OVEREXPRESSION OF HUMAN PROCESSING ALPHA-GLUCOSIDASE I AND FUNCTIONAL STUDY OF YEAST PROCESSING ALPHA-GLUCOSIDASE I

by

Zhong Peng

B.Sc., Northwest Agriculture and Forestry University, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Food Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2016

Abstract

Processing α-glucosidase I (Glu I) holds a key regulatory position in the N-glycosylation pathway. Glu I mutations in humans have severe or fatal outcomes. Despite the important role of human Glu I (MOGS), very little information is available regarding the function, mechanism and structure of the enzyme. The lack of methods to produce sufficient quantities of the enzyme is a major drawback in carrying out further structural and mechanistic studies. Therefore, this thesis was focused on establishing a *Pichia pastoris* expression system for MOGS. MOGS with an N-terminal histidine (6xHis) tag was overexpressed in *P. pastoris* and purified by FPLC using a His-trap column. The purified MOGS was N-linked glycosylated and was enzymatically active (specific activity of 735 U/mg) against the synthetic substrate α-D-Glc1,2α-D-Glc1,3α-D-Glc–OCH_3. However, the overexpression of MOGS in *P. pastoris* was not reproducible and it was overexpressed intracellularly in subsequent expression trials. Therefore, overexpression of MOGS in *E. coli* was attempted, but the expressed MOGS accumulated as inclusion bodies with either a MBP (maltose binding protein) tag or a 6xHis-tag. Two methods (urea denaturation-refolding and freeze-thawing) were used to purify MOGS from inclusion bodies. However, less than 0.5 mg from one litre culture was recovered and it was not enzymatically active. As the overexpression of active MOGS in either *P. pastoris* or *E. coli* was not successful, further characterization of Glu I was continued with the study of the N-terminal function of *Saccharomyces cerevisiae* Glu I (Cwh41p). The Cwh41p clones with N-terminal truncations were constructed and the proteins were overexpressed in *E. coli*. Among the six truncations, five truncations ($Cwh41_{\Delta 1\text{-}167}p$, $Cwh41_{\Delta 1\text{-}93}p$, $Cwh41_{\Delta 1\text{-}74}p$, $Cwh41_{\Delta 1\text{-}54}p$ and $Cwh41_{\Delta 1\text{-}45}p$) were expressed mainly in inclusion bodies and the eluted fractions from His-trap column were not

enzymatically active. However, the eluted fractions of Cwh41$_{\Delta 1\text{-}39}$p and the internal control (Cwh41$_{\Delta 1\text{-}34}$p) were active with a specific activity of 266 U/mg and 2000 U/mg, respectively. This indicates that the residues 35-45 of Cwh41p may be essential for the proper folding of Cwh41p. I propose that the α-helix (NH1) of Cwh41p plays a positive role in protein folding by interacting with C-domain of the enzyme.

Preface

Dr. Christine H. Scaman and I designed the research; I conducted the experiments and data analysis, as well as wrote the thesis. Dr. Christine H. Scaman critically reviewed the thesis, provided feedback, and made editorial changes.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AOX	alcohol oxidase
BMMY	buffered methanol-complex medium
bp	base pair
CAST	castanospermine
CDG	congenital disorder of glycosylation
CDG-IIb	congenital disorder of glycosylation type IIb
COS	CV-1 origin with SV40
CWH	calcofluor white hypersensitive
DEAE	diethylaminoethyl
DENV	dengue virus
DEPC	diethylpyrocarbonate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EC	enzyme commission
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
ERAD	endoplasmic reticulum associated degradation of protein
FPLC	fast protein liquid chromatography
Glc	glucose
GlcNAc	*N*-acetyl glucosamine

Glu I	processing α-glucosidase-I
GST	glutathione S-transferase
HBV	hepatitis B virus
HIV	human immunodeficiency virus
HRP	horseradish peroxidase
IPTG	isopropyl β-D-1-thiogalactopyranoside
kDa	kilodaltons
Ki	enzyme-inhibitor complex dissociation constant
Km	Michaelis constant
LB	Luria-Bertani medium
Man	mannose
MBP	maltose binding protein
MOGS	mannosyl-oligosaccharide glucosidase
μg	microgram
μL	microliter
mg	milligram
min	minute(s)
ml	milliliter
mM	millimolar
NBS	*N*-bromosuccinimide
nm	nanometer
nmole	nanomole
NMR	nuclear magnetic resonance

OD600	optical density at 600 nm
OST	oligosaccharyltransferase
PAGE	polyacrylamide gel electrophoresis
PAS	periodic acid-Schiff
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PGO	peroxidase and glucose oxidase
PMSF	phenylmethylsulfonyl fluoride
PNGase F	peptide-N-glycosidase F
P. pastoris	*Pichia pastoris*
PVDF	polyvinylidene fluoride
rpm	revolutions per minute
S. cerevisiae	*Saccharomyces cerevisiae*
SDS	sodium dodecyl sulfate
TNM	tetranitromethane
Tris	2-Amino-2-(hydoxymethyl)-1,3-propanediol
U	units of activity
UDP	uridine diphosphate
UGGT	UDP-Glc: glycoprotein glucosyltransferase
UV	ultraviolet
YPD	yeast peptone dextrose

Nucleotide base abbreviations

A	Adenine
C	Cytosine
G	Guanine
T	Thymine

Common amino acid abbreviations

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine

P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Acknowledgements

First of all, I would like to offer my sincere gratitude to my supervisor, Dr. Christine H. Scaman, for giving me the opportunity to study food science and joining in her research group. Thanks for her constant support and guidance throughout my master's study.

I would like to thank my supervisory committee members, Dr. Eunice Li-Chan, Dr. Jim Kronstad and Dr. Vivien Measday, for their time commitment and suggestions throughout my research.

I am extremely thankful for the financial support from funding agency NSERC (grant to Dr. Scaman) throughout my master's study and from scholarships provided by the Faculty of Land and Food Systems at UBC.

I thank my lab mate Venkatrao Konasani for his technical assistance and help.

Special thanks are owed to my parents, whose have supported me throughout my years of education, both morally and financially.

Dedication

I would like to dedicate this thesis to my family for their unconditional love, encouragement and support.

1. Introduction

N-linked glycosylation involves attachment of $Glc_3Man_9GlcNAc_2$ on nascent proteins during translation, followed by processing into a diverse array of glycan structures. Processing α-glucosidase I (Glu I) initiates these processing reactions by selectively catalyzing the first trimming step - removal of the α-1,2 linked glucose residue (Herscovics, 1999a). Then other glycosidases and glycosyltransferases continue with the processing steps (Herscovics, 1999a). Specifically, processing α-glucosidase II cleaves the two α-1,3 linked glucose residues from the $Glc_2Man_9GlcNAc_2$ precursor. Then endoplasmic reticulum (ER) resident and Golgi resident mannosidases remove up to four α-1,2 linked mannose residues. This pathway results in the transformation of simple high mannose N-linked glycans to complex and hybrid N-linked glycans (Herscovics, 1999a). Glu I enzyme has been found in virtually all eukaryotes, for example, in *Saccharomyces cerevisiae* (Bause *et al*.1986), *Aspergillus niger* (Miyazaki *et al*. 2011), mung bean seedlings (Zeng & Elbein, 1998) and different mammalian tissues (Hettkamp *et al*. 1984; Shailubhai *et al*. 1987; Bause *et al*. 1989; Shailubhai, 1991; Kalz-Fuller *et al*.1995).

Glu I, the first processing enzyme at the beginning of the trimming reactions, holds a key regulatory position in the N-glycosylation pathway. Glu I deficiency prohibits further glycoprotein processing and leads to impaired biological function related to protein folding, stabilizing protein conformation and cell-cell interaction (Helenius & Aebi, 2004). While an alternative pathway utilizes an endo α1,2-mannosidase that can remove Glc_3Man fragment from

the newly synthesized glycosylated polypeptide (Fujimoto and Kornfeld, 1991), it cannot compensate for the lack of Glu I and processing α-glucosidase II.

This enzyme is conserved among eukaryotes; the enzyme purified from different sources have similar substrate specificity, optimum pH and inhibitor sensitivity. Among the eukaryotic enzymes, *S. cerevisiae* Glu I has been studied the most (Faridmoayer & Scaman, 2005; Faridmoayer & Scaman, 2007; Barker & Rose, 2013). *S. cerevisiae* Glu I could be used as a suitable model to understand human Glu I. They share 37% amino acid sequence identity in the C-terminal catalytic domain region (Romaniouk & Vijay, 1997). It is practical to anticipate similar structure characteristics in catalytic domain between the two enzymes based on the identity level (Rost, 1999). They also have similar substrate specificity to the glucotriose oligosaccharide containing α(1,2) and α(1,3) linkages (Barker & Rose, 2013). Heterologous overexpression of *S. cerevisiae* Glu I without the N-terminus (amino acid residues 1-525) has been carried out (Faridmoayer & Scaman, 2007; Barker & Rose, 2013). However, there was no active protein expression, indicating that the N-terminus may be essential for the proper assembly and folding of *S. cerevisiae* Glu I.

There are few studies on human Glu I. Glu I mutations have been reported in humans resulting in a dysfunctional enzyme and disease termed congenital disorder of glycosylation type IIb (CDG-IIb), which has serious (Sadat *et al*. 2014) or fatal (Praeter *et al*. 2000) outcomes. The first CDG-IIb case was a heterozygote individual carrying two missense mutations in Glu I gene: one allele harbored a G to C transition at nucleotide 1587, resulting in the substitution of Arg at position 486 by Thr, the other allele was a T to C transition at nucleotide 2085, resulting in the

substitution of Phe at position 652 by Leu (Praeter *et al*. 2000). Sadat *et al* (2014) reported two CDG-IIb siblings with mutations in Glu I gene carrying a C to T transition at nucleotide 370, resulting in a substitution of Q124X, and two missense changes, C to A transition at nucleotide 65, resulting in A22E; and G to A transition at nucleotide 329, resulting in R110H. Interestingly, cells from the two siblings with CDG-IIb have an increased resistance to particular viral infections, such as human immunodeficiency virus (HIV) and influenza viruses, which are glycosylation-dependent enveloped viruses (Sadat *et al*. 2014). It has been observed that Glu I inhibitors possess potential therapeutic application to control viral amplification/diseases, since several viral proteins contain viral glycoproteins which are required in the viral life cycle, either in virus assembly and secretion and/ or infectivity (Mehta *et al*. 1998).

However, despite the important role of human Glu I and the potential therapeutic application of its specific inhibitors, very little information is available regarding the function, mechanism and structure of human Glu I. The lack of methods to produce sufficient quantities of the enzyme is a major drawback in carrying out further structural and mechanistic studies. Previous studies used mammalian cell line-COS cell (virus SV40 incorporated into CV-1 cells) as the host strain to express human Glu I (Kalz-Fuller *et al*. 1995; Völker *et al*. 2002). However, transformed COS cells can only produce the protein of interest transiently and this technology is relatively expensive (Yin *et al*. 2007). In contrast, overexpressing the eukaryotic protein in *P. pastoris* could be a more straightforward process. The methylotrophic yeast *P. pastoris* has been used for many human protein preparations (Lei *et al*. 2012; Huang *et al*. 2014) as well as for the *S. cerevisiae* Glu I overexpression (Barker *et al*. 2011). *P. pastoris* is easier to genetically manipulate and culture than mammalian cells (Daly & Hearn, 2005). As well, *P. pastoris*

expression system enables production of large amounts of the target proteins (Daly & Hearn, 2005). Another expression system by using *E. coli* has been used extensively for heterologous protein expression (including human proteins) because of its rapid growth rate and relatively low cost (Yin *et al*. 2007), so human Glu I can be also overexpressed in *E. coli*.

Research Hypotheses and Objectives

The first hypothesis is that *P. pastoris* expression system can be used to overexpress heterologous human Glu I.

Objective 1. To construct a human Glu I *P.* pastoris expression system and purify the protein

The second hypothesis is that *E. coli* expression system can be used to overexpress heterologous human Glu I.

Objective 2. To construct a human Glu I *E. coli* expression system and purify the protein

The third hypothesis is that N-terminal domain (residues 35-525) of *S. cerevisiae* Glu I is essential for its proper assembly and folding.

Objective 3. To construct *S. cerevisiae* Glu I truncations using the *E. coli* expression system and purify the truncations

Objective 4. To determine enzymatic activities of the truncated *S. cerevisiae* Glu I.

2. Literature Review

2.1. N-linked glycosylation

N-linked glycosylation is the most common covalent protein modification in eukaryotic cells (Helenius & Aebi, 2004). It is predicted that more than half of all eukaryotic proteins are glycoproteins and about 90% of these carry N-linked glycans (Apweiler, 1999). N-glycosylation is a complex process occurring within the secretory pathway in the endoplasmic reticulum and Golgi apparatus of eukaryotic cells that requires the participation of glycosidases as well as glycosyltransferases. The function of the process is to covalently link oligosaccharides to an asparagine residue of the nascent protein side chain via an N-glycosidic bond, then to modify the glycan for its biological role. The N-glycan biosynthesis process consists of three phases. (1) assembly of a high mannose oligosaccharide on the lipid carrier dolichyl-pyrophosphate; (2) transfer of this oligosaccharide to newly formed proteins; and (3) processing and trimming of the protein-bound oligosaccharide to complex and hybrid N-linked glycans (Herscovics, 1999a). Assembly of lipid linked oligosaccharide $Glc_3Man_9GlcNAc_2$ occurs via the dolichol cycle at the membrane of the endoplasmic reticulum (ER) and is highly conserved in virtually all eukaryotic cells (Kornfeld & Kornfeld, 1985). In eukaryotes, the lipid linked oligosaccharide is synthesized by a step-wise addition of sugars glucose (Glc), mannose (Man), and N-acetylglucosamine (GlcNAc) to the lipid carrier dolichyl-pyrophosphate catalyzed by oligosaccharyl transferase in the ER membrane (Verbert *et al*. 1987). Seven sugars ($Man_5GlcNAc_2$) are added on the cytosolic surface after which the sugar moiety is translocated and “flipped” across the membrane to the ER luminal side by ER flippase (Marolda *et al*. 2002). On the luminal side, the oligosaccharide chain

is further assembled into a triantennary structure. The final completed precursor has a sequence of $Glc_3Man_9GlcNAc_2$-pyrophosphate-dolichol.

With complete synthesis of the precursor, $Glc_3Man_9GlcNAc_2$ is transferred *en bloc* from the lipid donor onto nascent polypeptide chains by oligosaccharyl transferase (OST) (Parodi *et al*. 1972). Oligosaccharides are attached to the side chain of asparagine residues in Asn-X-Ser/Thr sites, or more rarely Asn-X-Cys/Val or Asn-Gly by oligosaccharyl transferase (Zielinska *et al*. 2010). Processing of the N-linked glycans begins immediately following the transfer of the $Glc_3Man_9GlcNAc_2$. During this stage, the N-linked glycans act as recognition "tags" that allow proteins to interact with a variety of molecular chaperones, glycosidases, and glycosyltranferases (Helenius & Aebi, 2004). The two glucose residues of glycan are trimmed by the processing α-glucosidases I and II to generate a monoglucosylated glycan ($Glc_1Man_9GlcNAc_2$). As a result, the monoglucosylated glycan facilitates binding of the protein to lectin chaperones calnexin/calreticulin (Ou *et al*. 1993), a process that is believed to promote proper folding (Hebert *et al.* 1995). Following interaction with lectins, the final glucose residue is removed by glucosidases II, and the protein no longer binds to the lectins. During this stage, UDP-glucose: glycoprotein glucosyltransferase (UGGT) inspects the structure or folding status of the proteins (Hebert & Molinari, 2012). If the glycoprotein is not properly folded, the UGGT adds a terminal glucose, which causes rebinding to the lectin chaperones (Hebert & Molinari, 2012) (Figure 1). Most of the glycoproteins are released from lectin chaperones in a native, transport competent state. At this stage, glucosidase II removes the final glucose and the substrates are partially demannosylated, and native glycoproteins are exported from the ER (Hebert & Molinari, 2007). However, extensively misfolded polypeptides are ignored by UGGT, and are prepared for

dislocation into the cytosol and degradation via the ER associated degradation (ERAD) pathway (Hebert & Molinari, 2007). When the glycoprotein moves to the Golgi complex, the glycan chains undergo further trimming of mannoses. It is at this point that the N-glycosylation pathway diverges between multicellular and unicellular organisms. Following the removal of glucose and mannose residues, Golgi glycosyltransferases add different sugars in a species and cell-specific manner. In yeast, additional mannose residues are added by Golgi mannosyltransferases to form mature high mannose N-glycans (Herscovics, 1999b); In mammalian cells, glycosyltransferases add different sugars, including galactose, fucose, sialic acid and GlcNAc, to form complex and hybrid N-glycans (Herscovics, 1999a).

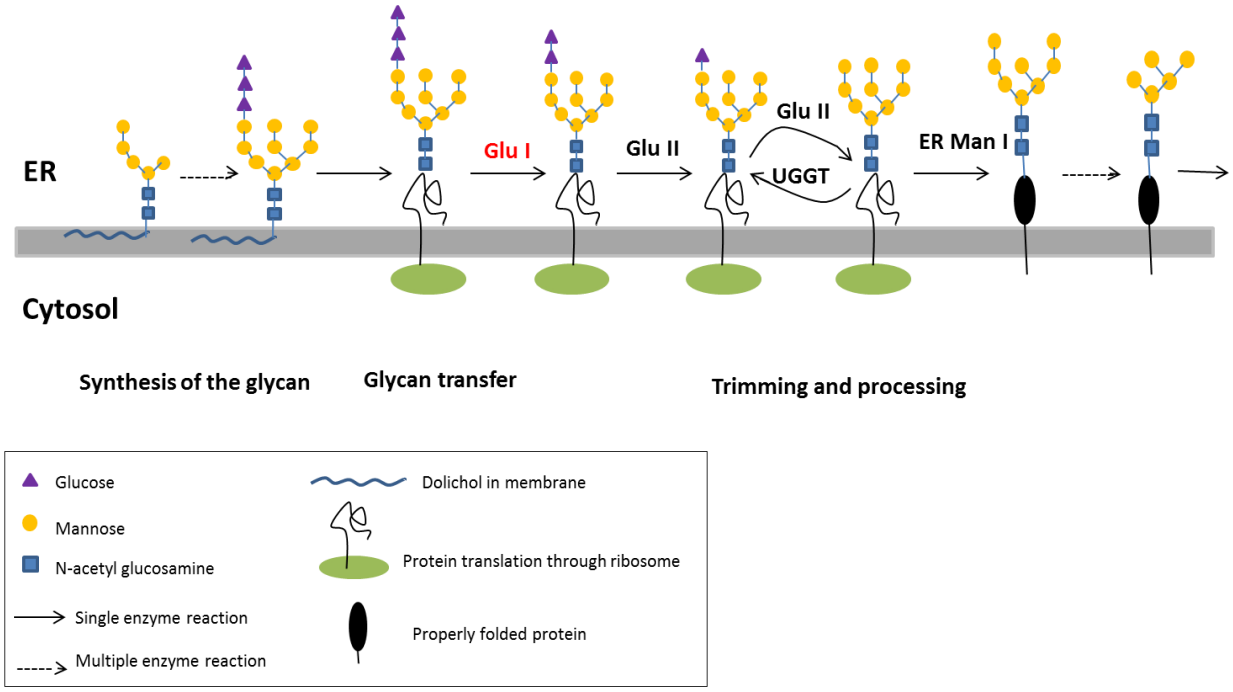


Figure 1. The eukaryotic N-linked glycosylation pathway in ER. Glycans are built on the dolichol, translocated into the ER and assembled to the final $Glc_3Man_9GlcNAc_2$ precursor. This precursor is transferred to a nascent protein. Subsequently, the glycan is trimmed and processed to a large variety of structures. Adapted from Helenius & Aebi (2001).

