *et al.*, 2001; Sinclair *et al.*, 2006).

Next, the use of the baculovirus expression vector system was examined as a means to express NAGLU and NTAT recombinant proteins in insect cells. Preliminary experiments with amino His tagged *NAGLU* constructs purified under native and denaturing conditions, indicated that the amino hexahistidine tag may be inaccessible to the purification matrix under native conditions, due to the conformational structure of the NAGLU protein (Nolla, 2007). Similar His tag accessibility problems had been overcome by the insertion of a linker sequence

to physically separate the purification tag from the recombinant protein (Pfeifer, *pers com*). Following this approach, a linker region of 78 bp (26 ααs) was designed and cloned downstream of the polyhistidine purification tag and upstream of the NTAT protein. Additionally, both *NAGLU* and *NTAT* constructs were designed with carboxyl terminal hexahistidine tags to test the accessibility of tags at each end of the NAGLU protein. These constructs were given to Dr. M. Boulanger (University of Victoria, BC) to perform small scale proof of principle studies to enable preliminary assessment of production of NAGLU and NTAT protein using an insect cell based lytic expression system.

2 Materials and Methods

2.1 Materials

Materials Contributed by Others

Anti-NAGLU primary antibody was provided by Drs. Neufeld and Zhao (University of California, CA) (Yu *et al.,* 2000). The humanized Pichia pastoris strain, pGlycoSwitchM5, was donated by Dr Contreras (Ghent University, Belgium) (Vervecken *et al.,* 2004). The pPICZα(NAGLU) template was supplied by C. Patrick (University of Victoria, BC), the p2ZOpTCXFegfp template by A. Vaags (University of Victoria, BC), and the p2ZOp2tcXF.NAGLU and p2ZOp2tcXF.NTAT templates by J. Bandsmer (University of Victoria, BC). The *Sf9* cell line produced CBD-NTAT NAGLU positive control was provided by J. Bandsmer (University of Victoria, BC). The *P. pastoris* produced His tagged H-protein positive control was given as a gift by A. Zay (University of Victoria, BC). Dr. Tom Pfeifer (University of British Columbia, BC) performed stable transfections in *Sf9* and *Tni* cell lines using p2ZOp2.THF.*NAGLU* and p2ZOp2.THF.*NTAT* clones. Construct design, cloning and analysis of transfection outcomes were performed by this author and described herein. Dr. Martin Boulanger (University of Victoria, BC) produced primary virus for constructs pAcGP67B.*NAGLU*.H_6, pAcGP67B.*NTAT*.H_6 and pAcGP67B.H_6.L.Fx.*NTAT* in *Sf9* cells, infected High Five™ cells and supplied supernatant aliquots for recombinant protein analysis. Construct design, cloning and recombinant protein analysis were performed by this author and described herein.

Chemicals and Reagents

ACP Chemicals, Montreal, QC: Na phosphate (monobasic), EDTA, glycerol, Na acetate. *BD Biosciences, Oakville, ON:* pAcGP67B, BaculoGold™ Bright Linearized Baculovirus DNA, 6 well tissue culture plates. *Bioneer, Alameda, CA:* 101mer oligos. *BioRad, Hercules, CA:* Bradford Protein Assay, Precision Plus dual colour marker. *Eastman Kodak, Rochester, NY*:

Kodak BioMax film. *EMB Biosciences, San Diego, CA:* pIEx-10, phenylmethylsulphonyl fluoride (PMSF), 0.2 mM 4-MU α-*N*-acetylglucosaminide, sorbitol. *Expression Systems, Woodland, CA:* Expressions Systems Formula Animal free Medium. *Fisher Scientific, Ottawa, ON*: Coomassie Brilliant Blue R-250. *GenScript Corporation, Piscataway, NJ*: pUC57 containing codon optimized insert. *Integrated DNA Technologies Inc., Coralville, IA*: custom primers. *Invitrogen, Burlington, ON:* 1x PCR Reaction Buffer (20 mM Tris-HCl (pH 8.4) 50 mM KCl), 20 mM $MgCl_2$, 100 mM dNTPs, Taq DNA Polymerase, 1x High Fidelity PCR Buffer (60 mM Tris-SO_4, pH 8.9, 18 mM $(NH_4)_2SO_4$), 50 mM $MgSO_4$, Platinum® Taq DNA Polymerase High Fidelity, isopropyl-beta-D-thiogalactopyranoside (IPTG), 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal), pPIC9K vector, *Pichia pastoris* Expression kit, 5'*AOX1* (F), 3'*AOX1* (R) primers, SuperScript™ II RNaseH$^-$ Reverse Transcriptase, Grace's Insect Cell Culture Medium, Cellfectin®, Zeocin™, SF-900 II SFM, High Five™ cells, Express Five® SFM, gentamicin sulphate. *InvivoGen, San Diego, USA:* QUANTI-Blue™ Assay. *MBI fermentas*, *Burlington, ON*: pUC Mix Marker 8. *New England Biolabs, Beverly, MA*: AvrII, BamHI, EcoRI, HindIII, SacI, SacII, SnaBI, SphI, NotI, XbaI restriction enzymes; 1x NEBuffer 1 (10 mM Bis-Tris-Propane-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.0), 1x NEBuffer 2 (10 mM Tris-HCl, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM dithiothreitol (DTT), pH 7.9), 1x EcoRI Buffer (100 mM Tris-HCl, 50 mM NaCl, 10 mM $MgCl_2$, 0.025% Triton X-100, pH 7.5), 100x BSA, 1x NEBuffer 4 (20 mM Tris-acetate, 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM dithiothreitol (DTT), pH 7.9), T4 DNA ligase, 1x T4 DNA ligase buffer (50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 25 µg/ml BSA), 3x SDS Sample buffer, 42 mM DTT, Broad Range Protein standard, 1 kb DNA standard. *Pierce Biotechnology, Rockford, IL:* SuperSignal® West Dura chemiluminescent reagent. *Promega, Madison, WI:* pGEM-T vector. *Qiagen, Mississauga, ON:* QIAquick® Gel Extraction Kit,

QIAquick® PCR Purification Kit, QIAprep® Spin Miniprep Kit, QIAGEN Plasmid Midi Kit®, RNeasy® Mini Kit, anti-His_4 antibody, RNase free DNase, Ni-NTA Agarose. *Seikagaku Corporation, Tokyo, Japan*: Zymolyase. *Sigma-Aldrich, Oakville*, *ON:* ampicillin, kanamycin, betaine, dimethyl sulphoxide (DMSO). *Stratagene, La Jolla, CA*: *E. coli* strain XL1 blue. *Stressgen Biotechnologies, Victoria, BC*: goat anti-rabbit horseradish peroxidase conjugated antibody. *Santa Cruz Biotechnology, Santa Cruz, CA:* goat anti-mouse horseradish peroxidase conjugated antibody.

Media and Prepared Solutions

Annealing buffer: 10 mM TrisHCl (pH 7.5), 50 mM NaCl. *Buffered glycerol complex medium (BMGY):* 1% (w/v) yeast extract, 2% (w/v) peptone, 0.1 M potassium phosphate buffer, pH 6.0, 1% (v/v) yeast nitrogen base with ammonium sulphate without amino acids, 1% (v/v) glycerol, 0.0004% (v/v) biotin. *Buffered methanol complex medium (BMMY):* 1% (w/v) yeast extract, 2% (w/v) peptone, 0.1 M potassium phosphate buffer, pH 6.0, 1% (v/v) yeast nitrogen base with ammonium sulphate without amino acids, 1.0% (v/v) methanol, 0.0004% (v/v) biotin. *Coomassie staining solutions: Coomassie concentrated stock:* 33.33 mg/ml Coomassie Brilliant Blue R-250 in 83.33% (v/v) methanol, 16.66% (v/v) acetic acid. *Destaining solution*: 45% (v/v) methanol, 10% (v/v) acetic acid. *Fixing solution:* 50% (v/v) methanol, 10% (v/v) acetic acid. *Cracking buffer*: 5 mM EDTA, 10% (w/v) sucrose, 0.25% (w/v) SDS, 100 mM NaOH, 60 mM KCl and 0.05% (w/v) bromophenol blue. *0.5 M Glycine-NaOH buffer:* 37.54 mg/ml glycine (pH 10.5). *Luria-Bertani broth (LB) plates:* 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% NaCl, pH 7.5, 1.5% (w/v) agar. *MD plates*: 1.34% (v/v) yeast nitrogen base with ammonium sulphate without amino acids, 0.0004% (v/v) biotin, 2% (v/v) dextrose and 1.5% agar. *MDH plates:* 1.34% (v/v) yeast nitrogen base with ammonium sulphate without amino acids, 0.0004% (v/v) biotin, 2% (v/v) dextrose, 0.004% (v/v) histidine and 1.5% agar.

MM plates: 1.34% (v/v) yeast nitrogen base with ammonium sulphate without amino acids, 0.0004% (v/v) biotin, 0.5% (v/v) methanol and 1.5% agar. *Silver Staining Solutions: Fixative*: 50% (v/v) methanol, 12% (v/v) acetic acid, 0.1% (v/v) formaldehyde*; Wash:* 50% (v/v) ethanol; *Pretreatment solution*: 0.02% (w/v) sodium thiosulfate-pentahydrate; *Stain*: 2 mg/ml silver nitrate in 0.075% (v/v) formaldehyde; *Developer:* 60 mg/ml sodium carbonate, 0.05% (v/v) formaldehyde, 0.002% (w/v) sodium thiosulfate pentahydrate; *Stop solution*: 50% (v/v) methanol. *SOC medium*: 2% tryptone (w/v), 0.5% yeast extract (w/v), 10 mM $NaCl_2$, 2.5 mM KCl, 10 mM $MgCl_2$-$6H_2O$, 10 mM $MgSO_4$-$7H_2O$. *0.1 M Na acetate buffer:* 8.203 mg/ml Na-acetate, (pH 4), 0.5 mg/ml BSA. *Transfer buffer:* 10% (v/v) methanol, 25 mM Tris-HCl, 0.2 M glycine. *TBS-T*: 20 mM Tris-HCl, pH 7.5, 0.05% (v/v) Tween-20, 500 mM NaCl. *Yeast breaking buffer (pH 7.4):* 50 mM sodium phosphate (monobasic), 1 mM PMSF, 1 mM EDTA and 5% glycerol. *YPD:* 1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) dextrose.

Equipment and Software

Amersham Biosciences, Piscataway, NJ: SE 260 mini-vertical gel electrophoresis unit, Hybond-P PVDF membrane. *Barnstead/Thermolyne, Dubuque, IA:* Sequoia-Turner Model 450 Digital fluorometer. *BioRad, Hercules, CA*: Genepulser electroporation machine, 0.2 cm gap electroporation cuvettes, Mini-protean II Electroblot apparatus. *Isis Pharmaceuticals, Carlsbad, CA*: BioEdit sequence alignment software. *Lazergene, Madison, Wisconsin:* Editseq. *Perkin Elmer, Wellesley, MA*: GeneAmp PCR thermal cycler. *Pharmacia LKB Biochrom Ltd, Cambridge, ENG*: Novaspec® visible spectrophotometer. *Stratagene, La Jolla, CA*: EagleEye II still video system.

2.2 Construction of Yeast Expression Vectors

2.2.1 Construct 1: pPIC9K.FL.WT.NAGLU

2.2.1a Isolation of full length WT NAGLU cDNA

FL.WT.*NAGLU* (FL, full length; WT, wildtype; *NAGLU*, α-*N*-acetylglucosaminidase) cDNA which encodes the mature peptide without the native signal peptide sequence, was obtained by EcoRI/XbaI restriction digest of a sequence verified pPICZα(*NAGLU*) construct (Patrick, 2006). Ten (10) μg of pPICZα(*NAGLU*) were digested concurrently with 40 units (U) of EcoRI and 40 U of XbaI in 1x NEBuffer 2 supplemented with 100 μg/ml bovine serum albumin (BSA). The reaction was incubated for 16 hours (hr) at 37 °C and then heat inactivated at 65 °C for 20 minutes (min). Correct digestion was confirmed by separation of products on a 1% (w/v) agarose gel followed by ethidium bromide (EtBr) staining and visualization using the EagleEye II still video system. The approximately (~) 2.2 kb insert band was excised and purified using a QIAquick® Gel Extraction Kit according to the manufacturer's protocol.

2.2.1b Cloning of full length WT NAGLU cDNA into pPIC9K

As the polylinker fragment of pPIC9K excised by EcoRI/AvrII digestion is too small to easily visualize, sequential digests of the pPIC9K vector were performed. EcoRI digestion followed by AvrII digestion is described below, but AvrII followed by EcoRI digestion was also performed to ensure complete digestion was successful with both enzymes. Initially 10 ug of pPIC9K DNA were digested with 40 U of EcoRI in 1x EcoRI Buffer for 16 hr at 37 °C to linearize the vector. Linearization was confirmed by electrophoresis of an aliquot of the digested product as described in section 2.2.1a. Addition of 20 U of AvrII was followed by

incubation at 37 °C for a further 16 hr. The now doubly digested vector was electrophoresed on a 1% (w/v) agarose gel, excised and purified as described in section 2.2.1a.

Ligation of 60 ng EcoRI/AvrII digested vector and 40 ng EcoRI/XbaI digested insert was carried out at room temperature (RT) for 1 hr using 1 U of T4 DNA ligase and 1x T4 DNA ligase buffer. Electrocompetent *E. coli* strain XL1 blue (40 µL) cells were transformed with 2 µl of the ligation reaction in 0.2 cm gap cuvettes using a Genepulser electroporation machine set at 2.5 kV and 200 Ω resistance. Following electroporation, 500 µl of SOC medium was added and the cells were incubated, shaking at 37 ^{o}C for 1 hr, plated on Luria-Bertani broth (LB) plates containing 50 µg/ml kanamycin and incubated at 37 ^{o}C for 16 hr.

Kanamycin resistant colonies were screened by colony cracking for the presence of plasmids containing the appropriate insert. Specifically, toothpick smears of individual colonies were mixed in 20 µl of cracking buffer, incubated at 55 °C for 1 hr, and centrifuged at 16,000 x g for 15 min. Plasmid sizes were then analysed by loading 20 µl of lysed cell supernatant onto a 1% (w/v) agarose gel and visualized as described in section 2.2.1a. Plasmids appearing to contain the correct insert were prepared with the QIAprep® Spin Miniprep Kit according to the manufacturer's protocol.

As the AvrII/XbaI site is destroyed through ligation, EcoRI/NotI digestions were performed to confirm correct size of insert in candidate clones. Plasmid DNA (1 µg) was incubated at 37 °C for 3 hr with 10 U EcoRI, 10 U NotI, 100 µg/ml BSA and 1x EcoRI Buffer and inactivated at 65 °C for 20 min. Digestion patterns were checked via 1% (w/v) agarose gel electrophoresis as described above and clones showing the correct digestion pattern were sent for sequencing by the DNA sequencing facility at the Center for Biomedical Research (University of Victoria, BC). DNA sequence data was analysed using BioEdit sequence alignment software. The QIAGEN Plasmid Midi Kit® was utilized to generate a sufficiently

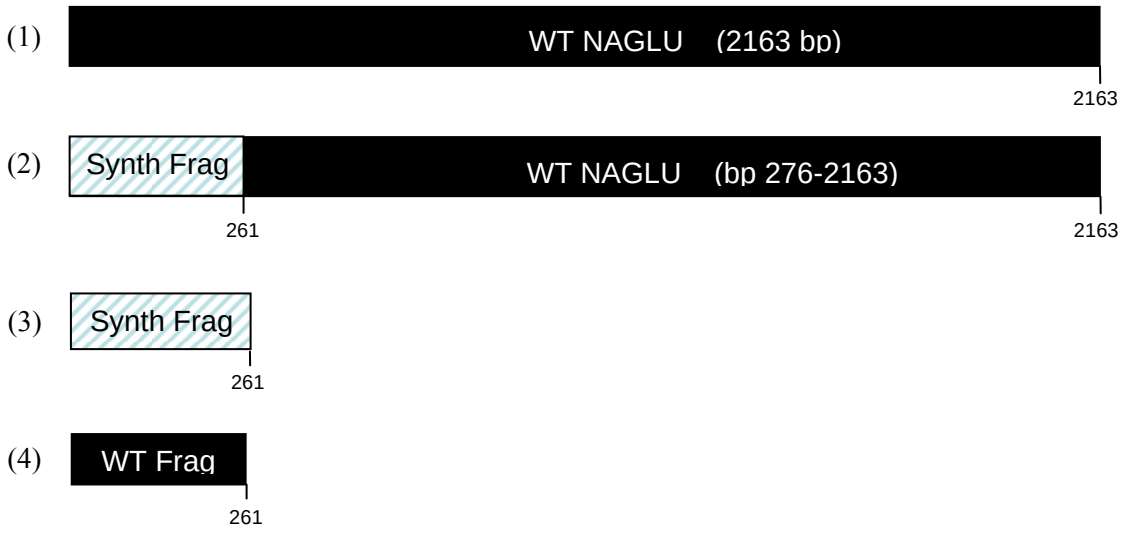
high quantity and quality of mutation free open reading frame (ORF) plasmid DNA for use in both the construction of pPIC9K.FL.(CO+WT).*NAGLU* (outlined below) and for future yeast transformation. A schematic representation of this final Construct 1: pPIC9K.FL.WT.*NAGLU* expression vector is shown in Figure 2.1. In this clone, an extra EcoRI site was created at the 3' end of the insert, through a cloning anomaly. As the extra site lies 3' of the stop codon, it causes no change in function of the protein and thus work was continued with this clone.

2.2.2 Construct 2: pPIC9K.FL.(CO+WT).NAGLU

2.2.2a Creation of the codon optimised insert

Synthesis of a 261 bp codon optimised fragment corresponding to the initial 261 bp of mature *NAGLU* cDNA was performed by GenScript (Figure 2.2). The synthetic 261 bp was designed to contain additional 5' and 3' EcoRI and SacII sites respectively and was supplied as a cloned fragment contained within the pUC57 vector. Polymerase chain reaction (PCR) amplification of this fragment was performed using 1x High Fidelity PCR Buffer, 3 mM $MgSO_4$, 0.25 mM dNTPs, 0.5 µM forward (SnaBI.CO.F) and reverse (SacII.CO.R) custom primers, and 1 U of Platinum® Taq DNA Polymerase High Fidelity in a 50 µl final volume. The anomalous EcoRI site created in Construct 1: pPIC9K.FL.WT.*NAGLU*, necessitated the design of the forward primer (SnaBI.CO.F) containing an additional 5' SnaBI site for use in cloning. Primer sequences are listed in Table 2.1. Amplification was performed using a GeneAmp PCR thermal cycler with an initial denaturation of 94 ^{o}C for 5 min, followed by 30 cycles of a 94 ^{o}C denaturation for 1 min, 58 ^{o}C annealing for 1 min, and 68 ^{o}C elongation for 2 min, with a final elongation for 10 min at 68 ^{o}C. Successful PCR amplification was confirmed by electrophoresis on a 1.5% agarose gel and visualized as described in section 2.2.1a. The remaining product was QIAquick® purified according to manufacturer's instructions.

a)



b)

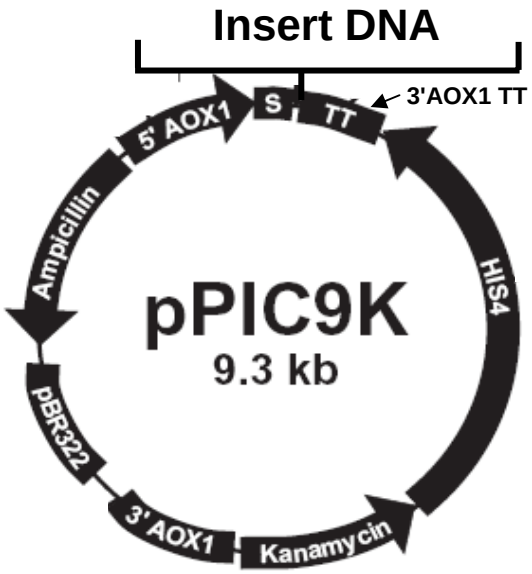


Figure 2.1 Schematic representation of inserts 1-4 cloned into the *P. pastoris* expression vector pPIC9K. a) (1) Construct 1: pPIC9K.FL.WT.*NAGLU*. (2) Construct 2: pPIC9K.FL.(CO+WT).*NAGLU*. (3) Construct 3: pPIC9K.control.HFxCO.F. (4) Construct 4: pPIC9K.control.HFxWT.F. **b)** pPIC9K encodes the highly inducible alcohol oxidase promoter (5' *AOX1*) and *S. cerevisiae* α-factor secretion signal (S) for secreted expression. Expression of the kanamycin/neomycin gene product (Kanamycin) allows for the selection of positive transformants via kanamycin selection in *E. coli* and geneticin selection in *P. pastoris*. Expression of the histidinol dehydrogenase (HIS4) gene provides an alternate means of selection in *P. pastoris.*

a)

```
Optimised    1 GATGAAGCTAGAGAAGCTGCTGCTGTTAGAGCTTTGGTTGCTAGATTGTTGGGTCCAGGT
Original     1 GACGAGGCCCGGGAGGCGGCGGCCGTGCGGGCGCTCGTGGCCCGGCTGCTGGGGCCAGGC

Optimised   61 CCAGCTGCTGATTTTTCTGTTTCTGTTGAAAGAGCTTTGGCTGCTAAGCCAGGTTTGGAT
Original    61 CCCGCGGCCGACTTCTCCGTGTCGGTGGAGCGCGCTCTGGCTGCCAAGCCGGGCTTGGAC


Optimised  121 ACTTACTCTTTGGGTGGTGGTGGTGCTGCTAGAGTTAGAGTTAGAGGTTCTACTGGTGTT
Original   121 ACCTACAGCCTGGGCGGCGGCGGCGCGGCGCGCGTGCGGGTGCGCGGCTCCACGGGCGTG

Optimised  181 GCTGCTGCTGCTGGTTTGCATAGATACTTGAGAGATTTTTGTGGTTGTCATGTTGCTTGG
Original   181 GCGGCCGCCGCGGGGCTGCACCGCTACCTGCGCGACTTCTGTGGCTGCCACGTGGCCTGG

Optimised  241 TCTGGTTCTCAATTGAGATTG
Original   241 TCCGGCTCTCAGCTGCGCCTG
```

b)

```
Optimised    1 DEAREAAAVRALVARLLGPGPAADFSVSVERALAAKPGLDTYSLGGGGAARVRVRGSTGV
Original     1 DEAREAAAVRALVARLLGPGPAADFSVSVERALAAKPGLDTYSLGGGGAARVRVRGSTGV

Optimised   61 AAAAGLHRYLRDFCGCHVAWSGSQLRL
Original    61 AAAAGLHRYLRDFCGCHVAWSGSQLRL
```

Figure 2.2 Codon optimisation of the initial 261 bp of *NAGLU* mature cDNA. a) DNA alignment of the optimised region. Codons were optimised to correspond to the most prevalent *P. pastoris* codon for a given amino acid. The changed codons are indicated in red. The optimised synthetic fragment has 45.59% GC content whereas the native sequence had 78.93% GC content. **b)** Protein alignment of the optimised region. All amino acids have remained unchanged.

Table 2.1 Oligonucleotide primers used to create *NAGLU* Constructs 2-4 for expression in *P. pastoris*.

Primer Name[d]	Nucleotide Sequence 5' to 3'
SnaBI.CO.F[a]	**TAC GTA** GAA TTC GAT GAA GCT AGA GAA G
SacII.CO.R[a]	**CCG CGG** CAA TCT CAA TTG AGA
Fx.WT.F[b]	*ATT GAA GGT CGT GTT* GAC GAG GCC CGG
AvrII.WT.R[a]	**CCTAGG** TTA CAG GCG CAG CTG AGA
Fx.CO.F[b]	*ATT GAA GGT CGT GTT* GAT GAA GCT AGA GAA G
NotI.CO.R[a]	**GCGGCCGC** TTA CAA TCT CAA TTG AGA
EcoRI.HFx.F[a,b,c]	**GAA TTC** CAT CAT CAT CAT CAT CAT *ATT GAA GGT CGT GTT*

[a] Restriction enzyme recognition sequences are presented in bold print.
[b] The Factor Xa cleavage site is presented in italic print.
[c] The hexahistidine tag is underlined.
[d] Primer orientation is indicated in the primer name as forward (F) or reverse (R).

2.2.2b Cloning of the codon optimised insert into pGEM-T

Amplicons of the SnaBI/SacII synthetic fragment were T/A ligated into the pGEM-T vector according to the manufacturer's instructions. Electrocompetent *E. coli* cells were transformed as described in section 2.2.1b. The transformation reaction was then plated on LB plates containing 0.4 mM isopropyl-beta-D-thiogalactopyranoside (IPTG), 0.04 mg/ml 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal), and 100 µg/ml ampicillin and incubated at 37 ^{o}C for 16 hr.

Positive clones were initially identified by blue-white colour screening and then screened for the presence of correct inserts by colony PCR. In short, using a sterile toothpick a small portion of each individual colony was first gridded onto an LB master plate containing 100 µg/ml ampicillin and grown at 37 ^{o}C for 16 hr. Each colony was then resuspended in an individual PCR reaction volume such that each PCR reaction tube contained an individual colony lysate, 1x PCR Reaction Buffer, 1.5 mM $MgCl_2$, 0.25 mM dNTPs, 0.5 µM forward (SnaB1.CO.F) and reverse (SacII.CO.R) custom primers, and 0.5 U of Taq DNA Polymerase in a 20 µl final volume. Primer sequences are listed in Table 2.1. PCR amplification was performed as described in section 2.2.2a, with the exception that the elongation temperature was increased to 72 ^{o}C. Several plasmids containing PCR products of expected size, as determined by separation on 2% (w/v) agarose gel, were grown up from the gridded master plate and isolated according to the manufacturer's instructions using a QIAprep® Spin Miniprep Kit. These were sequenced and the sequence data analysed as described in section 2.2.1b.

2.2.2c Cloning of the codon optimised insert into pPIC9K.FL.WT.NAGLU

Ten (10) µg of pGEM-T(CO) plasmid DNA and 10 µg of pPIC9K.FL.WT.*NAGLU* were each digested with 20 U SnaBI and 40 U SacII in 1x NEBuffer 4 and incubated for 16 hr at 37 °C. The reactions were heat inactivated at 80 ^{o}C for 20 min. The excised synthetic insert DNA was separated from the pGEM-T vector by electrophoresis on 2% (w/v) agarose gel. The pPIC9K.FL.WT.*NAGLU* (minus the initial 261 bp) was separated from the excised 261 WT bp of *NAGLU* mature cDNA by electrophoresis on 0.8% (w/v) agarose gel. Correct sized bands were excised and purified as described in section 2.2.1a.

Ligation of the vector and insert was carried out as described in section 2.2.1b except 75 ng of the digested pPIC9K.FL.WT.*NAGLU* (minus the initial 261 bp) vector and 5.6 ng of SnaBI/SacII digested codon optimised 261 bp insert was used to maintain a 1:3 molar ratio of vector:insert. Transformation of electrocompetent *E. coli* and selection of positive colonies was performed as outlined in section 2.2.1b. Colony PCR amplification, analysis and master plating of putative positive clones was performed as described in section 2.2.2b. Several plasmids containing PCR products of expected size were grown up in small scale preps from the master plate, isolated, sent for DNA sequencing and analysed as described in section 2.2.1b. A FL.(CO+WT).*NAGLU* clone (FL, full length; (CO+WT).*NAGLU*, codon optimised 261 initial bp + wildtype 262-2163 bp of *NAGLU*) containing a mutation free ORF was prepared for yeast transformation as described in section 2.2.1b. A schematic representation of this final Construct 2: pPIC9K.FL.(CO+WT).*NAGLU* vector is shown in Figure 2.1.

2.2.3 Construct 3: pPIC9K.control.HFxCO.F

2.2.3a PCR amplification of the control HFx codon optimised fragment

The HFxCO.F (H, Hexahistidine affinity tag; Fx, Factor Xa cleavage site; CO.F, codon optimised fragment: initial codon optimised bases 1-261 of *NAGLU*) control construct was created in two PCR steps. First, bases 1-261 of codon optimised *NAGLU* cDNA were PCR amplified from 30 ng of pUC57 containing the 261 bp codon optimised fragment. Amplification was performed as described in section 2.2.2a except different forward (Fx.CO.F) and reverse (NotI CO.R) primers were used (Table 2.1). Amplification generated a fragment containing bases 1-261 of codon optimised *NAGLU* cDNA with a sequence encoding the Factor Xa cleavage site incorporated on the 5' end of the PCR product (Fx) and a NotI site on the 3' end of the product. Upon confirmation of successful first round PCR, as described in section 2.2.1a, second round amplification was performed. One (1) µl of undiluted PCR product, a 1/10 diluted product and a 1/100 diluted product were each amplified under the conditions described above, but with the replacement of the Fx.CO.F forward primer with EcoRI.HFx.F forward custom primer, thus incorporating an EcoRI site for cloning followed by a hexahistidine tag to act as an antibody recognition site at the 5' end of the construct (Table 2.1). Successful PCR amplification of all three dilutions of first round PCR products was confirmed on a 1.5% agarose gel and gel purified as described in section 2.2.1a.

2.2.3b Cloning of the control HFx codon optimised fragment into pGEM-T

Amplicons of EcoRI.HFx.F/NotI CO.R synthetic fragment were T/A ligated into pGEM-T, transformed into electrocompetent *E. coli* cells, the transformation reaction plated and positive clones identified initially by blue-white colour screening and then by colony PCR,

as described in section 2.2.2b. The EcoRI.HFx.F/NotI CO.R second round product amplification primers (Table 2.1) were used for colony screening. Several plasmids containing PCR products of expected size, as determined by separation on 2% (w/v) agarose gel, were grown up from the gridded master plate, isolated, sent for DNA sequencing and analysed as described in section 2.2.1b.

2.2.3c Cloning of the control HFx codon optimised fragment into pPIC9K

Ten (10) µg of pGEM-T(HFxCO) and 10 µg of pPIC9K were each digested with 20 U NotI, 40 U EcoRI and 100 µg/ml BSA in 1x EcoRI Buffer for 16 hr at 37 ^{o}C. The reactions were then heat inactivated at 65 ^{o}C for 20 min. The excised codon optimised control fragment DNA and the linearized pPIC9K were separated on 2% (w/v) agarose gel and on 0.8% (w/v) agarose gels, respectively, and correct sized bands were selected and purified as described in section 2.2.1a.

Ligation of the vector and insert was carried out as described in section 2.2.1b with the exception of the ng quantities of constructs used. Approximately 62 ng of the EcoRI/NotI digested pPIC9K vector was used with 6.16 ng of EcoRI/NotI digested codon optimised control fragment. Transformation was performed as in 2.2.1b and colony PCR amplification, analysis and master plating of putative positive clones was performed as described in section 2.2.2b, with the exception of the antibiotic kanamycin for selection and the primers used in the colony PCR reaction. Amplification in this case was performed with 0.5 µM forward (5' *AOX1*) and reverse (3' *AOX1*) primers supplied with the *P. pastoris* expression kit (Table 2.1). Several plasmids containing PCR products of expected size were grown up, isolated, sent for DNA sequencing and analysed as described in section 2.2.1b. A mutation free clone was prepared as

described in section 2.2.1b for transformation into yeast. A schematic representation of this final Construct 3: pPIC9K.HFxCO.F expression vector is shown in Figure 2.1.

2.2.4 Construct 4: pPIC9K.control.HFxWT.F

2.2.4a PCR amplification of the control HFx wildtype fragment

The HFxWT.F (H, Hexahistidine affinity tag; Fx, Factor Xa cleavage site; WT.F, wildtype fragment: initial WT bases 1-261 of *NAGLU*) control construct was created in two PCR steps. First, bases 1-261 of WT *NAGLU* cDNA were PCR amplified from 30 ng of pPIC9K.FL.WT.*NAGLU* as described in 2.2.2a with the exception of the (Fx.WT.F) forward and the (AvrII.WT.R) reverse primers used and the addition of 1.3 M betaine and 8% dimethyl sulphoxide (DMSO) to the PCR mixture. Primer sequences are listed in Table 2.1. Amplification generated a fragment containing bases 1-261 of WT *NAGLU* cDNA with a sequence encoding the Factor Xa cleavage site (Fx) incorporated on the 5' end of the PCR product. Successful first round PCR amplification was confirmed by electrophoresis on a 1.5% agarose gel. In the second round PCR, 1 µl of undiluted first round PCR product, a 1/10 diluted product and a 1/100 diluted product were amplified under the conditions described above, but with the replacement of the Fx.WT.F forward primer with EcoRI.HFx.F forward primer (Table 2.1), thus incorporating an EcoRI site for cloning and a hexahistidine tag for construct purification. Successful PCR amplification of all three dilutions of first round PCR products was again confirmed by electrophoresis on a 1.5% agarose gel and visualized as described in section 2.2.1a. The remaining product was QIAquick® purified according to manufacturer's instructions.

2.2.4b Cloning of the control HFx wildtype fragment into pGEM-T

Amplicons of EcoRI.HFx.F/AvrII.WT.R wildtype fragment were T/A ligated into pGEM-T, transformed into electrocompetent *E. coli* cells, the transformation reaction plated and positive clones identified initially by blue-white colour screening and then by colony PCR, as described in section 2.2.2b, but with the following exceptions. Primers used for colony screening were the EcoRI.HFx.F/AvrII.WT.R primers used for second round amplification of the WT fragment (Table 2.1), and 1.3 M betaine and 8% DMSO was added to the PCR reaction mix. Several plasmids containing PCR products of expected size, as determined by separation on 2% (w/v) agarose gel, were grown up, isolated, sent for DNA sequencing and analysed as described in section 2.2.1b.

2.2.4c Cloning of the control HFx wildtype fragment into pPIC9K

Ten (10) µg of pGEM-T plasmid DNA containing the mutation free control HFxWT fragment were digested with, 40 U EcoRI, 20 U AvrII and 100 µg/ml BSA in 1x EcoRI Buffer for 16 hr at 37 ^{o}C. The excised fragment was separated from the pGEM-T vector by electrophoresis on a 2% (w/v) agarose gel, visualized and purified as described in section 2.2.1a.

Ligation of approximately 62 ng of the previously EcoRI/AvrII digested pPIC9K vector and 5.7 ng of the EcoRI/AvrII digested control HFxWT fragment was carried out as described in section 2.2.1b. Transformation, selection, master plating of putative positive colonies and colony PCR amplification was as described in section 2.2.3c with the addition of 1.3 M Betaine and 8% DMSO to the PCR reaction mix. Several plasmids containing PCR products of expected size were grown up, isolated, sent for DNA sequencing and analysed as described in

2.2.1b. A clone containing a mutation free ORF was prepared for yeast transformation as described in section 2.2.1b. A schematic representation of this final Construct 4: pPIC9K.HFx. WT expression vector is shown in Figure 2.1.

2.3 Transformation of *P. pastoris* with Yeast Vectors

2.3.1 Transformation of pGlycoSwitchM5 with expression vector Constructs 1-4

All four expression vector constructs and pPIC9K (vector only negative control) were linearized for yeast transformation using 10 µg of expression vector, 20 units of SacI, 1x NEBuffer 1 and 1x BSA at 37 ^{o}C for 16-18 hr. Linearization was confirmed by electrophoresis of a 1 µl aliquot of each digest on a 0.8% (w/v) agarose gel, visualized as described in section 2.2.1a. Humanized *P. pastoris* strain pGlycoSwitchM5, was provided as a gift by Dr. Contreras (Ghent University, Belgium). pGlycoSwitchM5 competent cells were prepared for transformation as described in the *Pichia* Expression Kit: A Manual of Methods for Expression of Recombinant Proteins in *P. pastoris,* with the following modifications. A 10 ml YPD culture was inoculated with an isolated pGlycoSwitchM5 yeast colony and grown overnight at 28 °C, shaking at 275 rpm. Fresh YPD (500 ml) was then inoculated with sufficient overnight culture (~250-500 µl) to give an optical density (OD) reading of 0.01 at 600 nm wavelength and grown, shaking at 275 rpm, at 28 °C until an OD_{600} reading of between 1.3-1.5 was attained (~17 hr). Cells were then processed as outlined in the protocol cited above. Competent yeast cells (80 µl) were mixed with 5 µg of linearized vector, transferred to ice cold 0.2 cm gap electroporation cuvettes and pulsed using a Genepulser electroporation machine with settings of 1.5 kV and 400 Ω resistance. One (1) ml of ice cold 1 M sorbitol was

immediately added to the electroporated cells, followed by 1 hr incubation at 30 ^{o}C, with subsequent plating on MD and MDH plates and incubation at 30 ^{o}C for 2-3 days.

2.3.2 Screening transformed pGlycoSwitchM5 for integration of constructs

Colonies able to grow on MD plates without histidine supplement were patched using a sterile toothpick onto a MD master plate and incubated for 2 days at 30 ^{o}C. Colonies were then screened for genomic integration of the various expression vectors (Constructs 1-4) by direct yeast PCR. In short, using a sterile toothpick, a small portion of a yeast colony was resuspended in 10 µl of sterile water. To lyse the cell walls, 4 U of zymolyase were added to the cells and incubated for 30 min at 37 ^{o}C, followed by 10 min incubation at 95 ^{o}C. The reaction mixture was then centrifuged at RT for 15 min at 16,000 x g and 1 µl of the supernatant was removed for use in PCR amplification. The PCR reaction mix and amplification was as described in section 2.2.2b with the exception of the 5' *AOX1* and 3' *AOX1* primers used (Table 2.1), the addition of 1.3 M betaine and 8% DMSO to the reaction mix, and an increase to 3 min for the elongation step. Successful PCR amplification was confirmed by electrophoresis on a 1.0% agarose gel as described in section 2.2.1a. As attaining equally efficient PCR amplification of both the endogenous *AOX1* gene and the introduced *AOX1* containing constructs proved difficult, new sets of primers specific for the amplification of the endogenous *AOX1* gene and each of the four integrated constructs were designed. Primer sequences are listed in Table 2.2. PCR buffer conditions for all primer pairs were as outlined above. Amplification was performed as outlined above with the exception of the cycled elongation time which was modified from 3 min to 1.5 min. Successful PCR amplification was confirmed by electrophoresis on a 1.5% agarose gel as described in section 2.2.1a.

Table 2.2 Oligonucleotide primers used to confirm integration of Constructs 1-4 and transcription of Constructs 1 and 2 in *P. pastoris.*

Primer Name[h]	Nucleotide Sequence 5' to 3'	Amplicon size[a] (bp)
WT/(CO+WT).1F[b]	TACGAGCAGAACAGCCGCTAC	440
WT.F.1F	CTACCTGCGCGACTTCTG	180
CO.F.1F	GCCAGGTTTGGATACTTACTC	278
5' *AOX1* (F)	GACTGGTTCCAATTGACAAGC	2200[c], 492[d], 2672[e], 782[f], 788[g]
AOX1 (F internal)	TGTCTGCTGGTCTTGCTCAC	493
3' *AOX1* (R)	GCAAATGGCATTCTGACATCC	

[a] Amplicon sizes result from the listed primer used in conjunction with the 3' *AOX1* primer
[b] The WT/(CO+WT).1F primer was used to amplify both integrated Construct 1 and 2 as it was designed to their common 3' portion.
[c] Amplicon size resulting from amplification of the endogenous *AOX1* gene.
[d] Amplicon size resulting from amplification of the integrated vector alone.
[e] Amplicon size resulting from amplification of the integrated Construct 1.
[f] Amplicon size resulting from amplification of the integrated Construct 3.
[g] Amplicon size resulting from amplification of the integrated Construct 4.
[h] Primer orientation is indicated in the primer name as forward (F) or reverse (R).

2.3.3 Replica plating pGlycoSwitchM5 transformants

To distinguish between Mut^+ and Mut^S transformants, master plated colonies were replica plated (using velvet squares and a replica plating block) in duplicate onto minimal methanol MM and fresh MD plates. After 2 days at 30 ^oC, the replica plates were scored. Patches that grew normally on the MD replica plate, but showed no growth on the MM replica plate were designated as Mut^S.

To detect multiple copy integrants, master plated colonies were replica plated onto MD plates containing increasing concentrations of geneticin in 0.25 mg/ml increments, from 0 mg/ml up to 1mg/ml geneticin. After 2-3 days at 30 ^oC, the replica plates were scored.

2.4 Growth and *NAGLU* Expression in Transformed *P. pastoris*

2.4.1 Protein expression in pGlycoSwitchM5 (Mut^+) strains

Growth and protein expression was assessed by performing small scale expression studies using positive clones and vector only (yeast cells transformed with pPIC9K vector not containing an insert) negative controls. Twenty (20) ml of buffered glycerol complex medium (BMGY) were inoculated with a match head size scrape of a patched single yeast colony and cultured at 28 ^oC, shaking at 275 rpm, until an OD_{600} of 2-6 (approximately 16 hr) was reached. The cells were harvested by centrifugation at 2,500 x g for 5 min at RT. The cell pellet was resuspended to an $OD_{600} = 1$ in buffered methanol complex medium (BMMY) and cultured at either 15 ^oC or 28 ^oC, shaking at 275 rpm, for 96 or 120 hr. The cultures were supplemented with 1.0% (v/v) methanol every 24 hr, and sampled for analysis of growth, protein production, and activity as described in sections 2.4.3-2.4.5. Immediately after obtaining the samples and removing an aliquot for growth analysis, the remaining sample was centrifuged at 16,000 x g

for 5 min at RT. The supernatant was transferred to a new sterile tube on ice and the two fractions (supernatant and pellet) were frozen in liquid nitrogen and then stored at -80 ^{o}C until analysis.

2.4.2 Protein expression in pGlycoSwitchM5 (Mut^{s}) strains

Using positive clones and the vector only controls described in section 2.4.1, 250 ml of BMGY were inoculated with a match head size scrape of a patched single yeast colony and cultured at 28 ^{o}C, shaking at 275 rpm, until an OD_{600} of 2-6 (approximately 16 hr) was reached. The cells were harvested as above and resuspended in 50 ml of BMMY and cultured, supplemented and sampled as above.

2.4.3 Sampling of culture supernatants and cell lysates

To ensure all cultures exhibited a healthy pattern of growth, samples were taken directly from small scale expression cultures of Mut^{+} and Mut^{S} pGlycoSwitchM5 strains at 0, 24, 48, 72, 96 and 120 hr post induction. Growth was assessed by OD measurements at 600 nm using a Novaspec® visible spectrophotometer. Measurements were taken in triplicate at each time point and averaged. An OD_{600} of 1 is designated as being equal to 1-2 x 10^{7} yeast cells/ml. Thus, cell concentration could be determined at all time points.

Both crude and concentrated culture supernatant samples were analysed for protein expression by activity assay and by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) followed by either silver stain or immunoblot analysis. Total protein concentration was determined using the Bradford Protein Assay and a BSA standard curve (0-2.0 mg/ml). Crude culture supernatants were prepared by centrifugal removal of yeast cells and stored at -80 ^{o}C as described in section 2.4.1. Concentrated culture supernatants were

prepared by the addition of ammonium sulphate at 50% (w/v) saturation to the culture medium. The precipitate was collected by centrifugation at 16,000 x g for 15 min, and then redissolved in a smaller volume to generate a 40x concentrated sample.

Cell pellets resulting from the centrifugation of 1 ml of crude culture medium, as described in section 2.4.1, were lysed in yeast breaking buffer (pH 7.4), releasing intracellular proteins. In an attempt to normalize for cell number, pellets from time points 0 and 24 hr (approximately 1-7 x 10^7 cells/ml) were resuspended in 100 µl yeast breaking buffer, whereas pellets from time points 48-120 hr (approximately 14-20 x 10^7 cells/ml) were resuspended in 500 µl yeast breaking buffer. An equal volume of 0.5 mm acid washed glass beads was added and the samples were sequentially vortexed for 30 seconds (sec) and incubated on ice for 30 sec. This process was repeated for a total of 8 cycles. The cell homogenate was then centrifuged at 4 °C at 16,000 x g for 10 min, and the supernatant containing the intracellular lysate fraction was retained for analysis by silver stain and immunoblot.

2.4.4 SDS-PAGE silver stain and immunoblot analysis

SDS-PAGE was performed using a combined gel consisting of a 10% (v/v) tris-glycine resolving gel and 4% (v/v) tris-glycine stacking gel. Samples were prepared with 3x SDS Sample buffer augmented with 42 mM DTT, boiled for 5 min and centrifuged at 16,000 x g for 5 min prior to loading. A NAGLU positive control, harvested and concentrated from stably transfected *Sf*9 cells known to produce low level expression of active recombinant CBD-NTAT, was supplied as a gift by J. Bandsmer (University of Victoria, BC). A His tagged H-protein positive control produced in *P. pastoris* cells was supplied as a gift by A. Zay (University of Victoria, BC). A Broad Range Protein standard or Precision Plus dual colour marker was used to estimate the molecular mass of proteins in the experimental samples. Gels

were run on an SE 260 mini-vertical gel electrophoresis unit at 25 mA per gel for approximately 1 hr.

For immunoblot analysis following SDS-PAGE, the resulting protein bands were electroblotted onto Hybond-P PVDF membrane overnight at 10 V in 1x Transfer buffer using a Mini-protean II Electroblot apparatus. The following steps were performed at RT with gentle agitation. PVDF membranes were blocked for 1 hr with 7.5% (w/v) skim milk powder in 1x TBS-T. To detect the NAGLU protein, a 1:7,500 dilution of primary (anti-NAGLU) antibody was added to the blocking solution and incubated for 1 hr. This rabbit polyclonal antibody was provided by Drs. Neufeld and Zhao (University of California, CA). To detect the hexahistidine tag, a 1:5,000 dilution of anti-His_4 antibody was added to the blocking solution and incubated as above. Membranes were briefly rinsed in TBS-T and then washed 3 times for 10 min (each time in fresh TBS-T), followed by a 1 hr incubation with a 1:25,000 dilution in blocking solution of goat anti-rabbit horseradish peroxidase conjugated secondary antibody or a 1:10,000 dilution of goat anti-mouse horseradish peroxidase conjugated antibody for detection of anti-NAGLU and anti-His_4, respectively. Membranes were again briefly rinsed and then washed 3 times for 10 min each in fresh TBS-T. The blots were then incubated with SuperSignal® West Dura chemiluminescent reagent according to the manufacturer's instructions, and visualized with Kodak BioMax film.

For silver stain analysis following electrophoresis, gels were placed in fixative and microwaved at maximum power for 90 sec. After removal of the fixative, gels were washed with 50% (v/v) ethanol by microwaving for 90 sec and then microwaved for a further 90 sec in a pretreatment solution. Gels were rinsed in ddH_2O for 90 sec at RT, and stained in 2 mg/ml silver nitrate in 0.075% (v/v) formaldehyde by microwaving for 40 sec, shaking at RT for 20 sec and returned to the microwave for a further 40 sec. Following another 90 second rinse in

ddH_2O, bands were resolved in developer, clarified by a 5 min incubation in 5% (v/v) acetic acid and the reaction stopped by the addition of a stop solution.

2.4.5 NAGLU activity assay

Samples from cell lysates and crude and concentrated samples from culture supernatants were assayed as described by Chow and Weissmann (1981) and Zhao and Neufeld (2000). In brief, samples were incubated for 1 hr at 37 °C with an equal volume (25 µl) of 0.2 mM 4-methylumbelliferyl (4-MU) α-*N*-acetylglucosaminide buffered in 0.1 M Na acetate buffer (pH 4) supplemented with 0.5 mg/ml BSA. The addition of 0.5 M Glycine-NaOH buffer (pH 10.5) to a final reaction volume of 1.5 ml was performed to stop the reaction. Activity of the NAGLU enzyme in the cell lysates and crude and concentrated supernatant samples was assessed in terms of level of 4-MU released in the above reaction and was measured in arbitrary fluorescent units (AFU). An arbitrary unit of NAGLU activity corresponds to the release of one nmol of 4-methylumbelliferone per hr. AFU levels were read using a Sequoia-Turner Model 450 Digital fluorometer with a 350 nm narrow band excitation filter and a 415 nm sharp cut emission detection filter. Readings were analysed relative to a vector only control pGlycoSwitchM5 strain which had been cultured concurrently with experimental samples.

2.4.6 RT-PCR analysis

Total RNA was extracted from cell pellets (harvested at T = 24 hr post induction and then stored at -80 °C) containing approximately 2.5 x 10^7 cells using an RNeasy® Mini Kit. RNA extraction was performed according to manufacturer's instructions, but with the following modification. On-column DNase digestion was omitted and replaced with an off-column DNase digestion, performed after the RNA extraction as follows: 1 µg RNA and 1ul

RNase free DNase were incubated in 1x PCR Reaction Buffer with 5 mM $MgCl_2$ at RT for 20 min. The reaction was stopped by heating the digest to 65 °C for 5 min, followed by addition of EDTA to a final concentration of 14 mM, and a subsequent 65 °C denaturation step for a further 5 min. The processed RNA was stored at -80 °C until use. First-strand cDNA synthesis was performed using an oligo dT primer and SuperScript™ II RNase H^- Reverse Transcriptase according to the manufacturer's protocol and the cDNA stored at -20 °C until use. To assess RNA expression of the four transformed constructs, each cDNA was used as a template for PCR. For all constructs, the PCR reactions contained 2 µl cDNA, 1x PCR Reaction Buffer, 1.5 mM $MgCl_2$, 0.25 µM dNTPs, 0.5 µM of each primer (Table 2.2), 1.3 M betaine, 8% DMSO and 1 U *Taq* DNA Polymerase in a final 50 µl volume. Amplification was performed using a GeneAmp PCR thermal cycler with an initial denaturation of 94 ^oC (5 min), followed by 35 cycles of 94 ^oC (1 min), 56 ^oC (1 min) and 72 ^oC (1 min), with a final elongation at 72 ^oC (10 min). Successful PCR amplification was confirmed by electrophoresis of an aliquot of the product on a 2.0% agarose gel as described in section 2.2.1a.

2.5 Construction of Insect Expression Vectors

2.5.1 Construction of expression vectors for stable expression

2.5.1a Amplification of the THF fragment

Creation of the 126 bp THF (T, human transferrin secretion signal; H, hexahistidine affinity tag; F, Factor Xa cleavage site) fragment was achieved by PCR amplification of a p2ZOpTCXFegfp construct (Vaags, 2004) with primers listed in Table 2.3 under conditions described in section 2.2.2a. This amplification generated a 126 bp THF fragment flanked by a

Table 2.3 Oligonucleotide primers used to create and screen for the THF fragment.

Primer Name[d]	Nucleotide Sequence 5' to 3'
HindIII.T.F[a]	ACCC **AAG CTT** ATG AGG CTC GCC GTG GGA GCC
EcoRI.H.Fx.T.R[a,b,c]	TC**GAA TTC** *AAC ACG ACC TTC AAT* ATG ATG ATG ATG ATG ATG TCC CGT CTT ATC AGG GAC
M13F	CGCCAGGGTTTTCCCAGTCACGAC
M13R	TCACACAGGAAACAGCTATGAC

[a] Restriction enzyme recognition sequences are presented in bold print.
[b] The Factor Xa cleavage site is presented in italic print.
[c] The hexahistidine tag is underlined.
[d] Primer orientation is indicated in the primer name as forward (F) or reverse (R).

5' HindIII site and a 3' EcoRI site. Successful PCR amplification was confirmed by electrophoresis of an aliquot of the product on a 1.5% (w/v) agarose gel and the remaining product was QIAquick® purified.

2.5.1b Cloning of the THF fragment into pGEM-T

The THF fragment was T/A ligated into pGEM-T vector according to manufacturer's instructions. Transformation was performed as outlined in section 2.2.1b, and positive colonies selected and screened by colony PCR as described in section 2.2.2b. The M13F and R primers used for colony screening are listed in Table 2.3. Several plasmids containing PCR products of expected size, as determined by separation on 2% (w/v) agarose gel, were prepared using a QIAprep® Spin Miniprep kit and sent for DNA sequencing and analysis as described in section 2.2.1b.

2.5.1c Creation of the p2ZOp2.THF.NAGLU and p2ZOp2.THF.NTAT clones

p2ZOp2*NAGLU* and p2ZOp2*NTAT* were obtained by HindIII/EcoRI digestion of sequence verified p2ZOp2tcXF.*NAGLU* and p2ZOp2tcXF.*NTAT* constructs (Bandsmer, 2004). This allows removal of the cellulose binding domain (CBD) tag and surrounding sequence, but leaves intact the p2ZOp2 insect vector sequence plus *NAGLU* and *NTAT* sequences, respectively. Ten (10) µg of each p2ZOp2tcXF.*NAGLU* and p2ZOp2tcXF.*NTAT* were concomitantly digested with 40 U HindIII, 40 U EcoRI in 1x EcoRI Buffer for 16 hr at 37 ^{o}C. Twenty five (25) µg of pGEM-T(THF) plasmid DNA were concomitantly digested with 80 U HindIII, 80 U EcoRI as above. Heat denaturation of enzymes was achieved by incubation at 65 ^{o}C for 20 min. The excised 114 bp THF fragment and the p2ZOp2*NAGLU* and p2ZOp2*NTAT* vectors were isolated by electrophoresis on a 2.5% (w/v) low melting point agarose gel and a

0.8% (w/v) agarose gel, respectively. The correct sized bands were excised and purified using the QIAquick® Gel Extraction kit, following the manufacturer's protocol.

Directional ligation of the vector and insert was carried out as described in section 2.2.1b with the exception of the ng quantities of constructs used. Approximately 40 ng of the digested THF fragment was ligated into 50 ng of each digested p2ZOp*NAGLU* and p2ZOp*NTAT*. Transformation was performed as outlined in section 2.2.1b, and positive colonies were selected by growth on LB plates containing 25 μg/ml Zeocin™. Master plating, colony PCR amplification and analysis of putative positive clones for each construct was performed as described in section 2.2.2b, with the HindIII.T.F and EcoRI.H.Fx.T.R primers listed in Table 2.3. Several plasmids containing PCR products of expected size for each construct were isolated and sent for DNA sequence analysis as described in section 2.2.1b. A clone containing a mutation free ORF was selected for each p2ZOp2.THF.*NAGLU* and p2ZOp2.THF.*NTAT* constructs. A schematic representation of these final constructs is shown in Figure 2.3. A bacterial stab was prepared for each clone. Briefly, 0.6 g of agar was added to 100 ml of LB, autoclaved to sterilize and allowed to cool to 55 °C. A 1.5 ml aliquot was pipetted into each of two sterile microfuge tubes and allowed to cool for 10 min. The agar was inoculated with a sequence verified colony from the appropriate master plate for each construct and grown overnight at 37 °C. The stabs were sent to our collaborator, Tom Pfeifer (University of British Columbia, BC), for plasmid isolation and subsequent transformation into *Sf9* cells.

2.5.2 Construction of expression vectors for baculovirus expression

2.5.2a pAcGP67B.NAGLU.H_6 and pAcGP67B.NTAT.H_6

Full length mature *NAGLU* cDNA with and without a fused TAT-PTD were obtained

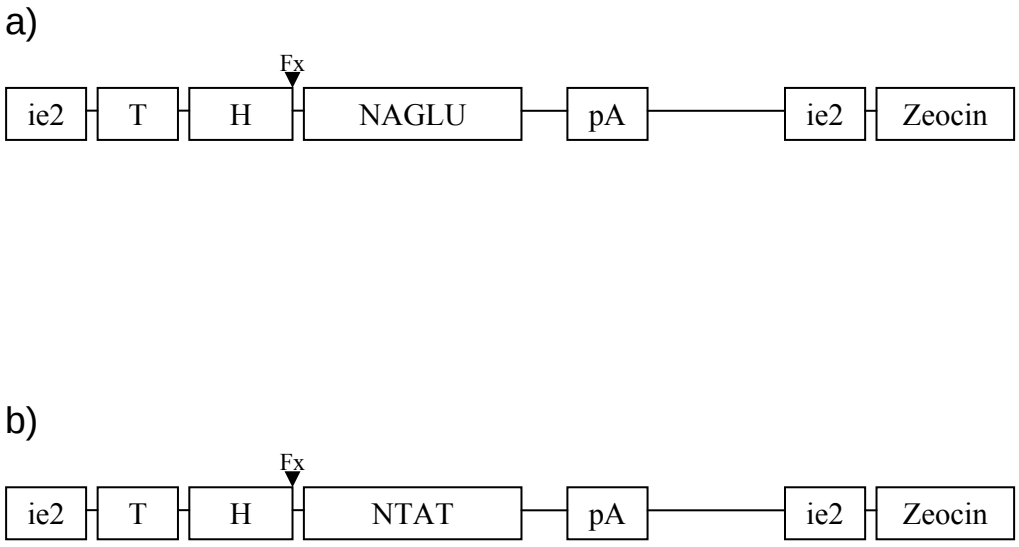


Figure 2.3 Schematic representation of p2ZOp2.THF.*NAGLU* and p2ZOp2.THF.*NTAT* constructs designed for stable expression in *Sf9* cells. a) p2ZOp2.THF.*NAGLU* contains *NAGLU* **b)** p2ZOp2.THF.*NTAT* contains *NTAT* (*NAGLU* with 3' in frame TAT-PTD). The p2ZOp2THF insect expression vector contains an immediate early 2 promoter (ie2) for stable expression of integrated recombinant genes, a human transferrin secretion signal (T) for secreted recombinant protein expression, a hexahistidine affinity tag (H) for purification, a Factor Xa (Ile-Asp-Gly-Arg) cleavage site (Fx), a poly A signal sequence from the ie2 gene (pA) and a gene producing the Zeocin resistance protein for antibiotic selection (Zeocin).

by respective PCR amplification from 30 ng of sequence verified pPICZα(*NAGLU*) and pPICZα(*NTAT*) constructs (Patrick, 2006). Amplification was performed as described in section 2.2.2a with the exception of the primers used. BamHI.*NAGLU*/*NTAT*.F was used in conjunction with *NAGLU*.H_6.EcoRI.R for amplification of *NAGLU*.H_6 and with *NTAT*.H_6.EcoR I.R for amplification of *NTAT*.H_6. Primer sequences are listed in Table 2.4. PCR cycling conditions were as described in section 2.2.2a with the exception that the extension time was increased to 3 min and 35 amplification cycles were performed. Successful amplification of each construct was confirmed by electrophoresis and visualization of an aliquot of the products as described in section 2.2.1a. The remaining products were QIAquick® purified according to manufacturer's instructions.

Approximately 3 μg of each *NAGLU*.H_6 and *NTAT*.H_6 were digested with 15 U of both BamHI and EcoRI in EcoRI Buffer supplemented with 100 μg/ml BSA for 16 hr at 37 °C and the sample QIAquick® purified according to manufacturer's instructions. Ten (10) μg of the baculovirus expression vector pAcGP67B was also concomitantly digested with 40 U of BamHI and EcoRI and purified as above. An aliquot of the digest was electrophoresed on a 1% (w/v) agarose to confirm linearization of the vector. The remaining products were QIAquick® purified according to manufacturer's instructions.

Directional ligation of approximately 44 ng of each BamHI/EcoRI digested *NAGLU* and *NTAT* inserts into approximately 65 ng BamHI/EcoRI digested pAcGP67B vector and transformation onto plates containing 100 μg/ml ampicillin was performed as described in section 2.2.1b. Positive clones were screened for the presence of correct inserts by colony PCR as described in section 2.2.2b with the exception of the pAcGP67B.F and pAcGP67B.R primers used (Table 2.4). For each construct, several plasmids containing PCR products of expected size were grown up, isolated and analysed, and clones containing mutation free ORFs

Table 2.4 Oligonucleotide primers used to amplify *NAGLU* and *NTAT* constructs for expression in the baculovirus vector pAcGP67B.

Primer Name[e]	Nucleotide Sequence 5' to 3'
BamHI.*NAGLU*/*NTAT*.F[a]	TAAT **GGA TCC** GAC GAG GCC CGG GAG G
NAGLU.H_6. EcoRI.R[a,b,c]	ATTA **GAA TTC** CTA ATG GTG ATG GTG ATG GTG CCA AGA GCC GGC CAC CCA
NTAT.H_6. EcoRI.R[a,b,c,]	ATTA **GAA TTC** CTA ATG GTG ATG GTG ATG GTG TCG TCG ACG CTG ACG TCG
BamHI.H_6.L.F[a,b]	TAAT **GGA TCC** CAC CAT CAC CAT CAC CAC AGC AGC GGC CTG GTG AG
L(*NAGLU/NTAT*).R[d]	CTC CCG GGC CTC GTC *CCT ACC CTC GAT* GCT AC
(L)*NAGLU/NTAT*.F[d]	GT AGC *ATC GAG GGT AGG* GAC GAG GCC CGG GAG
NTAT.EcoRI.R[c]	ATTA GAA TTC TCA TCG TCG ACG CTG ACG TCG
pAcGP67B.F	CAG TCA CAC CAA GGC TTC AA
pAcGP67B.R	TCATCGTGTCGGGTTTAACA

[a] Restriction enzyme recognition sequences are presented in bold print.
[b] The hexahistidine tag is underlined.
[c] Stop codons are boxed.
[d] Factor Xa cleavage sites are italicized.
[e] Primer orientation is indicated in the primer name as forward (F) or reverse (R).

were prepared as described in section 2.2.1b. A schematic representation of both pAcGP67B.*NAGLU*.H_6 and pAcGP67B.*NTAT*.H_6 constructs is shown in Figure 2.4.

2.5.2b pAcGP67B. H_6.L.Fx.NTAT

Prior to creation of the H_6.L.Fx.*NTAT* construct, the template for the 5' linker and Factor Xa fragment needed to be created. This L/Fx sequence was adapted from a linker sequence contained within the Novagen insect expression vector pIEx-10. Forward and reverse 101 base oligonucleotides corresponding to this adapted sequence (Figure 2.5) were resuspended to 10 µM in ddH_2O. The oligos were then annealed in 1x annealing buffer to give a final concentration of approximately 66 ng/µl of double stranded DNA.

The H_6.L.Fx.*NTAT* (H, hexahistidine tag; L, linker; Fx, Factor Xa cleavage site; *NTAT, NAGLU* fused with TAT-PTD) fusion gene was created in three PCR steps. First, H_6.L.Fx was PCR amplified from the annealed linker template under the conditions described in section 2.2.2a, except using primers BamHI.H_6.L.F and L(*NAGLU/NTAT*).R (Table 2.4). A sequence encoding the hexahistidine tag was thus incorporated on the 5' end of the PCR product and a sequence homologous to the first 15 bp of NTAT incorporated on the 3' end (H_6.L.Fx.). Second, *NTAT* cDNA was PCR amplified from 30 ng of sequence verified pPICZα(*NTAT*) (Patrick, 2006) using (L)*NAGLU*/*NTAT*.F and *NTAT*.EcoRI.R primers (Table 2.4) under the conditions described in section 2.2.2a. A sequence encoding the final 17 bp of the linker sequence was incorporated on the 5' end of the PCR product. PCR cycling conditions were as described in section 2.5.2a. Successful PCR amplification of both H_6.L.Fx and *NTAT* was confirmed by electrophoresis of an aliquot of the product on a 1.3% (w/v) agarose gel and visualized as described in section 2.2.1a. Overlap-extension PCR was then performed on 1 µl of undiluted H_6.L.Fx and *NTAT* PCR products under the same amplification conditions as

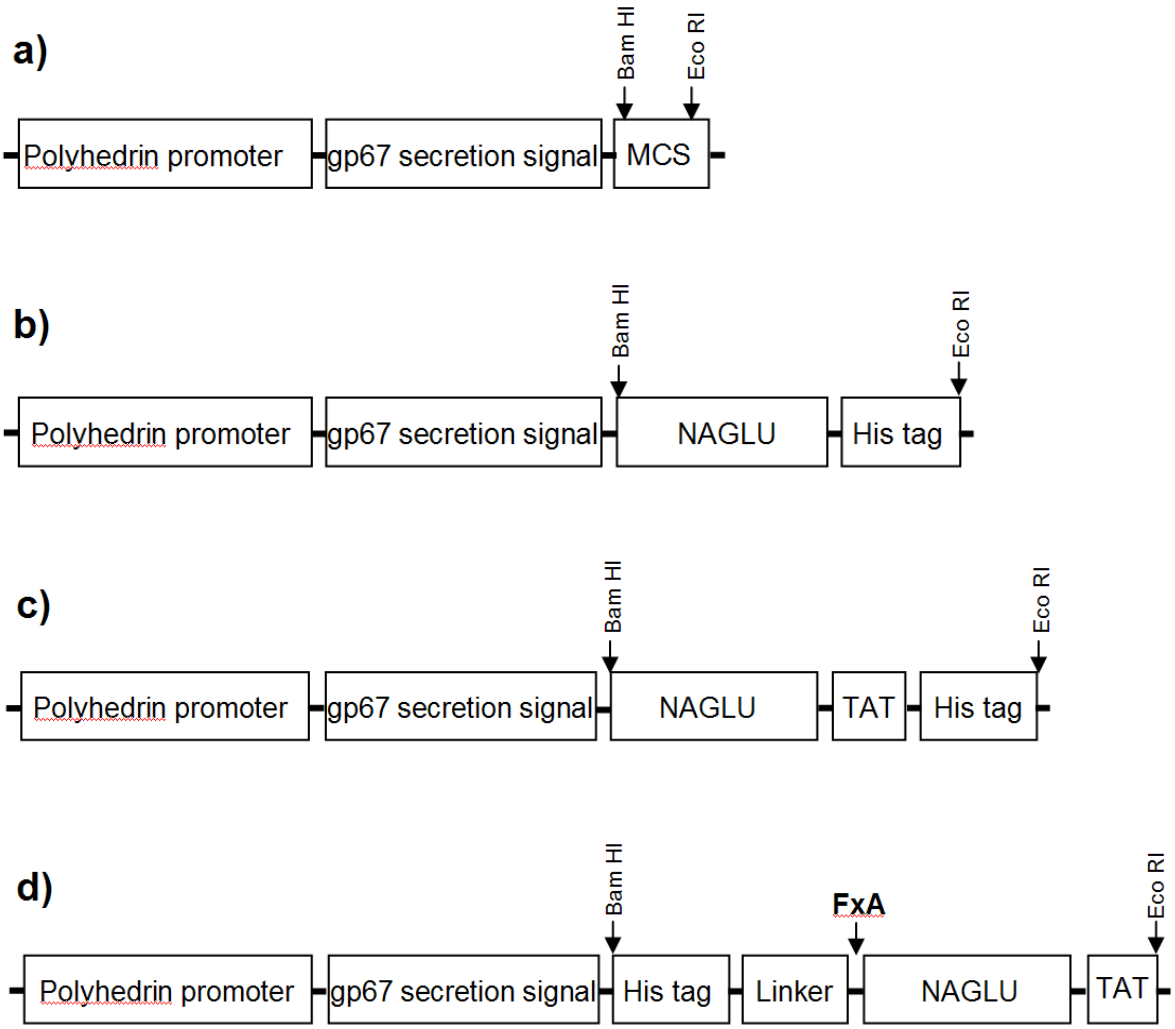


Figure 2.4 Schematic representation of the baculovirus expression vector and the *NAGLU* and *NTAT* constructs designed for secreted expression in insect cells. a) pAcGP67B empty expression vector. The baculovirus expression vector contains a strong very late polyhedrin promoter for high level recombinant gene expression, a gp67 secretion signal for secreted recombinant protein expression and a multiple cloning site for cloning of recombinant genes. **b)** pAcGP67B.*NAGLU*.H_6 contains *NAGLU* with a carboxyl fused hexahistidine tag for purification. **c)** pAcGP67B.*NTAT*.H_6 contains *NTAT* (*NAGLU* with 3' in frame TAT-PTD) with a carboxyl fused hexahistidine tag. **d)** pAcGP67B.H_6.L.Fx.*NTAT* contains an amino hexahistidine tag which is separated from NTAT by a 22 αα linker sequence and a Factor Xa (Ile-Asp-Gly-Arg) cleavage site.

a)

```
Forward    1   GAAGT GAA TTC AGC AGC GGC CTG GTG AGC AGC GGC AGC CAT ATG  44
Reverse    1  ACTTCA CTT AAG TCG TCG CCG GAC CAC TCG TCG CCG TCG GTA TAC  45

Forward   45 GCT AGC ATG ACT GGT GGA CAG CAA ATG GGT AGC ATC GAG GGT AGG  89
Reverse   46 CGA TCG TAC TGA CCA CCT GTC GTT TAC CCA TCG TAG CTC CCA TCC  90

Forward   90 GAA TTC CGAAGA  101
Reverse   91 CTT AAG GCTTC   101
```

b)

```
DNA          GAAGTGAATTC AGC AGC GGC CTG GTG AGC AGC GGC AGC CAT ATG GCT AGC
Amino acid               Ser Ser Gly Leu Val Ser Ser Gly Ser His Met Ala Ser

DNA          ATG ACT GGT GGA CAG CAA ATG GGT AGC ATC GAG GGT AGG GAA TTC CGAA
Amino acid   Met Thr Gly Gly Gln Gln Met Gly Ser Ile Glu Gly Arg
```

Figure 2.5 Linker Oligonucleotides: DNA and corresponding amino acid sequences. a) Double stranded annealed linker oligonucleotides. **b)** Protein sequence of the translated linker region.

described for *NTAT*, except primers BamHI.H_6.L.F and *NTAT*.EcoRI.R were used (Table 2.4). PCR cycling was performed as described above for *NTAT* with the exception that primers were added during the sixth cycle and a total of 40 cycles were performed. Amplification of the overlap product was confirmed by separation of products on a 1% (w/v) agarose gel. The approximately 2.3 kb insert band was excised without exposure to UV and purified using a QIAquick® Gel Extraction Kit according to the manufacturer's protocol.