2.2. Alternative glycan processing pathway

In vitro studies show that the deficiency of processing α-glucosidases I and II affects overall glycoprotein synthesis but does not have an effect on the maturation of certain glycoproteins. Formation of matured N-linked glycans has been detected in processing glucosidase deficient cells (Völker *et al*. 2002; Fujimoto & Kornfeld, 1991). Lubas & Spiro attributed this glucosidase independent deglucosylation to Golgi endo α1,2-mannosidase (Lubas & Spiro, 1987; Lubas & Spiro, 1988). Endo α1,2-mannosidase cleaves the α1,2 mannosidic bond between the glucose substituted mannose residue, and releases $Glc_{1\text{-}3}Man_1$ from $Glc_{1\text{-}3}Man_9GlcNAc_2$. A study found

that in the fibroblasts from a patient with processing α-glucosidases I deficiency, greater than 80% of synthesized N-glycoproteins were correctly processed, with only about 16% of the N-glycans being arrested at the $Glc_3Man_{7-9}GlcNAc_2$ stage. On the other hand, endo α1,2-mannosidase activity increased (Völker *et al*. 2002), indicating that the endo α1,2-mannosidase dependent deglucosylation pathway was able to bypass the glucosidase I defect to a considerable extent. Nevertheless, α-glucosidases I deficiency can still lead to impaired biological function as discussed below. Another study found that 40% of the oligosaccharides were de-glucosylated and processed correctly in the mouse cells with processing α-glucosidases II deficiency. (Fujimoto & Kornfeld, 1991).

2.3. N-glycosylation loss in congenital disorders of glycosylation

The congenital disorders of glycosylation (CDG) are a subset of genetic defects affecting primarily N-glycan assembly (Jaeken, 2003). Many of the causal mutations affect key biosynthetic enzymes. In addition, defects in chaperones and Golgi-trafficking complexes can also impair several glycosylation pathways (Freeze, 2006; Varki *et al*. 2009). The broad clinical features involve many organ systems and functions of the gastrointestinal, hepatic, visual, and immune systems (Varki *et al*. 2009). Since the first description of "Congenital Disorders of Glycosylation (CDGs)" in 1980, more than 100 different CDGs have been described affecting protein and lipid glycosylation. Nearly 60 different disorders in N-linked glycosylation have been reported by the year of 2013 (Freeze *et al*. 2014). For convenience, CDG defects are divided into types I and II. All defects in the biosynthesis of the dolichyl pyrophosphate-linked oligosaccharide *N*-glycan precursor and its transfer to proteins in the endoplasmic reticulum (ER) fall into type I; whereas defects in all subsequent processing steps are in type II (Freeze, 2006).

Hypoglycosylated proteins occur in all type I CDGs because of the synthesis of insufficient and/or incomplete glycan precursors. Since complete loss of *N*-glycosylation is lethal, all CDG-I patients have "leaky" mutations and some amount of $Glc_3Man_9GlcNAc_2$-pyrophosphate-dolichol could be found in cells (Schachter, 2001). The type II CDGs seem to be less frequent than type I disorders, but this might reflect a lack of awareness rather than the actual prevalence (Freeze, 2006).

In the past decade, our understanding of the complexity of the various glycosylation pathways suggests that many congenital diseases of unknown etiology will turn out to be CDGs (Schachter, 2001). One major reason is the diversity of clinical presentation and severity in different CDGs. Furthermore, improved diagnostic testing and technological advances in next generation sequencing has uncovered the existence of even more glycosylation disorders (Freeze *et al*. 2014). As in 2013, a new glycosylation disorder has been reported, on average, every 17 days (Freeze *et al*. 2014). Initially the nomenclature for glycosylation disorders was "CDG" followed by the type of transferrin isoelectric focusing pattern (I or II), and the lower case letter indicating different subtypes in chronological order of identification of the defective gene (ie, CDG-Ia or CDG-IIa). With the identification of many more types of glycosylation defects, this nomenclature was simplified with the gene name, followed by "CDG" (ie, PMM2-CDG, representing CDG-Ia), and sometimes both systems coexist with the older nomenclature in parenthesis (Jaeken *et al*. 2008).

2.4. Processing α-glucosidase I

2.4.1. Classification and Glu I orthologues

Processing α-glucosidases I **(**EC.3.2.1.106**)** belongs to the glycoside hydrolase family 63 and GH-G clan of $(\alpha/\alpha)_6$ barrel catalytic domain containing glycoside hydrolases (Henrissat *et al*.1991) (http://www.cazypedia.org/index.php/Glycoside_Hydrolase_Family_63). Processing α-glucosidases I (Glu I) catalyzes the onset of processing reactions by trimming the terminal α-1,2-linked glucose residue (Figure 1). It has been found to be present from lower to higher eukaryotes. *S. cerevisiae* Glu I gene (*CWH41*) was first identified by Romero *et al*. (1997). Disruption of the *CWH41* gene led to a K1 killer toxin-resistant phenotype and reduced level of cell wall β1,6-glucan (Jiang *et al*. 1996). A further study showed that only the cells obtained from *CWH41Δ* null mutants were unable to released glucose from the synthetic substrate $Glc_3O(CH_2)_8COOCH_3$ while the cells that contained the *CWH41* gene were able to release glucose residues. In addition, Cwh41p was the only gene product with high similarity to human Glu I among the complete *S. cerevisiae* genome, indicating that *CWH41* encodes yeast glucosidase I (Romero *et al*. 1997). Cwh41p consists of 833 amino acids with a molecular weight of 98 kDa (Romero *et al*. 1997). Cwh41p has a type II transmembrane topology and contains a short N-terminal cytoplasmic tail (residues 1-10), a transmembrane region (residues 11-28) and a major ER luminal C-terminal domain (residues 29-833) (Figure 2), it also has a signal peptide cleavage site between residues Ala 24 and Thr 25 of the transmembrane sequence (Dhanawansa *et al*. 2002).

Human Glu I that is often referred to as “MOGS” (mannosyl-oligosaccharide glucosidase), also has a type II transmembrane topology, consisting of a short N-terminal cytosolic tail (residues 1-38), a single transmembrane domain (residues 39-59, by sequence analysis of UniProtKB-Q13724), and a large C-terminal catalytic domain (residues 60-387, by sequence analysis of UniProtKB-Q13724) (Kalz-Fuller *et al*. 1995) (Figure 2). Further study revealed that the MOGS cDNA sequence comprised four exons separated by three introns. The first exon encodes the cytoplasmic tail and transmembrane domain; exon-2 and exon-3 encode for short polypeptides of 75 and 65 amino acid residues, respectively, the fourth encodes the putative catalytic domain of the enzyme (~64 kDa) (Völker *et al*. 2002). *MOGS* gene is located on chromosome 2 and the full length MOGS consists of 834 amino acids, corresponding to molecular mass of ~92 kDa (Kalz-Fuller *et al*. 1995). In the C-terminal catalytic domain region of yeast and human Glu I, the enzymes share 34% identity (Romaniouk & Vijay, 1997).

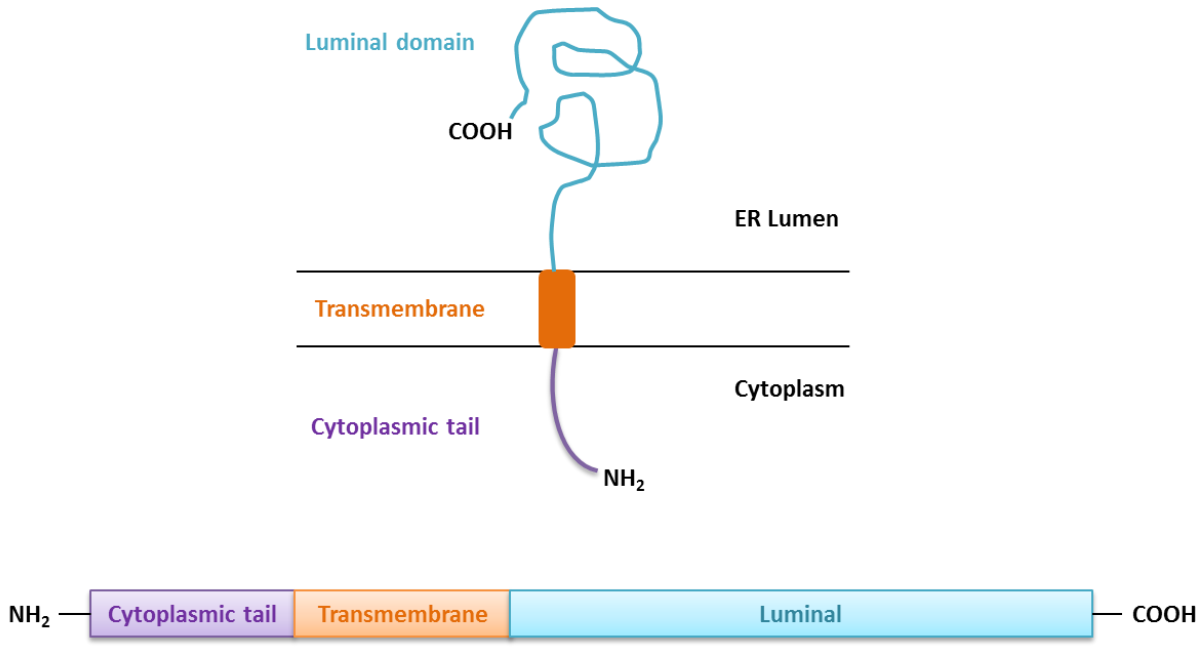


Figure 2. Topology of type II transmembrane protein.

2.4.2. Enzyme activity assay of Glu I

Two methods have been developed to determine Glu I activity. The first method is based on a radiolabeled natural substrate and the other method is a spectrophotometric approach using a synthetic trisaccharide substrate. The radiolabeled [^{14}C] or [^{3}H]glucose-oligosaccharide substrate has been used for measuring α-glucosidase I activity (Grinna & Robbins, 1979; Kilker *et al*. 1981). The mammalian microsomes were mixed with UDP-[U-^{14}C]glucose or UDP-[6-^{3}H]glucose and the labeled dolichyl pyrophosphate oligosaccharides were extracted with chloroform/methanol. The labeled lipid-oligosaccharide was hydrolyzed by HCl to cleave the pyrophosphate bond and neutralized by NaOH. Then, the water-soluble oligosaccharides were isolated and purified by a Bio-Gel P-4 column (Grinna & Robbins, 1979; Kilker *et al*. 1981). The α-glucosidase I released glucose from the labeled substrate and the enzyme activity was measured based on the quantity of released glucose. The released glucose was isolated from the remaining labeled mixture by paper chromatography and the radioactivity was measured. Any loss of radioactivity from the origin was accompanied by an equivalent amount of radioactivity migrating as glucose (Grinna & Robbins, 1979; Kilker *et al*. 1981). Although this approach is very sensitive and can measure very low quantities of activity, it is labor intensive and time consuming. Moreover, this method can result in inequity and inconsistency when measuring enzyme activity, because with each oligosaccharide substrate preparation, the glucose is labeled randomly (Neverova *et al*. 1994). To overcome the obstacle of the radiolabeled method, a spectrophotometric assay using a synthetic trisaccharide substrate α-D-Glc1,2α-D-Glc1,3α-D-Glc–O$(CH_2)_8COOCH_3$ (Neverova *et al*. 1994) or α-D-Glc1,2α-D-Glc1,3α-D-Glc–ManOCH_3 (Barker *et al*. 2011) was developed. The α-glucosidase I activity is measured based on the

amount of released glucose from the synthetic substrate. The terminal glucose is released by the enzyme and oxidized by glucose oxidase to generate glucuronic acid and hydrogen peroxide (H_2O_2). The peroxidase reacts with H_2O_2 and converts colorless o-dianisidine to the brown oxidized form (Neverova *et al*. 1994) (Figure 3). This method is rapid, convenient and can be used to determine enzyme kinetics. The Km value of α-D-Glc1,2α-D-Glc1,3α-D-Glc–$O(CH_2)_8COOCH_3$ for the yeast Glu I was determined to be 1.28 mM (Neverova *et al*. 1994). In comparison, the modified trisaccharide substrate, α-D-Glc1,2α-D-Glc1,3α-D-Glc–$ManOCH_3$, has a reported Km value of 1.26 mM with yeast Glu I (Barker *et al*. 2011). Most importantly, this method provides reproducible and consistent enzyme activity. This indicates that the synthetic substrate can provide necessary structural features for interacting with the enzyme active site.

The pH optimum of purified Glu I isolated from all orthologs sources is in the near neutral range. For example, the optimum for Glu I isolated from calf and pig microsomes is pH 6.2–6.4 (Hettkamp *et al*. 1984; Bause *et al*. 1989), and for Glu I isolated from bovine or rat mammary glands or yeast is pH 6.6–7.0 (Shailubhai *et al*. 1987; Shailubhai, 1991). No cation requirement was found for α-glucosidase I (Shailubhai *et al*. 1987; Bause *et al*. 1989).

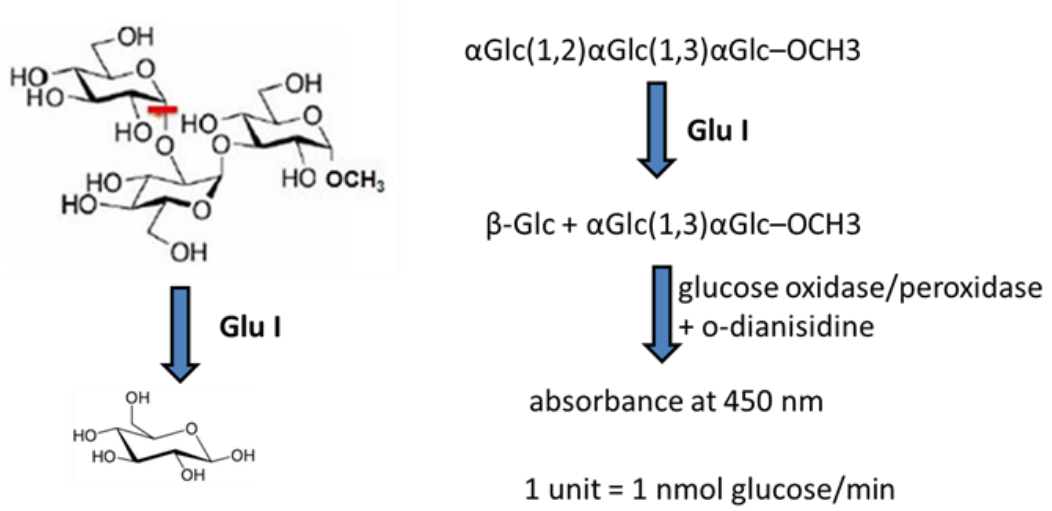


Figure 3. Glu I assay using a synthetic trisaccharide.

2.4.3. Substrate specificity of Glu I

Glu I in all eukaryotes catalyzes the removal of α1,2 linked glucose from the natural substrate $Glc_3Man_9GlcNAc_2$. However, this enzyme does not hydrolyze α1,2 linked glucose from the disaccharide kojibiose (α-D-Glc1,2α-D-Glc) (Neverova *et al*. 1994). On the contrary, kojibiose had an inhibitory effect on Glu I isolated from yeast, plant and mammalian sources when $Glc_3Man_9GlcNAc_2$ was used as substrate (Bause *et al*. 1986; Zeng & Elbein, 1998; Bause *et al*. 1989). Mammalian and yeast Glu I do not hydrolyze α1,3-glucosidic linkage of $Glc_3Man_9GlcNAc_2$ or simple sugar substrates such as *p*-nitrophenyl α-glucosides and aryl-α-glucosides (Bause *et al*. 1989; Bause *et al*. 1986). The minimal substrate requirement of Glu I is the tri-glucoside α-Glc1,2-α-Glc1,3-α-Glc (Neverova *et al*. 1994; Barker *et al*. 2011). Unlike mammalian and yeast Glu I, the *E. coli* Glu I ortholog Ygjk showed the highest activity for the α1,3-glucosidic linkage of nigerose, and it also hydrolyzed trehalose, kojibiose and

maltooligosaccharides with low activity, suggesting that YgjK is a glucosidase with relaxed specificity for sugars (Kurakata *et al*. 2008).