Ligation into pGEM-T and transformation and small scale plasmid preparations were performed as described in section 2.2.2b and 2.2.1b, respectively. Successful ligation was verified by concomitant digestion of approximately 1 µg of putative pGEM-T(H_6.L.Fx.*NTAT)* with 5 U SphI, 10 U SacI, in 1x NEBuffer 1 and 1x BSA for 16 hr at 37 ^{o}C. Enzymes were inactivated at 65 ^{o}C for 20 min. Plasmid preparations giving the correct digestion pattern upon electrophoresis on 1.0% (w/v) agarose gel were sent for DNA sequence analysis and the sequence data analysed as described in section 2.2.1b.

Approximately 10 µg of sequence verified pGEM-T(H_6.L.Fx.*NTAT*) was digested as described in section 2.5.2a, confirmed as correct by separation of products on a 1.3% (w/v) agarose gel, and the approximately 2.3 kb insert band excised and gel purified as previously described. Ligation of approximately 65 ng of BamHI/EcoRI digested pAcGP67B vector and approximately 48 ng of BamHI/EcoRI digested H_6.L.Fx.*NT* insert was carried out as described in section 2.2.1b. Transformation and screening by colony PCR was as described in section 2.5.2a. Plasmids generating a PCR product of the expected size using pAcGP67B.F and pAcGP67B.R primers (Table 2.4), were grown up from the gridded master plate, isolated and analysed as described in section 2.2.1b. A pAcGP67B.H_6.L.Fx.*NTAT* clone containing a mutation free ORF was prepared as described in section 2.2.1b to obtain sufficiently high

quantity and quality of plasmid DNA for *Sf9* cell transfection. A schematic representation of this construct is shown in Figure 2.4.

2.6 Transformation of Insect Vectors into *Sf9* and *Tni* cells

2.6.1 Transformation of insect vectors for stable expression *(performed by Dr. T. Pfeifer)*

Insect cells were grown at 27 °C in Expressions Systems Formula Animal free (ESF AF) medium. For transfections, 1-2 x 10^6 *Sf9* or *Tni* cells were seeded into a 6 well tissue culture plate in 3 ml Grace's Insect Cell Culture Medium and allowed to adhere for 1 hr prior to transfection. Standard quantities of 1 μg of DNA (0.9 μg mutation free construct DNA + 0.1 μg of the p2ZOp2F(SEAP) alkaline phosphatase vector (Dr. Pfeifer, University of British Columbia, BC)) and 10 μg of Cellfectin® reagent were thoroughly mixed in 1 ml of Grace's Insect Cell Culture Medium and incubated at RT for 30 min. Cell medium was replaced dropwise with the transfection mixture, and cells incubated at 27 °C for 4 hr. Two (2) ml of ESF AF media were then added and cells incubated for 48 hr. Standard transfection conditions were varied in attempts to optimise expression. Keeping standard the amount of transfected DNA, amounts of transfection reagent were varied between 5 μg, 10 μg and 15 μg. Keeping standard the amount of transfection reagent, amounts of DNA were varied from 0.75 μg, 1.0 μg, 1.25 μg. Transfections were performed in both *Sf9* and *Tni* cells. Success of transfection was determined by QUANTI-Blue™, a colourimetric assay for the detection of secreted embryonic alkaline phosphatase (SEAP). Cells were selected by the incremental addition of Zeocin™ up to 1.0 mg/ml. One (1) ml supernatant aliquots and approximately 2 x 10^6 cells were sampled for each of the different conditions and assayed for NAGLU and NTAT expression via immunoblot and activity assays.

2.6.2 Transformation of insect vectors for baculovirus expression (performed by Dr. M. Boulanger)

2.6.2a Transfection of pACGP67B constructs into Sf9 cells

Two (2) x 10^6 *Sf9* cells were seeded into a 6 well tissue culture plate in 2 ml SF-900 II SFM augmented with 10 μg/ml gentamicin sulphate and allowed to adhere for 1 hr. Experimental co-transfections were performed as follows: in a final 2 ml medium volume, 2 μg of pAcGP67B containing the recombinant genes of interest was combined with 0.2 μg of BaculoGold™ Bright Linearized Baculovirus DNA, mixed and incubated at RT for 5 min before the addition of 10 μg of Cellfectin® transfection reagent, mixing and incubation at RT for 30 min. A negative control, containing BaculoGold™ Bright Linearized Baculovirus DNA and transfection reagent only, was also prepared. The appropriate transfection mixture was added to aspirated cells and allowed to incubate at 27 °C for 5 hr. The cells were then incubated at 27 °C for 3-4 days in 2 ml of fresh medium. Fluorescent microscopy was utilized to assess the efficiency of transfection. At approximately day 4-5, primary virus was harvested from each well. Non adherent cells and cellular debris was pelleted by centrifugation at 10,000 x g and the primary virus containing supernatant of each sample retained and stored at 4 °C.

2.6.2b Infection of High Five™ cells with recombinant Sf9 virus

Two (2) x 10^6 non adherent High Five™ cells in 2 ml Express Five® SFM were added to each of the used wells and incubated, gently shaking, for 2-4 days to allow the remaining virus to infect the newly added High Five™ cells. The media was centrifuged to pellet the

cellular debris, and the 2 ml supernatant containing the secreted recombinant protein retained for analysis of recombinant protein expression.

2.7 NAGLU and NTAT Expression in Insect cells

2.7.1 Protein expression in stably transformed Sf9 cells

SDS polyacrylamide gel, sample preparation and gel electrophoresis were performed as described in section 2.4.4. The hexahistidine tagged protein used as a positive control and the immunoblot procedure followed, are described in section 2.4.4. Samples from cell lysates and crude culture supernatants were assayed for NAGLU activity as described in section 2.4.5 and analysed relative to cells transfected with p2ZOp2F(SEAP) alone.

2.7.2 Protein expression in baculovirus infected High Five™ cells

HEPES buffered saline (HBS) and imidazole were added to each transduction supernatant, to generate a final 20 mM HEPES, 150 mM NaCl, 20 mM imidazole, pH 8.0 solution. Fifty (50) µl of 50% nickel-nitrilotriacetic acid (Ni-NTA) slurry were prepared per transduction sample. Briefly, the Ni-NTA slurry was spun at 2,000 x g to remove the EtOH storage buffer, washed in the above 1x HBS solution, the beads resuspended to a 50% slurry and each transduction supernatant added to their respective slurry. The protein Ni-NTA bead mix was incubated at 4 °C with gentle agitation for 1 hr and then spun at 1,500 x g for 5 min. Ni-NTA bound proteins were eluted from the beads with 1 bed volume of 1x HBS containing 250 mM imidazole.

Fractions were analysed by immunoblot (as described in section 2.4.4) and by Coomassie Brilliant Blue R-250 staining described below. Gels were placed in fixative (50%

(v/v) methanol, 10% (v/v) acetic acid) for 30 min. After removal of fixative, gels were stained in a coomassie working solution of 50% (v/v) methanol, 10% (v/v) acetic acid, 3% (v/v) coomassie concentrated stock solution for 30 min, removed to destaining solution and incubated for 2-3 hr with gentle agitation at RT until background staining was minimal.

An aliquot of each transduction supernatant, taken prior to binding the Ni-NTA column, and each unbound protein fraction, taken post Ni-NTA binding were analysed by fluorometric assay as described in section 2.4.5.

3 Results

3.1 Construction of Yeast Expression Vectors

3.1.1 Construct 1: pPIC9K.FL.WT.NAGLU

An EcoRI/XbaI restriction digest was performed on sequence verified pPICZα(*NAGLU*) (Patrick, 2006) to obtain the mature lysosomal protein without its targeting sequence. The pPIC9K *P. pastoris* expression vector does not contain an XbaI site, but does contain an AvrII site which when cleaved, generates DNA with complementary restriction ends compatible with XbaI cloning. Sequential digests (both EcoRI followed by AvrII digestion and AvrII followed by EcoRI digestion) in addition to concomitant EcoRI/AvrII digests of the pPIC9K vector were conducted to ensure complete digestion with both enzymes.

EcoRI/XbaI digested *NAGLU* cDNA was ligated into EcoRI/AvrII digested pPIC9K and transformants were screened by colony cracking analysis followed by EcoRI/NotI restriction digestion of isolated plasmid DNA. In spite of the sequential and concomitant digests described above, cloning proved very difficult. A clone appearing to show the correct digestion pattern was eventually isolated. Sequence analysis confirmed this clone contained the desired full length WT NAGLU cDNA sequence cloned into the pPIC9K polylinker, however an extra EcoRI site, lying 3' of the stop codon, was identified. This introduced EcoRI site causes no change in function of the protein and thus work was continued with this clone. Illustrated in Figure 3.1 is an EcoRI digest of pPIC9K.FL.WT.*NAGLU* with bands corresponding to pPIC9K (~9.3 kb) with the full length WT *NAGLU* insert (~2.2 kb).

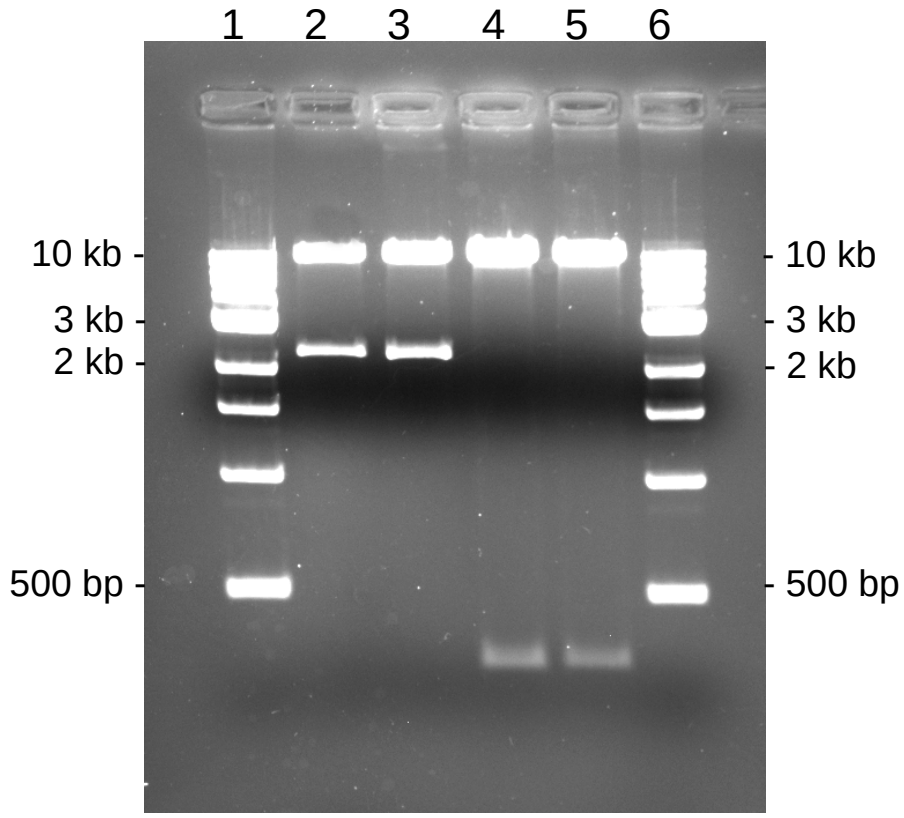


Figure 3.1 Agarose gel showing restriction digests of the final Constructs 1-4 illustrating correct insert sizes. Lanes 2 and 3 contain EcoRI digested pPIC9K.FL.(WT).*NAGLU* and pPIC9K.FL.(CO+WT).*NAGLU,* respectively. Lane 4 contains EcoRI/NotI digested pPIC9K.control.HFxCO.F. Lane 5 contains EcoRI/AvrII digested pPIC9K.control.HFxWT.F. Lanes 1 and 6 each contain a 1 kb DNA standard to allow size determination of the inserts.

3.1.2 Construct 2: pPIC9K.FL.(CO+WT).NAGLU

The 261 bp codon optimised fragment was designed flanked by a 5' EcoRI and a 3' SacII site. The anomalous EcoRI site created in pPIC9K.FL.WT.*NAGLU*, necessitated the design of a forward primer containing a 5' SnaBI site to replace the 5' EcoRI cloning site. The PCR amplified SnaBI/SacII fragment was directionally ligated into SnaBI/SacII digested pPIC9K.FL.WT.*NAGLU*. Transformed colonies containing pPIC9K.FL(CO+WT).*NAGLU* were initially screened by PCR, confirmed as mutation free by sequence analysis, and then propagated for large scale plasmid isolation. An EcoRI digest of pPIC9K.FL(CO+WT)*NAGLU NAGLU* was performed to release the insert. Figure 3.1 shows bands within a 1.5% (w/v) agarose gel corresponding to pPIC9K (~9.3 kb) with a FL.(CO+WT).*NAGLU* insert (~2.2 kb).

3.1.3 Construct 3: pPIC9K.control.HFxCO.F

The HFxCO.F control construct was created in two PCR steps. The first introduced a Factor Xa cleavage site upstream of the 261 bp codon optimised *NAGLU* cDNA. The second incorporated a 5' His antibody recognition site. An agarose gel showing the 2-step PCR amplification method used to generate the control construct is illustrated in Figure 3.2. The resultant 310 bp EcoRI/NotI flanked HFxCO.F product and pPIC9K were double digested with EcoRI and NotI, directionally ligated and transformed into *E. coli* cells. Kanamycin resistant colonies were PCR amplified and plasmids isolated from those generating the correct sized product. An EcoRI/NotI digest of this sequence verified, mutation free, control construct generated bands corresponding to pPIC9K (~9.3 kb) with the HFxCO.F insert (~300 bp) (Figure 3.1).

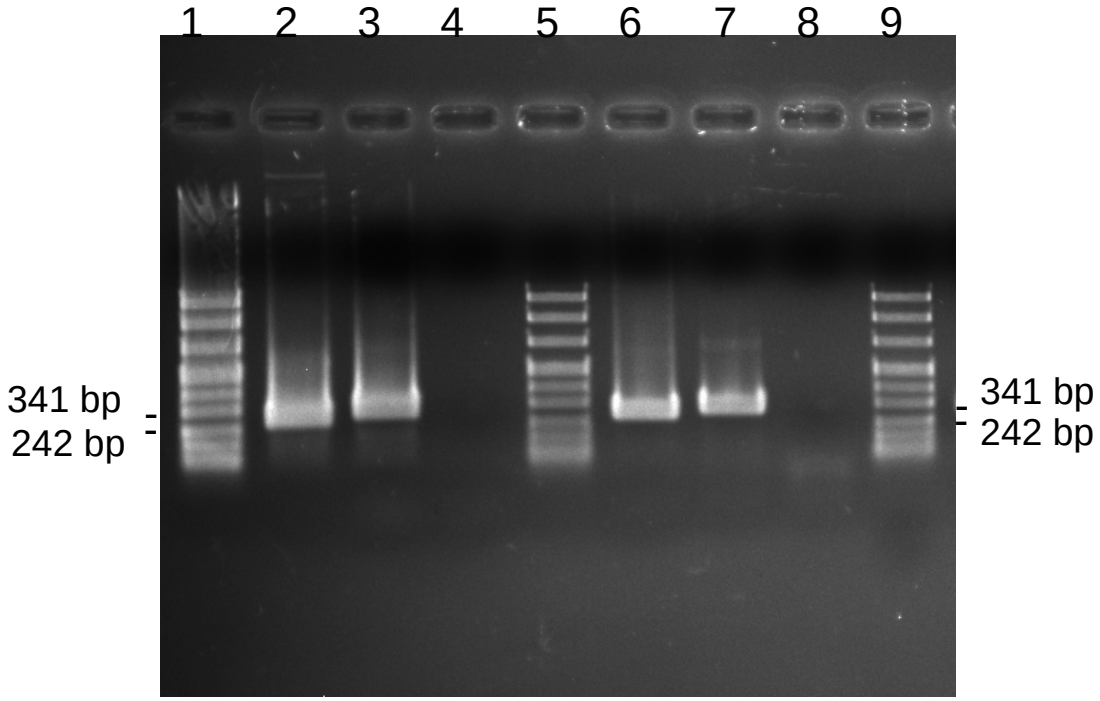


Figure 3.2 Agarose gel showing products from the 2-step PCR amplification to generate HFxCO.F and HFxWT.F inserts. The amplicon encoding the Factor Xa cleavage site followed by bases 1-261 of codon optimised *NAGLU* (FxCO.F) and a 3' NotI restriction site is shown in lane 2. Lane 3 contains the 2nd round product encoding the His antibody recognition site, the Factor Xa cleavage site and bases 1-261 of codon optimised *NAGLU* flanked by 5' EcoRI and 3' NotI restriction sites (HFxCO.F). The lane 6 product contains the Factor Xa cleavage site followed by bases 1-261 of wildtype *NAGLU* (FxWT.F) followed by an AvrII restriction site. The 2nd round product in lane 7 contains the His antibody recognition site, the Factor Xa cleavage site and bases 1-261 of wildtype *NAGLU* flanked by 5' EcoRI and 3' AvrII restriction sites (HFxWT.F). Lanes 4 and 8 contain a ddH_2O negative control for each PCR reaction. Lanes 1, 5 and 9 contain pUC Mix Marker 8 to allow size determination of the amplicons.

3.1.4 Construct 4: pPIC9K.control.HFxWT.F

The HFxWT.F control construct was similarly created in two PCR steps. First, a Factor Xa cleavage site was incorporated upstream of bases 1-261 of WT *NAGLU* cDNA. A second round PCR was then performed to incorporate a 5' His tag for construct purification. An agarose gel showing the 2-step PCR amplification method used to generate the control constructs is illustrated in Figure 3.2. The resultant 308 bp EcoRI/AvrII flanked product was digested with EcoRI and AvrII to allow directional ligation into EcoRI/AvrII digested pPIC9K vector. Following transformation, kanamycin resistant putative pPIC9K.control.HFxWT.F colonies were screened by PCR, confirmed as mutation free by sequence analysis and then propagated for large scale plasmid isolation. Figure 3.1 shows bands resulting from an EcoRI/AvrII digest of the final clone which correspond to pPIC9K (~9.3 kb) with a HFxWT.F insert (~300 bp).

3.2 Transformation of *P. pastoris* with Expression Vector Constructs 1-4

3.2.1 Confirmation of genomic integration in P. pastoris pGlycoSwitchM5

All four expression vector constructs as well as pPIC9K vector only (negative control) were linearized for pGlycoSwitchM5 transformation by SacI digestion. Linearization was confirmed by electrophoresis of a 1 µl aliquot of each digest on a 0.8% (w/v) agarose gel (Figure 3.3) prior to transformation. Selection of transformants was based on growth on MD (minimal dextrose media) versus MDH (minimal dextrose media with His supplement) plates. The *P. pastoris* host strain GS115, and consequently its derivative strain pGlycoSwitchM5, has a mutation in the histidinol dehydrogenase gene (*his4*) which prevents it from synthesizing histidine. The pPIC9K vector contains the *HIS4* gene which complements the *his4* deficiency

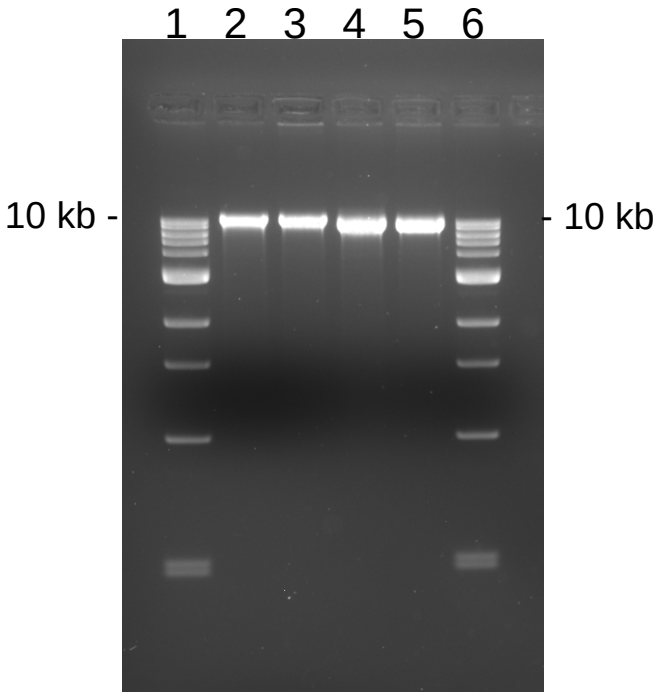


Figure 3.3 Agarose gel showing successful SacI linearization of Constructs 1-4. Lanes 2-5 contain SacI linearized pPIC9K.FL.(WT).*NAGLU*, pPIC9K.FL.(CO+WT).*NAGLU,* pPIC9K.control.HFxWT.F and pPIC9K.control.HFxCO.F, respectively. Lanes 1 and 6 each contain a 1 kb DNA standard to allow size determination of the linearized constructs.

in the host, allowing selection for transformants by their ability to grow on a histidine deficient medium. Colonies growing on MD plates were screened for genomic integration of various expression vectors (Constructs 1-4 and vector only) by direct yeast PCR. This was initially performed using 5' *AOX1* (F) and 3' *AOX1* (R) primers specific for the *AOX1* promoter. The PCR amplicons generated for Constructs 1, 3, 4 and vector only were resolved on 1% (w/v) agarose gel and correspond to the native *AOX1* gene (2200 bp), pPIC9K.FL.(WT).*NAGLU* (2672 bp), pPIC9K.control.HFx.CO.F (782 bp), pPIC9K.control.HFx.WT.F (788 bp) and pPIC9K vector alone (492 bp) (Figure 3.4).

As attaining equally efficient PCR amplification of both the endogenous *AOX1* gene and the introduced *AOX1* containing constructs proved difficult, new sets of primers specific for the amplification of the endogenous *AOX1* gene and each of the four integrated constructs were designed. Primer sequences are listed in Table 2.2 and amplification outcomes discussed in section 3.2.2.

3.2.2 Replica plating pGlycoSwitchM5 transformants

Transformation of pGlycoSwitchM5 with SacI linearized constructs favours recombination at the *HIS4* locus via a single crossover between the homologous sequences in the recombinant constructs and the *P. pastoris* genome resulting in a $His^{+}Mut^{+}$. However the presence of *AOX1* sequences in the plasmid also creates the possibility of recombination at the *AOX1* locus causing displacement of the wildtype *AOX1* gene and producing $His^{+}Mut^{S}$ transformants. Because Mut^{S} transformants are not producing alcohol oxidase from the displaced *AOX1* gene, they do not efficiently metabolise methanol as a carbon source and therefore grow poorly on minimal methanol (MM) medium. This can be used to distinguish

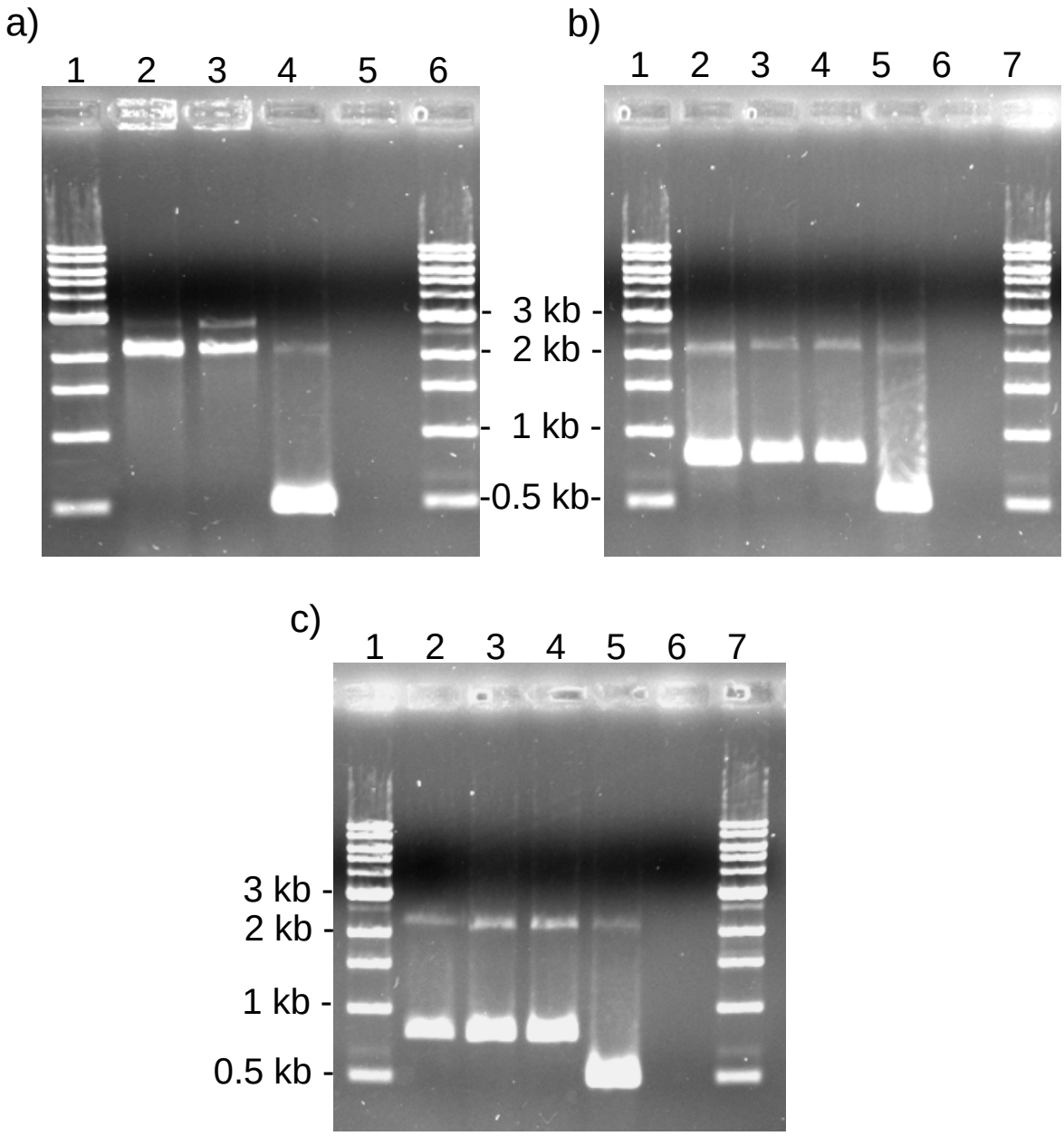


Figure 3.4 Agarose gels illustrating *AOX1* promoter specific direct yeast PCR confirming integration of a) pPIC9K.FL(WT).*NAGLU* b) pPIC9K.control.HFxWT.F c) pPIC9K.control.HFxCO.F into the *P. pastoris* genome. a) *P. pastoris* transformed with pPIC9K.FL(WT).*NAGLU* (lanes 2 and 3), transformed with pPIC9K vector alone (lane 4) and H_2O only negative control (lane 5). **b)** *P. pastoris* transformed with pPIC9K.control.HFx.WT. F (lanes 2, 3 and 4), transformed with pPIC9K vector alone (lane 5) and H_2O only negative control (lane 6). **c)** *P. pastoris* transformed with pPIC9K.control.HFx.CO.F (lanes 2, 3 and 4), transformed with pPIC9K vector alone (lane 5) and H_2O only negative control (lane 6). A 1 kb DNA standard is present in the external lanes of each gel to allow size determination of the amplicons.

His^+ transformants in which the *AOX1* gene has been disrupted (His^+Mut^S) from His^+ transformants with an intact *AOX1* gene (His^+Mut^+).

Forty (40) colonies per construct were replica plated onto MD versus MM plates to identify putative Mut^S colonies. Mut^S strains, growing slowly on methanol, were identified for Construct 1 (pPIC9K.FL.WT.*NAGLU*), Construct 2 (pPIC9K.FL.(CO+WT)*NAGLU*) and Construct 3 (pPIC9K.control.HFxCO.F), but not Construct 4 (pPIC9K.control.HFxWT.F). Replica plating of a further 80 colonies also failed to identify Mut^S colonies for Construct 4.

Absence of amplification of the endogenous *AOX1* gene was used to confirm the Mut^S status of colonies unable to grow on MM. However, attaining equally efficient PCR amplification of both the endogenous *AOX1* gene and the introduced *AOX1* containing constructs proved difficult, as discrepancies in amplicon size led to preferential amplification of smaller products. As such, new sets of primers specific for the amplification of the endogenous *AOX1* gene and each of the four integrated constructs were designed (Table 2.2). A new *AOX1* forward internal primer was designed to the coding region of the *AOX1* gene. As the *AOX1* and *AOX2* gene sequences are highly conserved, this forward primer will bind to both *AOX1* and *AOX2* genes. To maintain specificity of amplification for the *AOX1* gene, this forward primer was used in combination with the 3' *AOX1* (R) primer which was designed within the unique 3' untranslated region of the *AOX1* gene. Similar sized amplicons were designed for the specific amplification of each integrated construct. Lack of amplification from *AOX1* specific primers in conjunction with clear amplification with construct specific primers confirmed the Mut^S status of clones. Figure 3.5a illustrates this comparison for Construct 1. Endogenous *AOX1* specific bands (493 bp) are present for Mut^+ strains run in lanes 3 and 4 and the empty plasmid control run in lane 5, but are absent from the Mut^S strain in lane 2 and the H_2O only negative control.

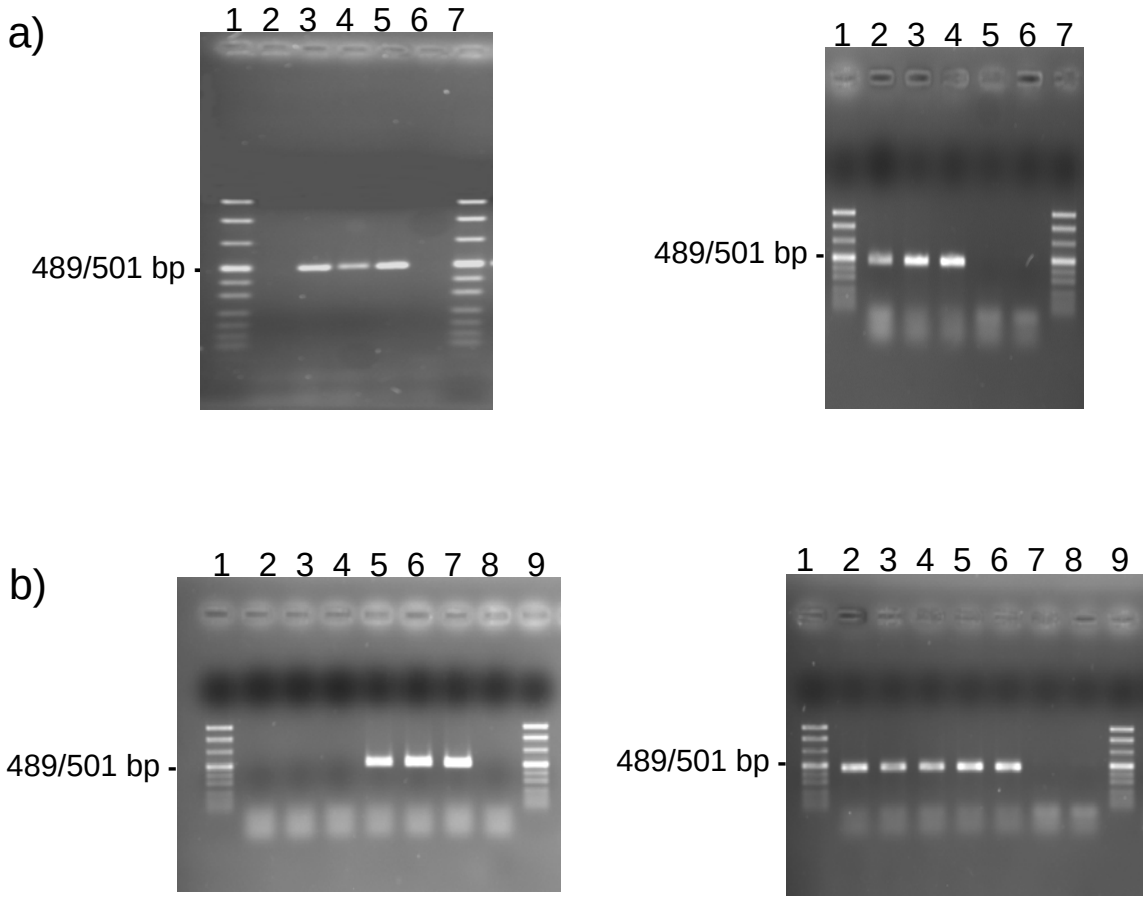


Figure 3.5 Agarose gel analysis of Mut^S and single and multiple copy Mut$^+$ strains for a) pPIC9K.FL(WT).*NAGLU* (Construct 1) and b) pPIC9K.FL(CO+WT).*NAGLU* (Construct 2). Duplicate samples have been amplified on the LHS gel by endogenous *AOX1* specific PCR and on the RHS gel by *NAGLU* integrant specific PCR. **a)** On each gel lane 2 contains a Mut^S strain, lane 3 contains a single integrant Mut$^+$ strain, lane 4 contains a multiple integrant Mut$^+$ strain, lane 5 contains a vector only Mut$^+$ strain, and lane 6 contains H_2O only PCR negative control. **b)** On each gel lanes 2, 3 and 4 contain Mut^S strains, lane 5 contains a single integrant Mut$^+$ strain, lane 6 contains a multiple integrant Mut$^+$ strain, lane 7 contains a vector only Mut$^+$ strain, and lane 8 contains H_2O only PCR negative control. The pUC Mix Marker 8 is present in the external lanes of each gel to allow size determination of the amplicons.

When these same strains are amplified using primers specific for integrated *NAGLU*, bands of the expected size (440 bp) are seen for all Construct 1 integrants, regardless of Mut^+/Mut^S status, and are absent in empty plasmid and H_2O controls. Similar results are seen in Figure 3.5b for Construct 2. Endogenous *AOX1* specific bands are absent from lanes 2, 3 and 4, confirming Mut^S status, and present in lanes 5, 6 and 7, confirming Mut^+ status. Mut^S strains were also isolated for Construct 3 and confirmed as such by the absence of *AOX1* specific bands in lanes 2 and 3 in Figure 3.6a. As seen in Figure 3.6b, all strains identified for Construct 4 were Mut^+ and hence *AOX1* specific PCR products were seen lanes 2-5 containing pPIC9K.control.HFxWT.F strains.

P. pastoris is capable of integrating multiple copies of transforming DNA into genomic sites of sequence homology. As successful integration of multiple copies of the gene of interest may increase the level of protein expressed, determining which Mut^+ strains have multiple integrants is highly desirable. This was achieved by selection on MD plates containing increasing concentrations of geneticin. Plates were augmented with geneticin from 0 mg/ml up to 1 mg/ml in 0.25 mg/ml increments. After 2-3 days at 30 ^oC, the replica plates were scored (Figure 3.7). Growth on 0.25 mg/ml, 0.5 mg/ml, 0.75 mg/ml and 1.0 mg/ml is postulated to correlate with 1, 2, 3, and 4 copies of the integrated vector, respectively. As illustrated in Figure 3.7, multiple copy integrants, assessed as those that grew on $\geq$ 0.5 mg/ml geneticin, were seen for all strains.

Genomic integration of both single and multiple integrants was confirmed by direct yeast PCR as described for Mut^S strains. Amplification with endogenous *AOX1* specific, and *NAGLU* integrant specific primers is shown in Figures 3.5 and 3.6. Both single and multiple integrant strains were identified for each of Constructs 1, 2, 3 and 4.

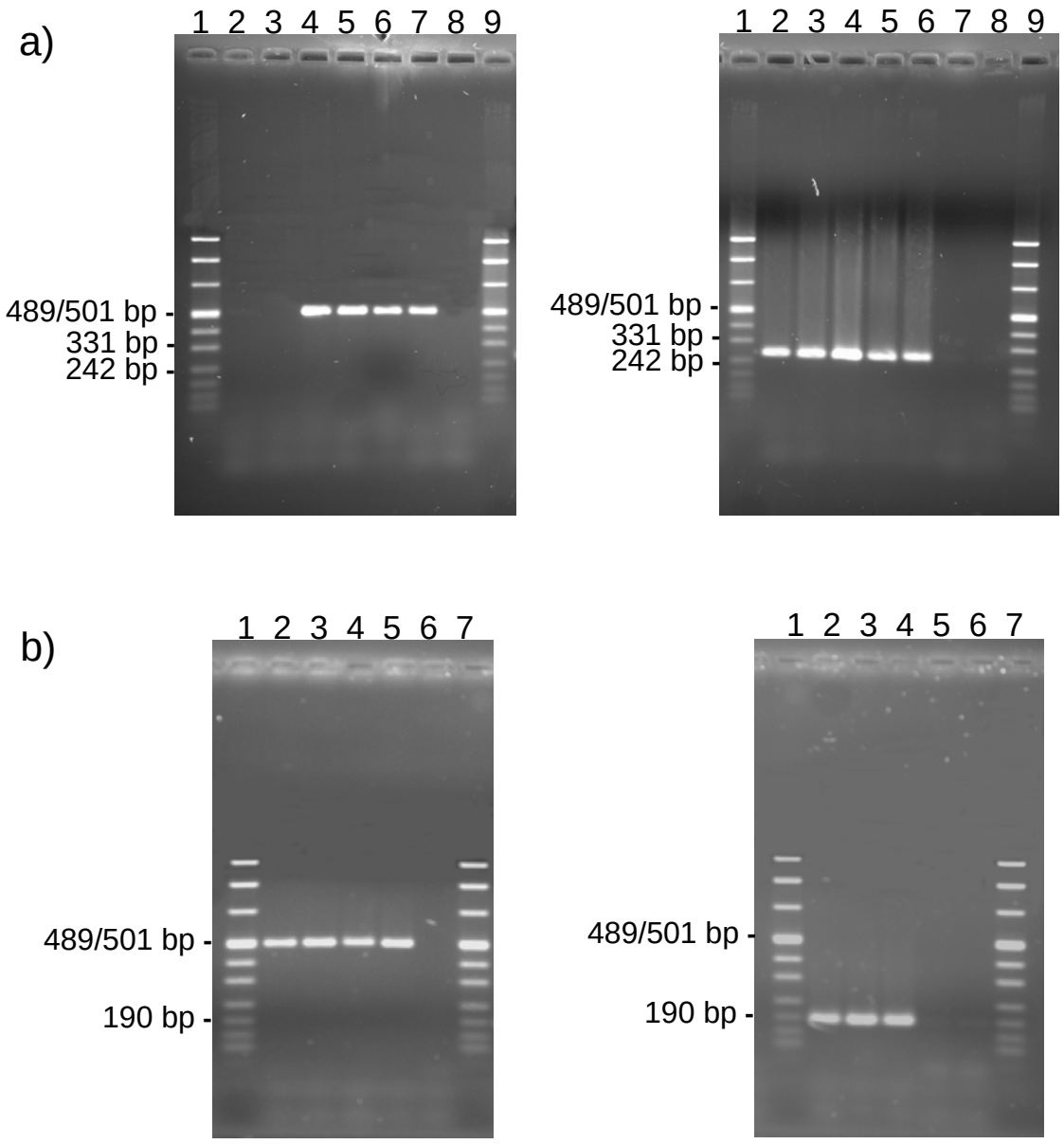


Figure 3.6 Agarose gel analysis of Mut^S and single and multiple copy Mut$^+$ strains for a) pPIC9K.control.HFxCO.F (Construct 3) and single and multiple copy Mut$^+$ strains for b) pPIC9K.control.HFxWT.F (Construct 4). Duplicate samples have been amplified on the LHS gel by endogenous *AOX1* specific PCR and on the RHS gel by *NAGLU* integrant specific PCR. **a)** On each gel lanes 2 and 3 contain Mut^S strains, lane 4 contains a single integrant Mut$^+$ strain, lanes 5 and 6 contain multiple integrant Mut$^+$ strains, lane 7 contains a vector only Mut$^+$ strain, and lane 8 contains H_2O only PCR negative control. **b)** On each gel lanes 2 contains a single integrant Mut$^+$ strain, lanes 3 and 4 contain multiple integrant Mut$^+$ strains, lane 5 contains a vector only Mut$^+$ strain, and lane 6 contains H_2O only PCR negative control. The pUC Mix Marker 8 is present in the external lanes of each gel to allow size determination of the amplicons.

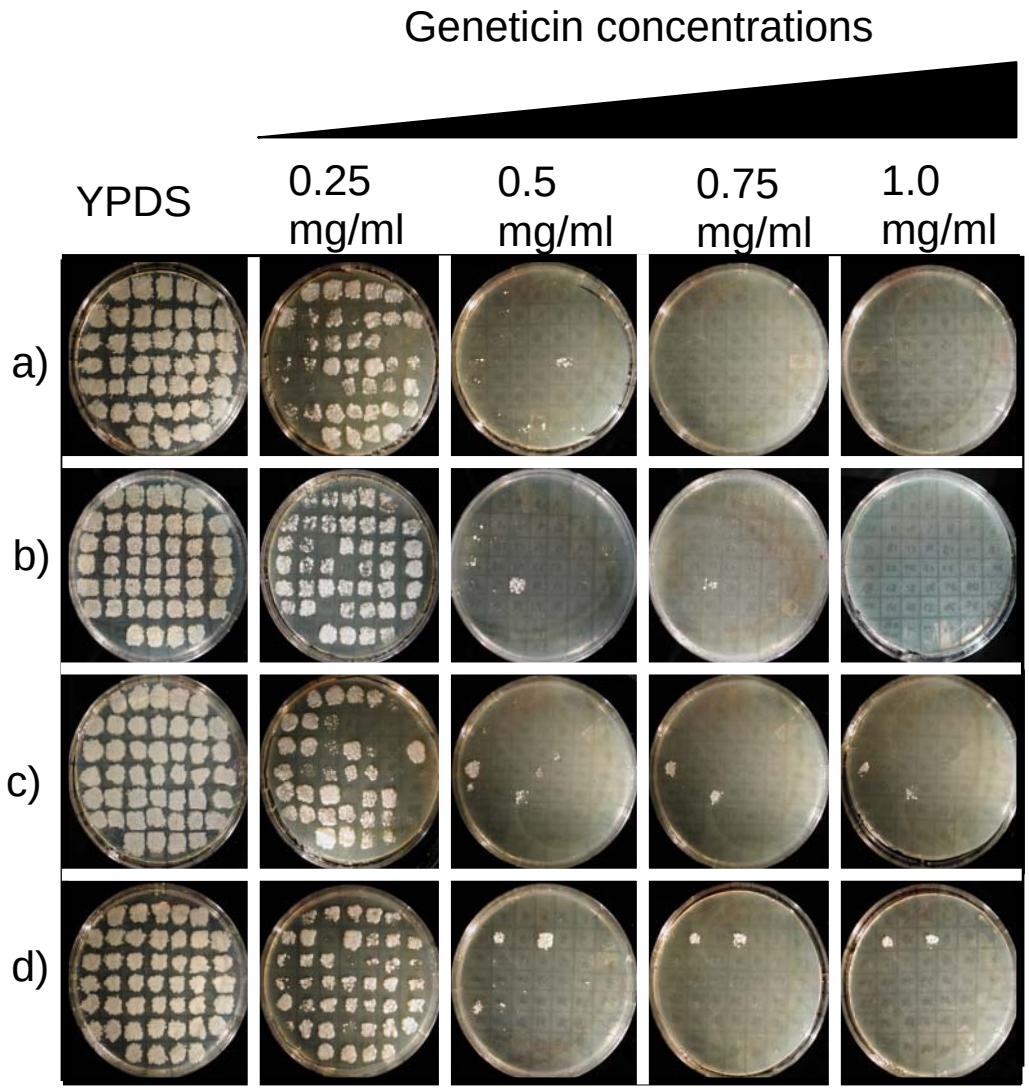


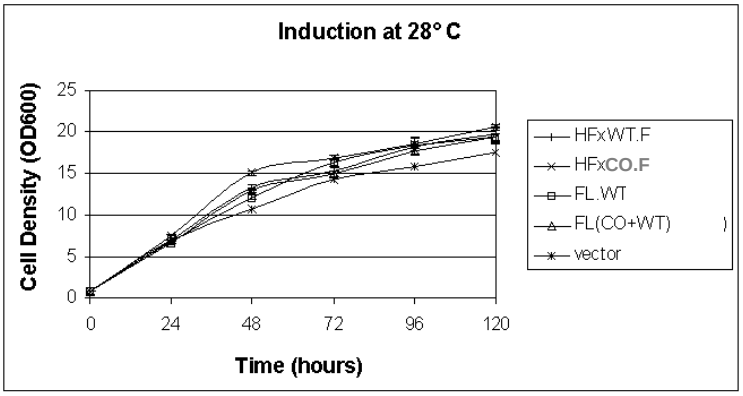
Figure 3.7 Identification of multiple copy integrants by growth on MD+geneticin plates. All 4 constructs were replica plated onto MD plates augmented with increasing geneticin concentrations from 0 mg/ml to 1 mg/ml in 0.25 mg/ml increments. Multiple copy integrants were seen for all strains. **a)** Growth on 0.5 mg/ml for Construct 1: pPIC9K.FL.WT.*NAGLU*. **b)** Growth on 0.75 mg/ml for Construct 2: pPIC9K.FL.(CO+WT).*NAGLU*. **c)** and **d)** Growth on 1 mg/ml for Construct 3: pPIC9K.control.HFxCO.F and Construct 4: pPIC9K.control.HFx WT.F, respectively.

3.3 Growth and *NAGLU* Expression in Transformed *P. pastoris*

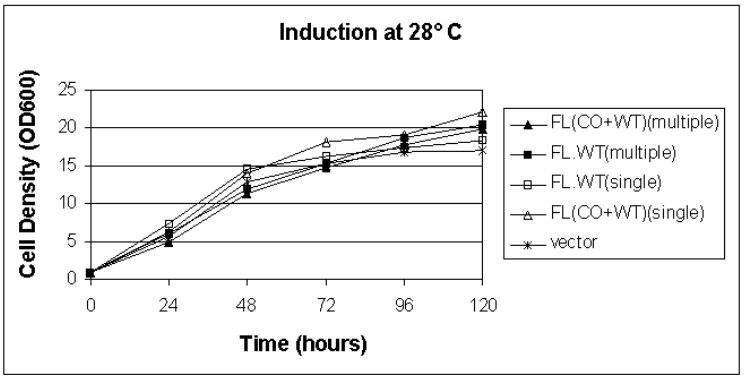
3.3.1 Induction of growth in pGlycoSwitchM5 (Mut$^+$ and Mut^S) strains

All strains described below were assessed for both pattern of growth and levels of recombinant protein expression. A vector only negative control was included in all trials. A variety of different strains were analysed. Assessment of strains grown at 28 °C was performed for single copy integrant Mut$^+$ strains of Constructs 1-4 and multiple copy integrant Mut$^+$ strains and Mut^S strains of full length Construct 1 and 2. Trials were similarly conducted on full length Construct 1 and 2 single copy integrant Mut$^+$ strain cultures grown at 15 °C. All strains under each of the conditions outlined above were cultured and assessed in an absolute minimum of two separate trials. Culture densities were monitored to demonstrate viability of cultures through consumption of the inducing agent methanol. This in turn indicates induction of the *AOX1* promoter by which the recombinant genes of interest are driven. Samples were taken directly from each of the cultures at 0, 24, 48, 72, 96 and 120 hr post induction (Figures 3.8 and 3.9). Measurements were taken in triplicate at each time point, and the averaged OD_{600} was used to calculate the number of yeast cells/ml over all time points tested. Due to their slow methanol utilization phenotype, Mut^S cultures were grown to a pre-induction density of approximately 1.5-3.0 x 10^8 cells/ml (OD_{600} ~15), approximately fifteen times higher than GS115 cultures which were grown to a pre-induction density of 1-2 x 10^7 cells/ml (OD_{600} = 1). Upon induction, Mut^S cultures acquired approximately 1.5 times biomass during the first 24 hr, followed by maintenance of cell density for the remaining 96 hr. Mut$^+$ cultures on the other hand, quickly multiplied to 5-6 times their initial density in the first 24 hr and continued to grow throughout the 120 hr induction period, reaching a biomass of approximately 20 times their initial value.

a)



b)



c)

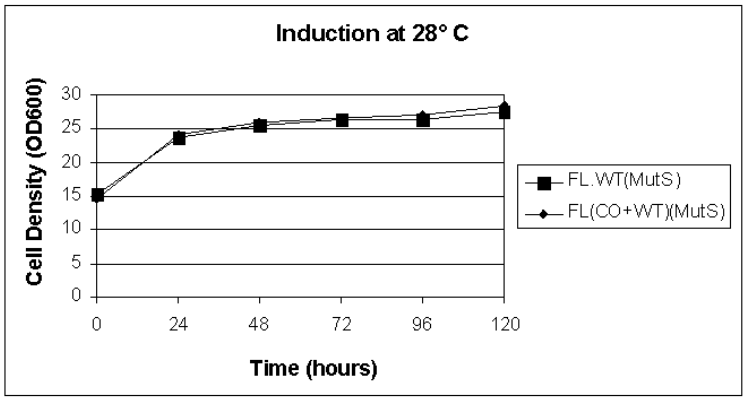


Figure 3.8 Growth curves for pGlycoSwitchM5 strains grown at 28 ^{o}C in small scale cultures following methanol induction. a) Single copy Mut^{+} transformants for Constructs 1-4. **b)** Multiple and single copy Mut^{+} transformants for Constructs 1-2. **c)** Mut^{S} transformants for Constructs 1-2. Vector only transformants were used as negative controls. Cells were initially grown in BMGY prior to transfer to BMMY to enable methanol induction of the *AOX1* promoter. Cultures were incubated at 28 ^{o}C, aerated at 275 rpm, and supplemented with 1.0% (v/v) methanol every 24 hr starting at T = 24. Samples were taken every 24 hr for 120 hr after methanol induction and analysed at OD_{600nm} to determine cell density.

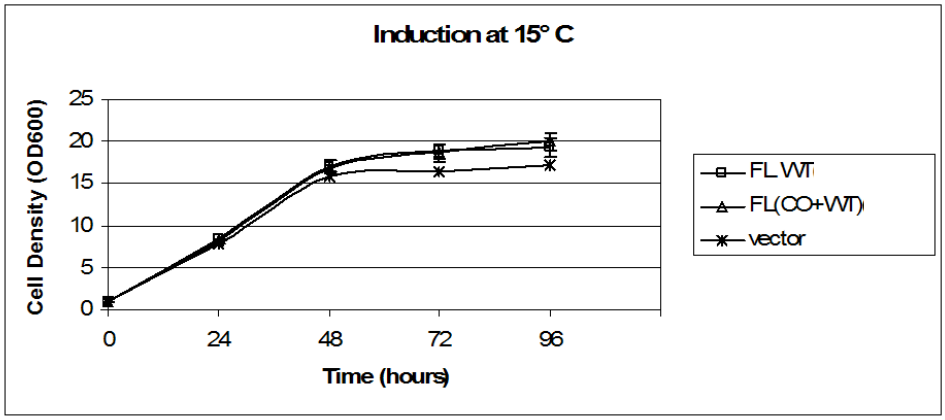


Figure 3.9 Growth curves for single copy integrant Mut$^+$ pGlycoSwitchM5 strains grown at 15 ^{o}C in small scale cultures following methanol induction. Illustrated are cultures of single copy Mut$^+$ transformants for Constructs 1-2 and a vector only transformant used as a negative control. Cells were initially grown in BMGY and transferred to BMMY for methanol induction of the *AOX1* promoter. Cultures were incubated at 15 ^{o}C, aerated at 275 rpm, and supplemented with 1.0% (v/v) methanol every 24 hr starting at T = 24. Samples were taken every 24 hr for 96 hr after methanol induction and analysed at OD_{600nm} to determine cell density.

3.3.2 SDS-PAGE immunoblot and silver stain analysis

3.3.2a Analysis of control Constructs 3 and 4

To analyse expression of the recombinant control proteins (Construct 3, pPIC9K.control.HFxCO.F and Construct 4, pPIC9K.control.HFxWT.F), supernatant samples were obtained every 24 hr post-induction over a time course of 120 hr from cultures grown at 28 ^{o}C. Both crude and 40x concentrated supernatant samples were assessed by SDS-PAGE followed by silver stain or immunoblot analysis. Neither control Construct 3 (pPIC9K.control. HFxCO.F), nor control Construct 4 (pPIC9K.control.HFxWT.F.) generated an induction specific band of the expected size when samples (both crude and concentrated) were analysed by silver stain analysis or immunoblot analysis using the anti-His_4 antibody. The recognition of a His tagged non NAGLU control construct served to indicate that the transfer and the anti-His_4 detection systems were both functioning successfully and thereby implied that lack of detection of protein from Construct 3 and 4 was due to absence of protein (data not included). As these fragment constructs failed to generate and/or maintain detectable levels of protein, they were omitted from subsequent experimental analysis. Analysis from this point focused on the full length constructs pPIC9K.FL.WT.*NAGLU* (Construct 1) and pPIC9K.FL(CO+WT).*NAGLU* (Construct 2).

3.3.2b Analysis of full length Constructs 1 and 2

To analyse expression of the full length recombinant proteins (produced by pPIC9K.FL.WT.*NAGLU* and pPIC9K.FL.(CO+WT).*NAGLU*), from both single and multiple Mut^{+} strains and Mut^{S} strains, supernatants were sampled, concentrated and electrophoresed as described above. For immunoblot analysis, blots were probed with anti-NAGLU primary

antibody followed by a goat anti-rabbit horseradish peroxidase conjugated secondary antibody as described in section 2.4.4. Analysis on unconcentrated supernatant fractions failed to detect bands for any samples (data not included). Initial immunoblot analysis on ammonium sulphate concentrated supernatant samples for Mut^S, as well as multiple and single copy Mut^+ strains, indicated there may be an approximately 60 kDa band produced by NAGLU constructs (Figure 3.10 and Figure 3.11a, b). However when further control samples were analysed, it was seen that this 60 kDa band also appeared in the empty vector only control samples (Figure 3.11c).

The full length wildtype and partially codon optimised proteins, under the direction of the α-MF secretion signal, were expected to be secreted into the culture media during induction of all strains. However, as the crude supernatant and the 40-fold ammonium sulphate concentrated supernatant appeared not to contain recombinant protein, cell pellets from all time periods were also analysed for the presence of NAGLU. Lysed cell pellets, roughly normalized for cell number, were prepared as described in section 2.4.3. Clear 60 kDa bands were seen at all induction time points for cell lysates, including empty vector only negative control cell lysates (Figure 3.12). To show that bands seen in the supernatant samples were identical in size to those generated from the pellet lysates, vector only negative control lysate samples were run along side T = 96 ammonium sulphate precipitated supernatant samples from strains containing both Constructs 1 and 2 (Figure 3.12 c).

To test whether low levels of NAGLU protein, beneath the detection limit of the immunoblot system using the anti-NAGLU antibody, could be visualized by alternative methods, the same supernatant samples were analysed by SDS-PAGE followed by silver staining. Silver stain analysis of 40-fold ammonium sulphate concentrated fractions showed a putative NAGLU specific approximately 90 kDa band present in the induced time points of all

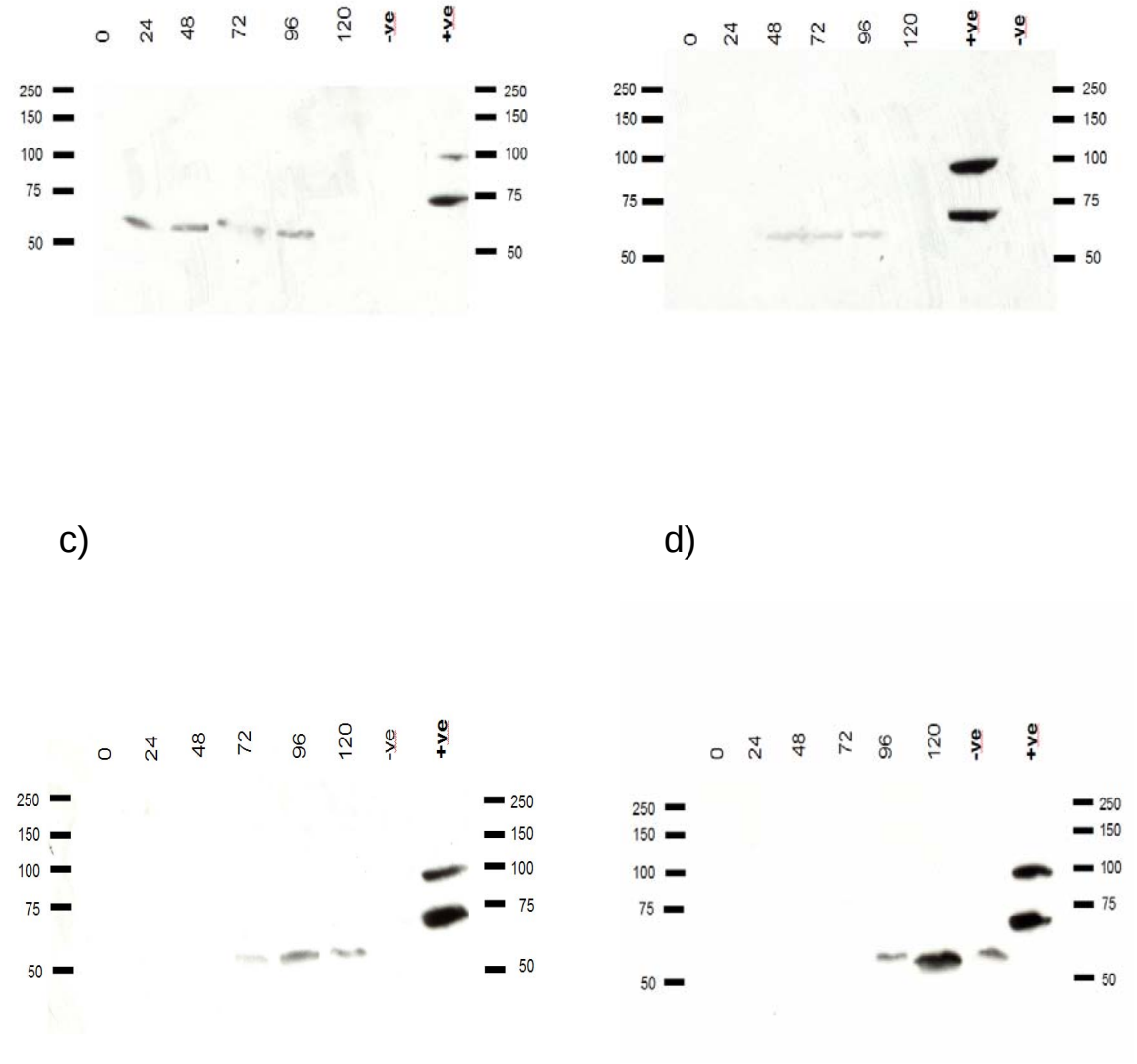
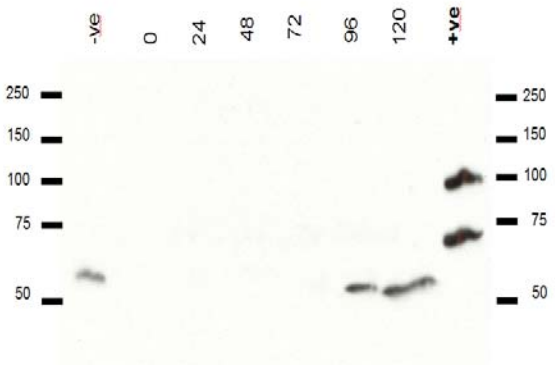


Figure 3.10 SDS-PAGE immunoblots of 40x concentrated supernatant samples from induced single and multiple integrant Mut$^+$ strains containing full length constructs pPIC9K.FL.WT.*NAGLU* (Construct 1) and pPIC9K.FL.(CO+WT).*NAGLU* (Construct 2). Strains were sampled at 24 hr intervals for 0-120 hr post induction. A negative vector only control and a positive *Sf9* CBD-NTAT control were included on each gel. a) and **b)** Samples from single copy integrants are shown for pPIC9K.FL.WT.*NAGLU* and pPIC9K.FL.(CO+WT).*NAGLU* respectively. **c**) and **d**) Samples from multiple copy integrants are shown for pPIC9K.FL.WT.*NAGLU* and pPIC9K.FL.(CO+WT).*NAGLU* respectively. Precision Plus dual colour marker was loaded in the outer lanes of all gels and used for size comparison.

a)



b)

c)

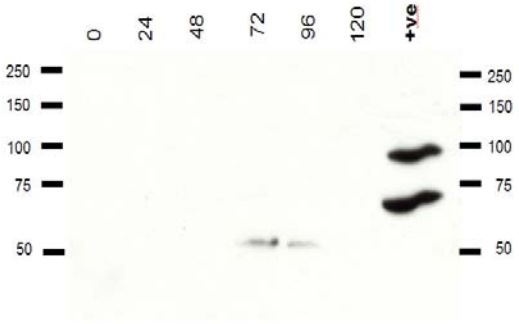


Figure 3.11 SDS-PAGE immunoblots of 40x concentrated supernatant samples from induced Mut^S strains containing full length constructs pPIC9K.FL.WT.*NAGLU* (Construct 1) and pPIC9K.FL.(CO+WT).*NAGLU* (Construct 2) and an induced vector only control strain. Strains were sampled at 24 hr intervals for 0-120 hr post induction. A positive CBD-NTAT control was included on each gel. a) Mut^S pPIC9K.FL.WT.*NAGLU*. **b**) Mut^S pPIC9K.FL.(CO+WT*).NAGLU*. **c)** Vector only control strain. A negative vector only control was included on gels **a)** and **b)**. Precision Plus dual colour marker was loaded in the outer lanes of all gels and used for size comparison.

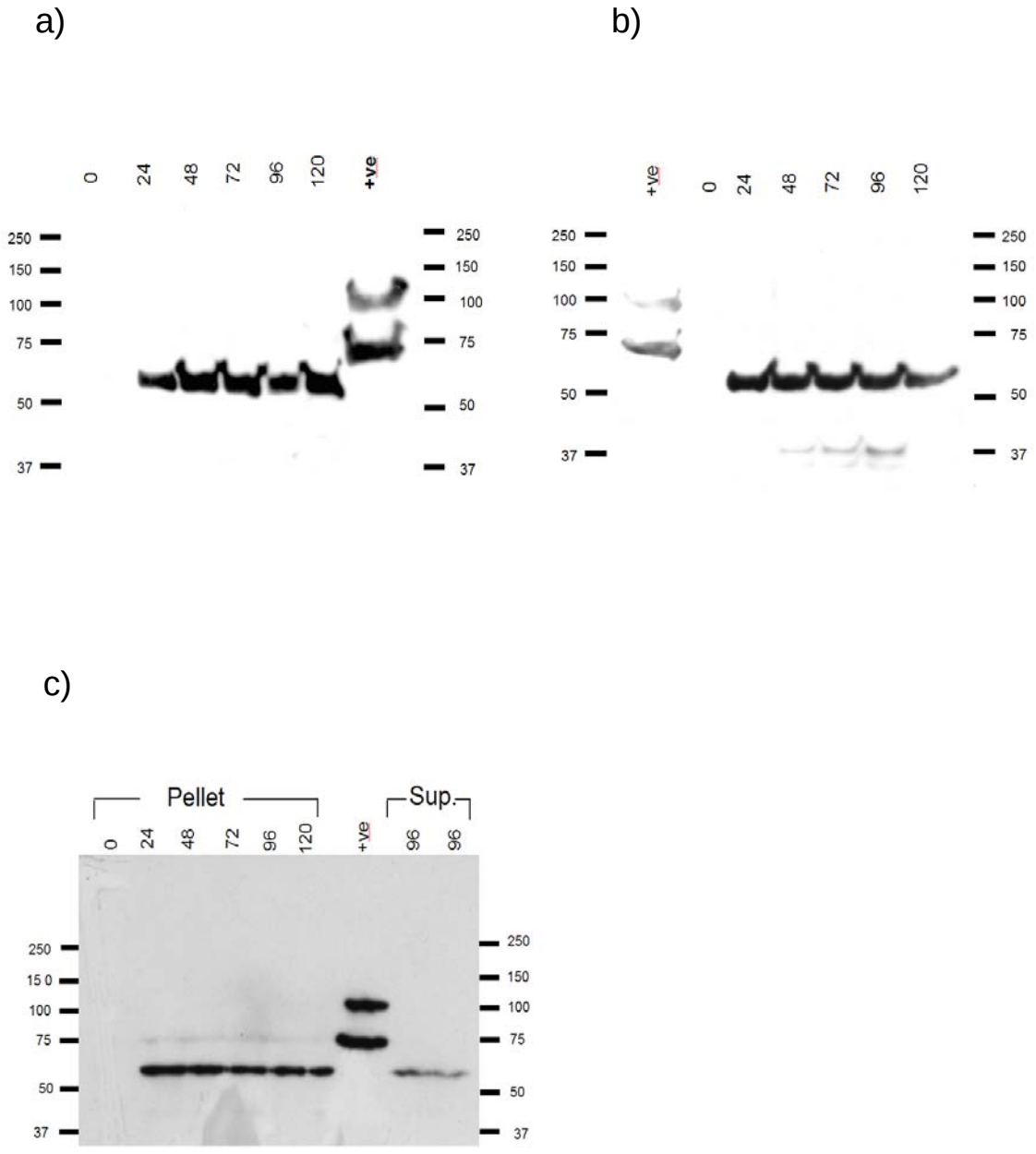


Figure 3.12 SDS-PAGE immunoblots of pellet lysate samples. Single integrant Mut$^+$ strains were sampled at 24 hr intervals for 0-120 hr post induction. A positive *Sf9* CBD-NTAT control was loaded on each gel. a) pPIC9K.FL.WT.*NAGLU* (Construct 1) **b**) pPIC9K.FL.(CO+WT).*NAGLU* (Construct 2). **c**) Vector only control. Supernatant samples (40x concentrated) from time point T = 96 of Mut$^+$ strains pPIC9K.FL.WT.*NAGLU* and pPIC9K.FL.(CO+WT).*NAGLU* were included for comparison between pellet lysate and supernatant samples. Precision Plus dual colour marker was loaded in the outer lanes of all gels and used for size comparison.