2.4.4. Inhibitors of α-glucosidases

Several inhibitors have been found to be able to inhibit α-glucosidase I, which impairs secretion and maturation of glycoproteins. Among these inhibitors, one compound is the most frequently used inhibitor: deoxynojirimycin. Deoxynojirimycin is a glucose analogue with a NH group substituting for the oxygen atom in the pyranose ring. Deoxynojirimycin produced by *Bacillus* species, is the reduced form of nojirimycin, an antibiotic synthesized by several strains of *Streptomyces* (Saunier *et al*. 1982). Deoxynojirimycin and its derivatives including N-butyl-deoxynojirimycin and N-nonyl-deoxynojirimycin inhibit both α-glucosidase I and α-glucosidase II (Mehta *et al*. 1998). Saunier *et al* (1982) observed that 20 μM deoxynojirimycin showed 50% inhibition (K_i) for yeast α-glucosidase I activity while the K_i value for α-glucosidase II activity was 2 μM. However, the reported K_i values of deoxynojirimycin were different for the two glucosidases in different organisms. For example, Hettkamp *et al* (1984) studied calf microsomes and reported a K_i value of 3 μM deoxynojirimycin for α-glucosidase I and 20 μM for α-glucosidase II. Except for the deoxynojirimycin, other Glu I inhibitors have also been studied. Pan *et al* (1983) reported a plant alkaloid castanospermine (CAST) that was found from the seeds of Australian plant *Castanospermum austral*, was able to inhibit Glu I activity. The K_i value was found about 1 μM for the Glu I from mung bean seedlings (Zeng & Elbein, 1998). Moreover, a disaccharide kojibiose (α-D-Glc1,2α-D-Glc) has been reported as an inhibitor for Glu I isolated from yeast, plant and mammalian sources (Bause *et al*. 1986; Zeng & Elbein, 1998; Bause *et al*. 1989)

2.4.5. Antiviral effects of α-glucosidase inhibitors

Many animal viruses are covered by envelope proteins, which are constituted of N-linked glycoproteins that play an important function in completing viral life cycle such as virus assembly and secretion and/ or infectivity (Asano, 2003). Therefore, inhibitors of N-glycosylation likely have potential anti-viral activity (Mehta *et al*. 1998). Several studies have shown that inhibiting the activity of ER processing α-glucosidases reduces viral infectivity. HIV infectivity was reduced in lymphocyte cells treated with α-glucosidase I inhibitor 6-O-butanoyl derivative of castanospermine (MDL 28584); reduction of virus attachment capacity and virus production were observed in the infected cells (Bridges *et al*. 1994). Similarly, in the lymphocyte cultures infected with HIV, the glycosylation pattern of the envelope glycoprotein was altered in the presence of the inhibitor deoxynojirimycin. The deoxynojirimycin treated envelope glycoprotein was able to bind to the cell surface antigen CD4, whereas its ability to induce membrane fusion was abolished, which further inhibited the HIV entry (Papandreou *et al*. 2002). The envelope glycoproteins of hepatitis B virus (HBV) contain only two glycosylation sites and its glycoproteins are sensitive to inhibitors of the N-linked glycosylation pathway (Mehta *et al*. 1998). Imino sugars are monosaccharide mimics that competitively inhibit glycoprotein processing enzymes α-glucosidases I and II (Chang *et al*. 2011). Imino sugars have been widely explored as antivirals against many enveloped viruses. *In vitro* treatment of HBV with imino sugar N-butyl-deoxynojirimycin resulted in a high proportion of virus particles being retained inside the cells (Block *et al*. 1994). *In vivo* treatment of dengue virus (DENV) with imino sugar UV-4, a deoxynojirimycin derivative, reduced mortality of mouse (Perry *et al*. 2013). This data provides the foundation for development of a DENV-specific antiviral compound in human

clinical trials. *N*-nonyl-deoxynojirimycin had an antiviral effect on flavivirus infection by reducing virus replication at the posttranslational modification level (Wu *et al*. 2002). Imino sugars have also been found as promising anti-viral candidates for controlling hemorrhagic fever caused by variety of viruses (Chang *et al*. 2013).

2.4.6. Isolation and purification of Glu I

The Glu I enzyme has been isolated from crude preparations from different mammalian sources. Affinity chromatography using the inhibitor, deoxynojirimycin has been used to purify the enzyme. For example, the enzyme was first isolated from calf liver microsomes and purified by affinity chromatography using AH-Sepharose 4B matrix coupled with N-5-carboxypentyl-dNM. The native enzyme was found to be a tetramer with a molecular mass of 320–350 kDa, the subunit molecular mass was 85 kDa under denaturation conditions (Hettkamp *et al*. 1984). Similarly, the Glu I enzyme purified from bovine mammary gland was a tetramer protein with a mass between 320-330 kDa by gel filtration. The enzyme was extracted by fractionation with detergents, purified by affinity chromatography coupled with N-5-carboxypentyldeoxynojirimycin and DEAE-Sepharose CL-6B (Shailubhai *et al*. 1987). Moreover, the enzyme was isolated from pig liver crude microsomes and purified by affinity chromatography (N-5-carboxypentyl-l-deoxynojirimycin coupled to AH-Sepharose 4B) and polyethylene glycol precipitation (Bause *et al*. 1989). Using the same method, Shailubhai *et al* (1991) reported the purification of the enzyme from rat mammary gland. However, the affinity chromatography with deoxynojirimycin requires multiple purification steps and results in insufficient protein yield.

Plant and yeast sources have also been used to isolate the enzyme. Glu I was purified from the microsomal fraction of mung bean seedlings by sequential chromatography on diethylaminoethyl cellulose, Concanavalin A-Sepharose, deoxynojirimycin resin and diethylaminoethyl cellulose. The enzyme had a molecular mass of 97 kDa and this was shifted to 94 kDa after treatment with Endo H or Endo F, suggesting that glucosidase I is an N-glycoprotein (Zeng & Elbein, 1998). Similarly, Glu I was purified from yeast microsomal preparations by diethylaminoethyl cellulose chromatography, affinity chromatography on AH-Sepharose 4B-linked N-S-carboxypentyl-1-deoxynojirimycin and Concanavalin A-Sepharose chromatography (Bause *et al*. 1986).

2.4.7. Recombinant overexpression of Glu I

With the development of molecular biology techniques on recombinant protein expression, there are several reports that involved expression and purification of recombinant Glu I. For example, Kalz-Fuller *et al* (1995) reported that transfection of COS 1 cells with human hippocampus glucosidase I cDNA resulted in overexpression (about fourfold) of enzymatic activity and the enzyme was found to be a type II membrane bound protein. Moreover, the gene encoding yeast Glu I (*CWH41*) was overexpressed in *S. cerevisiae* strain AH22 (Dhanawansa *et al*. 2002). The Glu I was purified by a combination of ammonium sulfate precipitation, anion exchange, concanavalin A and gel filtration chromatography (Dhanawansa *et al*. 2002). Faridmoayer & Scaman (2007) reported overexpression of the truncated form of Cwht41$_{\Delta 1\text{-}34}$p, with deletion of the 34 amino acids of the N-terminus and transmembrane domain, in *S. cerevisiae* strain AH22 as a catalytically active fragment. This truncated enzyme was purified as a soluble 94 kDa non-glycosylated protein with a specific activity (3,600 U/mg). A gene encoding Glu I from a filamentous fungus, *Aspergillus brasiliensis* ATCC 9642 was cloned and fused to glutathione S-

transferase (GST) and expressed in *E. coli* strain HMS174. The enzyme was purified by glutathione-agarose affinity column (Miyazaki *et al*. 2011).

2.4.8. Catalytic domain of Glu I

Trypsin hydrolysis of Glu I isolated from pig liver with a native molecular weight of 85 kDa yielded polypeptides with mass of 69 kDa, 45 kDa and 29 kDa after 40 min of hydrolysis. The 29 kDa polypeptide was the predominant fragment. Interestingly, the hydrolysis mixture retained 60% of the initial activity. However, it was not clear whether the mixture, individual polypeptides or aggregates of these polypeptides were responsible for the retained activity, since the separation of the individual polypeptides was not successful (Bause *et al*. 1989). A similar study observed that controlled trypsin digestion of rat mammary Glu I (85 kDa) produced an enzymatically active 39 kDa fragment, that was also observed to retain approximately 60% of the initial activity after 60 min of digestion. The 39 kDa fragment was isolated from the trypsin digested mixture using deoxynojirimycin affinity resin, indicating that the 39 kDa fragment contained a putative catalytic domain (Shailubhai *et al.* 1991). In *S. cerevisiae*, a controlled trypsin hydrolysis of the highly purified enzyme released 29, 32, 37 and 59 kDa polypeptides containing the Glu I catalytic domain (Faridmoayer & Scaman, 2004). The 37 kDa polypeptide was isolated as the smallest active fragment using affinity chromatography on resins of *N*-methyl-*N*-5'-carboxypentyl-deoxynojirimycin and *N*-5'-carboxypentyl-deoxynojirimycin. Interestingly, this 37 kDa polypeptide was 1.9 times more active than the full length Glu I when assayed with a synthetic trisaccharide, α-D-Glc1,2α-D-Glc1,3α-D-Glc–O$(CH_2)_8COOCH_3$ (Faridmoayer & Scaman, 2005). However, heterologous overexpression of this catalytic domain

of Cwht41$_{\Delta 1\text{-}525}$p lacking the N-terminal 525 residues was not successful in either *S. cerevisiae* or *P. pastoris* (Faridmoayer & Scaman, 2007; Barker & Rose, 2013).

2.4.9. Potential binding site residues of Glu I

Chemical modification using site selective reagents can be employed to identify amino acid residues that participate in the Glu I catalytic site. Specific modification of arginine, cysteine, tryptophan residue in rat Glu I resulted in almost abolished activity. However, the activity was protected in the presence of a Glu I inhibitor deoxynojirimycin (Pukazhenthi *et al*. 1993; Romaniouk & Vijay, 1997). This indicates that these amino acids may play an important role in Glu I catalytic activity. In addition, Pukazhenthi *et al* (1993) reported the presence of a cysteine residue in the putative catalytic domain of the bovine mammary gland Glu I. Based on the chemical modification studies of Glu I, a highly conserved region E_{594}RHLDLRCW$_{602}$ of rat Glu I was identified and proposed as the substrate binding motif (Romaniouk & Vijay, 1997). However, different key residues were found in yeast Glu I through chemical modification. Site-specific chemical modification of the *S. cerevisiae* Glu I using diethylpyrocarbonate (DEPC) and tetranitromethane (TNM) resulted in 80%-90% of activity loss; however, the activity was recovered in the presence of deoxynojirimycin. DEPC and TNM reacts preferentially with histidine residues and tyrosine residues, respectively, therefore, histidine and tyrosine were proposed to be involved in the catalytic activity of the yeast Glu I enzyme. Furthermore, deoxynojirimycin could not prevent inactivation of yeast Glu I treated with *N*-bromosuccinimide (NBS) used to modify tryptophan residues (Faridmoayer & Scaman, 2005).

2.4.10. Catalytic residues of Glu I

Hydrolysis of a glycosidic bond typically takes place via general acid catalysis and this process is likely catalyzed by a pair of carboxylic acid residues which act as general acid and base (Rye & Withers, 2000). There are two possible stereochemical outcomes for the hydrolysis of a glycosidic bond: inversion or retention of anomeric configuration. Inverting glycoside hydrolysis occurs via a one step, single-displacement mechanism, whereas retaining hydrolysis occurs through a two-step, double displacement mechanism (Rye & Withers, 2000). A ^{1}H NMR spectroscopy study showed that both yeast and mammalian Glu I released β-glucose, indicating that Glu I is an inverting glucosidase (Palcic *et al*. 1999). Sequence homology between Glu I orthologs revealed six highly conserved carboxylic residues (D601, D602, E613, D617, D670, and E804 of Cwh41p). The substitution of two carboxylic acid residues E613 and D617, with alanine inactivated the enzyme Cwh41p (Faridmoayer & Scaman, 2007). However, whether the enzyme inactivation was caused by affecting substrate binding/catalysis or overall enzyme structure was not clear due to the absence of substrate binding and crystal structure analysis. Researchers have determined the crystal structure of a truncated form of Cwht41$_{\Delta 1\text{-}34}$p (with a deletion of N-terminal 34 amino acids of Cwh41p) to 2 Å (Barker *et al*. 2011). Based on structural similarity with homologs *Escherichia coli* YgjK (PDB code 3D3I) and *T. thermophilus* TTHA0978 (PDB ID 2Z07), two catalytic site residues D568 and E771 of Cwht1p (with N-terminal 34 amino acids deletion of Cwh41p) , which correspond to D601 and E804 of Cwh41p in the center of the $(\alpha/\alpha)_6$ barrel were proposed to be the catalytic residues (Barker & Rose, 2013).

2.4.11. Structure of Glu I

Two processing α-glucosidase-I GH63 structures have been solved: the *Escherichia coli* Glu I homolog YgjK (PDB code 3D3I) at 1.78 Å resolution (Kurakata *et al*. 2008), and the yeast Glu I, Cwht1p (PDB code 4J5T) (Barker & Rose, 2013). They share structural similarities and contain an $(\alpha/\alpha)_6$ barrel. The C-terminal part of YgjK and *S. cerevisiae* Cwh41p share a 21% identity (Kurakata *et al*. 2008). YgjK consists of an N-domain (residues 1–231), a linker region (residues 232–276) and an A-domain (residues 277–759). The N-domain is composed of 18 antiparallel β-strands (a super β-sandwich); the A-domain contains of an $(\alpha/\alpha)_6$-barrel formed by 12 α-helices and an extra structural unit A′-region formed by 4 α-helices (Kurakata *et al*. 2008). Soluble Glu I from yeast consists of an N-domain (residues 1 to 278), a linker region and a C-domain (residues 320 to 808). The N-domain consists of an N-terminal α-helix (NH1), a super β-sandwich (NS1-NS13) and two α-helices (NH1,2) between NS11 and NS12; the C-domain consists of an $(\alpha/\alpha)_6$-barrel formed by 12 α-helices (CH1-CH12) and an extra structural unit C'-region formed by two α-helices (C'H1,2) and eight β-strands (C'S1–8) (Figure 4) (Barker & Rose, 2013) .

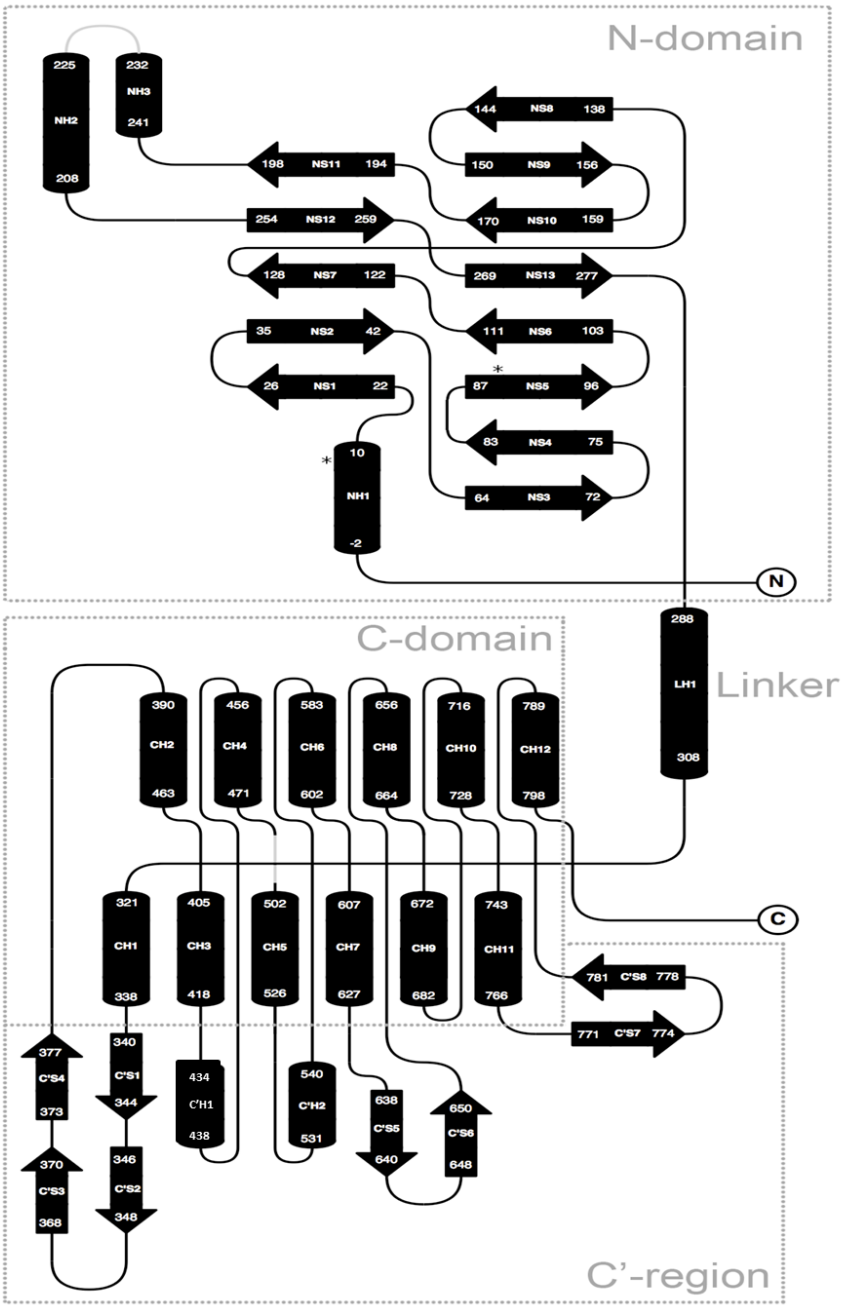


Figure 4. Topology of Cwht1p. α-helices are shown as cylinders; β-strands are shown as arrows. N-glycans are present at the starred locations. Dotted lines indicate domain/region boundaries. Reused with permission, https://s100.copyright.com/AppDispatchServlet#formTop (Barker & Rose, 2013).

2.4.12. Genetic mutants of Glu I

Various studies have reported genetic mutants of Glu I in different organisms. Hitt & Wolf (2004) identified a *DER7* gene (allelic to *CWH41*) that encodes *S. cerevisiae* Glu I; a substitution of the conserved glycine residue at position 725 by arginine (G725R) in the *DER7* gene led to the loss of Glu I activity and ER stress (Table 1). Ray *et al* (1991) identified a Chinese hamster ovary mutant, Lec23, that was defective in Glu I activity. Further analysis of cDNA and genomic DNA from Lec23 cells revealed a single point mutation (TCC to TTC; Ser to Phe) at amino acid position 440 (Hong *et al*. 2004). In humans, the genetic defects of human Glu I cause one kind of disease-congenital disorders of glycosylation type IIb (CDG-IIb) (Praeter *et al*. 2000), also known as MOGS-CDG (Sadat *et al*. 2014). However, the loss of Glu I in humans is extremely rare with only three patients documented. The first case reported an infant with a compound heterozygosity involving two missense mutations (R486T and F652L) in *MOGS* gene, leading to Glu I activity loss and death at the age of 74 days (Praeter *et al*. 2000). The second case reported two siblings with heterozygous *MOGS* genes carrying a maternally inherited nonsense mutation Q124X and two paternally inherited missense mutations A22E and R110H, resulting in undetectable Glu I expression and a severe hypogammaglobulinemia caused by a shortened immunoglobulin half-life (Sadat *et al*. 2014).

Table 1. Effects of genetic mutants of Glu I on phenotype of different organisms

Source	**Mutation**	**Effect**	**Reference**
Yeast	G725R	Loss of activity and ER stress, no phenotypic effect on yeast	Hitt & Wolf (2004)
Human	S440F	Loss of activity and lectin resistance	Ray *et al* (1991), Hong *et al* (2004)
Human, CDG-IIb	R486T & F652L	Loss of activity and fatal at age of 74 days	Praeter *et al* (2000)
Human, CDG-IIb	A22E, R110H & Q124X	Undetectable Glu I expression and a severe hypogammaglobulinemia	Sadat *et al* (2014)

3. Materials and Methods

3.1. Materials

Substrate (α-D-Glc1,2α-D-Glc1,3α-D-Glc–OCH_3) was provided by Dr. Todd Lowary and Dr. Akihiro Imamura from the Department of Chemistry, University of Alberta; *P. pastoris* strain X33 was kindly provided by Dr. Harry Brumer from Michael Smith Laboratories, University of British Columbia; pPICZαB-$MOGS_{\Delta 1\text{-}64}$ (with N-terminal 64 amino acids truncation) was generated by Venkatrao Konasani in Dr. Christine Scaman's lab at University of British Columbia. The BL21 (DE3) transformant with pET47b-$MOGS_{\Delta 1\text{-}64}$ was also generated by Venkatrao Konasani.