Mut^{+} single and multiple integrant strain and Mut^{S} strains and is absent from all time points of the empty vector negative control (Figures 3.13 and 3.14).

Comparison of pellet lysate and 40x concentrated supernatant samples from pPIC9K.FL.WT.*NAGLU* and pPIC9K.FL.(CO+WT).*NAGLU* single integrant Mut^{+} strains indicated that this protein was present in the secreted supernatant fraction only (Figure 3.15). Although this induced, secreted protein appears specific to strains containing *NAGLU* constructs, it consistently shows only extremely low levels of expression.

3.3.3 NAGLU activity assay

To determine the time point at which highest expression of functionally active protein occurred, time course supernatant samples from cultures transformed with full length recombinant proteins (Constructs 1 and 2) were obtained. Both crude and 40x concentrated supernatant samples from the singly and multiply integrated Mut^{+} strains as well as Mut^{S} strains were compared by assaying for NAGLU activity. Activity levels were measured relative to strains transformed with empty vector alone and reported in terms of arbitrary fluorescence units. Activity assays on unconcentrated supernatant fractions failed to detect activity levels above the empty vector negative control for any samples (data not included). Low levels of activity were however detected in the 40x concentrated samples of all strains. Slight differences in the level of activity were detected amongst the various strains, with the highest levels of expression of Construct 1 seen for the Mut^{S} strain and highest level of expression of Construct 2 seen for the singly integrated Mut^{+} strain (Figure 3.16). Although expression levels did vary minimally between trials, these patterns of expression remained consistent. Overall however, there was no appreciable or comparable increase in expression for both constructs from any of codon optimisation, multiple-integration, or expression in Mut^{S}

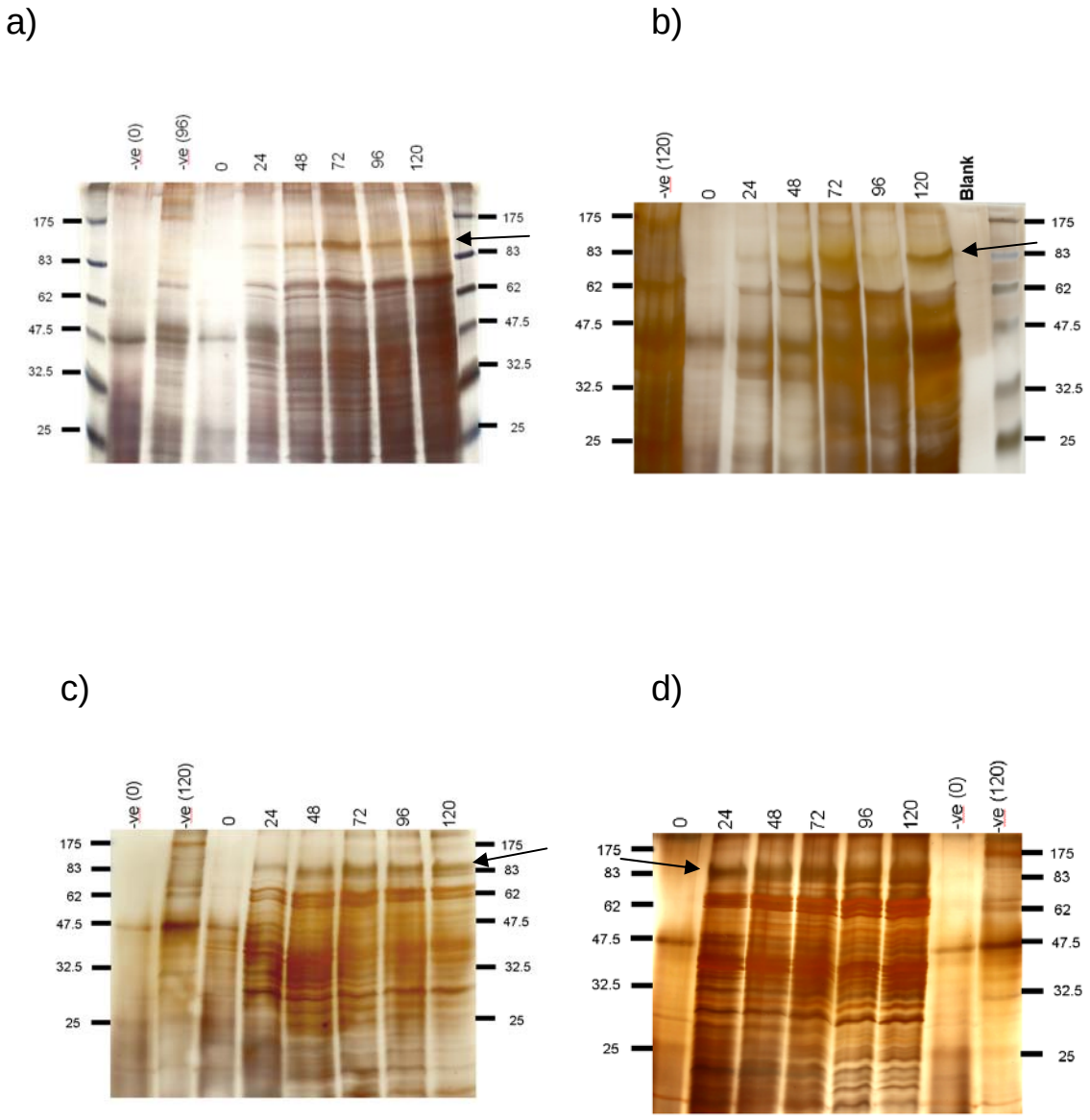


Figure 3.13 SDS-PAGE silver stain analysis of 40x concentrated supernatant samples from induced single and multiple integrant Mut$^+$ strains containing full length constructs pPIC9K.FL.WT.*NAGLU* (Construct 1) and pPIC9K.FL.(CO+WT).*NAGLU* (Construct 2). Strains were sampled at 24 hr intervals for 0-120 hr post induction. Negative vector only controls at T = 0 and T = 96 or 120 hr were included on each gel. Arrows indicate the approximately 90 kDa induced NAGLU specific band of interest. a) and **b)** Samples from single copy integrants are shown for pPIC9K.FL.WT.*NAGLU* and pPIC9K.FL.(CO+WT).*NAGLU* respectively. **c**) and **d**) Samples from multiple copy integrants are shown for pPIC9K.FL.WT.*NAGLU* and pPIC9K.FL.(CO+WT).*NAGLU,* respectively. A Broad Range Protein standard was loaded in the outer lanes of all gels and used for size comparison.

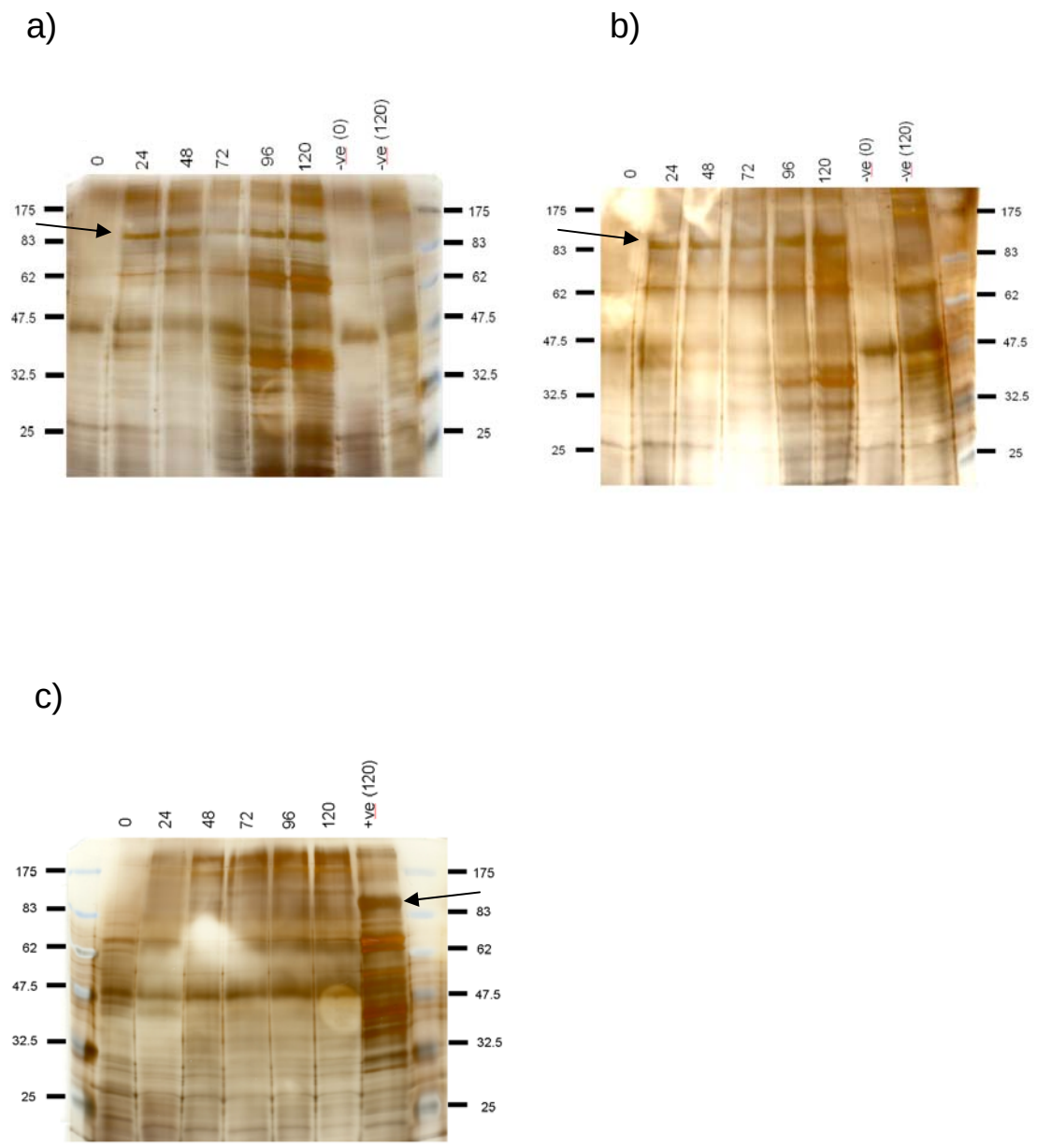
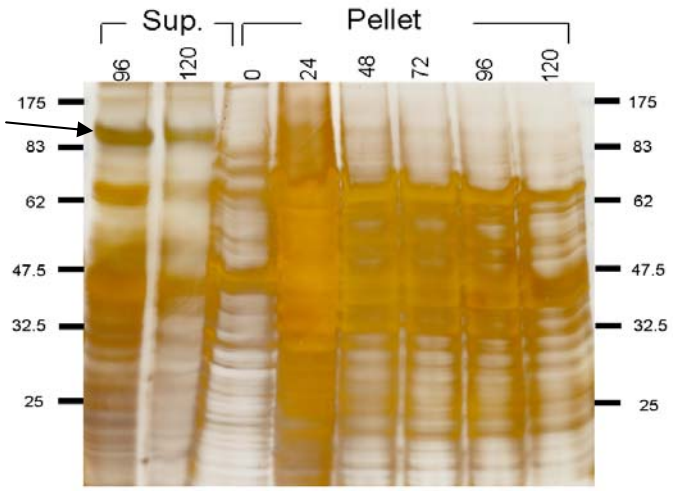


Figure 3.14 SDS-PAGE silver stain analysis of 40x concentrated supernatant samples from induced Mut^S strains containing full length constructs pPIC9K.FL.WT.*NAGLU* (Construct 1) and pPIC9K.FL.(CO+WT).*NAGLU* (Construct 2) and from an induced vector only control strain. Strains were sampled at 24 hr intervals for 0-120 hr post induction. Negative vector only controls at T = 0 and T = 120 hr were included on gels containing the Mut^S strains. A positive Construct 1 control at T = 120 hr was included on the vector only gel. Arrows indicate the approximately 90 kDa induced NAGLU specific band of interest. a) Mut^S pPIC9K.FL.WT.*NAGLU*. **b**) Mut^S pPIC9K.FL.(CO+WT).*NAGLU*. **c)** Vector only control strain. A Broad Range Protein standard was loaded in the outer lanes of all gels and used for size comparison.

a)



b)

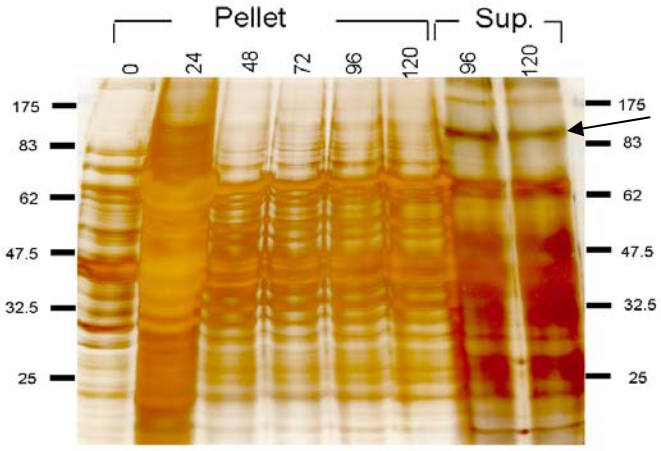


Figure 3.15 SDS-PAGE silver stain comparison of pellet lysate samples and of 40x concentrated supernatant samples from induced single integrant full length construct Mut$^+$ strains containing pPIC9K.FL.WT.*NAGLU* and pPIC9K.FL.(CO+WT).*NAGLU*. Strains were sampled at 24 hr intervals for 0-120 hr post induction. Arrows indicate the approximately 90 kDa supernantant specific band of interest. a) pPIC9K.FL.WT.*NAGLU* **b**) pPIC9K.FL.(CO+WT).*NAGLU.* Precision Plus dual colour marker was loaded in the outer lanes of all gels and used for size comparison.

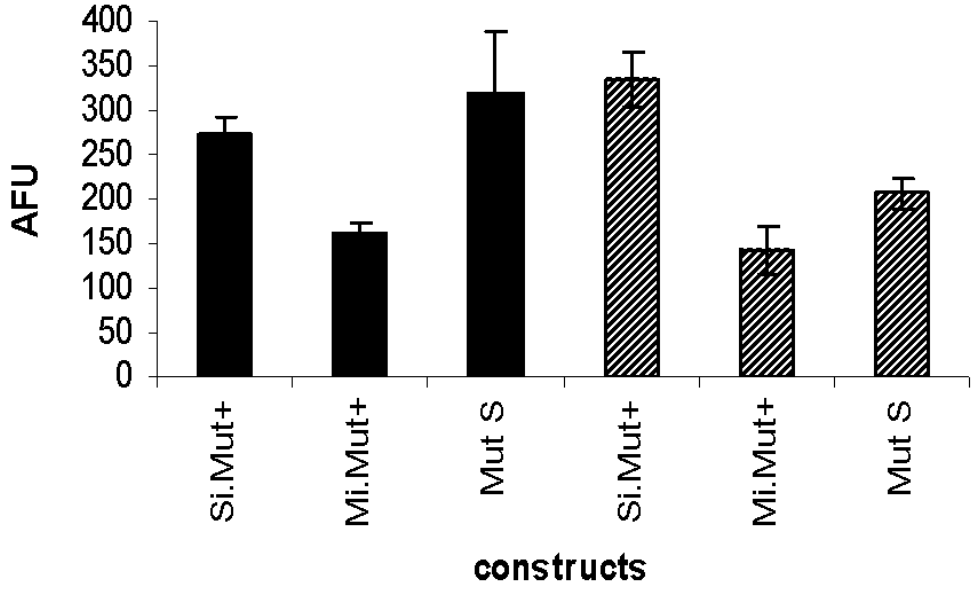


Figure 3.16 Comparison of enzyme activity levels in 40x concentrated culture supernatants for all full length construct strains. Strains sampled at 24 hr intervals for 0-120 hr post induction. Highest activity levels are represented. Activity was assayed in terms of arbitrary fluorescence units (AFU) and measured relative to strains transformed with empty vector alone. Filled columns represent readings from strains transformed by pPIC9K.FL.WT.*NAGLU* (Construct 1). Hatched columns represent readings from strains transformed by pPIC9K.FL.(CO+WT*).NAGLU* (Construct 2). For each, $SiMut^{+}$ values are from single integrant methanol utilization plus strains, $MiMut^{+}$ from multiple integrant Mut^{+} strains, and Mut^{S} from methanol utilization slow strains.

strains. Activity of singly integrated full length Construct 1 and Construct 2 Mut$^+$ strains grown at 15 °C as compared to 28 °C was also analysed to determine if degradation by proteases could be the cause of the low enzyme activity seen at 28 °C. Activity of strains grown at the lower temperature was repeatedly ≥ 2.5 times lower than that seen at 28 °C (Figure 3.17). In case enzymatically active NAGLU was being produced, but was being sequestered within the cell rather than being secreted into the supernatant, cell lysates from the single integrant full length Construct 1 and 2 Mut$^+$ strains analysed by immunoblot and silver stain procedures described above, were also tested for enzyme activity. Highest expression levels (for both pellet and supernatant samples) were seen at T = 120 hr for Construct 1 and T = 96 hr for Construct 2. Samples at these time points were assayed in triplicate for both pellets and supernatants as shown in Figure 3.18. Although low levels of enzyme activity were detected from the pellet samples, these levels were considerably lower than those seen for the corresponding supernatant samples (Figure 3.18).

3.3.4 RT-PCR analysis

To confirm that the transformed pGlycoSwitchM5 strains grown at 28 °C were transcribing the integrated DNA, reverse transcription PCR (RT-PCR) was performed. Construct 1 and 2 cDNA was amplified using a primer set common to their 3' region (Table 2.2). The 3' *AOX1 (*R) primer lies 3' of the insert DNA within the pPIC9K vector sequence. The *NAGLU* specific forward primer (WT/(CO+WT).1F) lies within the non codon optimised 3' region of the gene. Amplification with these primers of the very 3' end of the transcript indicates complete transcription of the inserted DNA. Figure 3.19a shows the expected 440 bp product amplifying in both the Construct 1 and Construct 2 (+) superscript RT lanes and absent from the (-) superscript RT lanes. Again, as expected there is a clean band of the correct size

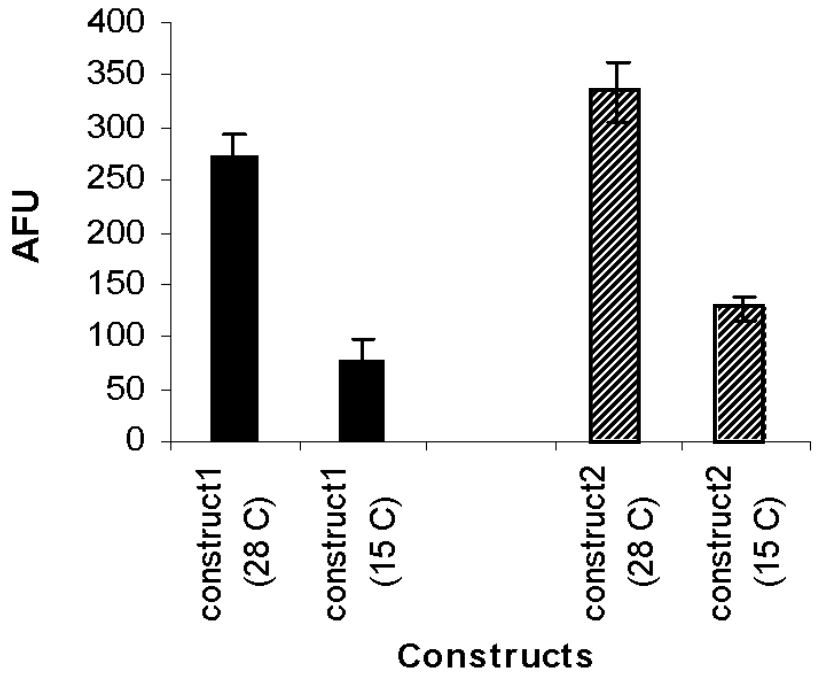


Figure 3.17 Comparison of enzyme activity levels of Construct 1 and Construct 2 single integrant Mut$^+$ strains grown at 28 °C and 15 °C. Culture supernatants sampled at 24 hr intervals for 0-120 hr post induction. Highest activity levels of 40x concentrated samples are represented. Activity was assayed in terms of arbitrary fluorescence units (AFU) and measured relative to strains transformed with empty vector alone. Filled columns represent readings from strains transformed by pPIC9K.FL.WT.*NAGLU* (Construct 1). Hatched columns represent readings from strains transformed by pPIC9K.FL.(CO+WT). *NAGLU* (Construct 2). For each, values from cultures grown at 28 °C precede those grown at 15 °C.

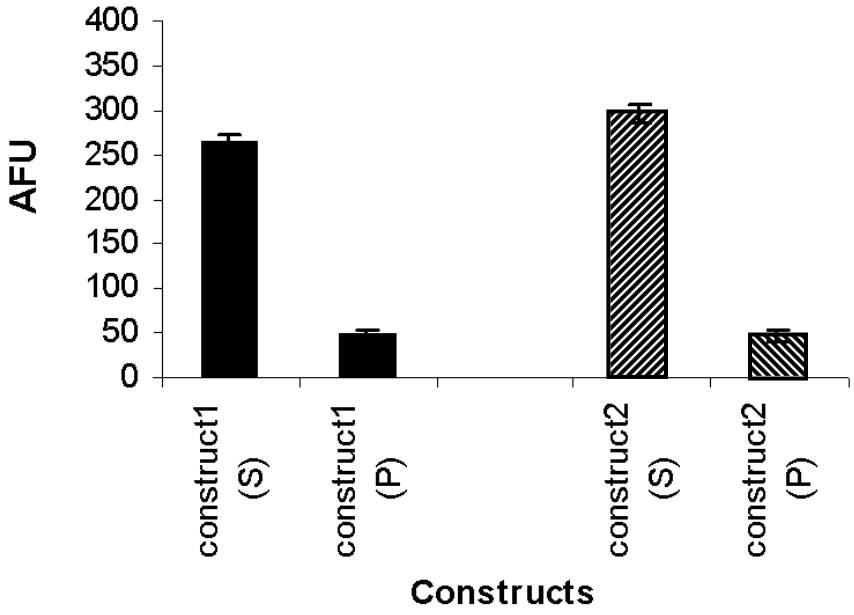


Figure 3.18 Comparison of enzyme activity levels of concentrated supernatant samples and pellet lysate samples roughly normalised for cell number from induced single integrant full length construct Mut$^+$ strains containing Construct 1 and Construct 2. Illustrated is the time point giving the highest reading for each strain, assayed in triplicate. Activity was assayed in terms of arbitrary fluorescence units (AFU) and measured relative to strains transformed with empty vector alone. Filled columns represent readings from strains transformed by pPIC9K.FL.WT.*NAGLU* (Construct 1). Hatched columns represent readings from strains transformed by pPIC9K.FL.(CO+WT*). NAGLU* (Construct 2). For each, readings from supernatants precede those from pellets.

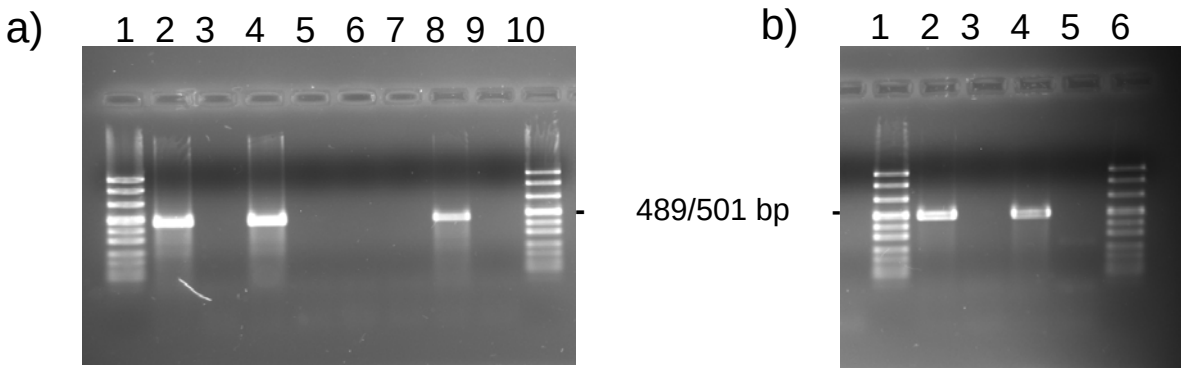


Figure 3.19 Agarose gel electrophoresis of RT-PCR amplified products from RNA extracted from induced pGlycoSwitchM5 strains transformed with Construct 1 (pPIC9K.FL.WT.*NAGLU*), Construct 2 (pPIC9K.FL.(CO+WT).*NAGLU*) and pPIC9K vector only control, all grown at 28 ºC. a) PCR amplification using WT/(CO+WT).1F and 3' *AOX1* (R) primers generates a 440 bp product. Lanes 2 and 3 contain Construct 1, lanes 4 and 5 contain Construct 2, lanes 6 and 7 contain empty vector; each with and without superscript RT, respectively. Lane 8, contains plasmid DNA as a positive control. Lane 9 contains a H_2O only negative PCR control. **b)** PCR amplification using 5' *AOX1* (F) and 3' *AOX1* (R) primers generates a 492 bp product. Lanes 2 and 3 contain empty vector; with and without superscript RT respectively. Lane 4 contains vector only plasmid DNA. Lane 5 contains a H_2O only negative PCR control. The pUC Mix Marker 8 is present in the external lanes of each gel to allow size determination of the amplicons.

from the positive plasmid DNA control and there is no amplification from the vector only templates. To confirm the absence of amplification in the vector only samples is due to specificity of Construct 1/Construct 2 primers rather than the absence of sample, amplification was performed on these same samples using 5' *AOX1* (F) and 3' *AOX1* (R) primers (Table 2.2). Short 1 min PCR extension times optimised conditions for the amplification of the smaller vector cDNA product rather than the endogenous 2.2 kb endogenous *AOX1* product. The expected 492 bp product is seen in the plasmid DNA lane and the vector (+) superscript RT lane and is absent from the (-) superscript RT lane (Figure 3.19b). For all PCRs, bands are absent from the water only negative control lanes. Taken together, these results indicate Constructs 1 and 2 are being successfully transcribed.

The above analysis was repeated on transformed pGlycoSwitchM5 strains grown at 15 °C (Figure 3.20). Again, specific amplification generating correct sized product was seen for all (+) superscript and plasmid DNA samples and was absent from all (-) superscript RT samples and water only negative controls. This indicates that Constructs 1 and 2 are also being successfully transcribed at 15 °C.

3.4 Construction of Insect Expression Vectors

3.4.1 Construction of expression vectors for stable expression

3.4.1a p2ZOp2.THF.NAGLU and p2ZOp2.THF.NTAT

Creation of the HindIII/EcoRI flanked THF fragment was achieved by PCR amplification from the p2ZOpTCXFegfp construct (Vaags, 2004). The 126 bp amplicon was T/A ligated into pGEM-T, sequenced to confirm the presence of mutation free THF inserts, and HindIII/EcoRI digested to allow directional ligation into each HindIII/EcoRI digested p2ZOp2-

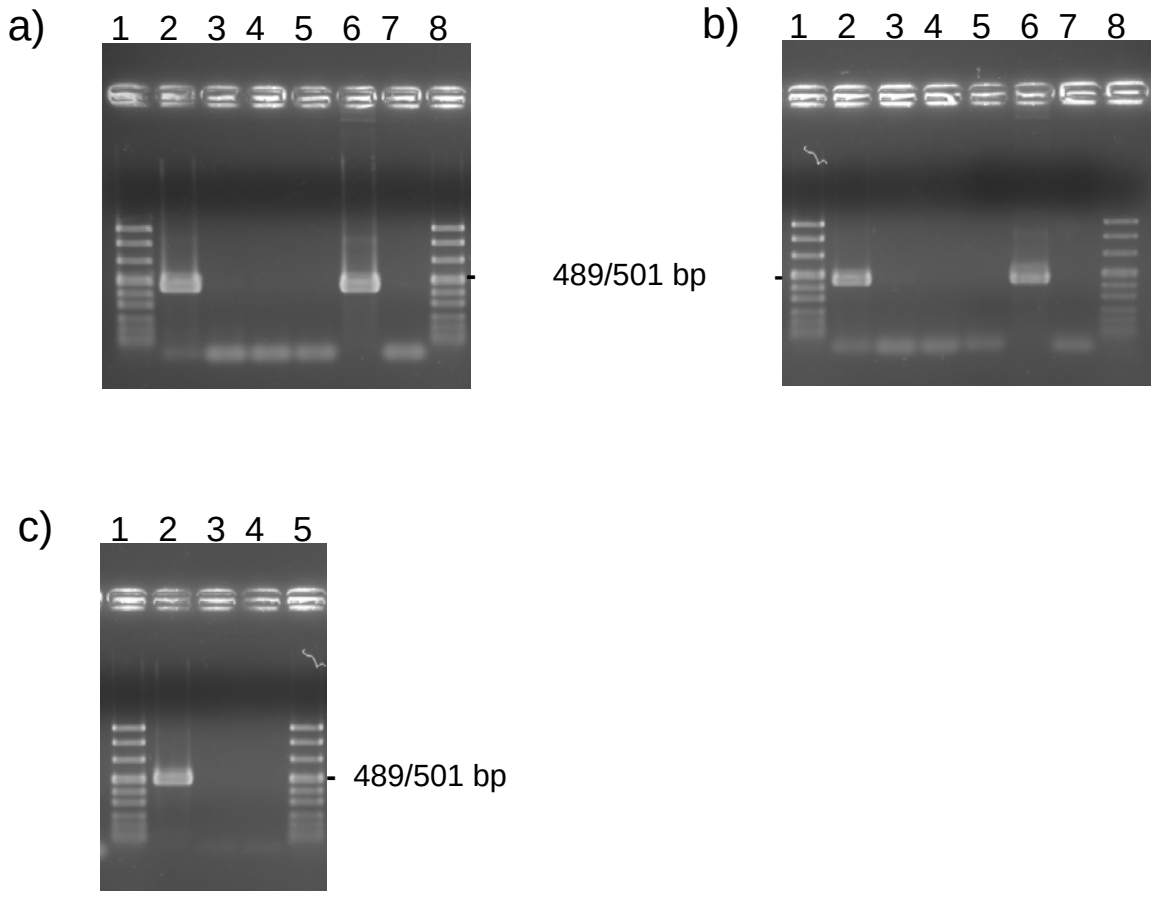


Figure 3.20 Agarose gel electrophoresis of RT-PCR amplified products from RNA extracted from induced pGlycoSwitchM5 strains transformed with Construct 1 (pPIC9K.FL.WT.*NAGLU*), Construct 2 (pPIC9K.FL.(CO+WT).*NAGLU*) and pPIC9K vector only control, all grown at 15 ºC. a) PCR amplification using WT/(CO+WT).1F and 3' *AOX1*(R) primers generates a 440 bp product. Lanes 2 and 3 contain Construct 1; lanes 4 and 5 contain empty vector; each with and without superscript RT, respectively. Lane 6 contains plasmid DNA as a positive control. Lane 7 contains a H_2O only negative PCR control. **b)** PCR amplification using WT/(CO+WT).1F and 3' *AOX1* (R) primers generates a 440 bp product. Lanes and templates are identical to those described in a) with the exception of lanes 2 and 3 which instead contain Construct 2 with and without superscript RT respectively **c)** PCR amplification using 5' *AOX1* (F) and 3' *AOX1* (R) primers generates a 492 bp product. Lanes 2 and 3 contain empty vector; with and without superscript RT respectively. Lane 4 contains a H_2O only negative PCR control. The pUC Mix Marker 8 is present in the external lanes of each gel to allow size determination of the amplicons.

NAGLU and p2ZOp2*NTAT*. Transformed, zeocin resistant colonies were screened by PCR, and plasmids isolated from colonies generating the correct sized product. Upon confirmation of mutation free ORFs, a bacterial stab was prepared for each p2ZOp2.THF.*NAGLU* and p2ZOp2.THF.*NTAT* clone and sent to our collaborator Dr. Pfeifer (University of British Columbia, BC), for plasmid isolation and subsequent stable transformation of *Sf9* and *Tni* cells. The initial PCR amplification of the THF fragment is illustrated in Figure 3.21a, HindIII/EcoRI digested and undigested pGEM-T(THF) constructs are shown in Figure 3.21b and HindIII/EcoRI digested p2ZOp2.THF.*NAGLU* and p2ZOp2.THF.*NTAT* are depicted in Figure 3.21c.

3.4.2 Construction of expression vectors for baculovirus expression

3.4.2a pAcGP67B.NAGLU.H_6 and pAcGP67B.NTAT.H_6

Full length mature *NAGLU* cDNA with and without a carboxyl fused TAT protein transduction domain followed by an inframe hexahistidine tag was obtained by respective PCR amplification of sequence verified pPICZα(*NTAT*) and pPICZα(*NAGLU*) (Patrick, 2006). An agarose gel showing the PCR amplification of each BamHI/EcoRI flanked *NTAT*.H_6 and *NAGLU*.H_6 is illustrated in Figure 3.22.

The baculovirus expression vector (pAcGP67B) and each PCR amplicon described above were concomitantly digested with BamHI and EcoRI. Directional ligation of each *NTAT*.H_6 and *NAGLU*.H_6 into pAcGP67B was performed and, following transformation, ampicillin resistant colonies were screened by PCR and plasmids isolated from colonies generating the correct sized product. Sequence analysis of their entire ORFs confirmed the

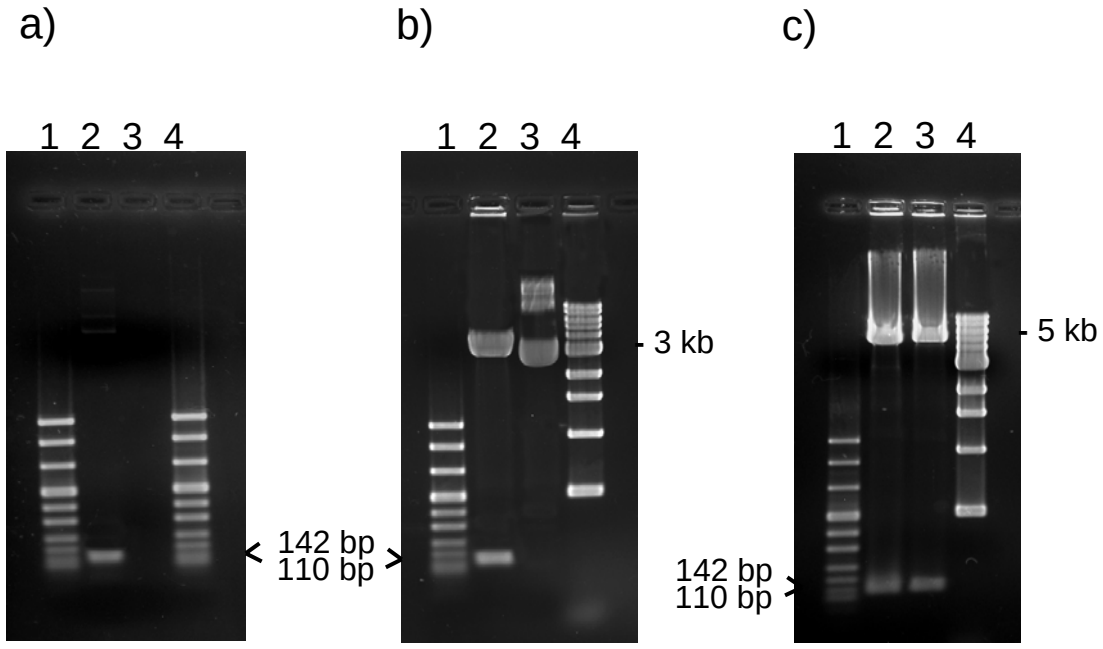


Figure 3.21 Agarose gels showing the various steps in production of the p2ZOp2.THF.*NAGLU* and p2ZOp2.THF.*NTAT* insect expression vectors designed for stable expression in *Sf9* and *Tni* cells. a) Initial PCR amplification of the 126 bp THF fragment is shown in lane 2. Lane 3 contains a H_2O only negative control. **b)** Lane 2 contains HindIII/EcoRI digested pGEM-T(THF) and lane 3 contains the undigested pGEM-T(THF) construct **c)** Lanes 2 and 3 contain HindIII/EcoRI digested p2ZOp2.THF.*NAGLU* and p2ZOp2.THF.*NTAT,* respectively. To allow size determination of the amplicons, the pUC Mix Marker 8 is present in both external lanes of gel a), and the left hand external lane of gels b) and c). A 1 kb DNA standard is present in the right hand external lane of gels b) and c).

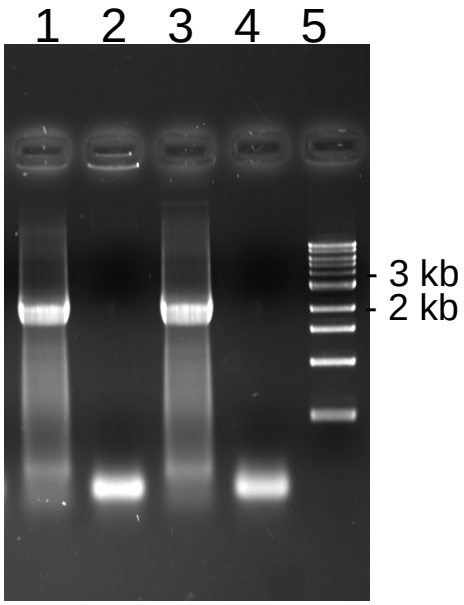


Figure 3.22 Agarose gels showing PCR amplification of *NAGLU*.H_6 and *NTAT*.H_6 for expression in the baculovirus transfer vector pAcGP67B. Lanes 1 and 3 contain *NAGLU* and *NTAT* amplicons, respectively. Lanes 2 and 4 contain H_2O only negative controls for each positive PCR reaction. A 1 kb DNA standard is present in the right hand external lane.

absence of mutations in both constructs. A schematic representation of each of the pAcGP67B.*NAGLU*.H_6 and pAcGP67B.*NTAT*.H_6 final constructs is shown in Figure 2.4.

3.4.2b pAcGP67B. H_6.L.Fx.NTAT

Complimentary forward and reverse oligos corresponding to a linker sequence adapted from the pIEx-10 vector, which includes a Factor Xa cleavage site, were annealed to generate a doubled stranded linker DNA template. Amplification of the H_6.L.Fx.*NTAT* fusion gene was created in three PCR steps. First, a 121 bp product containing a 5' hexahistidine tag followed by the linker sequence containing a Factor Xa cleavage site, with a 3' sequence homologous to the first 15 bp of NTAT was amplified from the template above to generate H_6.L.Fx. Second, *NTAT* with a 5' sequence homologous to the final 17 bp of the linker was amplified to generate an ~2.2 kb product (*NTAT*). The presence of the identical linker/*NTAT* 32 bp on both H_6.L.Fx and *NTAT* templates allowed the two amplicons to anneal via this complementary sequence and serve as a single template for overlap extension PCR amplification. This third, overlap extension PCR produced an ~2.3 kb product encoding H_6.L.Fx.*NTAT*. Figure 3.23 illustrates these three amplification steps.

The final PCR product containing a 5' BamHI site and a 3' EcoRI site was T/A ligated into pGEM-T and confirmed as mutation free prior to digestion and directional ligation into BamHI/EcoRI digested pAcGP67B. A pAcGP67B.H_6.L.Fx.*NTAT* clone with a mutation free ORF was identified by sequence analysis. Figure 3.24 shows digests to excise the inserts from all three baculovirus constructs. Bands within the gel correspond to the pAcGP67B transfer vector (9.76 kb) with an ~2.2 kb *NAGLU*.H_6 insert, an ~2.2 kb *NTAT*.H_6 insert and an ~2.3 kb H_6.L.Fx.*NTAT* insert.

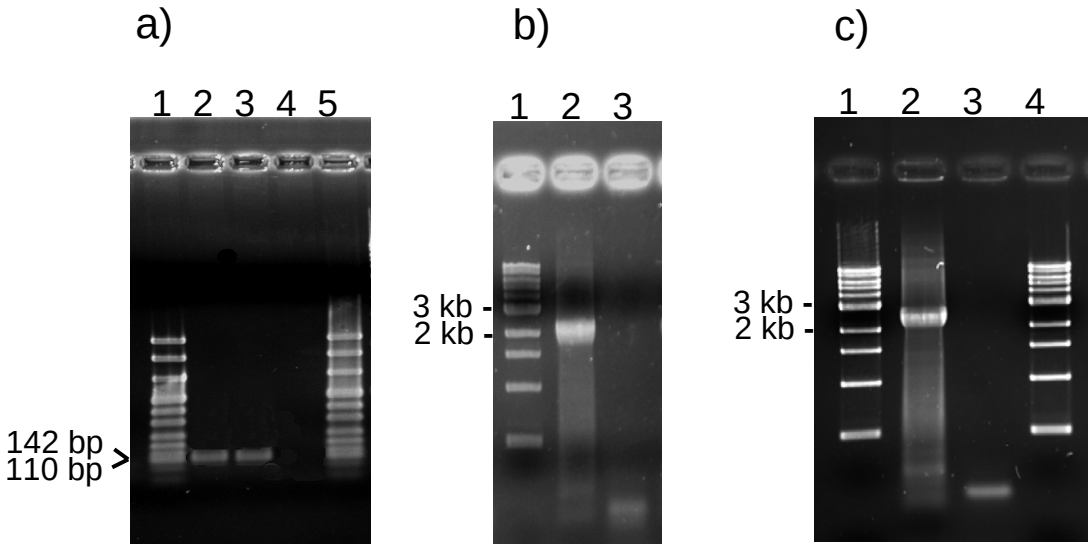


Figure 3.23 Agarose gels showing the three amplification steps for the production of the H_6.L.Fx.*NTAT* sequence designed for expression in the pAcGP67B baculovirus expression vector. a) Initial PCR amplification of the 121 bp 5' His tagged, linker, factor Xa fragment is shown in lanes 2 and 3. Lane 4 contains a H_2O only negative PCR control. **b)** Initial PCR amplification of the approximately 2.2 kb 3' NTAT fragment is shown in lane 2. The H_2O only negative PCR control is seen in lane 3. **c)** Overlap extension PCR to generate H_6.L.Fx.*NTAT* is shown in lane 2. The H_2O only negative PCR control is seen in lane 3. To allow size determination of the amplicons, the pUC Mix Marker 8 is present in both external lanes of gel a). A 1 kb DNA standard is present in the left most external lane of gel b) and both external lanes of gel c).

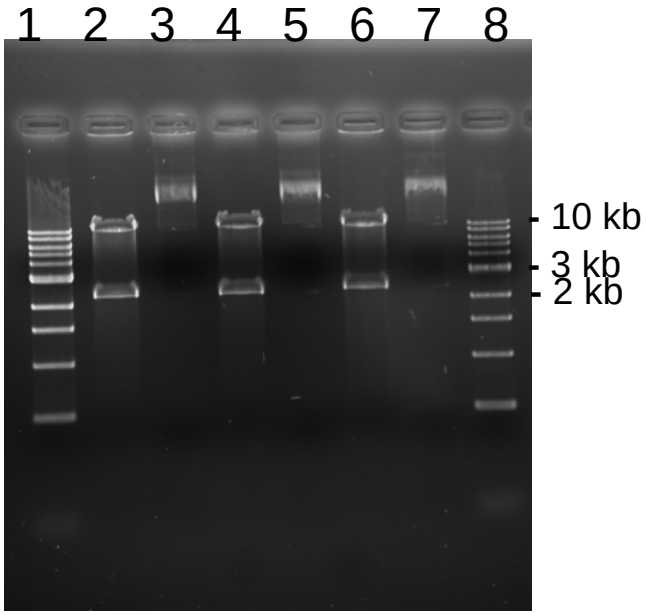


Figure 3.24 Agarose gel showing BamHI/EcoRI restriction digests of the final baculovirus constructs illustrating correct insert sizes. Lanes 2 and 3 contain digested and undigested pAcGP67B.*NAGLU*.H_6, respectively. Lanes 4 and 5 contain digested and undigested pAcGP67B.*NTAT*.H_6, respectively. Lanes 6 and 7 contain digested and undigested pAcGP67B.H_6.L.Fx.*NTAT*, respectively. Lanes 1 and 8 each contain a 1 kb DNA standard to allow size determination of the inserts.

3.5 NAGLU and NTAT Expression in Insect Cells

3.5.1 Protein expression in stably transformed Sf9 and Tni cells

Co-transfection of the p2ZOp2F(SEAP) expression vector with the p2ZOp2.THF.*NAGLU* or p2ZOp2.THF.*NTAT* allowed transfection success to be determined by assaying for SEAP. Positive results were obtained for all cultures prior to performance of immunoblot and activity assay (data not shown). With the aim of achieving high expression levels of the NAGLU and NTAT proteins, the standard transfection conditions were varied. Keeping constant the amount of transfected DNA (1 μg), amounts of transfection reagent were varied between 5 μg, 10 μg and 15 μg. Transfections under these conditions were performed in both *Sf9* and *Tni* cells. Keeping constant the amount of transfection reagent (10 μg), amounts of DNA were varied from 0.75 μg, 1.0 μg, 1.25 μg, again in both *Tni* and *Sf9* cells.

Crude supernatant and lysed pellet samples from all conditions were analysed by immunoblot and activity assay to identify the highest expressing transfectants. Immunoblot analysis using the anti-His_4 antibody failed to detect expression from all samples (data not included). Inclusion of a His tagged control construct on each gel served to confirm that the transfer and the anti-His_4 detection systems were both functioning successfully. Activity levels were measured relative to strains transformed with the p2ZOp2F(SEAP) reporter vector alone and reported in AFU. Positive activity readings were obtained, but levels of activity (AFU) were low or negligible in all cases (data not included).

3.5.2 Protein expression in baculovirus infected High Five™ cells

Recombination, within the transfected *Sf9* cells, of each of the three baculovirus transfer vector constructs with the BD BaculoGold™Bright DNA to successfully create primary virus, was demonstrated by visualization of GFP by fluorescence microscopy (Figure 3.25). Addition of High Five™ cells to each of the virus containing wells followed by 2-4 day incubation allowed the virus to infect the newly added cells. Prior to cell lysis, the transduction supernatants containing the secreted heterologous proteins were collected for analysis of recombinant protein expression.

Approximately 2 ml of each transduction supernatant were separately bound to a Ni-NTA affinity column and competitively eluted by application of elution buffer containing 250 mM imidazole. When resolved on a 10% (w/v) SDS-PAGE followed by coomassie staining there was no visible band specific to the construct samples when compared to cells alone. Immunoblot using the anti-NAGLU antibody also failed to detect a NAGLU specific band, indicating the Ni-NTA pull down was unsuccessful. The activity of the *NAGLU*.H_6, *NTAT*.H_6, H_6.L.Fx.*NTAT* transduction supernatants assayed prior to Ni-NTA purification was consistent with those of the unbound Ni-NTA column protein flow through, further indicating all three His tagged constructs failed to bind the Ni-NTA column. Activity levels of unbound Ni-NTA column protein flow through for the *NAGLU*.H_6, *NTAT*.H_6 and H_6.L.Fx.*NTAT* constructs and two controls (cells only and a non NAGLU construct also expressed in the pAcGP67B transfer vector) are shown (Figure 3.26). AFU levels range from 430-630 AFU for experimental constructs as compared to ~4 AFU for controls. A Bradford assay was performed to assess total protein concentration and thus allow calculation of the specific activity (U/μg) in these crude samples. Specific activities were similar for each construct with values of 20, 16 and 17 U/mg for *NAGLU*.H_6, *NTAT*.H_6 and H_6.L.Fx.*NTAT* constructs, respectively. Immunoblots of

the *NAGLU*.H_6, *NTAT*.H_6 and H_6.L.Fx.*NTAT* constructs probed with the anti-NAGLU primary antibody show an approximately 80-85 kDa band for each construct which correlates with the expected size band for the NAGLU protein. A blot of duplicate samples probed with the anti-His primary antibody failed to detect bands for any of the constructs (Figure 3.27).

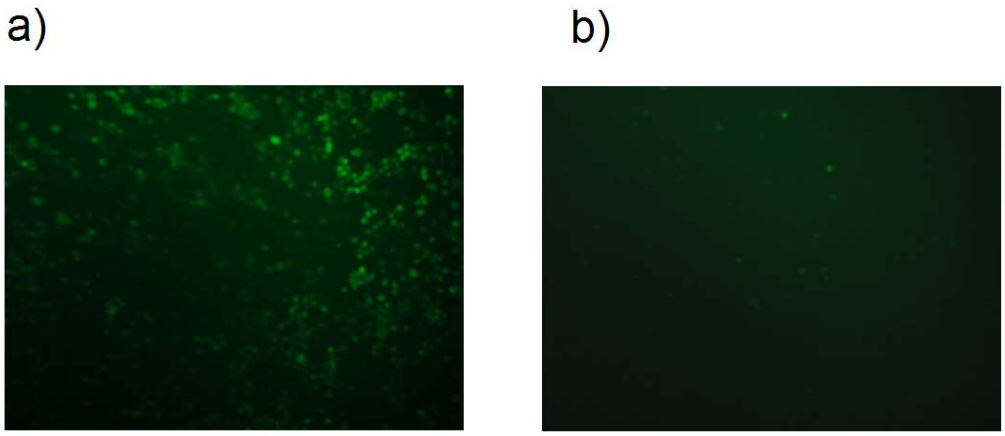


Figure 3.25 Visualization of GFP in co-transfected (transfer vector + baculovirus DNA) and singly transfected (baculovirus only) *Sf9* cells by fluorescence microscopy. a) The green fluorescent signal apparent throughout the well indicates successful recombination of baculovirus transfer vector constructs with the BD BaculoGold™Bright DNA to create GFP expressing primary virus. **b)** Sparse GFP signal indicates BD BaculoGold™Bright DNA alone is unable to produce primary virus.

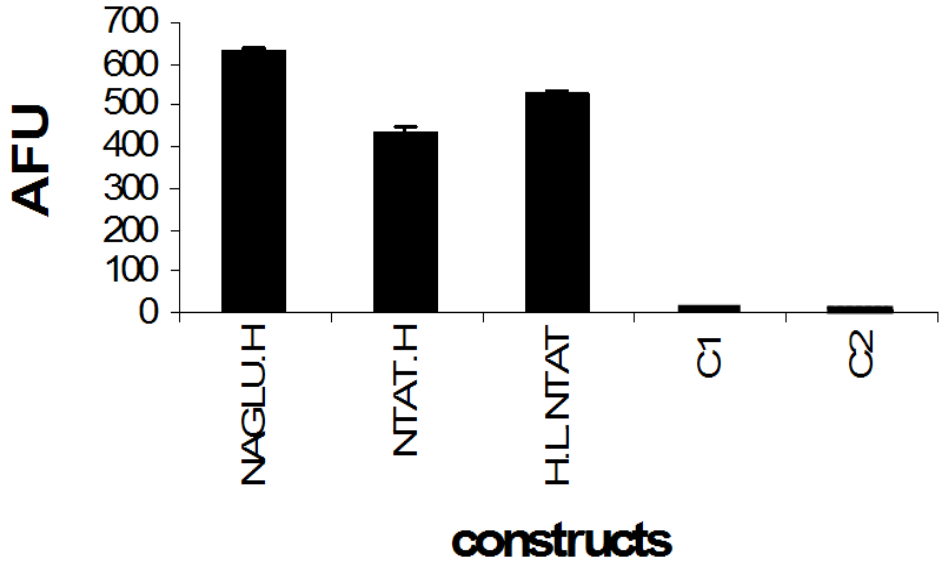


Figure 3.26 Comparison of enzyme activity levels in crude transduction supernatants from primary virus infected High Five™ cells. Activity was assayed in triplicate in terms of arbitrary fluorescence units (AFU). NAGLU.H is cells infected with the *NAGLU*.H_6 construct; NTAT.H is cells infected with the *NTAT*.H_6 construct; H.L NTAT is cells infected with the H_6.L.Fx.*NTAT* construct; C1 is High Five™ cells alone and C2 is cells infected with a non-*NAGLU* construct.

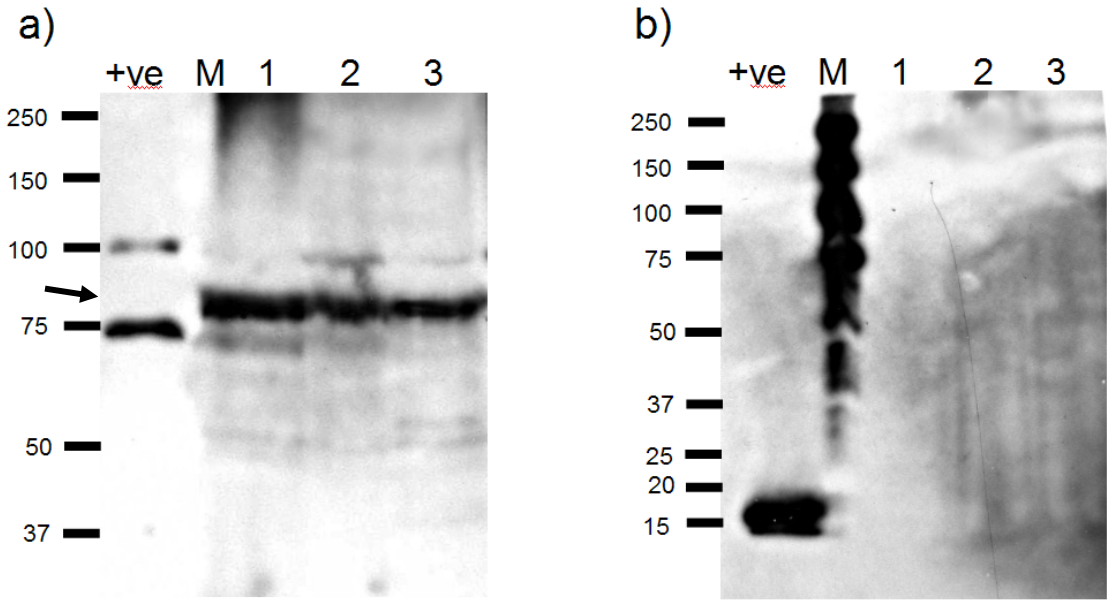


Figure 3.27 SDS-PAGE immunoblots of crude transduction supernatant samples from primary virus infected High Five™ cells. Duplicate immunoblots of lanes 1-3 containing *NAGLU*.H_6, *NTAT*.H_6 and H_6.L.Fx.*NTAT,* respectively were probed with **a)** the anti-NAGLU antibody and **b)** the anti-His antibody. A positive control for each antibody was included in the far left lane of each blot (*Sf9* produced CBD-NTAT and *P. pastoris* produced hexahistidine tagged H-protein, respectively). The expected 80-85 kDa product is indicated by an arrow. Precision Plus dual colour marker (M) was loaded on each gel for size comparison.

4 Discussion

The aim of this thesis was to express recombinant human α-*N*-acetylglucosaminidase in a heterologous gene expression system, ultimately for use in enzyme replacement therapy for treatment of patients with MPS IIIB. Identification of a simple eukaryotic system able to produce high levels of stable, functional enzyme is the first step in this process. To this end, constructs for NAGLU expression in the methylotrophic yeast *P. pastoris*, and NAGLU and NTAT expression in both stably transfected and baculovirus infected *Sf9* and *Tni* cell lines were produced, and expression levels assessed.

4.1 Design of Yeast Expression Constructs

The coding region of endogenous *NAGLU* consists of 743 $\alpha\alpha$s with a signal-peptidase cleavage consensus site at position 23 (Weber *et al.*, 1996). The 23 $\alpha\alpha$ signal peptide sequence at the amino terminal of the protein is cleaved *in vivo* to generate the 720 $\alpha\alpha$ sequence that give rise to the mature protein (Zhao *et al.*, 1996). Therefore, the 2160 bp DNA product that encodes only the mature lysosomal protein without its targeting sequence was cloned into the *P. pastoris* vector pPIC9K (Construct 1). A 261 bp codon optimised 5' fragment was PCR amplified, restriction digested and ligated into restriction digested pPIC9K*FL.WT.NAGLU* to generate pPIC9K.FL.(CO+WT).*NAGLU* (Construct 2) to assess whether codon optimisation of a portion of the gene would affect the level of recombinant protein expression (see section 4.2.2).

Previous NAGLU expression studies conducted in our laboratory using *P. pastoris* predominantly focused on expression of NTAT proteins, NAGLU constructs with carboxyl terminal PTD domains, with and without an additional hexahistidine tag at the amino end of the protein. As expression levels of these previously studied proteins were found to be low, full

length constructs for this study (pPIC9K.FL.WT.*NAGLU* and pPIC9K.FL.(CO+WT).*NAGLU*) were initially designed without additional tags (i.e., without a hexahistidine tag for purification or PTD tag for cellular uptake) in case these domains were leading to instability of protein. Had we obtained significant levels of recombinant expression without these domains, attempts would then have been made to sequentially include the tags and repeat the expression studies with these tagged constructs.

Two control hexahistidine tagged constructs, pPIC9K*HFxCO.F* (Construct 3) and pPIC9K*HFxWT.F* (Construct 4), were also created. Control constructs were designed to assess whether any difference in expression would be seen between a completely codon optimised and a non optimised *NAGLU* gene fragment. Polyhistidine tags were included on these constructs as a means of verifying protein expression via immunodetection.

All Constructs 1-4, with sequence verified mutation free ORFs, were successfully transformed into the *P. pastoris* strain pGlycoSwitchM5 for expression and secretion of the resultant recombinant proteins. A variety of different strains for both codon optimised and wildtype NAGLU were produced in an attempt to increase levels of protein expression.

4.2 Mechanisms for Increased Expression of Recombinant NAGLU in *P. pastoris*

4.2.1 Selection of a glycoengineered humanized strain of P. pastoris

A variety of proteins that could not be expressed in prokaryotic organisms due to a lack of correct posttranslational maturation have subsequently been produced in the methylotrophic yeast, *P. pastoris* (Daly and Hearn, 2005). This is presumably due to the ability of *P. pastoris* to perform posttranslational modifications such as *O*- and *N*-linked glycosylation, which allow correct protein folding and thus confer stability on the heterologous protein. Although it is unknown whether specific glycosylation is necessary for NAGLU activity, the presence of six

putative glycosylation sites within the protein indicates that this modification may be essential (Zhao *et al.*, 1996).

Glycoengineered strains of *P. pastoris*, such as pGlycoSwitchM5, have additional advantages over the wildtype *P. pastoris* strains. Unlike mammalian cells which produce an inherently heterogeneous mixture of glycoforms, humanized *P. pastoris* strains can produce proteins with homogeneous glycosylation patterns. The advantages of producing recombinant proteins with purely mannosylated *N*-glycans has already been illustrated by the production of a modified recombinant form of glucocerebrosidase (GBA), the enzyme deficient in Gaucher disease (Furbish *et al.*, 1981). This modified protein has proven to be more successful in ERT than the endogenously glycosylated protein.

Glycosylation plays an essential role in the function of many proteins intended for therapeutic use in humans. The humanized *P. pastoris* strain utilized in this study, pGlycoSwitchM5, has both an inactivated α-1,6-mannosyltransferase Och1p (*OCH1*) gene to eliminate hypermannosylation, and an overexpressed α-1,2-mannosidase to trim mannose residues from eight to a total of five (Vervecken *et al.*, 2004). The putative advantages of controlling glycosylation by using pGlycoSwitchM5 to produce a homogeneous Man_5GlnAc_2 high mannose form of recombinant NAGLU are twofold. Firstly, the more "mammalian" glycosylation pattern will presumably aid the correct localization of this recombinant protein and prevent its degradation and clearance from the blood. Secondly, the high mannose glycosylation pattern may enable more efficient mannose 6-phosphorylation of these residues, circumventing the uptake problems experienced by researchers expressing recombinant NAGLU from CHO cells, where the recombinant proteins contain a variety of oligosaccharide chains. As demonstrated by the modified GBA enzyme, the ability to produce a homogeneously glycosylated NAGLU recombinant could result in an improvement of

therapeutic function and concurrently enable the possible elimination of undesirable functions. Unfortunately in the case of recombinant NAGLU produced in pGlycoSwitchM5, these advantages failed to be realized due to negligible levels of protein expression.

4.2.2 Codon optimisation of recombinant NAGLU

Many genetic and physiological factors determine the productivity of a recombinant system. One potential bottleneck is codon usage of the expressed gene. It has been well documented that different organisms show a preference for different subpopulations of the 61 possible codons, with the genes coding for abundant proteins using these major codons almost exclusively throughout their sequences. This concept of codon bias, where the codons utilized to encode a protein in one organism are substantially different from those used to code for the same protein in a different species, has long been recognized to have a profound impact on the expression of heterologous proteins (Grantham *et al.*, 1980; Gouy and Gautier, 1982). Previous studies undertaken in this laboratory focused on the expression of full length wildtype *NTAT* in *P. pastoris*. Results indicated that although the gene was correctly inserted without mutation and full length cDNA transcripts were successfully PCR amplified, protein levels were negligible (Patrick, 2006).

To evaluate the possible benefit of codon optimisation for increasing NAGLU expression, we initially designed a codon optimised synthetic fragment which encodes only the N-terminal 87 amino acids of the protein. Entering the non codon optimised sequence into the Graphical Codon Usage Analyser program (http://gcua.schoedl.de/) indicated that 18/87 amino acid residues in this N-terminal fragment are encoded by codons that are used at a frequency of 15% or less in *P pastoris*. The occurrence of infrequently used codons in the remainder of the

gene is much lower with only 26/634 such codons identified, with these codons distributed relatively evenly throughout the gene.

As has been noted for codon optimisation of many human genes in *P. pastoris,* optimisation of the initial 261 bp of *NAGLU* resulted in an accompanying decrease in the overall G+C content in this portion of the gene. Sinclair and Choy (2002) have previously reported that when G+C content was optimised independently of codon usage, increased levels of protein expression were also observed, leading to the hypothesis that these two effects may be additive. The optimised synthetic fragment has 45.59% G+C content as compared to the native sequence which had 78.93% G+C content. This reduction in G+C content accompanying codon optimisation was much more pronounced for the initial 261 bases of the gene. Optimising the sequence for the remaining 1905 bp would have minimally affected the G+C content of this portion, decreasing it from the original 61.99% to an optimised 60.31%. Additionally, Hoekema *et al.* (1987) showed that replacing an increasing number of major codons with synonymous minor ones just at the 5' end of the coding sequence of the PGK1 gene caused a dramatic decline in the level of protein expression. Extrapolating from that, we may have expected to see an increase in protein level by focusing on just the 5' end of the coding sequence of *NAGLU*.

These observations, in addition to cost considerations, led us to codon optimise just the 5' portion of the gene. Had this partial codon optimisation led to an increase in expression level, optimisation of the entire gene would have been considered. However, constructs containing the codon optimised portion (full length construct pPIC9K.FL.(CO+WT).*NAGLU* (Construct 2) and control hexahistidine tagged construct *HFxCO.F* (Construct 3)) failed to generate levels above those seen for wildtype.

A recent study describing the expression of 79 different human cDNAs in *P. pastoris* failed to detect a correlation between codon usage or G+C content on protein yield (Boettner *et al.*, 2007). In most studies where codon optimisation is reported to increase protein expression, detectable levels of expression from unconcentrated samples have already been achieved by the wildtype recombinant protein (Sinclair and Choy, 2002; Yadava and Ockenhouse, 2003; Hu *et al.*, 2006; Xiong *et al.*, 2006). As NAGLU could not be detected from crude supernatant by silver stain, immunoblot or activity assay, it is perhaps not surprising that codon optimisation did not deliver the desired increase in expression levels.

4.2.3 Comparison of NAGLU expression in Mut^S versus Mut$^+$ Strains

P. pastoris strains can exhibit different phenotypes with regard to methanol utilization. The majority of heterologous protein expression currently performed uses the wildtype methanol utilization plus phenotype (Mut$^+$) because of the rapid growth of this strain and its ability to produce large quantities of foreign protein when methanol is the sole carbon and energy source (Macauley-Patrick *et al.*, 2005). However, strains with a slow methanol utilization phenotype (Mut^S), which do not grow rapidly in the presence of methanol, may increase the stability of recombinant proteins that are difficult to synthesize, slow to fold, or must undergo other complex posttranslational modifications (Daly and Hearn, 2005). As NAGLU is a large protein of approximately 80 kDa, contains 6 putative *N*-linked glycosylation sites and must undergo phosphorylation to allow MPR-mediated uptake into cells, it appeared that this protein may benefit from expression under Mut^S, in addition to Mut$^+$ conditions.

Although transformation of pGlycoSwitchM5 with SacI linearized constructs favours recombination at the *HIS4* locus (generating His$^+$Mut$^+$ transformants), recombination at the *AOX1* locus (producing His$^+$Mut^S transformants) can also occur. Slow growth on methanol,

due to disruption of the *AOX1* gene, was used to distinguish Mut^S and Mut^+ transformants and allow selection of each strain. Observations of a difference in protein expression levels between strains exhibiting the two methanol utilization phenotypes could provide information as to the importance of posttranslational modification of recombinant NAGLU. However, even after 40x concentration, only minimal expression was detected for both the fast growing Mut^+ and slower growing Mut^S strains.