All the materials used in this study were of reagent grade. The glucose assay kit (GAGO-20) was obtained from Sigma (St. Louis, Missouri). Restriction enzymes, DNA ladder, protein ladder and dialysis tubing (MWCO 12 kDa-14 kDa) were obtained from Thermo Fisher Scientific (Waltham, Massachusetts), T4 DNA ligase, amylose resin was obtained from New England Biolabs (Whitby, Ontario), plasmid extraction kit and DNA gel extraction kit were obtained from Qiagen (Valencia, California), anti 6xHis-tag monoclonal antibody was obtained from Santa Cruz Biotechnology (Santa Cruz, CA)

3.2. Microbial strains and media

Escherichia coli DH5α and BL21(DE3) were used for transformation. Methylotrophic *P. pastoris* strains GS115 (*his4*) and wild-type strain X33 were used. Transformed *E. coli* strains were grown in LB medium (0.5% yeast extract, 1% peptone, 0.5% NaCl). *P. pastoris* strains

were grown in YPD medium (1% yeast extract, 2% peptone, 2% glucose). The pPICZαB vector carries a C-terminal 6xHis tag, the pMALC5E vector carries an N-terminal MBP (maltose binding protein) tag, and the pET47b(+) and pET30a(+) vector carries an N-terminal 6xHis tag. Antibiotics of Zeocin (25μg/mL) was used for *P. pastoris* transformation, ampicillin (100μg/mL) was used for the pMALC5E constructs, and kanamycin (50μg/mL) was used for the pET47b(+) constructs.

3.3. Human Glu I (MOGS) expression and purification in *P. pastoris*

3.3.1. Molecular cloning of pPICZ*α*B-*MOGS*$_{\Delta 1\text{-}59}$ / pPICZ*α*B-*MOGS*$_{\Delta 1\text{-}71}$

The genes that encode the truncated form of *MOGS*$_{\Delta 1\text{-}59}$ (with N-terminal 59 amino acids truncation) and *MOGS*$_{\Delta 1\text{-}71}$ (with N-terminal 71 amino acids truncation) were amplified by PCR using the primers in Table 2. The PCR products were digested using SacII/XbaI and KpnI/XbaI, respectively. They were then inserted into the pPICZαB vector with T4 DNA ligase. Ligation mixtures were transformed into *E. coli* DH5α. Colony PCR was used to test the insertion of *MOGS* gene for five randomly picked colonies from 20 transformants. Recombinant plasmids were isolated from positive transformants and sequenced using the primers listed in Table 2.

Table 2. Primers used for cloning of pPICZαB-*MOGS*$_{\Delta 1\text{-}59}$/pPICZαB-*MOGS*$_{\Delta 1\text{-}71}$

	Primer	**Sequence (5' to 3')**
Amplification	pPICZαB-*MOGS* F-SacII	AAATATCCGCGGCGCTGGGTGCTGGCGTGGTAC
	pPICZαB-*MOGS* F-KpnI	TATGGTACCTCGTCACGCTGCACTCCGCGCCT
	pPICZαB-*MOGS* R-XbaI	GCTCTAGAAAGTAGTCTTCAGCCATGGCCAGTAAGA CAAGGCT
Sequencing	5' AOX	GACTGGTTCCAATTGACAAGC
	3' AOX	GCAAATGGCATTCTGACATCC
	MOGS Internal 1	TCCCTATGGCTGGGAGTT
	MOGS Internal 2	GAGGAAATCAAGCCCTGC

3.3.2. Transformation of *MOGS* constructs into *P. pastoris*

Positive pPICZαB-*MOGS* plasmids were transformed into the *P. pastoris* using a PEG transformation method based on the Invitrogen manual (Invitrogen, 2013). Integration into yeast genomic DNA was verified by PCR using the sequencing primers in the Table 2, followed by agarose gel electrophoresis. Yeast genomic DNA was extracted and used as template in the PCR. The genomic DNA was isolated according to the method described by Harju *et al* (2004). Repeated freeze-thawing of yeast cells was carried out in lysis buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0) coupled with glass beads to break them, the released genomic DNA was extracted by chloroform and subsequently precipitated by ethanol.

3.3.3. Secreted MOGS expression in *P. pastoris*

Protein expression followed the *P. pastoris* expression manual recommendations (Invitrogen, 2013). BMGY media with 0.1 mg/mL Zeocin antibiotic was inoculated with frozen glycerol cell stocks and the yeast was allowed to grow at 30 ^{o}C for 2 days. The cells were then re-suspended to an OD_{600} of 1.0 in BMMY with 0.5% (v/v) methanol for induction and incubated at 30 ^{o}C in baffled flasks with shaking at 250 rpm. Methanol was added to 0.5% concentration every 12 hours until harvesting at the end of log phase, at approximately 96 h. Over the course of the expression, protein expression level in the supernatant was monitored by SDS-PAGE.

3.3.4. Purification of secreted MOGS using His-Trap HP column

Extracellular secreted protein was harvested by centrifuging the culture twice at 4182xg and the supernatant was collected for purification; the pellet was used to isolate intracellular protein. After adjusting to pH 7.4 with phosphoric acid and disodium phosphate, the supernatant was filtered using a 0.45 um syringe filter (Millipore), and was loaded onto a 5 mL His-Trap HP column (GE Healthcare) attached to an AKTA Purifier. Protein purification followed the method described by Faridmoayer & Scaman (2007). In this system, the column was equilibrated with binding buffer (20 mM sodium phosphate buffer, pH 7.4, 20 mM imidazole and 500 mM NaCl). Then the sample was loaded, the column was washed thoroughly with binding buffer, and bound protein was eluted with elution buffer (20 mM phosphate buffer, pH 7.4, 500 mM imidazole and 500 mM NaCl). Fractions were collected and the imidazole was removed by exchanging with 20 mM phosphate buffer pH 6.8 using a Amicon Centricon 30kDa membrane filter (Millipore) at least for three times, the final concentrated protein volume was 0.5-1 mL.

3.3.5. Isolation of intracellular protein expressed in *P. pastoris*

The cell pellet from Section 3.3.4 was resuspended with 8 volumes of binding buffer. Then lysozyme (final concentration 150 µg/mL), PMSF (final concentration 1mM) and 100x protease inhibitor cocktail (final concentration 1x, Thermo Scientific catalogue number 78430, USA) were added into the mixture. Cell lysis was carried out using ultra-sonication at 30% amplitude for 30 cycles using a Qsonica™ ultrasonicator (50 W), alternating pulses and cooling on ice every 30 seconds. The sample was kept on ice during the sonication. The pellet was separated by centrifugation at 12,000 x g for 20 min and kept for SDS-PAGE gel.

3.3.6. Glu I enzyme assay

Glu I activity was measured as described by Neverova *et al* (1994). In this study, the synthetic trisaccharide, α-D-Glc1,2α-D-Glc1,3α-D-Glc–OCH_3 was used as the substrate. The released glucose was measured using PGO enzyme preparation (Sigma-Aldrich chemicals cat# P7119). In brief, a 4 µL aliquot of enzyme was mixed with 1 µL of 10 mM synthetic trisaccharide and incubated at 37 ^{o}C. The reaction was quenched by adding 45 µL of 1.25 M Tris–HCl, pH 7.6. The solution was then mixed with 250 µL PGO enzyme solution, prepared according to the kit instructions. After 30 min incubation at 37 ^{o}C in the dark, the absorbance at 450 nm was measured. One unit of enzyme activity corresponds to the amount of enzyme that produces 1 nmol of glucose per min at 37 ^{o}C, pH 6.8. Enzyme was dissolved with 20 mM phosphate buffer, pH 6.8.

3.3.7. Periodic acid-Schiff staining

To determine whether protein was glycosylated, a staining method of Periodic acid-Schiff (PAS) staining was used based on Egito *et al* (2001). Protein (5 µg) was blotted onto the PVDF membrane. All the further treatments of the membrane were performed in the dark. At first, the membrane was washed in 50 mL of 1% (w/v) of periodic acid for 15 min. The membrane was then treated by 50 mL of 15% acetic acid for 15 min and was repeated with three times. A volume of 50 mL of Schiff's reagent (basic fuchsin 10 g/L, sodium metabisulfite 18 g/L, 10 mL HCl) was then added for 30 min and the membrane was washed six times with 100 mL of 7.5% acetic acid for 1 hour. Finally, the membrane was dried and stored at room temperature.

3.3.8. De-glycosylation with PNGase F

To determine whether the protein was N-glycosylated, PNGase F was used according to the manufacturer's protocol (New England Biolabs). A 5 µL aliquot of protein (1 mg/mL) was mixed with Glycoprotein Denaturing Buffer (0.5% SDS, 40 mM DTT) and water. The glycoprotein was denatured by heating at 100°C for 10 minutes and chilled on ice for 1 minute. Then 2 µL of 10x Reaction Buffer, 2 µL 10% NP-40 and 2 µL of PNGase F were added into the mixture, the reaction was shaken at 37°C overnight. The change in molecular weight was monitored by SDS-PAGE gels. A positive control of chicken egg white conalbumin (Sigma) was subjected to deglycosylation.

3.3.9. Protein assay and Western blotting

Protein concentration was determined by the Bradford method (Bradford, 1976) and was performed with the Bio-Rad protein assay kit using bovine serum albumin as the standard protein. All of the sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels were performed using 12% polyacrylamide gels and stained using Coomassie blue (Laemmli, 1970). One millilitre of whole cell culture was centrifuged at 19600 x g for 10 minutes, the pellet was resuspended by 50 µL of 2x SDS-PAGE loading buffer (100 mM Tris-HCl, pH6.8, 4% (w/v) SDS, 0.2% glycerol, 200 mM DTT), the sample was boiled for 10 minutes and cooled on ice for 10 minutes, the sample was centrifuged at 19600 x g for 10 minutes, 10 µL of the supernatant was loaded onto the SDS-PAGE gel. The ChemiDoc MP system (Bio-Rad) was used to take gel images and to analyze SDS-PAGE gel bands and intensity. The Bio-Rad Mini Trans-Blot system was used for protein blotting onto PVDF membranes following methods by Faridmoayer & Scaman, (2007). In brief, electro-blotting was conducted at 100 V for 1 h in transfer buffer (25 mM Tris base, 192 mM glycine, 20% methanol and 0.025% SDS). Western blotting was then carried out using rabbit anti 6xHis-tag monoclonal antibody according to the manufacturer's procedure with detection using HRP conjugated anti-rabbit IgG antibodies and Immun-Star™ WesternC™ chemiluminiscent kit (BioRad).

3.4. Human Glu I (MOGS) expression in *E. coli* and purification

3.4.1. Molecular cloning of $MOGS_{\Delta 1\text{-}63}$ fused with MBP tag

The truncated version of $MOGS_{\Delta 1\text{-}63}$ was amplified using a pair of primers (forward 5'- GCGTCGACCTGGCGTGGTACCGTGCG-3',

reverse 5'- GCTCCCAAGCCTTCAGTAGTCTTCAGCCATGGCCAGTAAGAC-3'). The PCR product was restriction digested by Sal I and Hind III. It was then inserted into the pMALC5E vector with T4 DNA ligase. Ligation mixtures were transformed into *E. coli* DH5α. Recombinant plasmids were isolated from positive transformants and sequenced using the primers listed in Table 3.

Table 3. Primers used for sequencing of pMALC5E-*MOGS*$_{\Delta 1\text{-}63}$

Primer	**Sequence (5' to 3')**
pMALC5E Forward	GGTCGTCAGACTGTCGATGAAGCC
pMALC5E Reverse	TGTCCTACTCAGGAGAGCGTTCAC
MOGS Internal 1	TCCCTATGGCTGGGAGTT
MOGS Internal 2	GAGGAAATCAAGCCCTGC

3.4.2. Expression of MBP-MOGS$_{\Delta 1\text{-}63}$ in *E. coli*

Recombinant plasmid (pMALC5E-*MOGS*$_{\Delta 1\text{-}63}$) was transformed into *E. coli* BL21(DE3). Three colonies were picked randomly from the LB agar plate from more than 50 transformants to screen for the expression level. The colony with the highest expression level was selected for further investigation. The selected transformant was cultured at 37 °C in 2 mL of LB medium (100 μg/mL ampicillin) overnight. This was transferred into 200 mL of LB medium until the absorbance at 600 nm reached 0.6, and then 0.2 mM of isopropyl-D-thiogalactopyranoside (IPTG) was added to induce protein expression at 25°C for 16 hours.

3.4.3. Purification of MBP-MOGS$_{\Delta 1\text{-}63}$

The pMALC5E vector expresses the MBP fusion protein in the cytoplasm. Overnight cells from 3.4.2 were pelleted by centrifugation at 5,000 x g for 10 min. The pellet was resuspended in 8 volumes of MBP-Column Buffer (20 mM phosphate buffer pH7.4, 1 mM EDTA and 200 mM NaCl), then lysozyme (final concentration 150 µg/mL), PMSF (final concentration 1mM) and 100x protease inhibitor cocktail (final concentration 1x, Thermo Scientific catalogue number 78430, USA) were added into the mixture. Cell lysis was carried out using ultra-sonication at 30 % amplitude for 30 cycles using a Qsonica™ ultrasonicator (50 W), alternating pulses and cooling on ice every 30 seconds, the sample was kept on ice during the sonication. The supernatant was separated by centrifugation at 12,000 x g for 20 min and was filtered by a 0.45 um syringe filter (Millipore), it was labeled as soluble fraction on SDS-PAGE gel; the pellet by centrifugation was labeled as insoluble fraction on SDS-PAGE gel. The supernatant was loaded into the amylose resin column at a flow rate of no more than 80 µL/min, washed with 20 mL of MBP-Column Buffer, and the fusion protein was eluted with MBP-Column Buffer containing 10 mM maltose. Protein containing fractions were collected and the buffer was exchanged with 20 mM phosphate buffer pH 6.8 using a Amicon Centricon 30kDa membrane filter (Millipore) for at least three times. The final concentrated protein volume was 0.5-1 mL.

3.4.4. Purification of MOGS$_{\Delta 1\text{-}64}$ (pET47b) from inclusion bodies

3.4.4.1. Urea-denaturation and refolding method

The protein expression of *E. coli* BL21(DE3) with pET47b-*MOGS*$_{\Delta 1\text{-}64}$ was carried out as described in 3.4.2. The inclusion bodies were purified by a modified method (Santos *et al*. 2012).

The cells (100 mL volume) were harvested by centrifugation at 5,000 x g for 10 min. The pellet was resuspended in 8 volumes (3 mL) of Buffer A (1 mM DTT, 1 mM PMSF, 0.2 mM EDTA, 1M NaCl, 50 mM Tris, pH 8). The mixture was sonicated on ice at 30 % amplitude for 30 cycles. The resulting cell lysate was centrifuged at 12,000 g for 20 minutes. The supernatant was labeled as soluble fraction on SDS-PAGE gel; the pellet by centrifugation was labeled as inclusion body on SDS-PAGE gel. The inclusion bodies were resuspended in 3 mL of Buffer A containing 8 M of urea. The mixture was rotated at 4°C for 12 hours, and centrifuged at 12,000 x g for 10 minutes. The supernatant was placed into dialysis tubing membrane (MWCO 12 kDa-14 kDa), prepared by washing with ddH_2O for 30 minutes. Then the urea-solubilized protein was dialyzed in 500 mL of refolding Buffer B (1 mM DTT, 1 mM PMSF, 0.2 mM EDTA, 100 mM NaCl, 10 % (v/v) glycerol, 50 mM Tris, pH 8) with sequentially decreasing concentrations of urea (6, 5, 4, 3, 2, 1, 0.5 and 0 M urea) for at least 4 hours per concentration.

3.4.4.2. Freeze-thawing method

The inclusion bodies were also purified by the method of Qi *et al* (2015). The protein expression of *E. coli* BL21(DE3) with pET47b-$MOGS_{\Delta 1\text{-}64}$ was carried out as described in Section 3.4.2. The cells were harvested by centrifugation at 5,000 x g for 10 min and the pellet was stored at −20°C overnight. The pellet was thawed briefly at room temperature and resuspended in 10 mL of PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4 \cdot 12H_2O$, 2 mM KH_2PO_4, pH 8.0), and then lysed by sonication on ice. The resulting cell lysate was centrifuged at 12,000 g for 20 minutes. The supernatant was labeled as soluble fraction on SDS-PAGE gel; the pellet by centrifugation was labeled as inclusion body on SDS-PAGE gel. The inclusion body was washed extensively three times with 10 mL of washing buffer (20 mM Tris,

300 mM NaCl, 1 mM EDTA, 1% Triton X-100, 1 M urea, pH 8.0). Finally, the pure inclusion bodies were resuspended in 2 mL of PBS buffer with 2M urea. The suspension was frozen at −20 °C overnight and thawed at room temperature. The supernatant containing solubilized protein was collected by centrifuge at 12,000 x g for 15 minutes at 4 °C. The solubilized protein was recovered by dialysis into 500 mL of recovery buffer (20 mM Tris, 100 mM NaCl, 10 mM CaCl2, 0.2 mM ZnCl2, pH 7.5) for overnight at 4 °C. The recovered solution was applied into a Amicon Centricon 30kDa membrane filter (Millipore) by exchanging with 20 mM phosphate buffer pH 6.8, the final concentrated protein volume was 0.5-1 mL.

3.5. Expression of Cwh41p truncations in *E. coli* and purification

3.5.1. Molecular cloning of *CWH41* truncations

The truncations of *CWH41*$_{(\Delta 1\text{-}39,\ \Delta 1\text{-}45,\ \Delta 1\text{-}54,\ \Delta 1\text{-}74,\ \Delta 1\text{-}93,\ \Delta 1\text{-}167)}$ with cut sites of EcoRI (5’ end) and XhoI (3’ end) were amplified using the primers listed in Table 3. The pET30a vector was double digested with EcoRI and XhoI, purified with the gel extraction kit. The double digested gene and vector were ligated by T4 DNA ligase and transformed into *E. coli* DH5α. Colony PCR was used to test the insertion of each *CWH41* gene for five randomly picked colonies from 20 transformants. Recombinant plasmids were isolated from positive transformants and sequenced using the corresponding primers listed in Table 4.