4.2.4 Multiple copy integrants

Both the copy number of the cassette and the site and mode of chromosomal integration of the expression cassette are factors that have been found to drastically influence protein production (Sreekrishna *et al.*, 1997). Single crossover events (insertions) in a *his4* strain such as pGlycoSwitchM5 are the most frequent means of recombination and led to His^+Mut^+ strains. Gene replacement, resulting from a rarer double crossover event with the endogenous *AOX1* gene, gives rise to a His^+Mut^S phenotype. Hence assaying expression from the Mut^+ and Mut^S strains described above, allows analysis of the affect of mode of chromosomal insertion. Influence of the copy number of the cassette was also investigated.

P. pastoris is capable of integrating multiple copies of transforming DNA into genomic sites of sequence homology. The exact mechanism by which this recombination occurs is not fully understood, however such events are recognized to occur quite frequently amongst His^+ selected transformants. Gene copy number has been identified as a rate limiting step in the production of recombinant proteins from *P. pastoris* (Clare *et al.*, 1991). Successful integration of multiple copies of the gene of interest into the *AOX1* locus may be necessary to generate sufficiently high levels of expression of a given recombinant gene.

The pPIC9K vector contains the kanamycin/neomycin gene which confers resistance to kanamycin in bacteria and geneticin in yeast. The presence of this gene allows the detection of multiple copy yeast integrants by selection on increasing concentrations of geneticin. Growth on 0.25 mg/ml, 0.5 mg/ml, 0.75 mg/ml and 1.0 mg/ml is postulated to correlate with 1, 2, 3, and 4 copies of the integrated vector, respectively. Growth on 0.5 mg/ml for a single strain containing Construct 1 indicated incorporation of two copies of the cassette and growth on 0.75 mg/ml for Construct 2 indicated a single strain had incorporated three copies of the cassette.

It has been demonstrated by a number of groups that multiple integration events have led to increased heterologous protein production in *P. pastoris* (Vassileva *et al.*, 2001; Macauley-Patrick *et al.*, 2005; Vad *et al.*, 2005). Therefore, determining the presence of strains that have multiple integrants is highly desirable. Unfortunately, neither of the multiple integrant strains tested expressed recombinant NAGLU at a level that could be detected in an activity assay. Even after 40x concentration, the levels of activity were negligible, and were even lower than those seen for single integrant strains. Other groups have also reported no increase in expression level with multiple integrants. Woo *et al*. (2002) found no difference in expression level between single and double copies of a scFv-fused immunotoxin. Hu *et al.* (2006) confirmed integration of greater than five copies of their anti-erbB2 single chain antibody by Southern blot, but did not see a corresponding increase in level of protein expression. Hohenblum *et al.* (2004) compared the *AOX1* and *GAP* promoters at different gene copy numbers. No effect on the expression level was observed with the constitutive *GAP* promoter, whereas a strong negative effect on expression was seen with the *AOX1* promoter with any increase in copy number above two (Hohenblum *et al.*, 2004). The authors hypothesized that the lower gene dosage limit seen for genes under the control of the *AOX1*

promoter may be accounted for by the increase strength of the *AOX1* as compared to the *GAP* promoter (Cereghino and Cregg, 2000).

It should be recognized that for our study, selection for increased antibiotic resistance does not conclusively indicate that colonies contain multiple copies of the vector (Cereghino and Cregg, 2000). Had an increase in expression level of NAGLU recombinant protein been detected in these strains showing increased geneticin resistance, the actual copy number could have been determined by quantitative dot blot analysis or Southern blot techniques.

4.2.5 Induction at 15 ℃

The optimum growth temperature for *P. pastoris* is usually defined as 28-30 °C, however it is well established that reduction in temperature can significantly improve recombinant protein productivity without hampering growth (Li *et al.*, 2001; Andre *et al.*, 2006; Gasser *et al.*, 2007; Toikkanen *et al.*, 2007). Toikkanen *et al.* (2007) reported an improvement in the yield and the quality of two recombinantly produced enzymes after decreasing the cultivation temperature from 30 °C to 22 °C. This is in agreement with the findings of Li *et al.* (2001), who showed that lowering the temperature from 30 °C to 23 °C approximately tripled the yield of their proteins and also increased cell viability. Two groups have recently reported increased expression levels of recombinant proteins upon reduction of the growth temperature to 20 °C (Andre *et al.*, 2006; Gasser *et al.*, 2007).

Lowering of the temperature in the induction phase of cell growth has been reported to lead to increased protein levels due to both an increase in level of recombinant protein expression and decrease in the degree of proteolysis (Daly and Hearn, 2005; Macauley-Patrick *et al.*, 2005). The slowing down of protein production allows more time for correct protein processing and has been hypothesized to upregulate cold-shock proteins such as chaperones,

which enable the correct folding and assembly of polypeptide chains. Although little is known about chaperone proteins in *P. pastoris*, it has been postulated that lowering the temperature in this yeast species may cause a similar increase in the level of chaperones as has been seen in *S. cerevisiae* (Andre *et al.*, 2006). Proteolytic degradation of heterologous proteins has long been recognized as problematic in recombinant expression systems. The known yeast proteases are expressed intracellularly and are thought to be released upon cell lysis due to stress in culture (Jahic *et al.*, 2003). The increased cell viability noted at lower temperatures could thereby reduce the rate of protease release from within the cells into the supernatant.

In a previous study undertaken in this laboratory, *P. pastoris* strains were induced at 25 °C to evaluate the effect on NAGLU expression (Patrick, 2006). No notable increase in expression was observed, however other groups have reported that a more extreme drop in temperature may be necessary to increase expression to detectable levels. The expression of human mu-opioid receptor fusion protein was examined at a number of different temperatures and the best level of expression was seen at 15-20 °C. Expression at these low temperatures gave a twofold increase in that obtained at 25 °C and a fourfold increase over that at 30 °C (Sarramegna *et al.*, 2002).

Based on these findings, we compared NAGLU expression levels in strains induced at 15 °C and 28 °C. Promisingly, the healthy cell growth exhibited at 28 °C was also seen for growth at the reduced temperature (Figures 3.8 and 3.9). However, NAGLU expression levels were not increased by reduction in temperature. A 40x concentration of the supernatant was necessary as activity levels in crude supernatant could not be detected (data not shown). As can be seen in Figure 3.17, enzyme activity levels for both Construct 1 (pPIC9K.FL.WT.*NAGLU*) and Construct 2 (pPIC9K.FL.(CO+WT).*NAGLU*) single integrant

Mut^+ strains show an approximately threefold reduction in NAGLU activity levels for growth at 15 °C as compared to growth at 28 °C.

4.3 Detection of Recombinant NAGLU Expressed in *P. pastoris*

As immunoblot analysis on crude supernatant samples from both single and multiple Mut^+ and Mut^S strains transformed by Construct 1, 2 and vector alone failed to elicit bands, immunoblot analysis on 40x concentrated samples was undertaken. Samples concentrated from culture medium taken every 24 hr post-induction over a time course of 120 hr were assessed. Immunoblot analysis detected a faint band of approximately 60 kDa which appeared in some time point samples for most strains. To determine whether the appearance of this band correlated with a greater concentration of total protein loaded, samples were quantitated and an equal concentration loaded for analysis by immunoblot (Figure 3.11). No obvious correlation between total protein concentration and presence of a 60 kDa band was detected. As this intermittently present band also appeared in samples from strains transformed with vector alone, it was concluded that the detected protein was not NAGLU or a derivative thereof. Immunoblot analysis of lysed cell pellets from strains transformed by full length Construct 1, Construct 2 and vector alone showed clear 60 kDa bands for all induction time points, including induction time points for empty vector only negative control cell lysates. To establish that the bands seen in the supernatant samples were identical in size to those generated from the pellet lysates, vector only negative control lysate samples were run along side T = 96 ammonium sulphate precipitated supernatant samples from both Constructs 1 and 2 (Figure 3.12c). These results imply that the bands seen in the supernatant fraction were merely contaminating non specific bands from cell lysates.

The presence of the 60 kDa lysate band described above is in agreement with findings of a previous Master's student in our laboratory. Serum obtained from a non-immunized New Zealand Buck White rabbit (the same rabbit strain as was used to produce the NAGLU antibody) when used as the primary antibody in a blot of cell lysates from vector only and NTAT transformed strains detected this same 60 kDa band in all samples (Patrick, unpublished data). As such, we conclude that this cell lysate protein is non NAGLU specific.

SDS-PAGE followed by silver stain analysis of 40-fold ammonium sulphate concentrated fractions shows a putative NAGLU specific band of approximately 90 kDa present in the induced time points of all Mut^+ single and multiple integrant strains and Mut^S strains and is absent from all time points of the empty vector negative control (Figures 3.13 and 3.14). Comparison of pellet lysate and 40x concentrated supernatant samples from pPIC9K.FL.WT.*NAGLU* and pPIC9K.FL.(CO+WT).*NAGLU* single integrant Mut^+ strains indicated that this protein is specific to the supernatant fraction (Figure 3.15). Although this induced, secreted protein appears specific to strains containing NAGLU constructs, it consistently shows only extremely low levels of expression and is presumably beneath the detection level of immunoblot analysis when using our anti-NAGLU antibody as the primary antibody for detection. These putative NAGLU specific silver stained bands lend some support to the activity levels detected. Assays for NAGLU activity did detect low AFU levels in the 40x concentrated samples of all strains. However, as there was no appreciable increase in expression from any of codon optimisation, multiple integration, expression in Mut^S strains or growth at a reduced temperature, it is difficult to imagine how this putative minimal level of NAGLU expression could be improved sufficiently to make ERT using *P. pastoris* as the recombinant expression system a feasible choice.

4.4 Confirmation of mRNA Expression in *P. pastoris*

Prior to transformation, the entire ORF of all constructs was sequence verified as correct. Growth of transformants on appropriate media then showed successful integration had occurred for all strains. To confirm this successful integration, extensive PCR analysis, described in Figures 3.4, 3.5 and 3.6, was conducted. However, the negligible levels of recombinant protein detected necessitated analysis of expression at the RNA level. *P. pastoris* has been reported to produce truncated transcripts resulting in low or no protein expression, although this is usually ascribed to RNAs containing AT-rich stretches or genes with codon biases very different from those of the organism in which the recombinant protein is being expressed (Scorer *et al.*, 1993; Sinclair and Choy, 2002). Although neither of these problems applies to the codon optimised NAGLU transcript, it was still necessary to establish the point at which expression was being blocked.

To ensure that our gene of interest was being actively transcribed, total RNA extracted from cell pellets of single integrant Mut^+ strains transformed by both Construct 1 and 2 was DNase treated and then oligo dT reverse transcribed to generate a cDNA template. Primers chosen for amplification lie within the extreme 3' end of the transcript, thus ensuring that only when complete transcription of the gene has successfully occurred will a product be generated. Figure 3.19 shows correct sized, robustly amplified products for all positively transformed samples and plasmid DNA controls, and an absence of bands in all (–) superscript RT lanes, vector only templates and water negative controls. NAGLU specific amplification was evident for cultures grown at both 15 ºC and 28 ºC. The presence of NAGLU transcripts confirms that the *P. pastoris* expression system is effectively transcribing the integrated DNA. Thus the observed low levels of recombinant protein expression indicate translational or post-translational difficulties.

4.5 Detection of rNAGLU and rNTAT Expressed in Insect Cells

4.5.1 Stable expression in insect cells

Replacement of the large 114 αα CBD domain with a small hexahistidine purification tag was achieved to create the p2ZOp2.THF.*NAGLU* and p2ZOp2.THF.*NTAT* constructs for stable expression in insect cells. Inclusion of a Factor Xa protease cleavage site allowed for removal of the hexahistidine tag after purification. Polyhistidine tags have been successfully employed for the affinity purification of recombinant proteins from a variety of expression systems including bacteria (Araujo *et al.*, 2000; Hefti *et al.*, 2001), yeast (Naested *et al.*, 2006; Toonkool *et al.*, 2006) and insect cell systems (Khoo *et al.*, 2005; Olczak and Olczak, 2005).

Successful transfection of *Sf9* and *Tni* cells with the p2ZOp2.THF.*NAGLU* and p2ZOp2.THF.*NTAT* vectors was inferred from positive SEAP assay results. Confirmation of uptake of one type of exogenous plasmid DNA (p2ZOp2F(SEAP)) suggests, but does not confirm, the successful co-uptake of the second vector (p2ZOp2.THF.*NAGLU* or p2ZOp2.THF .*NTAT*). The observation of positive, albeit low, activity readings support successful transfection. It is well recognized that optimisation of transfection efficiency is critical in obtaining maximum production yield of recombinant proteins. A variety of gene transfer methods are utilized, but lipofection is often the method of choice for transfection of eukaryotic cells. Lipid mediated transfection was shown to be 5- to 100-fold more efficient than methods using calcium phosphate precipitation or DEAE dextran transfection (Felgner *et al.*, 1987). Varying amounts and ratios of DNA and transfection reagent, as performed in this study, are frequently the first steps in optimising transfection efficiency. Other commonly investigated factors are length of incubation for the lipofection reagent in the media, the lipofection reagent

with the DNA, the lipofection mixture with the cells and the cell density at which the transfection is conducted (Keith *et al.*, 2000; Shin and Cha, 2002). Further optimisation of these factors in repeated transfection trials, may have enabled optimal production of NAGLU and NTAT.

Substantial levels of recombinant glycoprotein expression have been achieved using stably transformed lepidopteran lines (Farrell *et al.*, 1998; Sinclair *et al.*, 2006). Similar levels were achieved for recombinant glucocerebrosidase (10-15 mg/l) and a modified Factor X protein (9 mg/ml) when the respective genes were cloned into the same insect cell expression vector (Pfeifer *et al.*, 2001; Sinclair *et al.*, 2006). As the Factor X protein was fused with a CBD tag, this may indicate that the intrinsic properties of each heterologous protein have a greater influence on expression levels than the selected purification tag.

4.5.2 Baculovirus expression in insect cells

Full length mature *NAGLU* cDNA with and without a carboxyl-fused TAT-PTD was successfully PCR amplified to produce *NAGLU* or *NTAT* cDNA fused with an inframe 3' hexahistidine tag. Both constructs were cloned into the pAcGP67B baculovirus transfer vector and sequence verified as mutation free clones prior to their expression in the baculovirus system. Carboxyl tagging of the recombinant proteins was utilized, as previous studies indicated amino-fused NTAT hexahistidine tags may be inaccessible for Ni-NTA protein purification under native conditions, due to the conformational structure of the NAGLU protein (Nolla, 2007).

The H_6.L.Fx.*NTAT* fusion construct was successfully amplified using overlap extension PCR, and cloned into the pAcGP67B baculovirus transfer vector for expression and secretion by the BEVS. Linker sequences, such as L.Fx, can facilitate protein purification by enabling

the spatial separation of the purification tag from the recombinant protein of interest (Arai *et al.*, 2001). Linkers have been reported to both increase the binding capacity of the affinity tag and to frequently aid effective endoprotease cleavage (Kwon *et al.*, 2005; Arnau *et al.*, 2006). His tag accessibility problems were overcome in a study of Pfeifer *et al.* (*pers com*) by the insertion of linker residues. The 26 αα L.Fx sequence was adapted from a linker sequence contained within the commercially available insect expression vector pIEx-10.

The baculovirus expression system has been successfully used for the expression of a variety of alleles of sulfamidase, another enzyme required for the degradation of HS, the deficiency of which leads to MPS IIIA. Both precursor and mature sulfamidase proteins were produced, indicating insect cells are able to achieve normal processing of this glycoprotein (Montfort *et al.*, 2004). His tagged proteins, including glycoproteins, have been successfully expressed and purified from the *Trichoplusia ni* derived cell line (commercially available as High Five™ cells) (Lindsey-Boltz *et al.*, 2001; Hirohata *et al.*, 2002; Marchand *et al.*, 2002). However, Coomassie blue detection and immunoblot analysis indicated Ni-NTA affinity purification failed for all baculovirus constructs expressed in our study. To determine whether this was due to problems with ligand binding or due to lack of protein expression, activity assays were conducted on transduced supernatant aliquots taken prior to Ni-NTA purification and on unbound protein fractions.

Similar activity in the pre Ni-NTA and unbound aliquots indicates a complete lack of His tag binding to the affinity column. Activity assay results outlined in Figure 3.26 show all NAGLU constructs generate levels of activity clearly above those of both untransfected cells alone and cells co-transfected with baculovirus and a non-*NAGLU* gene. Immunoblot analysis of unbound protein fractions further indicate that all three constructs are being successfully expressed and generate bands of the expected approximately 80-85 kDa size.

Partial folding inside the tertiary structure of the protein may render the hexahistidine tags either partially or completely inaccessible to interaction with the Ni-NTA agarose (Schmitt *et al.*, 1993). Tag length and/or position can also affect affinity binding (Fu and Maloney, 1997). In an attempt to overcome these potential problems, we tested both C- and N-hexahistidine tags, the latter designed to include an intervening linker region. However, these measures had no apparent affect on affinity or accessibility.

Proteolytic degradation or cleavage of the polyhistidine tag from the recombinant protein could also explain lack of binding. This has been reported by other groups attempting to express and purify glycoproteins. Insertion of N-terminal and C-terminal polyhistidine tags failed to assist affinity purification of the β-glucosidase protein due to posttranslational processing at both termini (Toonkool *et al.*, 2006).

Lack of Ni-NTA binding can occur as a consequence of low level recombinant protein expression and this is perhaps the most probable explanation of our results. Proteins are retained on Ni-NTA columns according to the number of accessible histidines. As such, certain endogenous proteins with naturally occurring contiguous histidine residues will bind the affinity column. When recombinant proteins of interest are expressed at sufficiently high levels, they are able to out-compete these naturally expressed proteins. However, if heterologous protein expression is low, successful competition may no longer occur (Schmitt *et al.*, 1993). It is frequently necessary to concentrate the polyhistidine tagged recombinant protein to enable successful affinity purification when using a Ni-NTA column. The QIAexpressionist™ handbook (QIAGEN, 2003) states that a 50- to 100-fold concentration is recommended to purify significant amounts of hexahistidine tagged protein.

Activity assay and immunoblot using the anti-NAGLU antibody indicate we have successfully expressed all three recombinant proteins using the baculovirus expression system.

The inability to detect recombinant protein by immunoblot using anti-His_4 may again be explained by tag removal or insufficient levels of expression. If the latter is the case, then it would appear the anti-His_4 antibody has lower affinity for the hexahistidine tag than the anti-NAGLU antibody has for our recombinant NAGLU proteins. The selected hexahistidine tagged anti-His_4 antibody positive control is a highly expressed protein, amicon concentrated from *P. pastoris* culture medium containing few other contaminating proteins, as evidenced by a single dominant band on a silver stained gel (Zay, 2006). Logically it follows that there is a vast excess of polyhistidine tag in this control sample as compared to the crude, unconcentrated, small scale baculovirus preparations, and hence it is not surprising that a signal should be attained for the positive control sample when it is absent from the experimental samples. These observations are consistent with those reported by Nolla (2007) where hexahistidine tagged recombinant protein expressed in insect cells could not be detected in unconcentrated samples even under denaturing conditions, but could be detected in amicon concentrated samples under the same conditions.

Having established that baculovirus infection of High Five™ cells does give rise to assayable levels of recombinant protein expression, many avenues can be explored to increase heterologous protein levels. The factor likely to have greatest effect is amplification of the primary virus. A routine step in recombinant protein production using BEVS is amplification of the primary virus (typically performed 2-3 times) to produce a suitably high titre stock (Pharmingen, 1999). Determining the ideal multiplicity of infection (MOI), when using this stock, is also important in attaining maximal recombinant protein production. High MOIs are usually required to guarantee all cells are infected at the same time, thus enabling synchronous initiation of viral protein expression (Morais *et al.*, 2001). Lack of concomitant infection can result in non-uniformity of timing of cell lysis, thereby resulting in both reduced levels of

heterologous protein expression and contamination of secreted recombinant protein with intracellular proteins, increasing the difficulty of purification. This fits with our observations. We obtained relatively high AFU expression levels (produced by a baculovirus infected 2 ml culture of approximately 5 x 10^6 cells/ml) which are roughly 3-4 times those produced by 60 ml stably transformed shaker culture of 5 x 10^6 cells/ml (Nolla, unpublished data). However, the total protein in the cell media from baculovirus expression appears much higher than that produced by stably transformed cells, leading to specific activity readings from crude baculovirus cultures of approximately half that of stably transformed cultures. This may well be an artifact of this proof of principal experiment conducted on a small scale under non optimised conditions leading to increased cell lysis. It should be noted that another factor established to affect protein expression is culture size. As this initial experiment was performed in a 6 well culture dish, there is much room for increase in culture volume. The effect of cell density at time of infection and harvesting time post infection are also variables that warrant optimisation (Gao *et al.*, 2005; Meghrous *et al.*, 2005).

5 Conclusions and Future Directions

5.1 NAGLU Expression in *P. pastoris*

There are many factors influencing the expression of target recombinant proteins in *P. pastoris* and many studies have outlined specific features enabling an increase in expression level of a particular protein of interest. It seems universally accepted that there is no one solution applicable to the expression of all, or even the majority of recombinant proteins, and the reasons behind the low expression level of some proteins remain unclear. In attempts to express NAGLU in this methylotrophic yeast system, we have tried many approaches: selection of a humanized strain of *P. pastoris* to more closely imitate the posttranslational modifications possible in mammals, codon optimisation of a portion of the gene to overcome any potential bottle neck caused by unavailability of rarer tRNA molecules in the host system, assessing the impact of the alcohol oxidase pathways in terms of Mut^+ and Mut^S phenotypes, the affect of multiple integration events and decreased temperature.

The impact of upstream sequences in the pPIC9K vector, such as the *S. cerevisiae* alpha mating factor (α-MF) secretion signal or polylinker cloning site, on NAGLU expression have not been determined. The non native α-MF in pPIC9K is the secretion signal sequence that has been used with most success for recombinant expression in *P. pastoris* (Macauley-Patrick *et al.,* 2005). However, Xiong *et al*. (2006) achieved a large increase in expression after codon optimisation of this secretion signal. Outchkourov *et al.* (2002) found that removal of a short stretch of 12 nucleotides (containing the SnaBI and EcoRI sites) from the MCS of the pPIC9 vector led to a 4- to 10-fold increase in protein expression. The 5' EcoRI site was utilized for cloning full length NAGLU, and hence these same 12 bases are present in our constructs. The authors postulate that the SnaBI portion of this sequence has homology to a sequence known to

affect translation in yeast and suggest the presence of this sequence upstream of their gene may have inhibited protein translation (Outchkourov *et al.*, 2002). These potential problems could be overcome by cloning *NAGLU* with the native *P. pastoris* secretion signal, or by expressing the NAGLU protein without a secretion signal, to give intracellular expression. However, intracellular expression could create difficulties with purifying the heterologous protein away from native yeast proteins. Extracellular localization of recombinant protein avoids these problems as the secreted heterologous protein constitutes the vast majority of the protein in the medium.

The effect on NAGLU expression from cultures grown in an acidic environment, rather than the relatively neutral environment created by BMMY media buffered to pH 6.0, has not been determined. As cells grow, other cells die, releasing proteases into the media. If the recombinant protein of interest is susceptible to cleavage by neutral proteases, it may be advantageous to express in a more acidic media (and conversely, to express in a neutral environment if the protein of interest is susceptible to acid proteases). Although addition of casamino acids and use of the protease deficient strain SMD1163 did not improve NAGLU expression (Patrick, 2006), several other groups have reported a lack of effect with these measures but an increased expression upon reduction of pH to 3 (Werten *et al.*, 1999; Curvers *et al.*, 2001). Although this very low pH may be too extreme to allow expression of functionally active NAGLU enzyme, purified recombinant NAGLU is stable in a broad range of pH from pH 8.1 down to a pH 4.6, and as such growth in an acidic environment could be considered (Zhao and Neufeld, 2000).

A recent paper assessing the sequence based factors influencing the expression of 79 heterologous genes in *P. pastoris* found that only three factors had a statistically significant association with expression level. High levels of expression were associated with low

abundance of AT-rich regions and the occurrence of a protein homologue in yeast. Failure of expression was associated with a comparatively high isoelectric point. At pI = 6.65, NAGLU has a reasonably low isoelectric point and also has a low abundance of AT-rich regions. The *P. pastoris* genome has either not been completely sequenced, or the sequence has not yet been deposited in public databases. However, DNA BLAST (Basic Local Alignment Search Tool) searches through the NCBI database failed to reveal any significant homology when searched against all sequenced Fungi genomes. A protein BLAST of these same genomes identified a putative *α-N*-acetylglucosaminidase in *Aspergillus fumigatus* (Accession number XP_755275), but failed to identify any homologue in the more closely related *S. cerevisiae*. The lack of an apparent protein homologue in *P. pastoris* may mean the specific chaperone proteins needed for correct folding may also be absent in this yeast, resulting in degradation of the misfolded protein.

It has been suggested that the maximum level of protein secretion is ultimately determined by the protein folding capacity of the ER (Parekh *et al.*, 1995; Parekh and Wittrup, 1997). The basic tenets of this ER-associated degradation pathway are conserved amongst all eukaryotes (Spear and Ng, 2005). Secretory proteins enter the ER in an unfolded state and are modified and folded to acquire their functional conformations by ER-resident chaperones. Improperly folded proteins are then transported back to the cytosol and degraded by the ubiquitin-proteasome machinery to prevent toxification by the accumulation of aberrant proteins (Kostova and Wolf, 2003; Hirsch *et al.*, 2004; Romisch, 2005). High level expression of recombinant proteins could mean the components of the folding machinery become rate limiting, leading to inefficient folding of heterologous proteins (Inan *et al.*, 2006). Incorrect posttranslational modification or lack of necessary chaperones for a specific recombinant protein could also give rise to incorrect folding of recombinant proteins, thus eliciting the

cellular stress response leading to ER-associated degradation and resulting in the observed lack of secreted heterologous protein.

In conclusion, numerous experimental approaches were attempted to increase the level of heterologous expression of α-*N*-acetylglucosaminidase in *P. pastoris*, however none culminated in the production of anything over low level expression of NAGLU as assessed by activity assay. To have confidence in achievement of even this low level of expression, confirmation of expression by immunoblot detection would be desirable. It may be possible to increase the level of protein expression by adjustment of culturing conditions (pH and media components), however poor translation, or protein instability due to incorrect folding suggest that *P. pastoris* may not be a suitable expression host for the production of recombinant NAGLU.

5.2 NAGLU and NTAT Expression in Insect Cells

Preliminary experiments utilizing baculovirus expression of NAGLU and NTAT proteins have yielded promising results. Immunoblot and activity assays indicate functionally active NAGLU and NTAT protein of the expected approximately 80 to 85 kDa size has been produced. Current inability to purify the polyhistidine tagged proteins will likely be overcome by the measures outlined in section 4.5.2 to improve recombinant protein expression levels. This is in keeping with reports from other researchers who failed to purify a weakly expressed His tagged protein but had success in purifying a highly expressed protein using the same affinity system (Lichty *et al.*, 2005). If further optimisation is necessary, there are a number of approaches to increase specificity of binding and binding efficiency. Optimisation of low concentrations of imidazole and high concentrations of NaCl in the binding buffer enabled a respective competitive inhibition of non specific binding and inhibition of binding to charged

groups on the gel matrix via an ion exchange effect (Wojczyk *et al.*, 1996). A double His_6 tag, in which two hexahistidines are separated by a short intervening linker sequence, was shown to bind more efficiently to the Ni-NTA column than a single His_6 tag and enabled purification to near homogeneity (Khan *et al.*, 2006). Increasing the number of histidines from 6 to 10 also led to increased binding (Mohanty and Wiener, 2004).

This preliminary success with a baculovirus-based recombinant protein expression system allows many new research avenues to be explored. If glycosylation patterns from expression in High Five™ cells are found to be immunogenic or generate unstable NAGLU protein, the option of "humanized" insect cells can be investigated. The SfSWT-3 transgenic cell line is able to efficiently mimic the sialylation patterns seen in mammalian cells. The absence of terminal sialic acids on recombinant glycoproteins produced in traditional insect cell systems can lead to extremely short half lives *in vivo*, but transduction of the SfSWT-3 cell line may overcome this problem (Jarvis, 2003). Another option is the use of a non-lytic baculovirus system. Cells infected with conventional baculovirus are eventually disrupted at the late stage of viral infection leading to the loss of cellular homeostasis which may result in inefficient processing of glycoproteins. Random mutagenesis of viral genomes has led to the creation of an essentially non-lytic BEVS. At 5 days post infection, non-lytic BEVS showed only 7% cell lysis, whereas lytic BEVS showed 60% cell lysis at this same stage (Ho *et al.*, 2004). Significantly less cell lysis is more likely to enable the continuance of the protein quality control mechanisms characteristic of eukaryotic cells, enabling the posttranslational modifications necessary for the maturation of secretory proteins.

The aims in undertaking this proof of principle experiment were firstly to determine if recombinant NAGLU and NTAT proteins can be expressed using the BEVS. Secondly, to establish if expressed, whether the recombinant proteins were active. Thirdly, to investigate

whether the expressed recombinant proteins could be Ni-NTA purified by their affinity tags. This preliminary work has shown, using the BEVS, the first two of these three aims have been successfully achieved. Unfortunately, due to the small scale of this initial study, we were unable to affinity purify the expressed proteins. These encouraging results indicate the BEVS is an attractive option for the large scale production of biologically active recombinant NAGLU and NTAT protein.

6 Bibliography

Ailor E, Betenbaugh MJ (1999) Modifying secretion and post-translational processing in insect cells. Current Opinion in Biotechnology 10:142-145.

Akashi H (1994) Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. Genetics 136:927-935.

Altmann F, Staudacher E, Wilson IB, Marz L (1999) Insect cells as hosts for the expression of recombinant glycoproteins. Glycoconjugate Journal 16:109-123.

Andersen SS (2004) Expression and purification of recombinant vesicular glutamate transporter VGLUT1 using PC12 cells and High Five insect cells. Biological Procedures Online 6:105-112.

Andre N, Cherouati N, Prual C, Steffan T, Zeder-Lutz G, Magnin T, Pattus F, Michel H, Wagner R, Reinhart C (2006) Enhancing functional production of G protein-coupled receptors in *Pichia pastoris* to levels required for structural studies via a single expression screen. Protein Science 15:1115-1126.

Applegarth DA, Toone JR, Rolland MO, Black SH, Yim DK, Bemis G (2000) Non-concordance of CVS and liver glycine cleavage enzyme in three families with non-ketotic hyperglycinaemia (NKH) leading to false negative prenatal diagnoses. Prenatal Diagnosis 20:367-370.

Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T (2001) Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein Engineering 14:529-532.

Araujo AP, Oliva G, Henrique-Silva F, Garratt RC, Caceres O, Beltramini LM (2000) Influence of the histidine tail on the structure and activity of recombinant chlorocatechol 1,2-dioxygenase. Biochemical and Biophysical Research Communications 272:480-484.

Arnau J, Lauritzen C, Petersen GE, Pedersen J (2006) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expression and Purification 48:1-13.

Aumiller JJ, Hollister JR, Jarvis DL (2003) A transgenic insect cell line engineered to produce CMP-sialic acid and sialylated glycoproteins. Glycobiology 13:497-507.

Bandsmer, J (2004) Expression of Alpha-N-Acetylglucosaminidase Fused to the HIV-1 Protein Transduction Domain and a Modified Protein Transduction Domain. Master's Thesis. Department of Biology. University of Victoria, Victoria, BC, Canada.

Bandsmer J, Campbell T, Cheyn I, Choy F (2006) Expression of active alpha-N-acetyl-glucosaminidase/TAT Chimerae in cultured *Spodoptera frugiperda* cells. Protein and Peptide Letters 13:353-356.

Bax MC, Colville GA (1995) Behaviour in Mucopolysaccharide disorders. Archives of Disease in Childhood 73:77-81.

Becker-Hapak M, McAllister SS, Dowdy SF (2001) TAT-mediated protein transduction into mammalian cells. Methods 24:247-256.

Beesley C, Jackson M, Young E, Vellodi A, Winchester B (2005) Molecular defects in Sanfilippo syndrome type B (mucopolysaccharidosis IIIB). Journal of Inherited Metabolic Disease 28:759-767.

Beesley C, Moraitou M, Winchester B, Schulpis K, Dimitriou E, Michelakakis H (2004) Sanfilippo B syndrome: molecular defects in Greek patients. Clinical Genetics 65:143-149.

Beesley CE, Young EP, Vellodi A, Winchester BG (1998) Identification of 12 novel mutations in the alpha-N-acetylglucosaminidase gene in 14 patients with Sanfilippo syndrome type B (mucopolysaccharidosis type IIIB). Journal of Medical Genetics 35:910-914.

Bernfield M, Gotte M, Park PW, Reizes O, Fitzgerald ML, Lincecum J, Zako M (1999) Functions of cell surface heparan sulfate proteoglycans. Annual Review of Biochemistry 68:729-777.

Bobrowicz P, Davidson RC, Li H, Potgieter TI, Nett JH, Hamilton SR, Stadheim TA, Miele RG, Bobrowicz B, Mitchell T, Rausch S, Renfer E, Wildt S (2004) Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast *Pichia pastoris*: production of complex humanized glycoproteins with terminal galactose. Glycobiology 14:757-766.

Boettner M, Steffens C, von Mering C, Bork P, Stahl U, Lang C (2007) Sequence-based factors influencing the expression of heterologous genes in the yeast *Pichia pastoris* - A comparative view on 79 human genes. Journal of Biotechnology 130:1-10.

Bulmer M (1991) The selection-mutation-drift theory of synonymous codon usage. Genetics 129:897-907.

Bunge S, Knigge A, Steglich C, Kleijer WJ, van Diggelen OP, Beck M, Gal A (1999) Mucopolysaccharidosis type IIIB (Sanfilippo B): identification of 18 novel alpha-N-acetylglucosaminidase gene mutations. Journal of Medical Genetics 36:28-31.