Table 4. Primers used for cloning of *CWH41* truncations

	Primer	**Sequence (5' to 3')**
Amplification	$CWH41_{\Delta 1\text{-}167}$ Forward	CCGGAATTCGGAAAATCTTCCTTAGCAATGATAGGTC
	$CWH41_{\Delta 1\text{-}93}$ Forward	CCGGAATTCGATAAATTGCAAAAGTATGGTTGGGAAG
	$CWH41_{\Delta 1\text{-}74}$ Forward	CCGGAATTCTTCAACAGTTTGAGTCAGGATGG
	$CWH41_{\Delta 1\text{-}54}$ Forward	CCGGAATTCTACTTTGGTATGAGGCCCAG
	$CWH41_{\Delta 1\text{-}45}$ Forward	CCGGAATTCCTGTGGGCACCGTATAGAT
	$CWH41_{\Delta 1\text{-}39}$ Forward	CCGCCGGAATTCTTCACGAATGAATCTTTACTG
	CWH41 Reverse	CCGCTCGAGTCAGAAGCGTCCAAG
Sequencing	T7 Forward	TAATACGACTCACTATAGGG
	T7 Reverse	GCTAGTTATTGCTCAGCGG
	RSP1	TAAAGAATTAGGCGAGTATC
	RSP2	AGTATGATTTTGACCTTGCC

3.5.2. Expression of Cwh41p truncations in *E. coli*

The protein expression of *E. coli* BL21(DE3) with truncated Cwh41p was carried out as described in Section 3.4.2. The expression of Glu-I was analyzed on SDS-PAGE.

3.5.3. Purification of Cwh41p truncations

The cells were harvested by centrifugation at 5,000 x g for 10 min. The pellet was resuspended in 8 volumes of binding buffer (20 mM phosphate buffer pH7.4, 20 mM imidazole and 500 mM NaCl), then lysozyme (150 µg/mL), PMSF (1mM) and 100x protease inhibitor cocktail (final

concentration 1x, Thermo Scientific catalogue number 78430, USA) were added into the mixture. Then the cell lysis was carried out using ultra-sonication at 30 % amplitude for 30 cycles using a Qsonica™ ultrasonicator (50 W), alternating pulses and cooling on ice every 30 seconds, the sample was kept on ice during the sonication. The supernatant was separated by centrifugation at 12,000 x g for 20 min and was filtered through a 0.45 um syringe filter (Millipore), it was labeled as soluble fraction on SDS-PAGE gel; the pellet by centrifugation was labeled as insoluble fraction on SDS-PAGE gel. The supernatant was loaded onto a 1 mL/5 mL His-Trap HP column (GE Healthcare) attached to an AKTA Purifier and purification was carried out as described in Section 3.3.4.

4. Results

4.1. MOGS overexpression in *P. pastoris*

4.1.1. Expression and purification of $MOGS_{\Delta 1\text{-}64}$

The construct of pPICZαB-*MOGS*$_{\Delta 1\text{-}64}$ was transformed into *P. pastoris* strain GS115. Yeast genomic DNA was isolated from one positive transformant and was identified by colony PCR (Appendix Figure 1). This clone was further used for the expression and purification. The $MOGS_{\Delta 1\text{-}64}$ protein, purified by FPLC using a His-trap column, showed two bands on SDS-PAGE gel; based on the molecular weight standards, one band was estimated as 101 kDa and the other band at estimated as 105 kDa (Figure 5A). Anti-His Western blotting analysis further identified the molecular mass of the purified $MOGS_{\Delta 1\text{-}64}$ at 101 kDa (Figure 5B), indicating that the 105 kDa band was a His-trap non-specific binding protein. Most importantly, the $MOGS_{\Delta 1\text{-}64}$ was active (specific activity was 735 U/mg) against the synthetic trisaccharide substrate α-D-Glc1,2α-D-Glc1,3α-D-Glc–OCH_3.

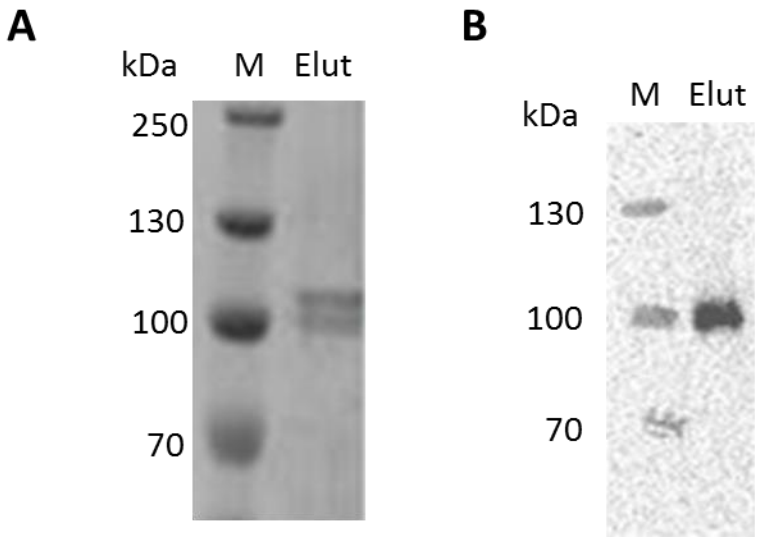


Figure 5. Analysis of the purified $MOGS_{\Delta 1\text{-}64}$. (A) $MOGS_{\Delta 1\text{-}64}$ had two bands at 101 kDa and 105 kDa on SDS-PAGE gel (12%) with Coomassie blue staining; (B) $MOGS_{\Delta 1\text{-}64}$ was detected as the 101 kDa band by anti-His Western blotting. M, molecular weight standards. Elut, eluted fraction from His-trap column. Aliquots of 5 µg proteins were loaded on the gel.

4.1.2. Glycosylation analysis of the $MOGS_{\Delta 1\text{-}64}$

In PAS staining, periodic acid oxidizes two vicinal diol groups to form an aldehyde, which reacts with the Schiff reagent to give a pink to magenta color. Pink color was observed for the $MOGS_{\Delta 1\text{-}64}$ (dot 1) on the PAS stained membrane as well as the positive control, chicken egg white conalbumin (dot 2) indicating the presence of a glycoprotein, while the negative control of BSA did not show pink color (dot 3) (Figure 6).

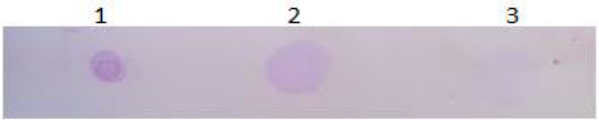


Figure 6. Periodic acid-Schiff (PAS) staining. Dot 1, $MOGS_{\Delta 1\text{-}64}$; Dot 2, chicken egg white conalbumin (Sigma), positive control; Dot 3, bovine serum albumin, negative control. Aliquots of 5 µg proteins were loaded on the membrane.

Following de-glycosylation treatment by PNGase F, the molecular mass of $MOGS_{\Delta 1\text{-}64}$ was reduced from 101 kDa to 97 kDa (the expected mass) on SDS-PAGE gel (Figure 7, lane 2), indicating that $MOGS_{\Delta 1\text{-}64}$ was N-linked glycosylated. The molecular mass of positive control (chicken egg white conalbumin) is 76 kDa. There was a slight molecular mass decrease by PNGase F treatment (Figure 7, lane 4). The molecular mass of negative control (bovine serum albumin) is 69 kDa. There was no molecular mass decrease by PNGase F treatment (Figure 7, lane 8). Interestingly, the His-trap non-specific binding 105 kDa protein (shown in Figure 5) was not observed on the gel (Figure 7, lane 1) probably due to degradation during storage for 12 days. Some faint degradation bands at ~56 kDa and ~37 kDa were also observed on the gel.

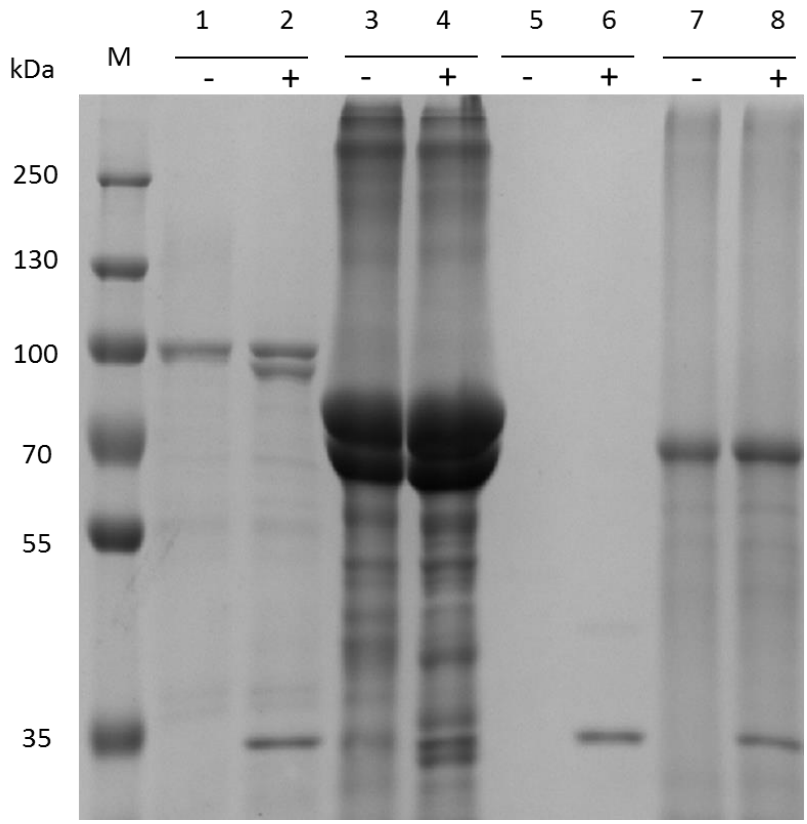


Figure 7. SDS-PAGE (12%) analysis of de-glycosylation with PNGase F treatment. The SDS-PAGE gel was stained by Coomassie blue. Lane 1&2, $MOGS_{\Delta 1\text{-}64.}$ Lane 3&4, chicken egg white conalbumin (Sigma), positive control; Lane 5&6, phosphate buffer, blank control; Lane 7&8, bovine serum albumin, negative control. "-", protein untreated; "+", protein treated with 2 µL of PNGase F (New England Labs catalog number P0704S). Molecular mass of $MOGS_{\Delta 1\text{-}64}$ was decreased from 101 kDa to 97 kD. The 35 kDa band was PNGase F enzyme. M, molecular weight standard. Aliquots of 5 µg proteins of $MOGS_{\Delta 1\text{-}64}$ and bovine serum albumin were loaded on the gel; Aliquots of 20 µg proteins of chicken egg white conalbumin were loaded on the gel.

4.1.3. Subsequent expression and purification of $MOGS_{\Delta 1\text{-}64}$

In three subsequent trials, $MOGS_{\Delta 1\text{-}64}$ was expressed as an intracellular protein (Figure 8). The medium supernatant was applied to the FPLC, with no protein detected in the eluted fractions from the column, indicating the absence of soluble secreted protein. The cell pellet was treated by lysozyme and ultra-sonication, the mixture was centrifuged to obtain the intracellular content. It was observed that the molecular size of intracellular $MOGS_{\Delta 1\text{-}64}$ was 101 kDa (Figure 8B), indicating that intracellular $MOGS_{\Delta 1\text{-}64}$ was N-linked glycosylated.

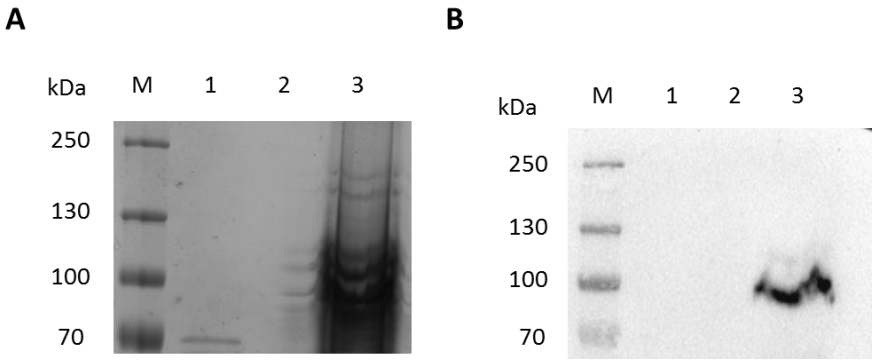


Figure 8. $MOGS_{\Delta 1\text{-}64}$ was expressed predominantly as an insoluble protein. (A) SDS-PAGE gel (12%); (B) anti-His Western blotting. The SDS-PAGE gel was stained by Coomassie blue. Lane 1, eluted fraction from His-trap column; Lane 2, medium supernatant; Lane 3, intracellular content.

4.1.4. Expression and purification of $MOGS_{\Delta 1\text{-}59}$ and $MOGS_{\Delta 1\text{-}71}$

Two different constructs of pPICZαB-*$MOGS_{\Delta 1\text{-}59}$* and pPICZαB-*$MOGS_{\Delta 1\text{-}71}$* were transformed into *P. pastoris* strain GS115 as well as wild-type strain X33. The yeast genomic DNAs were isolated from positive transformants and were identified by colony PCR (Appendix Figure 2). The positive clones were further used for protein expression and purification. The supernatants

from the positive transformants were applied to the FPLC, but no protein was detected in the eluted fractions from the column.

4.2. MOGS overexpression in *E. coli*

4.2.1. MBP-MOGS$_{\Delta 1\text{-}63}$ purification

The expression of MBP-MOGS$_{\Delta 1\text{-}63}$ was induced in the presence of isopropyl-D-thiogalactopyranoside (IPTG) (Appendix Figure 3). After cell lysis, MBP-MOGS$_{\Delta 1\text{-}63}$ was observed mainly in inclusion bodies (Figure 9, lane 1) with a minor fraction of soluble protein in the supernatant (Figure 9, lane 2). The supernatant was applied onto the amylase resin column to isolate MBP fusion protein. However, the eluted fraction from the amylase resin column underwent a severe degradation (Figure 9, lane 3). Many small fractions were observed between the size of MBP (42.5 kDa) and the size expected for the full-length fusion MOGS$_{\Delta 1\text{-}63}$ (predicted mass of 130 kDa). Even though a small amount of the MBP-MOGS$_{\Delta 63}$ (130 kDa) was observed from the eluted fraction, it did not show enzyme activity with the trisaccharide substrate when 10 μg of the protein was incubated with 10 nmoles of substrate for 30 min at 37 ˚C.

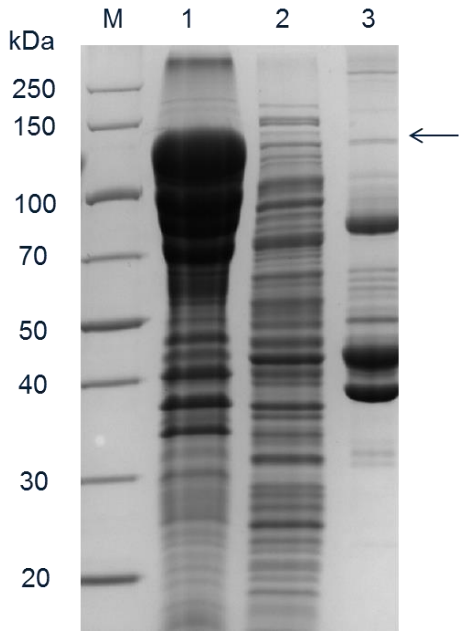


Figure 9. SDS-PAGE gel (12 %) analysis of overexpression of MBP-MOGS$_{\Delta 1\text{-}63}$. The SDS-PAGE gel was stained by Coomassie blue. The fusion protein at 130 kDa was degraded into small fragments including MBP at 42.5 kDa. M, molecular weight standard; Lane 1, insoluble fraction; Lane 2, soluble fraction; Lane 3, eluted fraction from amylase resin column. The arrow indicates the expected protein band. An aliquot of 15 µg protein of eluted fraction was loaded on the gel.

4.2.2. Purification of MOGS$_{\Delta 1\text{-}64}$ from inclusion bodies

4.2.2.1. Urea-denaturation and refolding

After cell lysis, the supernatant (Figure 10A, lane 1) was applied onto the His-trap affinity column to isolate any soluble enzyme. No protein was detected in the eluted fractions from the column (data not shown). The MOGS$_{\Delta 1\text{-}64}$ (pET47b) was observed predominantly as insoluble

inclusion bodies (Figure 10A, lane 2). When treated by 8M urea at 4°C for 12 hours, all of the inclusion bodies were solubilized (Figure 10A, lane 3). The solubilized protein was treated with step-wise dialysis to gradually decrease the urea concentration. However, at the end of dialysis, very little soluble protein was recovered (about 3% recovery) (Figure 10B, lane 1); almost all of the protein precipitated (Figure 10B, lane 2). The intensity and distribution of the precipitated protein (Figure 10B, lane 2) was very close to that of the original inclusion body isolation (Figure 10A, lane 2). Furthermore, the recovered protein did not show enzyme activity with the trisaccharide substrate when 1 μg of the protein was incubated with 10 nmoles of substrate for 30 min at 37 ˚C.

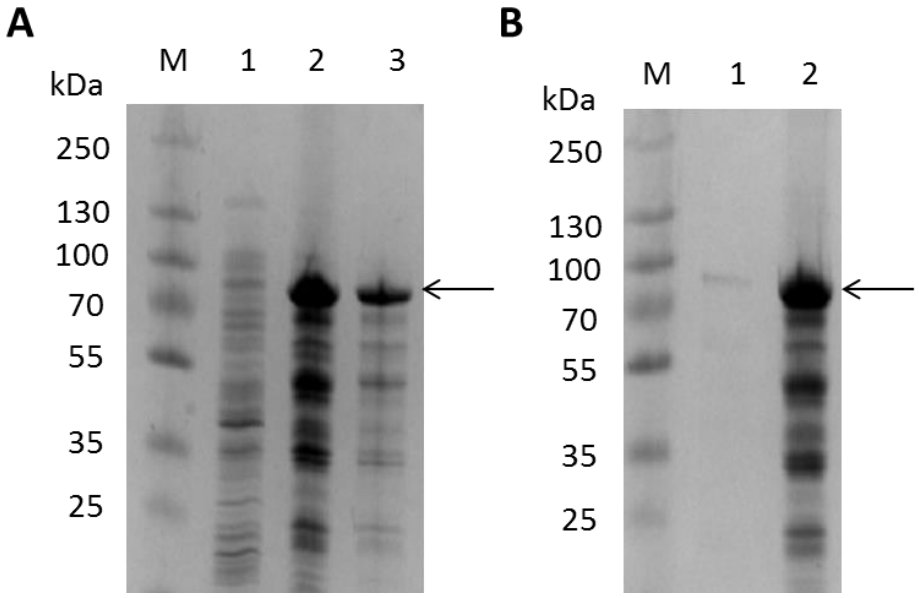


Figure 10. SDS-PAGE gel (10-15%, PhastGel gradient, GE Healthcare) analysis of inclusion body purification of $MOGS_{\Delta 1\text{-}64}$ (pET47b) by urea-denaturation and refolding. The SDS-PAGE gels were stained by Coomassie blue. M, molecular weight standards. (A) Lane 1, soluble fraction with cell lysis and sonication; Lane 2, inclusion body with cell lysis and sonication; Lane 3, solubilized inclusion body treated by 8M urea; (B). Lane 1, recovered soluble protein after dialysis; Lane 2, precipitation after dialysis. The arrows indicate the expected protein band of $MOGS_{\Delta 1\text{-}64}$ (pET47b).