Callewaert N, Laroy W, Cadirgi H, Geysens S, Saelens X, Min Jou W, Contreras R (2001) Use of HDEL-tagged *Trichoderma reesei* mannosyl oligosaccharide 1,2-alpha-D-mannosidase for N-glycan engineering in *Pichia pastoris*. FEBS Letters 503:173-178.

Cao G, Pei W, Ge H, Liang Q, Luo Y, Sharp FR, Lu A, Ran R, Graham SH, Chen J (2002) In Vivo Delivery of a Bcl-xL Fusion Protein Containing the TAT Protein Transduction Domain Protects against Ischemic Brain Injury and Neuronal Apoptosis. Journal of Neuroscience 22:5423-5431.

Cereghino L, Cregg J (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiology Review 24: 45-66.

Chinen Y, Tohma T, Izumikawa Y, Uehara H, Ohta T (2005) Sanfilippo type B syndrome: five patients with an R565P homozygous mutation in the alpha-N-acetylglucosaminidase gene from the Okinawa islands in Japan. Journal of Human Genetics 50:357-359.

Choi BK, Bobrowicz P, Davidson RC, Hamilton SR, Kung DH, Li H, Miele RG, Nett JH, Wildt S, Gerngross TU (2003) Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast *Pichia pastoris*. Proceedings of the National Academy of Sciences USA 100:5022-5027.

Chow P, Weissmann B (1981) 4-Methylumbelliferyl 2-Acetamido-2-deoxy-alpha-D-glucopyranoside, a fluorogenic substrate for *N*-acetyl-α-D-glucosaminidase. Carbohydrate Research 96: 87-93.

Clare JJ, Romanos MA, Rayment FB, Rowedder JE, Smith MA, Payne MM, Sreekrishna K, Henwood CA (1991) Production of mouse epidermal growth factor in yeast: high-level secretion using *Pichia pastoris* strains containing multiple gene copies. Gene 105:205-212.

Cleary M, Wraith J (1993) Management of Mucopolysaccharidosis type III. Archives of Disease in Childhood 69:403-406.

Coll MJ, Anton C, Chabas A (2001) Allelic heterogeneity in Spanish patients with Sanfilippo disease type B. Identification of eight new mutations. Journal of Inherited Metabolic Disease 24:83-84.

Cooper DN, Youssoufian H (1988) The CpG dinucleotide and human genetic disease. Human Genetics 78:151-155.

Cregg JM, Cereghino JL, Shi J, Higgins DR (2000) Recombinant protein expression in *Pichia pastoris*. Molecular Biotechnology 16:23-52.

Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA (1989) Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. Molecular and Cellular Biology 9:1316-1323.

Curvers S, Brixius P, Klauser T, Thommes J, Weuster-Botz D, Takors R, Wandrey C (2001) Human chymotrypsinogen B production with *Pichia pastoris* by integrated development of fermentation and downstream processing. Part 1. Fermentation. Biotechnology Progress 17:495-502.

Dahms NM, Lobel P, Kornfeld S (1989) Mannose 6-phosphate receptors and lysosomal enzyme targeting. Journal of Biological Chemistry 264:12115-12118.

Daly R, Hearn MT (2005) Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. Journal of Molecular Recognition 18:119-138.

Davis TR, Wickham TJ, McKenna KA, Granados RR, Shuler ML, Wood HA (1993) Comparative recombinant protein production of eight insect cell lines. In Vitro Cellular and Developmental Biology Animal 29A:388-390.

Delroisse JM, Dannau M, Gilsoul JJ, El Mejdoub T, Destain J, Portetelle D, Thonart P, Haubruge E, Vandenbol M (2005) Expression of a synthetic gene encoding a Tribolium castaneum carboxylesterase in *Pichia pastoris*. Protein Expression and Purification 42:286-294.

Di Natale P (1991) Sanfilippo B disease: a re-examination of a particular sibship after 12 years. Journal of Inherited Metabolic Disease 14:23-28.

Di Natale P, Salvatore D, Daniele A, Bonatti S (1985) Biosynthesis of alpha-N-acetylglucosaminidase in cultured human kidney carcinoma cells. Enzyme 33:75-83.

Dong H, Nilsson L, Kurland CG (1996) Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. Journal of Molecular Biology 260:649-663.

Duchardt F, Fotin-Mleczek M, Schwarz H, Fischer R, Brock R (2007) A comprehensive model for the cellular uptake of cationic cell-penetrating peptides. Traffic 8:848-866.

Elliger SS, Elliger CA, Lang C, Watson GL (2002) Enhanced secretion and uptake of beta-glucuronidase improves adeno-associated viral-mediated gene therapy of mucopoly-saccharidosis type VII mice. Molecular Therapy 5:617-626.

Emre S, Terzioglu M, Tokatli A, Coskun T, Ozalp I, Weber B, Hopwood JJ (2002) Sanfilippo syndrome in Turkey: Identification of novel mutations in subtypes A and B. Human Mutation 19:184-185.

Ezhevsky SA, Nagahara H, Vocero-Akbani AM, Gius DR, Wei MC, Dowdy SF (1997) Hypo-phosphorylation of the retinoblastoma protein (pRb) by cyclin D:Cdk4/6 complexes results in active pRb. Proceedings of the National Academy of Sciences USA 94:10699-10704.

Fan X, Zhang H, Zhang S, Bagshaw RD, Tropak MB, Callahan JW, Mahuran DJ (2006) Identification of the gene encoding the enzyme deficient in mucopolysaccharidosis IIIC (Sanfilippo disease type C). American Journal of Human Genetics 79: 738-744.

Farrell PJ, Lu M, Prevost J, Brown C, Behie L, Iatrou K (1998) High-level expression of secreted glycoproteins in transformed lepidopteran insect cells using a novel expression vector. Biotechnology and Bioengineering 60:656-663.

Fawell S, Seery J, Daikh Y, Moore C, Chen LL, Pepinsky B, Barsoum J (1994) Tat-mediated delivery of heterologous proteins into cells. Proceedings of the National Academy of Sciences USA 91:664-668.

Felgner PL, Gadek TR, Holm M, Roman R, Chan HW, Wenz M, Northrop JP, Ringold GM, Danielsen M (1987) Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. Proceedings of the National Academy of Sciences USA 84:7413-7417.

Frankel AD, Pabo C (1988) Cellular uptake of the tat protein from human immunodeficiency virus. Cell 55:1189-1193.

Fu D, Maloney PC (1997) Evaluation of secondary structure of OxlT, the oxalate transporter of Oxalobacter formigenes, by circular dichroism spectroscopy. Journal of Biological Chemistry 272:2129-2135.

Furbish FS, Steer CJ, Krett NL, Barranger JA (1981) Uptake and distribution of placental glucocerebrosidase in rat hepatic cells and effects of sequential deglycosylation. Biochimica and Biophysica Acta 673:425-434.

Gao X, Yo P, Harris TK (2005) Improved yields for baculovirus-mediated expression of human His(6)-PDK1 and His(6)-PKBbeta/Akt2 and characterization of phospho-specific isoforms for design of inhibitors that stabilize inactive conformations. Protein Expression and Purification 43:44-56.

Gasser B, Maurer M, Rautio J, Sauer M, Bhattacharyya A, Saloheimo M, Penttila M, Mattanovich D (2007) Monitoring of transcriptional regulation in *Pichia pastoris* under protein production conditions. BMC Genomics 8:179.

Gellissen G (2000) Heterologous protein production in methylotrophic yeasts. Applied Microbiology and Biotechnology 54:741-750.

Gerngross TU (2004) Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. Nature Biotechnology 22:1409-1414.

Gouy M, Gautier C (1982) Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Research 10:7055-7074.

Grantham R, Gautier C, Gouy M, Mercier R, Pave A (1980) Codon catalog usage and the genome hypothesis. Nucleic Acids Research 8:49-62.

Green M, Loewenstein P (1988) Autonomous functional domains of chemically synthesized human immunodeficiency virus tat trans-activator protein. Cell 55:1179-1188.

Groves ML, McKeon R, Werner E, Nagarsheth M, Meador W, English AW (2005) Axon regeneration in peripheral nerves is enhanced by proteoglycan degradation. Experimental Neurology 195:278-292.

Gustafsson C, Govindarajan S, Minshull J (2004) Codon bias and heterologous protein expression. Trends in Biotechnology 22:346-353.

Hamilton SR, Bobrowicz P, Bobrowicz B, Davidson RC, Li H, Mitchell T, Nett JH, Rausch S, Stadheim TA, Wischnewski H, Wildt S, Gerngross TU (2003) Production of complex human glycoproteins in yeast. Science 301:1244-1246.

Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, Bobrowicz P, Stadheim TA, Li H, Choi BK, Hopkins D, Wischnewski H, Roser J, Mitchell T, Strawbridge RR, Hoopes J, Wildt S, Gerngross TU (2006) Humanization of yeast to produce complex terminally sialylated glycoproteins. Science 313:1441-1443.

Hefti MH, Van Vugt-Van der Toorn CJ, Dixon R, Vervoort J (2001) A novel purification method for histidine-tagged proteins containing a thrombin cleavage site. Analytical Biochemistry 295:180-185.

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

Hirohata S, Wang LW, Miyagi M, Yan L, Seldin MF, Keene DR, Crabb JW, Apte SS (2002) Punctin, a novel ADAMTS-like molecule, ADAMTSL-1, in extracellular matrix. Journal of Biological Chemistry 277:12182-12189.

Hirsch C, Jarosch E, Sommer T, Wolf DH (2004) Endoplasmic reticulum-associated protein degradation--one model fits all? Biochimica and Biophysica Acta 1695:215-223.

Ho Y, Lo HR, Lee TC, Wu CP, Chao YC (2004) Enhancement of correct protein folding in vivo by a non-lytic baculovirus. Biochemical Journal 382:695-702.

Hoekema A, Kastelein RA, Vasser M, de Boer HA (1987) Codon replacement in the PGK1 gene of *Saccharomyces cerevisiae*: experimental approach to study the role of biased codon usage in gene expression. Molecular and Cellular Biology 7:2914-2924.

Hohenblum H, Gasser B, Maurer M, Borth N, Mattanovich D (2004) Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*. Biotechnology and Bioengineering 85:367-375.

Hopwood JJ (2005) Prenatal diagnosis of Sanfilippo syndrome. Prenatal Diagnosis 25:148-150.

Hu S, Li L, Qiao J, Guo Y, Cheng L, Liu J (2006) Codon optimization, expression, and characterization of an internalizing anti-ErbB2 single-chain antibody in *Pichia pastoris*. Protein Expression and Purification 47:249-257.

Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. Molecular Biology and Evolution 2:13-34.

Inan M, Aryasomayajula D, Sinha J, Meagher MM (2006) Enhancement of protein secretion in *Pichia pastoris* by overexpression of protein disulfide isomerase. Biotechnology and Bioengineering 93:771-778.

Invitrogen (2002) *Pichia* Expression Kit: A Manual of Methods for Expression of Recombinant Proteins in *P. pastoris.* Version M.

Jahic M, Wallberg F, Bollok M, Garcia P, Enfors SO (2003) Temperature limited fed-batch technique for control of proteolysis in *Pichia pastoris* bioreactor cultures. Microbial Cell Factories 2:6.

Jarvis DL (2003) Developing baculovirus-insect cell expression systems for humanized recombinant glycoprotein production. Virology 310:1-7.

Kakkis ED, McEntee M, Schmidtchen A, Neufeld E, Ward D, Gompf R, Kania S, Bedolla C, Chien S, Shull R (1996) Long-term and high-dose trials of enzyme replacement therapy in the canine model of mucopolysaccharidosis I. Biochemical and Molecular Medicine 58:156-167.

Keith MB, Farrell PJ, Iatrou K, Behie LA (2000) Use of flow cytometry to rapidly optimize the transfection of animal cells. Biotechniques 28:148-154.

Khan F, He M, Taussig MJ (2006) Double-hexahistidine tag with high-affinity binding for protein immobilization, purification, and detection on ni-nitrilotriacetic acid surfaces. Analytical Chemistry 78:3072-3079.

Khoo KM, Chang CF, Schubert J, Wondrak E, Chng HH (2005) Expression and purification of the recombinant His-tagged GST-CD38 fusion protein using the baculovirus/insect cell expression system. Protein Expression and Purification 40:396-403.

Kitts PA, Possee RD (1993) A method for producing recombinant baculovirus expression vectors at high frequency. Biotechniques 14:810-817.

Kost TA, Condreay JP, Jarvis DL (2005) Baculovirus as versatile vectors for protein expression in insect and mammalian cells. Nature Biotechnology 23:567-575.

Kostova Z, Wolf DH (2003) For whom the bell tolls: protein quality control of the endoplasmic reticulum and the ubiquitin-proteasome connection. EMBO Journal 22:2309-2317.

Kukuruzinska MA, Lennon K (1998) Protein N-glycosylation: molecular genetics and functional significance. Critical Reviews in Oral Biology and Medicine 9:415-448.

Kwon SY, Choi YJ, Kang TH, Lee KH, Cha SS, Kim GH, Lee HS, Kim KT, Kim KJ (2005) Highly efficient protein expression and purification using bacterial hemoglobin fusion vector. Plasmid 53:274-282.

Lee-Chen GJ, Lin SP, Lin SZ, Chuang CK, Hsiao KT, Huang CF, Lien WC (2002) Identification and characterisation of mutations underlying Sanfilippo syndrome type B (mucopolysaccharidosis type IIIB). Journal of Medical Genetics 39:E3.

Lee KO, Luu N, Kaneski CR, Schiffmann R, Brady RO, Murray GJ (2005) Improved intracellular delivery of glucocerebrosidase mediated by the HIV-1 TAT protein transduction domain. Biochemical and Biophysical Research Communications 337:701-707.

Li H, Sethuraman N, Stadheim TA, Zha D, Prinz B, Ballew N, Bobrowicz P, Choi BK, Cook WJ, Cukan M, Houston-Cummings NR, Davidson R, Gong B, Hamilton SR, Hoopes JP, Jiang Y, Kim N, Mansfield R, Nett JH, Rios S, Strawbridge R, Wildt S, Gerngross TU (2006) Optimization of humanized IgGs in glycoengineered *Pichia pastoris*. Nature Biotechnology 24:210-215.

Li HH, Zhao HZ, Neufeld EF, Cai Y, Gomez-Pinilla F (2002) Attenuated plasticity in neurons and astrocytes in the mouse model of Sanfilippo syndrome type B. Journal of Neuroscience Research 69:30-38.

Li Z, Xiong F, Lin Q, d'Anjou M, Daugulis AJ, Yang DS, Hew CL (2001) Low-temperature increases the yield of biologically active herring antifreeze protein in *Pichia pastoris*. Protein Expression and Purification 21:438-445.

Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S (2005) Comparison of affinity tags for protein purification. Protein Expression and Purification 41:98-105.

Lindsey-Boltz LA, Bermudez VP, Hurwitz J, Sancar A (2001) Purification and characterization of human DNA damage checkpoint Rad complexes. Proceedings of the National Academy of Sciences USA 98:11236-11241.

Liu Y, Jones M, Hingtgen CM, Bu G, Laribee N, Tanzi RE, Moir RD, Nath A, He JJ (2000) Uptake of HIV-1 tat protein mediated by low-density lipoprotein receptor-related protein disrupts the neuronal metabolic balance of the receptor ligands. Nature Medicine 6:1380-1387.

Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM (2005) Heterologous protein production using the *Pichia pastoris* expression system. Yeast 22:249-270.

Mann DA, Frankel AD (1991) Endocytosis and targeting of exogenous HIV-1 Tat protein. EMBO Journal 10:1733-1739.

Marchand C, Pourquier P, Laco GS, Jing N, Pommier Y (2002) Interaction of human nuclear topoisomerase I with guanosine quartet-forming and guanosine-rich single-stranded DNA and RNA oligonucleotides. Journal of Biological Chemistry 277:8906-8911.

McDowell GA, Cowan TM, Blitzer MG, Greene CL (1993) Intrafamilial variability in Hurler syndrome and Sanfilippo syndrome type A: implications for evaluation of new therapies. American Journal of Medical Genetics 47:1092-1095.

Meghrous J, Aucoin MG, Jacob D, Chahal PS, Arcand N, Kamen AA (2005) Production of recombinant adeno-associated viral vectors using a baculovirus/insect cell suspension culture system: from shake flasks to a 20-L bioreactor. Biotechnology Progress 21:154-160.

Meikle PJ, Hopwood JJ, Clague AE, Carey WF (1999) Prevalence of lysosomal storage disorders. Journal of the American Medical Association 281:249-254.

Mohanty AK, Wiener MC (2004) Membrane protein expression and production: effects of polyhistidine tag length and position. Protein Expression and Purification 33:311-325.

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

Montfort M, Garrido E, Hopwood JJ, Grinberg D, Chabas A, Vilageliu L (2004) Expression and functional characterization of human mutant sulfamidase in insect cells. Molecular Genetics and Metabolism 83:246-251.

Morais VA, Serpa J, Palma AS, Costa T, Maranga L, Costa J (2001) Expression and characterization of recombinant human alpha-3/4-fucosyltransferase III from *Spodoptera frugiperda* (*Sf9*) and *Trichoplusia ni* (*Tn*) cells using the baculovirus expression system. Biochemical Journal 353:719-725.

Mukherjee B, Burma S, Talwar GP, Hasnain SE (1995) Transcriptional regulation of cell line-dependent, baculovirus-mediated expression of foreign genes. DNA and Cell Biology 14:7-14.

Naested H, Kramhoft B, Lok F, Bojsen K, Yu S, Svensson B (2006) Production of enzymatically active recombinant full-length barley high pI alpha-glucosidase of glycoside family 31 by high cell-density fermentation of *Pichia pastoris* and affinity purification. Protein Expression and Purification 46:56-63.

Nelson J (1997) Incidence of the mucopolysaccharidoses in Northern Ireland. Human Genetics 101:355-358.

Neufeld E, Muenzer J (1995) The Mucopolysaccharidoses. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) The Metabolic and Molecular Bases of Inherited Disease. McGraw-Hill, New York, pp 2465-2494.

Neufeld E, Muenzer J (2001) The Mucopolysaccharidoses. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) The Metabolic and Molecular Bases of Inherited Disease. McGraw-Hill, New York, pp 3421-3452.

Nolla N (2007) The Expression of a-N-acetylglucosaminidase in *Sf9* cells. Honour's Thesis. Department of Biology. University of Victoria, Victoria, BC, Canada.

Olczak M, Olczak T (2005) Expression and purification of active plant diphosphonucleotide phosphatase/phosphodiesterase from baculovirus-infected insect cells. Protein Expression and Purification 39:116-123.

Ong CE, Miners JO, Birkett DJ, Bhasker CR (1998) Baculovirus-mediated expression of cytochrome P4502C8 and human NADPH-cytochrome P450 reductase: optimization of protein expression. Xenobiotica 28:137-152.

Orii K, Grubb J, Vogler C, Levy B, Tan Y, Markova K, Davidson B, Mao Q, Orri T, Kondo N, Sly W (2005) Defining the pathway for Tat-mediated delivery of beta-glucuronidase in cultured cells and MPS VII mice. Molecular Therapy 12:345-352.

Outchkourov NS, Stiekema WJ, Jongsma MA (2002) Optimization of the expression of equistatin in *Pichia pastoris*. Protein Expression and Purification 24:18-24.

Pan C, Nelson MS, Reyes M, Koodie L, Brazil JJ, Stephenson EJ, Zhao RC, Peters C, Selleck SB, Stringer SE, Gupta P (2005) Functional abnormalities of heparan sulfate in mucopoly-saccharidosis-I are associated with defective biologic activity of FGF-2 on human multipotent progenitor cells. Blood 106:1956-1964.

Parekh R, Forrester K, Wittrup D (1995) Multicopy overexpression of bovine pancreatic trypsin inhibitor saturates the protein folding and secretory capacity of *Saccharomyces cerevisiae*. Protein Expression and Purification 6:537-545.

Parekh RN, Wittrup KD (1997) Expression level tuning for optimal heterologous protein secretion in *Saccharomyces cerevisiae*. Biotechnology Progress13:117-122.

Patrick C (2006) The Expression of a-N-acetylglucosaminidase in the methylotrophic yeast *Pichia pastoris*. Master's Thesis. Department of Biology. University of Victoria, Victoria, BC, Canada.

Patterson R, Selkirk J, Merrick B (1995) Baculovirus and Insect Cell Gene Expression: Review of Baculovirus Biotechnology. Environmental Health Perspectives 103:756.

Pfeifer TA, Hegedus DD, Grigliatti TA, Theilmann DA (1997) Baculovirus immediate-early promoter-mediated expression of the Zeocin resistance gene for use as a dominant selectable marker in dipteran and lepidopteran insect cell lines. Gene 188:183-190.

Pfeifer TA (1998) Expression of heterologous proteins in stable insect cell culture. Current Opinion in Biotechnology 9:518-521.

Pfeifer TA, Guarna MM, Kwan EM, Lesnicki G, Theilmann DA, Grigliatti TA, Kilburn DG (2001) Expression analysis of a modified factor X in stably transformed insect cell lines. Protein Expression and Purification 23:233-241.

Pharmingen (1999) Baculovirus Expression Vector System Manual. 6th edition.

QIAGEN (2003) The QIAexpressionist™. A handbook for high-level expression and purification of 6x His-tagged proteins. 5th edition.

Qian P, Li X, Tong G, Chen H (2003) High-level expression of the ORF6 gene of porcine reproductive and respiratory syndrome virus (PRRSV) in *Pichia pastoris*. Virus Genes 27:189-196.

Qian X, Davis AA, Goderie SK, Temple S (1997) FGF2 concentration regulates the generation of neurons and glia from multipotent cortical stem cells. Neuron 18:81-93.

Reuss B, von Bohlen und Halbach O (2003) Fibroblast growth factors and their receptors in the central nervous system. Cell and Tissue Research 313:139-157.

Richard JP, Melikov K, Vives E, Ramos C, Verbeure B, Gait MJ, Chernomordik LV, Lebleu B (2003) Cell-penetrating peptides. A reevaluation of the mechanism of cellular uptake. Journal of Biological Chemistry 278:585-590.

Rocha EP (2004) Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Research 14:2279-2286.

Romisch K (2005) Endoplasmic reticulum-associated degradation. Annual Review of Cell and Developmental Biology 21:435-456.

Salvatore D, Bonatti S, Di Natale P (1984) Lysosomal alpha-N-acetylglucosaminidase: purification and characterisation of the human enzyme. Bulletin of Molecular Biology and Medicine 9:111-121.

Sanfilippo S, Podosin R, Langer L, Good R (1963) Mental retardation associated with acid mucopolysaccharidurie (heparitin sulfate type). Journal of Pediatrics 63:837-838.

Sarramegna V, Demange P, Milon A, Talmont F (2002) Optimizing functional versus total expression of the human mu-opioid receptor in *Pichia pastoris*. Protein Expression and Purification 24:212-220.

Sasaki T, Sukegawa K, Masue M, Fukuda S, Tomatsu S, Orii T (1991) Purification and partial characterization of alpha-N-acetylglucosaminidase from human liver. Journal of Biochemistry (Tokyo) 110:842-846.

Schiffmann R, Brady RO (2002) New prospects for the treatment of lysosomal storage diseases. Drugs 62:733-742.

Schmitt J, Hess H, Stunnenberg HG (1993) Affinity purification of histidine-tagged proteins. Molecular Biology Reports 18:223-230.

Schwarze SR, Ho A, Vocero-Akbani A, Dowdy SF (1999) In vivo protein transduction: delivery of a biologically active protein into the mouse. Science 285:1569-1572.

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

Sharp PM, Stenico M, Peden JF, Lloyd AT (1993) Codon usage: mutational bias, translational selection, or both? Biochemical Society Transactions 21:835-841.

Sharp PM, Tuohy TM, Mosurski KR (1986) Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Research 14:5125-5143.

Shin HS, Cha HJ (2002) Facile and statistical optimization of transfection conditions for secretion of foreign proteins from insect Drosophila S2 cells using green fluorescent protein reporter. Biotechnology Progress 18:1187-1194.

Sinclair G, Choy FY (2002) Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, *Pichia pastoris*. Protein Expression and Purification 26:96-105.

Sinclair G, Pfeifer TA, Grigliatti TA, Choy FY (2006) Secretion of human glucocerebrosidase from stable transformed insect cells using native signal sequences. Biochemistry and Cell Biology 84:148-156.

Smith GE, Summers MD, Fraser MJ (1983) Production of human beta interferon in insect cells infected with a baculovirus expression vector. Molecular and Cellular Biology 3:2156-2165.

Sorensen MA, Kurland CG, Pedersen S (1989) Codon usage determines translation rate in *Escherichia coli*. Journal of Molecular Biology 207:365-377.

Spear ED, Ng DT (2005) Single, context-specific glycans can target misfolded glycoproteins for ER-associated degradation. Journal of Cell Biology 169:73-82.

Sreekrishna K, Brankamp RG, Kropp KE, Blankenship DT, Tsay JT, Smith PL, Wierschke JD, Subramaniam A, Birkenberger LA (1997) Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*. Gene 190:55-62.

Tamagawa K, Morimatsu Y, Fujisawa K, Hara A, Taketomi T (1985) Neuropathological study and chemico-pathological correlation in sibling cases of Sanfilippo syndrome type B. Brain Development 7:599-609.

Tanaka A, Kimura M, Lan HT, Takaura N, Yamano T (2002) Molecular analysis of the alpha-N-acetylglucosaminidase gene in seven Japanese patients from six unrelated families with mucopolysaccharidosis IIIB (Sanfilippo type B), including two novel mutations. J Human Genetics 47:484-487.

Tessitore A, Villani GR, Di Domenico C, Filocamo M, Gatti R, Di Natale P (2000) Molecular defects in the alpha-N-acetylglucosaminidase gene in Italian Sanfilippo type B patients. Human Genetics 107:568-576.

Theilmann DA, Stewart S (1992) Molecular analysis of the trans-activating IE-2 gene of *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus. Virology 187:84-96.

Toikkanen JH, Niku-Paavola ML, Bailey M, Immanen J, Rintala E, Elomaa P, Helariutta Y, Teeri TH, Fagerstrom R (2007) Expression of xyloglucan endotransglycosylases of *Gerbera hybrida* and *Betula pendula* in *Pichia pastoris*. Journal of Biotechnology 130:161-170.

Tomiya N, Narang S, Lee YC, Betenbaugh MJ (2004) Comparing N-glycan processing in mammalian cell lines to native and engineered lepidopteran insect cell lines. Glycoconjugate Journal 21:343-360.

Toonkool P, Metheenukul P, Sujiwattanarat P, Paiboon P, Tongtubtim N, Ketudat-Cairns M, Ketudat-Cairns J, Svasti J (2006) Expression and purification of dalcochinase, a beta-glucosidase from *Dalbergia cochinchinensis Pierre*, in yeast and bacterial hosts. Protein Expression and Purification 48:195-204.

Tyagi M, Rusnati M, Presta M, Giacca M (2001) Internalization of HIV-1 tat requires cell surface heparan sulfate proteoglycans. Journal of Biological Chemistry 276:3254-3261.

Vaags A (2004) Expression and purification of HIV-1 TAT protein transduction domain fused with acid B-glucosidase and enhanced green fluorescent protein. Master's Thesis. Department of Biology. University of Victoria, Victoria, BC, Canada.

Vad R, Nafstad E, Dahl LA, Gabrielsen OS (2005) Engineering of a *Pichia pastoris* expression system for secretion of high amounts of intact human parathyroid hormone. Journal of Biotechnology 116:251-260.

van de Kamp J, Niermeijer M, von Figura K, Giesberts M (1981) Genetic heterogeneity and clinical variability in the Sanfilippo syndrome (types A, B, and C). Clinical Genetics 20:152-160.

van Vactor D, Wall DP, Johnson KG (2006) Heparan sulfate proteoglycans and the emergence of neuronal connectivity. Current Opinion in Neurobiology 16:40-51.

Varenne S, Buc J, Lloubes R, Lazdunski C (1984) Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. Journal of Molecular Biology 180:549-576.

Vassileva A, Chugh DA, Swaminathan S, Khanna N (2001) Effect of copy number on the expression levels of hepatitis B surface antigen in the methylotrophic yeast *Pichia pastoris*. Protein Expression and Purification 21:71-80.

Vervecken W, Kaigorodov V, Callewaert N, Geysens S, De Vusser K, Contreras R (2004) In vivo synthesis of mammalian-like, hybrid-type N-glycans in *Pichia pastoris*. Applied and Environmental Microbiology 70:2639-2646.

Vives E, Brodin P, Lebleu B (1997) A truncated HIV-1 Tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus. Journal of Biological Chemistry 272:16010-16017.

Wang L, Denburg JL (1992) A role for proteoglycans in the guidance of a subset of pioneer axons in cultured embryos of the cockroach. Neuron 8:701-714.

Weber B, Blanch L, Clements PR, Scott HS, Hopwood JJ (1996) Cloning and expression of the gene involved in Sanfilippo B syndrome (mucopolysaccharidosis III B). Human Molecular Genetics 5:771-777.

Weber B, Guo XH, Kleijer WJ, van de Kamp JJ, Poorthuis BJ, Hopwood JJ (1999) Sanfilippo type B syndrome (mucopolysaccharidosis III B): allelic heterogeneity corresponds to the wide spectrum of clinical phenotypes. European Journal of Human Genetics 7:34-44.

Weber B, Hopwood JJ, Yogalingam G (2001) Expression and characterization of human recombinant and alpha-N-acetylglucosaminidase. Protein Expression and Purification 21:251-259.

Wegner G, Harder W (1986) Methylotrophic yeasts - Microbial growth on C1 compounds. In: van Verseveld H, Duine J (eds) Proceedings of the 5th International Symposium. Martinus and Nijhoff, Dordrecht, pp 139-149.

Werten MW, van den Bosch TJ, Wind RD, Mooibroek H, de Wolf FA (1999) High-yield secretion of recombinant gelatins by *Pichia pastoris*. Yeast 15:1087-1096.

Whitelock JM, Iozzo RV (2005) Heparan sulfate: a complex polymer charged with biological activity. Chemical Reviews 105:2745-2764.

Wildt S, Gerngross TU (2005) The humanization of N-glycosylation pathways in yeast. Nature Reviews Microbiology 3:119-128.

Wojczyk BS, Czerwinski M, Stwora-Wojczyk MM, Siegel DL, Abrams WR, Wunner WH, Spitalnik SL (1996) Purification of a secreted form of recombinant rabies virus glycoprotein: comparison of two affinity tags. Protein Expression and Purification 7:183-193.

Woo JH, Liu YY, Mathias A, Stavrou S, Wang Z, Thompson J, Neville DM, Jr (2002) Gene optimization is necessary to express a bivalent anti-human anti-T cell immunotoxin in *Pichia pastoris*. Protein Expression and Purification 25:270-282.

Wraith JE (2006) Limitations of enzyme replacement therapy: Current and future. Journal of Inherited Metabolic Disease 29:442-447.

Xia H, Mao Q, Davidson BL (2001) The HIV Tat protein transduction domain improves the biodistribution of beta-glucuronidase expressed from recombinant viral vectors. Nature Biotechnology 19:640-644.

Xiong AS, Yao QH, Peng RH, Zhang Z, Xu F, Liu JG, Han PL, Chen JM (2006) High level expression of a synthetic gene encoding Peniophora lycii phytase in methylotrophic yeast *Pichia pastoris*. Applied Microbiology and Biotechnology 72:1039-1047.

Yadava A, Ockenhouse CF (2003) Effect of codon optimization on expression levels of a functionally folded malaria vaccine candidate in prokaryotic and eukaryotic expression systems. Infection and Immunity 71:4961-4969.

Yamaguchi Y (2001) Heparan sulfate proteoglycans in the nervous system: their diverse roles in neurogenesis, axon guidance, and synaptogenesis. Seminars in Cell and Developmental Biology 12:99-106.

Yogalingam G, Hopwood JJ (2001) Molecular genetics of mucopolysaccharidosis type IIIA and IIIB: diagnostic, clinical, and biological implications. Human Mutation 18:264-281.

Yogalingam G, Weber B, Meehan J, Rogers J, Hopwood JJ (2000) Mucopolysaccharidosis type IIIB: characterisation and expression of wild-type and mutant recombinant alpha-N-acetylglucosaminidase and relationship with sanfilippo phenotype in an attenuated patient. Biochimica and Biophysica Acta 1502:415-425.

Yu WH, Zhao KW, Ryazantsev S, Rozengurt N, Neufeld EF (2000) Short-term enzyme replacement in the murine model of Sanfilippo syndrome type B. Molecular Genetics and Metabolism 71:573-80.

Zay A (2006) Expression, purification, and characterization of human H-Protein, a member of the glycine cleavage system. Master's Thesis. Department of Biology. University of Victoria, Victoria, BC Canada.

Zhang F, Murhammer DW, Linhardt RJ (2002) Enzyme kinetics and glycan structural characterization of secreted alkaline phosphatase prepared using the baculovirus expression vector system. Applied Biochemistry and Biotechnology 101:197-210.

Zhao HG, Li HH, Bach G, Schmidtchen A, Neufeld EF (1996) The molecular basis of Sanfilippo syndrome type B. Proceedings of the National Academy of Sciences USA 93:6101-6105.

Zhao KW, Neufeld EF (2000) Purification and characterization of recombinant human alpha-N-acetylglucosaminidase secreted by Chinese hamster ovary cells. Protein Expression and Purification 19:202-211.

Ziegler A, Nervi P, Durrenberger M, Seelig J (2005) The cationic cell-penetrating peptide CPP(TAT) derived from the HIV-1 protein TAT is rapidly transported into living fibroblasts: optical, biophysical, and metabolic evidence. Biochemistry 44:138-148.