4.2.2.2. Freeze-thawing

After cell lysis and buffer washing, the inclusion body of $MOGS_{\mathbf{\Delta 1\text{-}64}}$ (pET47b) (Figure 11, lane 2) was resuspended in PBS buffer with 2M urea. The suspension was frozen at −20°C overnight and thawed at room temperature. After the freeze-thawing cycle, large amount of insoluble protein (Figure 11, lane 4) and a small amount of soluble protein was observed (Figure 11, lane

5). The solubilized protein was recovered by dialysis. Even though no precipitation was observed during dialysis, very little protein was recovered (6% recovery). The recovered protein did not show enzyme activity with the trisaccharide substrate when 0.6 μg of the protein was incubated with 10 nmoles of substrate for 60 min at 37 ˚C.

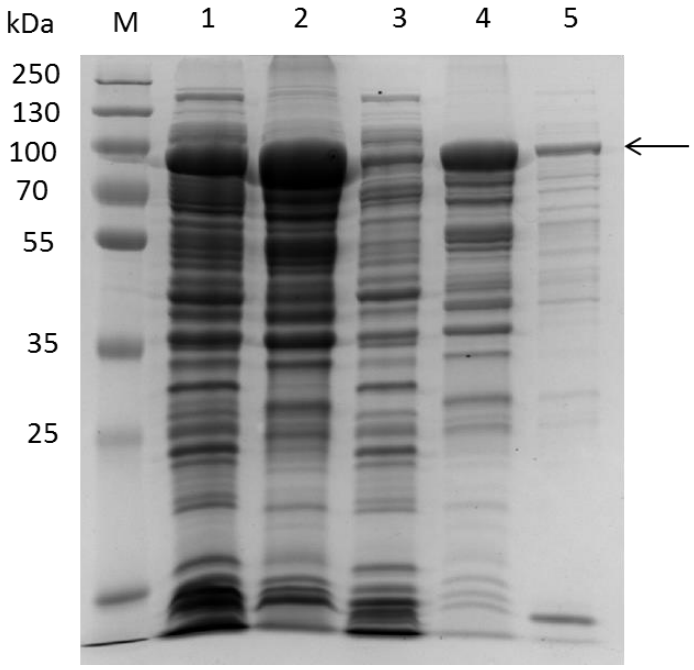


Figure 11. SDS-PAGE gel (12 %) analysis of inclusion body purification of $MOGS_{\Delta 1\text{-}64}$ (pET47b) using the freeze-thawing method. The SDS-PAGE gel was stained by Coomassie blue. M, molecular weight standard; Lane 1, whole cell homogenate; Lane 2, inclusion body with cell lysis and sonication; Lane 3, soluble fraction with cell lysis and sonication; Lane 4, centrifuge pellet after freeze-thawing cycle; Lane 5, centrifuge supernatant after freeze-thawing cycle. The arrow indicates the expected protein band of $MOGS_{\Delta 1\text{-}64}$ (pET47b).

4.3. Truncated Cwh41p overexpression in *E. coli*

Truncated Cwh41$_{\Delta 1\text{-}167}$p was expressed in *E. coli* BL21(DE3). The expressed proteins accumulated as inclusion bodies (Figure 12A, lane 3). After cell lysis, the supernatant was applied onto the His-trap affinity column to isolate any soluble fraction. No protein was detected in the eluted fractions from the column (data not shown). Truncated Cwh41$_{\Delta 1\text{-}93}$p was also overexpressed mainly in inclusion bodies with a minor fraction of soluble protein in the supernatant after cell lysis (Figure 12B).

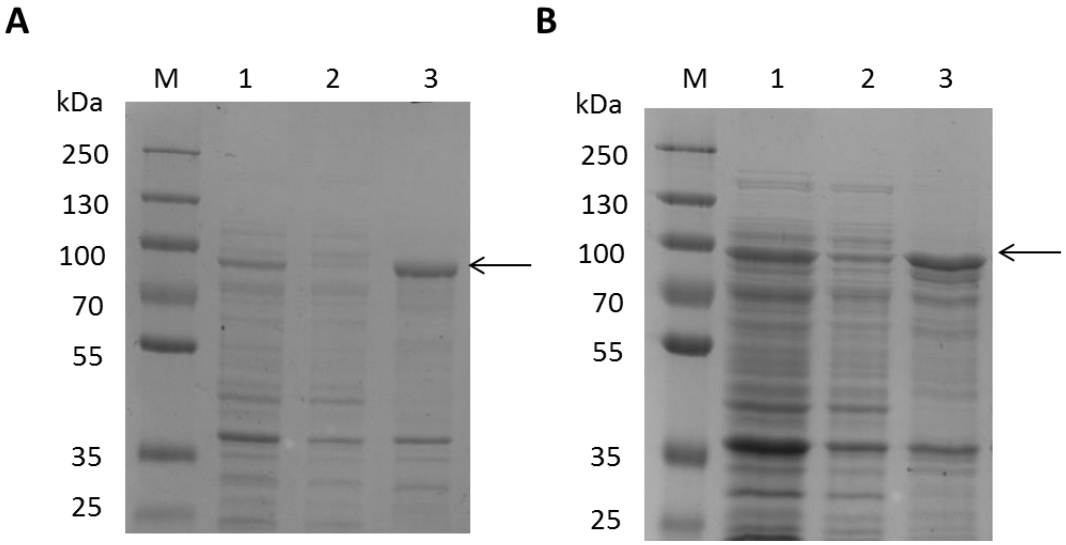


Figure 12. SDS-PAGE gel (12 %) analysis of expression of (A) Cwh41$_{\Delta 1\text{-}167}$p and (B) Cwh41$_{\Delta 1\text{-}93}$p. The SDS-PAGE gels were stained by Coomassie blue. M, molecular weight standard; Lane 1, whole cell; Lane 2, soluble fraction; Lane 3, insoluble fraction. The arrows indicate the expected protein bands.

The soluble fraction of Cwh41$_{\Delta 1\text{-}93}$p was applied to the His-trap column, the eluted fraction from the column (Figure 13A, lane 1) showed the presence of 6xHis tag by a Western blotting with anti-His antibodies (Figure 13B). However, this eluted fraction did not show enzyme activity with the trisaccharide substrate when 2 μg of the protein was incubated with 10 nmoles of substrate for 30 min at 37 ˚C.

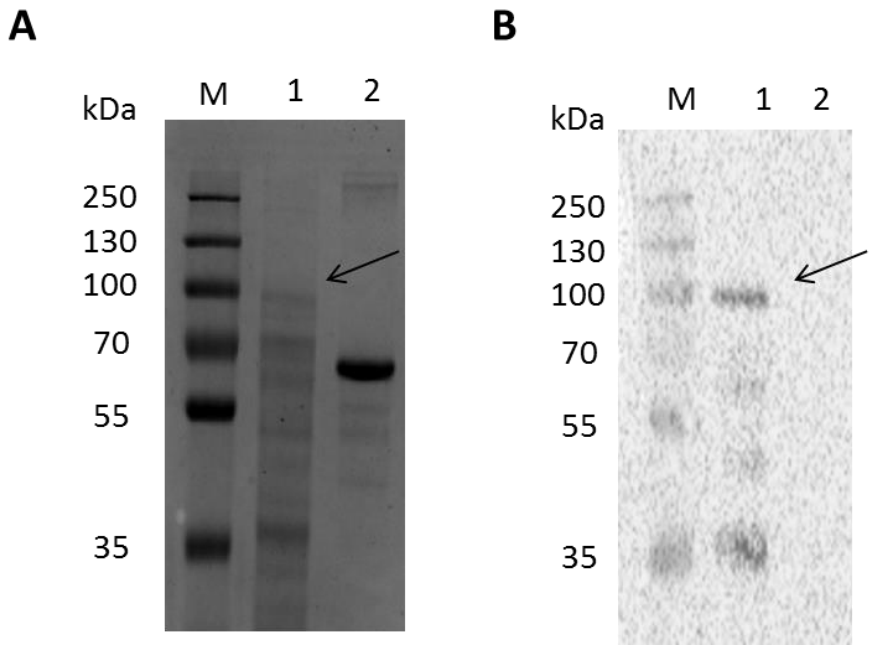


Figure 13. SDS-PAGE (12%) analysis (A) and anti-His Western blotting (B) of eluted fraction of Cwh41$_{\Delta 1\text{-}93}$p from His-trap column (Lane 1), BSA (Lane 2). The SDS-PAGE gel was stained by Coomassie blue. M, molecular weight standard. The arrows indicate the expected protein bands. Aliquots of 5μg of protein were loaded on the gel.

Similar to other truncations, Cwh41$_{\Delta 1\text{-}74}$p, Cwh41$_{\Delta 1\text{-}54}$p and Cwh41$_{\Delta 1\text{-}45}$p were overexpressed mainly in inclusion bodies (Figure 14, Figure 15A). The soluble fraction of each preparation

was applied to the His-trap column, and the eluted fractions did not show enzyme activity with the trisaccharide substrate. Interestingly, the eluted fraction of Cwh41$_{\Delta 1\text{-}39}$p from the column (Figure 15B) showed enzyme activity of 266 U/mg with the trisaccharide substrate. However, there was a large amount of small proteolytic fragments in the eluted fraction (Figure 15B, lane 3), which may have occurred during the cell lysis.

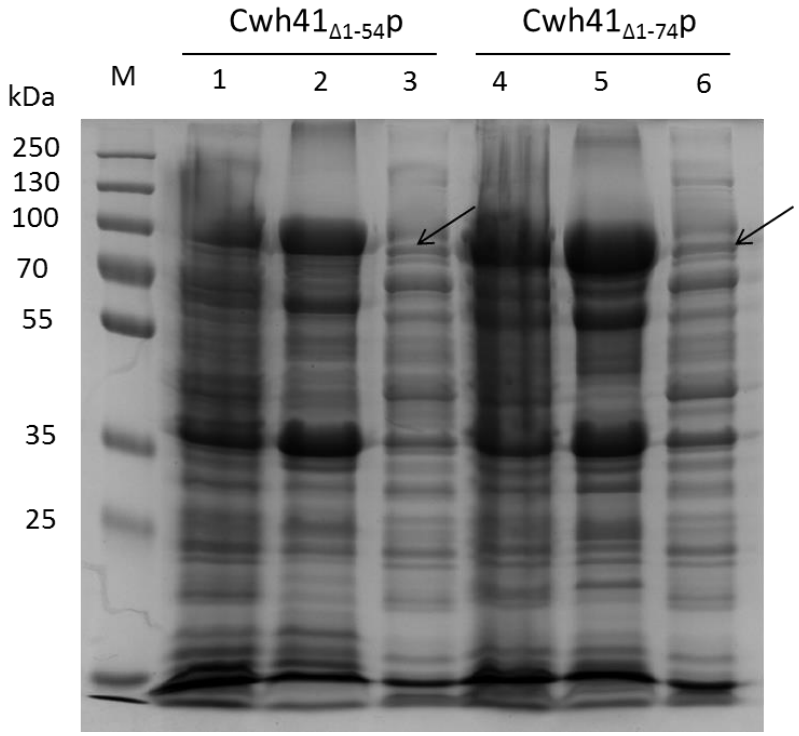


Figure 14. SDS-PAGE gel (12 %) analysis of expression of Cwh41$_{\Delta 1\text{-}54}$p and Cwh41$_{\Delta 1\text{-}74}$p. The SDS-PAGE gel was stained by Coomassie blue. M, molecular weight standard; Lane 1&4, whole cell; Lane 2&5, insoluble fraction; Lane 3&6, soluble fraction. The arrows indicate the expected protein bands.

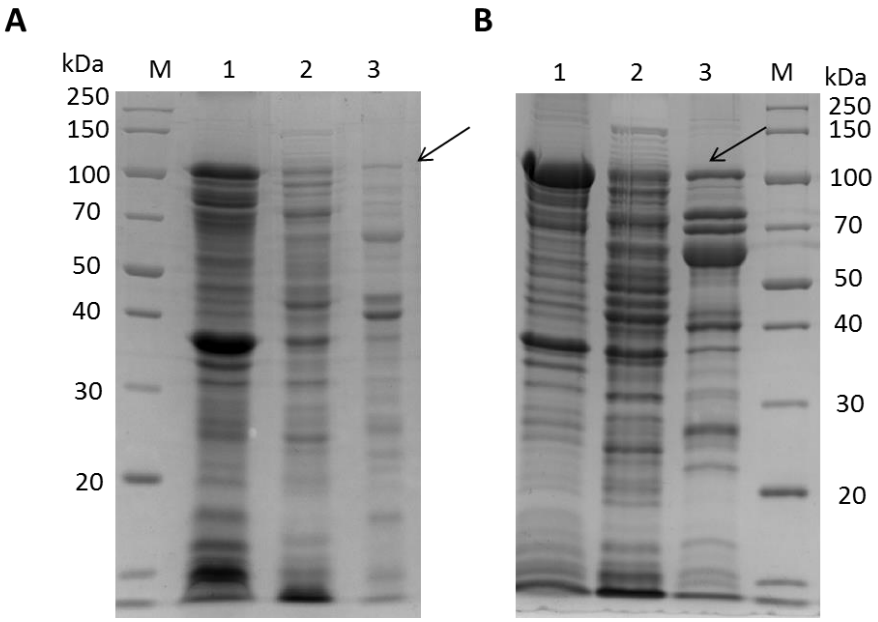


Figure 15. SDS-PAGE gel (12 %) analysis of expression and purification of (A) Cwh41$_{\Delta 1\text{-}45}$p and (B) Cwh41$_{\Delta 1\text{-}39}$p. The SDS-PAGE gels were stained by Coomassie blue. M, molecular weight standard; Lane 1, insoluble fraction; Lane 2, soluble fraction; Lane 3, eluted fraction from His-trap column. The arrows indicate the expected protein bands. Aliquots of 5µg Cwh41$_{\Delta 1\text{-}45}$p eluted fraction and 10µg Cwh41$_{\Delta 1\text{-}39}$p eluted fraction were loaded on the gel.

Finally, although inclusion bodies were also observed in the overexpression of Cwh41$_{\Delta 1\text{-}34}$p, there was a large amount of soluble protein after cell lysis (Figure 16, lane 3). Most importantly, the eluted fraction of Cwh41$_{\Delta 1\text{-}34}$p from the His-trap column (Figure 16, lane 4) showed enzyme activity of about 2000 U/mg with the trisaccharide substrate.

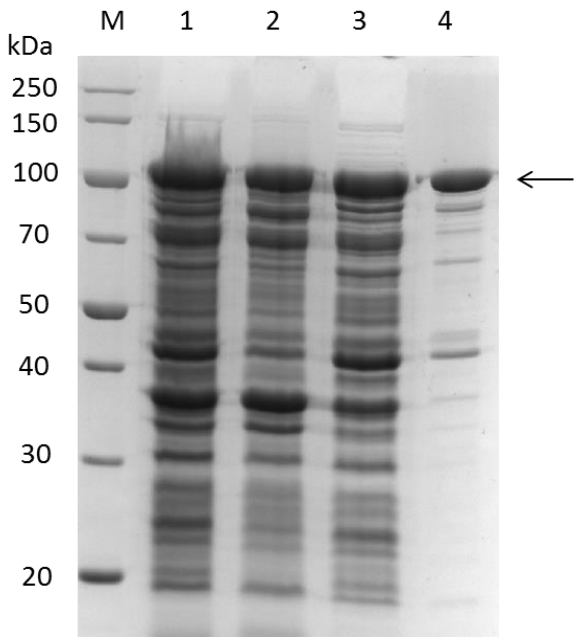


Figure 16. SDS-PAGE gel (12 %) analysis of expression and purification of Cwh41$_{\Delta 1\text{-}34}$p. The SDS-PAGE gel was stained by Coomassie blue. M, molecular weight standard; Lane 1, whole cell; Lane 2, insoluble fraction; Lane 3, soluble fraction; Lane 4, eluted fraction from His-trap column. The arrow indicates the expected protein band. An aliquot of 5µg Cwh41$_{\Delta 1\text{-}34}$p eluted fraction was loaded on the gel.

5. Discussion

5.1. MOGS expression and characterization

5.1.1. *MOGS* gene constructs

As previously described, MOGS has a type II transmembrane topology, consisting of a short N-terminal cytosolic tail (residues 1-38), a single transmembrane domain (residues 39-59), and a large C-terminal catalytic domain (residues 60-837). To obtain soluble protein expression, three different *MOGS* gene clones pPICZαB-$MOGS_{\Delta 1\text{-}59,}$ pPICZαB-$MOGS_{\Delta 1\text{-}64}$ and pPICZαB-$MOGS_{\Delta 1\text{-}71}$ with deletions of the N-terminus and transmembrane domain were constructed using the same strategy as two studies of *S. cerevisiae* Glu I(Faridmoayer & Scaman 2007; Barker *et al*. 2011).

5.1.2. The expression system selection for MOGS

In this project, both *P. pastoris* and *E. coli* expression systems were used to overexpress MOGS. The primary objective was to overexpress soluble MOGS protein. Initially, I tried to establish a *P. pastoris* expression system for MOGS as *P. pastoris* expression system has been used for many human protein preparations (Lei *et al*. 2012; Huang *et al*. 2014) as well as for the *S. cerevisiae* Glu I overexpression (Barker *et al*. 2011). *P. pastoris* expression system can also enable production of large amounts of the target proteins (Daly & Hearn, 2005). In the first expression trial, soluble $MOGS_{\Delta 1\text{-}64}$ was purified by His-trap column from the *P. pastoris* medium (Figure 5) and $MOGS_{\Delta 1\text{-}64}$ was enzymatically active. However, the extracellular expression was not reproducible, and in subsequent expression trials, the $MOGS_{\Delta 1\text{-}64}$ was

expressed as an inactive intracellular protein (Figure 8), although the expression condition was same as the first expression trial. The factors that affected the protein expression and led to un-reproducible result could not be identified in this project. Similarly, two different MOGS constructs of $MOGS_{\Delta 1\text{-}59}$ and $MOGS_{\Delta 1\text{-}71}$ were also expressed intracellularly in *P. pastoris* and no proteins could be eluted from the His-trap column.

Then *E. coli* was used to overexpress MOGS, since *E. coli* has been used extensively for heterologous protein expression due to its rapid growth rate and relatively low cost. Human proteins can be overexpressed in *E. coli* even though the expression rate of soluble proteins is not high. For example, the humanized antibody fragment anti-CD18 F(ab')2 was expressed in the periplasm in *E. coli* (Chen *et al*. 2004). However, MOGS was expressed in inclusion bodies (Figure 9). The *E. coli* expression system has several disadvantages particularly for eukaryotic protein expression. Firstly, N-glycosylation has not been found in *E. coli*. In *E. coli*, only two glycoproteins have been identified, AIDA-I and TibA, which are O-glycosylated (Charbonneau *et al*. 2007). Therefore, Glu I expressed in *E. coli* would not have the N-glycosylation that may be required for correct protein folding. Secondly, the lack of eukaryotic chaperones in *E. coli* can result in protein misfolding and aggregation (Dyson *et al*. 2004). One study using 54 human genes as test targets for *E. coli* expression reported that 35 clones were expressed as recombinant proteins, of which 12 were soluble and only four were able to be purified to homogeneity (Ding *et al*. 2002). To obtain soluble protein from the inclusion bodies, two different methods (urea denaturation-refolding and freeze-thawing) were used (Figure 10 &11). However, the recovery rate of MOGS from inclusion bodies was very low (3% for urea denaturation-refolding and 6% for freeze-thawing), and the recovered protein did not show any enzyme activity. As a result, the

attempt of overexpressing soluble and active MOGS in either *P. pastoris* or *E. coli* was not successful.

5.1.3. N-linked glycosylation of MOGS overexpressed in *P. pastoris*

$MOGS_{\Delta 1\text{-}64}$ (pPICZαB) has an expected mass of 97 kDa, but based on the molecular weight standards, the estimated molecular weight was calculated at 101 kDa on SDS-PAGE (Figure 5B). Periodic acid-Schiff staining indicated that eluted fraction from the His-trap column contained sugar chains (Figure 6). However, this method is not able to determine whether the target $MOGS_{\Delta 1\text{-}64}$ (pPICZαB), degradation fragments during 12 days storage (Figure 7), or a contaminating protein contained the sugar chains. Therefore, the more specific method of de-glycosylation by PNGase F was used. According to the SDS-PAGE gel with PNGase F treatment (Figure 7), the N-glycosylation was confirmed in the 101 kDa target band, which accounted for the increase in molecular weight. Similarly, MOGS expressed in mammalian COS 1 cell showed a molecular mass decrease from 95 kDa to 93 kDa after endoglycosidase H treatment, indicating a high-mannose glycosylation at the single N-glycosylation site of Asn655 (Kalz-Fuller *et al.* 1995). Glycosylation may occur if there are recognized glycosylation sites (Asn-X-Ser/Thr) in protein's primary sequence when expressed in *P. pastoris*. There is one N-glycosylation site at MOGS at Asn655, so it is not surprising to observe glycosylation in MOGS when expressed in *P. pastoris*. Furthermore, $MOGS_{\Delta 1\text{-}64}$ that expressed in either extracellular soluble form (Figure 5) or intracellular insoluble form (Figure 8) was N-linked glycosylated, suggesting that glycosylation was not a determining factor for enzyme solubility. Therefore, it may be that glycosylation is not required for the correct conformation/folding of MOGS. However, the de-glycosylation of $MOGS_{\Delta 1\text{-}64}$ was incomplete according to the SDS-PAGE gel and ChemiDoc

tool, 35% of the protein was calculated as de-glycosylated based on the band intensities at 97 kDa and 101 kDa. The positive control of chicken egg white conalbumin showed molecular mass decrease (Figure 7, lane 4) after de-glycosylation but not for the negative control of bovine serum albumin (Figure 7, lane 8). Therefore, the PNGase F treatment was likely to be efficient. The reason of incomplete de-glycosylation for MOGS remains unknown.

S. cerevisiae Glu I that expressed in *P. pastoris* was also N-linked glycosylated, its mass decreased 2-4 kDa after de-glycosylation by PNGase F treatment (Barker *et al*. 2011). Furthermore, two N-glycans were observed at Asn 9 and 89 (two predicted N-glycosylation sites) of Cwht1p based on its crystal structure (Barker & Rose, 2013). The C-terminal 37 kDa of Cwh41p released by trypsin digestion was catalytically active even without glycosylation modification (Faridmoayer & Scaman, 2005). Similar specific activity (~3000 U/mg) was observed between Cwht1p (without glycosylation) and the Cwh41p (with glycosylation) that expressed in *S. cerevisiae* (Faridmoayer & Scaman, 2007). Therefore, the glycosylation may not be required for catalytic activity of yeast Glu I.

5.1.4. Intracellular MOGS overexpression in *P. pastoris*

$MOGS_{\Delta 1\text{-}64}$ (pPICZαB) was expressed as insoluble protein in subsequent expression trials, even though the same expression conditions were used as in the first trial. In this experiment, the pPICZαB vector is fused with the α-factor secretion signal, which facilitates extracellular protein expression. However, the expressed protein could not be secreted into the medium but was constrained inside of the cell. One possible reason is that internal dibasic amino acids of Lys-Arg in MOGS affected the secretion signal. The cleavage of the "prepro" α mating factor signal

sequence by Kex2 protease occurs between arginine and glutamic acid in the sequence Lys-Arg-$(Glu\text{-}Ala)_3$, resulting in the secretion of high levels of proteins to the medium (Brake *et al*. 1984). However, proteins containing the internal accessible dibasic amino acids such as Lys-Arg and Arg-Arg in the coding region cannot be expressed using this signal sequence, as the protein will be fragmented (Govindappa *et al*. 2013). There is a Lys-Arg at position of 174-175 in full length MOGS, which could lead to intracellular protein expression instead of secretion.

Another possible reason is that the native signal sequence of MOGS is required for protein secretion. *P. pastoris* have low specificity requirement for signal sequence recognition (Cregg & Higgins, 1995). While different secretion signal sequences including the native secretion signal and *S. cerevisiae* "prepro" α mating factor have resulted in variable success. The secretion signal sequence from the *S. cerevisiae* "prepro" α mating factor has been used successfully to express human interleukin-2-serum albumin in *P. pastoris* (Lei *et al*. 2012). The native signal sequence of bovine lysozyme has also been successfully utilized to express the lysozyme in *P. pastoris* (Digan *et al*. 1989). There is no definitive means to pre-determine whether a native signal or α-secretion factor will result in successful secretion (Daly & Hearn, 2005). Therefore, the native signal sequence of MOGS may be required for the protein secretion when expressed in *P. pastoris*.

5.1.5. MOGS purification from *E. coli*

Two different methods of urea-denaturation and refolding as well as freeze-thawing were used to purify $MOGS_{\Delta 1\text{-}64}$ from *E. coli* inclusion bodies. A high concentration of denaturant (8M urea) was first used to solubilize the inclusion body (Figure 10A, lane 3), with the assumption that the

protein may be refolded into its native state by dialysis with gradual removal of urea (Singh & Panda, 2005), however, the protein was poorly recovered during the step-wise dialysis (Figure 10B). Since the denaturant causes the loss of secondary structure, leading to random coil and exposure of hydrophobic surfaces (Singh *et al*. 2015), at medium denaturant concentrations (2-3 M urea), the proteins often re-form into aggregates during the refolding process (Tsumoto *et al*. 2003). In this experiment, the solubilized protein aggregated during the 2M urea denaturant concentration. Some studies have found that protein aggregation predominantly occurs at medium concentrations of denaturant (Ho *et al*. 2003, Liu *et al*. 2005).

The method of freeze-thawing was used to overcome the obstacle of aggregation. It has been reported that freeze-thawing of protein solutions can cause protein denaturation by destabilizing the protein tertiary structure with retention of secondary structure (Cao *et al*. 2003; Qi *et al*. 2015). This mild solubilization method has been reported to preserve the existing native secondary structure of proteins, reducing protein aggregation during refolding (Singh *et al*. 2015); however, this was not the case in this experiment. The low level of solubilized protein after freeze-thawing led to low recovery efficiency of MOGS.

5.2. Yeast Glu I expression and characterization

5.2.1. *CWH41* truncation gene constructs

Heterologous overexpression of Cwh41p truncations (Cwht41$_{\Delta 1\text{-}319}$p and Cwht41$_{\Delta 1\text{-}525}$p) in *S. cerevisiae* (Faridmoayer & Scaman, 2007) and *P. pastoris* (Barker & Rose, 2013) have been carried out. However, there was no active protein expression. Based on the Cwht1p crystal

structure, Barker & Rose (2013) proposed that the truncation of Cwht41$_{\Delta 1\text{-}319}$p starting at the random coil linker region prior to LH1 led to exposure of LH1 hydrophobic patch and misfolding, degradation or aggregation, while truncation of Cwht41$_{\Delta 1\text{-}525}$p started at the long disorder loop between CH4 and CH5 led to incomplete translation and incorrect folding. This indicates that the residues 35-525 may be essential for the proper assembly and folding of Cwh41p. Furthermore, it has been shown that the trypsin-hydrolyzed C-terminal catalytic domain (37 kDa) of Cwh41p was catalytically active without the residues 35-525 (Faridmoayer & Scaman, 2005), which suggests that residues 35-525 do not play a direct role in catalysis. To determine the function of residues 35-525, two truncated proteins Cwh41$_{\Delta 1\text{-}314}$ and Cwh41$_{\Delta 1\text{-}349}$ were constructed in our lab based on the structure of Cwht1p, which resulted in truncations with a complete N-domain and with the N-domain and linker region (LH1), respectively (unpublished data from Venkatrao Konasani thesis). However, both were expressed as inclusion bodies without enzyme activity (data not shown).

To continue the study, six longer Cwh41p truncation products (Cwh41$_{\Delta 1\text{-}167}$p, Cwh41$_{\Delta 1\text{-}93}$p, Cwh41$_{\Delta 1\text{-}74}$p, Cwh41$_{\Delta 1\text{-}54}$p, Cwh41$_{\Delta 1\text{-}45}$p and Cwh41$_{\Delta 1\text{-}39}$p) were constructed in this project. In order to achieve functional protein expression, all of these constructs contained the $(\alpha/\alpha)_6$ barrel and linker region (LH1) and additional N-terminal residues were included comparing with the two studies by Faridmoayer & Scaman (2007) and Barker & Rose (2013). The two truncations (pET30a-*CWH41*$_{\Delta 1\text{-}93}$ and pET30a-*CWH41*$_{\Delta 1\text{-}167}$) were constructed initially based on the protein structure of Cwht1p (Figure 4). The beginning point of these two gene products (Cwh41$_{\Delta 1\text{-}93}$p, Cwh41$_{\Delta 1\text{-}167}$p) started at residue 94 and residue 168 of Cwh41p, which corresponded to residue 60 and residue 134 of Cwht1p. Therefore, Cwh41$_{\Delta 1\text{-}93}$p would result in the deletion of one α-

helix (NH1) and two β-strands (NS1,NS2) while Cwh41$_{\Delta 1\text{-}167}$p would result in deletion of one α-helix (NH1) and seven β-strands (NS1-NS7). The two truncations were expected to possess a complete secondary structure unit (α-helix and β-strand). Two longer clones (pET30a-*CWH41*$_{\Delta 1\text{-}54}$ and pET30a-*CWH41*$_{\Delta 1\text{-}74}$) were constructed. The beginning point of these two gene products started at residue 55 and residue 75 of Cwh41p, which were corresponded to residue 21 and residue 41 of Cwht1p. Therefore, Cwh41 $_{\Delta 1\text{-}54}$ would result in one α-helix (NH1) deletion while Cwh41 $_{\Delta 1\text{-}74}$ would result in one α-helix (NH1) deletion as well as part of the NH2 helix deletion. Finally, the two longest clones (pET30a-*CWH41*$_{\Delta 1\text{-}39}$ and pET30a-*CWH41*$_{\Delta 1\text{-}45}$) were constructed based on the protein structure of Cwht1p (Figure 4). The beginning points of these two truncations started at residue 40 and residue 46 of Cwh41p, which are corresponded to residue 6 and residue 12 of Cwht1p. Therefore, Cwh41$_{\Delta 1\text{-}39}$p would result in half α-helix (NH1) deletion while Cwh41$_{\Delta 1\text{-}45}$p would result in the α-helix (NH1) deletion.

5.2.2. Truncated Cwh41p overexpression in *E. coli*

The third objective of this work was to determine the function of N-terminus of Glu I. Since the attempt of overexpressing soluble and active MOGS in either *P. pastoris* or *E. coli* was not successful, *S. cerevisiae* Glu I was used to study the N-terminal function of Glu I. *E. coli* system was used to overexpress Cwh41p with N-terminal truncations. The *E. coli* expression system used for this work was previously established to overexpress Cwh41$_{\Delta 1\text{-}34}$p (unpublished data from Venkatrao Konasani thesis).

These proteins were expressed mainly in inclusion bodies (Figure 12 & 14) and the eluted fraction from the His-trap column had no enzyme activity. The two truncations of Cwh41$_{\Delta 1\text{-}39}$p

and Cwh41$_{\Delta 1\text{-}45}$p lack 5 amino acids and 11 amino acids, respectively, were also expressed mainly in inclusion bodies (Figure 15). However, the eluted fraction of Cwh41$_{\Delta 1\text{-}45}$p from the His-trap column was not active but the eluted fraction of Cwh41$_{\Delta 1\text{-}39}$p was active with a specific activity of 266 U/mg. This data indicates that the residues 40-45 of Cwh41p may be essential for the proper assembly and folding of Cwh41p. Furthermore, the internal positive control, eluted fraction of Cwh41$_{\Delta 1\text{-}34}$p from His-trap column, had a specific activity of 2000 U/mg against the trisaccharide substrate. The elimination of the residues 35-39 of Cwh41p caused a reduction in specific activity from 2000 U/mg to 266 U/mg. So the residues 35-39 of Cwh41p may also play an important role in the proper assembly and folding of Cwh41p.

Based on SDS-PAGE analysis using ChemiDoc software, in the eluted fraction, ~ 70% of protein was Cwh41$_{\Delta 1\text{-}34}$p by comparing the relative band intensity (Figure 16, lane 4); ~ 21% of protein was Cwh41$_{\Delta 1\text{-}39}$p (Figure 15B, lane 3). This may explain the lower enzyme activity since the degradation fragments are not likely to contribute to the enzyme activity. Based on SDS-PAGE band intensity analysis (ChemiDoc software, equation: soluble target protein intensity/soluble target protein intensity + insoluble target protein intensity), a significant percentage (~56%) of Cwh41$_{\Delta 1\text{-}34}$p was soluble, while in comparison, ~28% of Cwh41$_{\Delta 1\text{-}39}$p was soluble. This suggests that the absence of residues 35-39 of Cwh41p caused the solubility to decrease significantly in Cwh41$_{\Delta 1\text{-}39}$p. In fact, when comparing the solubility percentage among the other Cwh41p truncations (Cwh41$_{\Delta 1\text{-}93}$p, Cwh41$_{\Delta 1\text{-}74}$p, Cwh41$_{\Delta 1\text{-}54}$p and Cwh41$_{\Delta 1\text{-}45}$p), they all have solubility percentage less than 25%. There was no evidence for a trend of increasing solubility among the four Cwh41p truncations mentioned above. However, it is evident that the Cwh41$_{\Delta 1\text{-}34}$p and Cwh41$_{\Delta 1\text{-}39}$p have higher solubility percentage than the four truncations. It was previously

reported that the soluble yeast Glu I (95% homogeneity) had specific activity of 3130 U/mg against the a trisaccharide substrate α-D-Glc1,2α-D-Glc1,3α-D-Glc–$O(CH2)_8COOCH_3$ (Faridmoayer & Scaman, 2004); Cwh41$_{\Delta 1\text{-}34}$p expressed in *S. cerevisiae* was purified with 90% homogeneity, and its specific activity was 3600 U/mg (Faridmoayer & Scaman, 2007). Furthermore, Cwh41$_{\Delta 1\text{-}34}$p expressed in *E. coli* has been purified to 95% homogeneity and had specific activity of 3370 U/mg with the trisaccharide substrate α-D-Glc1,2α-D-Glc1,3α-D-Glc–OCH_3 used in this project (unpublished data from Venkatrao Konasani thesis). Therefore, the measured specific activity of Cwh41$_{\Delta 1\text{-}34}$p in this project was comparable to the previous studies. Based on the crystal structure of Cwht1p as shown in Figure 17, the first N-terminal α-helix (NH1) is close to the C-domain. The first glycan located at the N-terminal α-helix was found at the distal end of the crystal stacking model and there was no major protein-sugar contact in the modeled residues (Barker & Rose, 2013). However, the N-terminal α-helix (NH1) may still interact with the C-domain and affect the steric structure of the enzyme. Considering that there was no active protein expression in truncations (Cwht41$_{\Delta 1\text{-}319}$p and Cwht41$_{\Delta 1\text{-}525}$p) (Faridmoayer & Scaman, 2007; Barker & Rose, 2013), and that it was shown in this work that Cwh41$_{\Delta 1\text{-}34}$p and Cwh41$_{\Delta 1\text{-}39}$p have higher solubility percentage than the four shorter proteins (Cwh41$_{\Delta 1\text{-}93}$p, Cwh41$_{\Delta 1\text{-}74}$p, Cwh41$_{\Delta 1\text{-}54}$p and Cwh41$_{\Delta 1\text{-}45}$p), I propose here that the α-helix (NH1) plays a positive role in protein folding by interacting with the C-domain of the enzyme during protein expression.

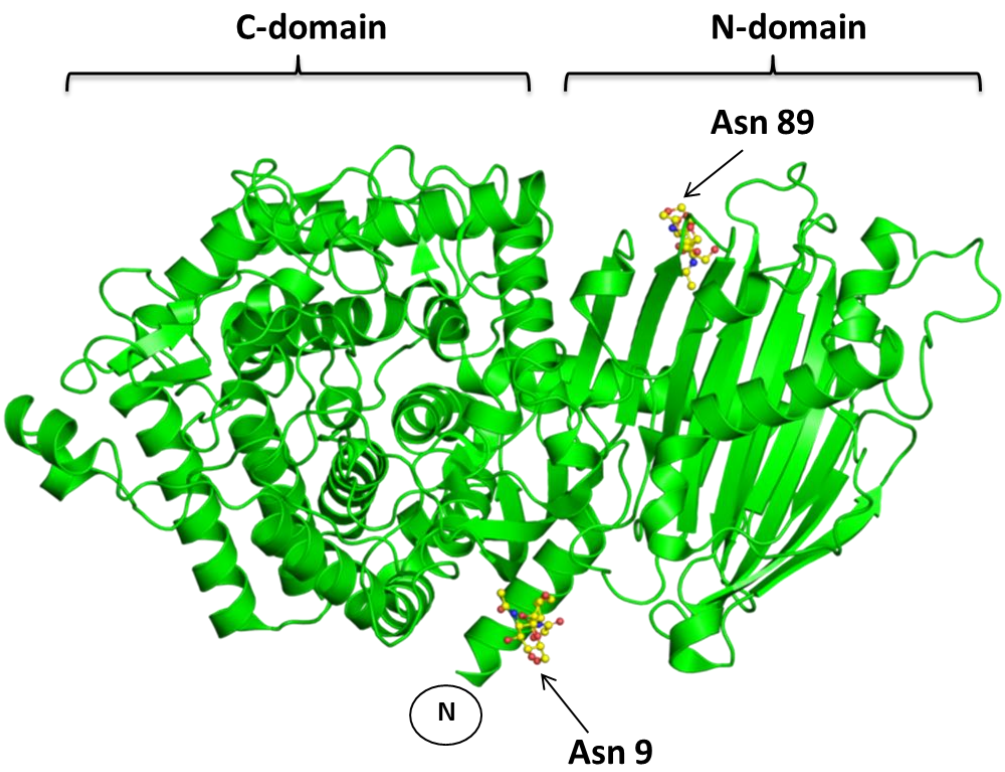


Figure 17. Crystal structure of Cwht1p (PDB number 4J5T). This is the side view of the structure, source: Protein Data Bank, http://www.ebi.ac.uk/pdbe/entry/pdb/4J5T. Two N-glycans link to the asparagine residues 9 and 89. This figure is adapted from Barker & Rose (2013).

6. Conclusions and Future Studies

6.1. Conclusions

In this project, the construct of pPICZαB-*MOGS*$_{\Delta 1\text{-}64}$ was transformed into *P. pastoris.* The MOGS$_{\Delta 1\text{-}64}$ (pPICZαB) was purified by FPLC using His-trap column from the first expression trial. The purified MOGS was N-linked glycosylated with a higher molecular size (101kDa) than the expected size (97 kDa). The purified MOGS was active (specific activity was 735 U/mg) against the synthetic trisaccharide substrate α-D-Glc1,2α-D-Glc1,3α-D-Glc–OCH3. However, the expression of MOGS$_{\Delta 1\text{-}64}$ (pPICZαB) was not reproducible and it was expressed as an insoluble protein in subsequent expression trials. Two proteins MOGS$_{\Delta 1\text{-}59}$ and MOGS$_{\Delta 1\text{-}71}$ were also expressed as insoluble proteins in *P. pastoris*.

The MBP fusion MOGS was expressed in inclusion bodies and the eluted fraction from the amylase resin column was not active. Two different methods were used to purify MOGS$_{\Delta 1\text{-}64}$ (pET47b) from *E. coli* inclusion bodies. However, very little protein was recovered by the two methods. Moreover, the recovered MOGS$_{\Delta 1\text{-}64}$ (pET47b) by the two methods had no enzyme activity. As the MOGS overexpression in either *P. pastoris* or *E. coli* was not successful, more research to find a suitable expression system for MOGS is needed in the future.

Expression of six Cwh41p truncations was carried out using a previously established *E. coli* system. Among the truncations, Cwh41$_{\Delta 1\text{-}167}$p, Cwh41$_{\Delta 1\text{-}93}$p, Cwh41$_{\Delta 1\text{-}74}$p, Cwh41$_{\Delta 1\text{-}54}$p and Cwh41$_{\Delta 1\text{-}45}$p was expressed mainly in inclusion bodies and the eluted fractions had no enzyme activity. The eluted fraction of Cwh41$_{\Delta 1\text{-}39}$p was active with a specific activity of 266 U/mg,

indicating that the residues 40-45 of Cwh41p may be essential for the proper assembly and folding of Cwh41p. However, the eluted fraction of $Cwh41_{\Delta 1\text{-}34}p$ had a specific activity of 2000 U/mg, indicating the residues 35-39 of Cwh41p may also play an important role in the proper assembly and folding of Cwh41p.

In conclusion, the attempt to establish *P. pastoris* or *E. coli* expression for human Glu I was not successful. However, the N-terminus in *S. cerevisiae* Glu I was found to be important for proper protein folding. I propose that the α-helix (NH1) of yeast Glu I plays a positive role in protein folding by interacting with C-domain of the enzyme.

6.2. Future Studies

6.2.1. Future perspectives for MOGS

In this project, the N-terminus and the transmembrane domain of the three proteins $MOGS_{\Delta 1\text{-}59}$, $MOGS_{\Delta 1\text{-}64}$ and $MOGS_{\Delta 1\text{-}71}$ were deleted in order to achieve solubilized expression, similar to the soluble expression of Cwht41p. However, the MOGS was trapped inside the yeast cell without secretion. One possible reason is that internal dibasic amino acids of Lys-Arg in MOGS affected the secretion efficiency when using α-factor signal (Govindappa *et al*. 2013). In contrast, the native signal sequence of MOGS could be fused to the recombinant proteins, in order to improve the secretion efficiency. Furthermore, the transmembrane domain (residues 39-59) acts biochemically as a signal sequence of MOGS and the successful secretion may need this native signal sequence. So the whole N-terminus (residues 1-59) of MOGS can be fused in front of the gene to obtain secreted MOGS expression in *P. pastoris* in the future.

On the other hand, the first N-terminal α-helix of yeast Glu I has been found to be important for the enzyme expression and activity. Based on the secondary structure prediction results for MOGS (Appendix Figure 4) by using PSIPRED from UCL Department of Computer Science Bioinformatics Group (http://bioinf.cs.ucl.ac.uk/psipred/), the residues between Ala45 to Arg67 are predicted as the first N-terminal α-helix of MOGS. It is reasonable to predict that the residues 45-59 of MOGS are also important for enzyme folding. So it is also important to investigate whether the inclusion of residues 45-67 of MOGS can improve protein native folding and soluble protein expression when overexpressed either in *E. coli* or *P. pastoris* in the future.

The 6xHis tag and MBP tag were not able to enhance the solubility of heterologous MOGS when overexpressed in *E. coli* in this project. Instead, MOGS was accumulated as inclusion bodies. However, *E. coli* expression system possesses advantages of rapid growth rate and relatively low cost in comparison with other expression systems. Furthermore, the glycosylation may not be required for the correct conformation/folding of MOGS. To obtain large quantities of MOGS protein, *E. coli* expression system with adjustments and improvements can be carried out in the future. Previous study in our laboratory has found that heterologous expression of the yeast catalytic domain Cwh41$_{\Delta 1\text{-}525}$p in *E. coli* was in inclusion bodies (unpublished data from Venkatrao Konasani thesis). To improve its solubility, the molecular chaperones and solubility-enhancement tag, Glutathione S-transferase (GST) tag were used, but there was no solubility improvement (unpublished data from Venkatrao Konasani thesis). Disulfide bond formation may be required for proper folding of the eukaryotic proteins (Yin *et al*. 2007). Disulfide bond formation has been found in prokaryotes of *E. coli* and the formation occurs in periplasm

(Hatahet *et al*. 2014). However, in *E. coli* expression system, recombinant proteins are expressed in the cytoplasm. So recombinant proteins which require disulfide bonds for their folding can be poorly expressed, misfolded, and are not active when expressed in the *E. coli* cytoplasm (Lobstein *et al*. 2012). However, *E. coli*-engineered host strain, for example, Shuffle strain (available from New England Biolabs, https://www.neb.com/products/c3026-shuffle-t7-competent-e-coli) that is derived from the thioredoxin reductase/glutathione reductase suppressor strain SMG96 with diminished cytoplasmic reductive pathways, allows for the formation of disulfide bonds in the cytoplasm (Lobstein *et al*. 2012). Moreover, there is one predicted disulfide-bond in MOGS according to a web server DiANNA (http://clavius.bc.edu/~clotelab/DiANNA/)which has been described by Ferrè & Clote (2005). So it is important to examine if the Shuffle strain could be used to achieve soluble MOGS expression in the future.

6.2.2. Future perspectives for yeast Glu I

Gene expression is a multi-step process, which proceeds from DNA through RNA to protein and RNA molecules are the central effectors of the genetic information flow. Regulation of the messenger RNA (mRNA) levels has emerged as a crucial event in controlling the physiological levels of proteins. The rate of mRNA production (transcription level), mRNA stability and translation efficiency control protein levels (Cheneval *et al*. 2010). No protein overexpression of Cwh41p truncations (Cwht41$_{\Delta 1\text{-}319}$p and Cwht41$_{\Delta 1\text{-}525}$p) in *S. cerevisiae* (Faridmoayer & Scaman, 2007) was detected. The authors proposed that the deletion of residues 35-525 may have an effect on gene transcription or stability of mRNA (Faridmoayer & Scaman, 2007). However, in this project, the expression of six Cwh41p truncations could be detected, but most of them were

not enzymatically active except for Cwht41$_{\Delta 1\text{-}39}$p and the internal control Cwht41$_{\Delta 1\text{-}34}$p. Therefore, it is important to further investigate whether the protein expression level in these truncations is related to gene transcription level in the future. To quantify the transcription level (mRNA production), reverse transcription polymerase chain reaction (RT-PCR) can be used. In addition, stability of mRNA is also related to the protein level, the analysis of the decay rates of the mRNA over time by real time RT-PCR has been reported in study (Leclerc *et al*. 2002). To further investigate the stability of mRNA in these truncations can provide a clue of protein expression levels. Furthermore, based on the data obtained in this project, N-terminal residues 35-45 of Cwh41p are essential for the proper assembly and folding. It is essential to further investigate the function of the 11 residues individually. Site mutation for each residue can be carried out in the future to examine the protein expression and solubility.

The soluble form of yeast Glu I (Cwht1p) has been purified in this project and it was enzymatically active against the synthetic substrate. Moreover, the measured specific activity of Cwht1p in this project was comparable to the previous studies as described above. However, it is not conclusive to report that the Cwht1p in this project has same characteristics with yeast Glu I from previous studies due to the lack of enzyme kinetic study. An enzyme's kinetics is important to reveal the catalytic mechanism and its role in metabolism. The Michaelis constant value (Km) is an important parameter for a specific enzyme. For the yeast Glu I enzyme, its Km value of α-D-Glc1,2α-D-Glc1,3α-D-Glc–O(CH2)$_8$COOCH$_3$ has been determined to be 1.28 mM (Neverova *et al*. 1994); Km value of α-D-Glc1,2α-D-Glc1,3α-D-Glc–OCH$_3$ was determined to be 1.27 mM (unpublished data from Venkatrao Konasani thesis). To further characterize the yeast Glu I enzyme in this project, kinetic study is required in the future.

References

Apweiler, R. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1473*(1), 4–8. http://doi.org/10.1016/S0304-4165(99)00165-8

Asano, N. (2003). Glycosidase inhibitors: Update and perspectives on practical use. *Glycobiology, 13*(10), 93-104. doi:10.1093/glycob/cwg090

Barker, M. K., & Rose, D. R. (2013). Specificity of processing α-glucosidase I is guided by the substrate conformation: crystallographic and in silico studies. *The Journal of Biological Chemistry*, *288*(19), 13563–74. http://doi.org/10.1074/jbc.M113.460436

Barker, M. K., Wilkinson, B. L., Faridmoayer, A., Scaman, C. H., Fairbanks, A. J., & Rose, D. R. (2011). Production and crystallization of processing α-glucosidase I: *Pichia pastoris* expression and a two-step purification toward structural determination. *Protein Expression and Purification*, *79*(1), 96–101. http://doi.org/10.1016/j.pep.2011.05.015

Bause, E., Erkens, R., Schweden, J., & Jaenicke, L. (1986). Purification and characterization of trimming glucosidase I from *Saccharomyces cerevisiae*. *FEBS Letters*, *206*(2), 208–212. http://doi.org/10.1016/0014-5793(86)80982-6

Bause, E., Schweden, J., Gross, A., & Orthen, B. (1989). Purification and characterization of trimming glucosidase I from pig liver. *European Journal of Biochemistry*, *183*(3), 661–669. http://doi.org/10.1111/j.1432-1033.1989.tb21096.x

Block, T. M., Lu, X., Platt, F. M., Foster, G. R., Gerlich, W. H., Blumberg, B. S., & Dwek, R. A. (1994). Secretion of human hepatitis B virus is inhibited by the imino sugar N-butyldeoxynojirimycin. *Proceedings of the National Academy of Sciences of the United States of America, 91*(6), 2235-2239. doi:10.1073/pnas.91.6.2235

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry, 72*(1), 248-254. doi:10.1016/0003-2697(76)90527-3

Brake, A. J., Merryweather, J. P., Coit, D. G., Heberlein, U. A., Masiarz, F. R., Mullenbach, G. T.. . Barr, P. J. (1984). α -factor-directed synthesis and secretion of mature foreign proteins in saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America, 81*(15), 4642-4646. doi:10.1073/pnas.81.15.4642

Bridges, C. G., Brennan, T. M., Taylor, D. L., McPherson, M., & Tyms, A. S. (1994). The prevention of cell adhesion and the cell-to-cell spread of HIV-1 in vitro by the α-glucosidase 1 inhibitor, 6- O-butanoyl castanospermine (MDL 28574). *Antiviral Research, 25*(2), 169-175. doi:10.1016/0166-3542(94)90105-8

Cao, E., Chen, Y., Cui, Z., & Foster, P. R. (2003). Effect of freezing and thawing rates on denaturation of proteins in aqueous solutions. *Biotechnology and Bioengineering*, *82*(6), 684–690. http://doi.org/10.1002/bit.10612

Chang, J., Schul, W., Butters, T. D., Yip, A., Liu, B., Goh, A.. . Block, T. M. (2011). Combination of α-glucosidase inhibitor and ribavirin for the treatment of dengue virus infection in vitro and in vivo. *Antiviral Research, 89*(1), 26-34. doi:10.1016/j.antiviral.2010.11.002

Chang, J., Warren, T. K., Zhao, X., Gill, T., Guo, F., Wang, L., Block, T. M. (2013). Small molecule inhibitors of ER α-glucosidases are active against multiple hemorrhagic fever viruses. *Antiviral Research, 98*(3), 432-440. doi:10.1016/j.antiviral.2013.03.023

Cheneval, D., Kastelic, T., Fuerst, P., & Parker, C. N. (2010). A review of methods to monitor the modulation of mRNA stability: A novel approach to drug discovery and therapeutic intervention. *Journal of Biomolecular Screening, 15*(6), 609-622. doi:10.1177/1087057110365897

Charbonneau, M., Girard, V., Nikolakakis, A., Campos, M., Berthiaume, F., Dumas, F., Mourez, M. (2007). O-linked glycosylation ensures the normal conformation of the autotransporter adhesin involved in diffuse adherence. *Journal of Bacteriology, 189*(24), 8880-8889. doi:10.1128/JB.00969-07h

Daly, R., & Hearn, M. T. W. (2005). Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *Journal of Molecular Recognition : JMR*, *18*(2), 119–38. http://doi.org/10.1002/jmr.687

Dhanawansa, R., Faridmoayer, a., van der Merwe, G., Li, Y. X., & Scaman, C. H. (2002). Overexpression, purification, and partial characterization of *Saccharomyces cerevisiae* processing alpha glucosidase I. *Glycobiology*, *12*(3), 229–234. http://doi.org/10.1093/glycob/12.3.229

Digan, M., Lair, S., Brierley, R., Siegel, R., Williams, M., Ellis, S. Craig, W. (1989). Continuous production of a novel lysozyme via secretion from the yeast, *pichia pastoris*. *Bio/technology*, *7*(2), 160-164.

Ding, H., Ren, H., Chen, Q., Fang, G., Li, L., Li, R.. . Lund University. (2002). Parallel cloning, expression, purification and crystallization of human proteins for structural genomics. *Acta Crystallographica Section D*, *58*(12), 2102-2108. doi:10.1107/S0907444902016359

Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L., & McCafferty, J. (2004). Production of soluble mammalian proteins in escherichia coli: Identification of protein features that correlate with successful expression. *BMC Biotechnology*, *4*(1), 32-32. doi:10.1186/1472-6750-4-32

Egito, A. S., Girardet, J.-M., Miclo, L., & Gaillard, J.L. (2001). Highly sensitive periodic acid/Schiff detection of bovine milk glycoproteins electrotransferred after nondenaturing

electrophoresis, urea electrophoresis, and isoelectric focusing. *Le Lait*, *81*(6), 775–785. http://doi.org/10.1051/lait:2001104

Faridmoayer, A., & Scaman, C. H. (2004). An improved purification procedure for soluble processing alpha-glucosidase I from *Saccharomyces cerevisiae* overexpressing CWH41. *Protein Expression and Purification*, *33*(1), 11–8. http://doi.org/10.1016/j.pep.2003.09.013

Faridmoayer, A., & Scaman, C. H. (2005). Binding residues and catalytic domain of soluble *Saccharomyces cerevisiae* processing alpha-glucosidase I. *Glycobiology*, *15*(12), 1341–8. http://doi.org/10.1093/glycob/cwj009

Faridmoayer, A., & Scaman, C. H. (2007). Truncations and functional carboxylic acid residues of yeast processing alpha-glucosidase I. *Glycoconjugate Journal*, *24*(8), 429–37. http://doi.org/10.1007/s10719-007-9035-2

Ferrè, F., & Clote, P. (2005). Disulfide connectivity prediction using secondary structure information and diresidue frequencies. *Bioinformatics, 21*(10), 2336-2346. doi:10.1093/bioinformatics/bti328

Freeze, H. H. (2006). Genetic defects in the human glycome. *Nature Reviews. Genetics*, *7*(7), 537–51. http://doi.org/10.1038/nrg1894

Freeze, H. H., Chong, J. X., Bamshad, M. J., & Ng, B. G. (2014). Solving glycosylation disorders: Fundamental approaches reveal complicated pathways. *American Journal of Human Genetics*, *94*(2), 161–175. http://doi.org/10.1016/j.ajhg.2013.10.024

Fujimoto,K. and Kornfeld,S. (1991) α-Glucosidase II-deficient cells use endo α-mannosidase as a bypass route for N-linked oligosaccharide processing. *J. Biol. Chem*, 266, 3571–3578.

Govindappa, N., Hanumanthappa, M., Venkatarangaiah, K., Periyasamy, S., Sreenivas, S., Soni, R., & Sastry, K. (2014). A new signal sequence for recombinant protein secretion in pichia pastoris. *Journal of Microbiology and Biotechnology, 24*(3), 337-345. doi:10.4014/jmb.1308.08085

Grinna, L. S., & Robbins, P. W. (1979). Glycoprotein biosynthesis. rat liver microsomal glucosidases which process oligosaccharides. *Journal of Biological Chemistry, 254*(18), 8814.

Harju, S., Fedosyuk, H., & Peterson, K. R. (2004). Rapid isolation of yeast genomic DNA: Bust n' grab. *BMC Biotechnology*, *4*(1), 8-8. doi:10.1186/1472-6750-4-8

Hatahet, F., Boyd, D., & Beckwith, J. (2014). Disulfide bond formation in prokaryotes: History, diversity and design. *BBA - Proteins and Proteomics*, *1844*(8), 1402. doi:10.1016/j.bbapap.2014.02.014

Hebert, D. N., Foellmer, B., & Helenius, A. (1995). Glucose trimming and reglucosylation determine glycoprotein association with calnexin in the endoplasmic reticulum. *Cell*, 81, 425-43.

Hebert, D. N., & Molinari, M. (2007). In and out of the ER: protein folding, quality control, degradation, and related human diseases. *Physiological Reviews*, *87*(4), 1377–408. http://doi.org/10.1152/physrev.00050.2006

Hebert, D. N., & Molinari, M. (2012). Flagging and docking: dual roles for N-glycans in protein quality control and cellular proteostasis. *Trends in Biochemical Sciences*, *37*(10), 404–10. http://doi.org/10.1016/j.tibs.2012.07.005

Helenius, A., & Aebi, M. (2001). Intracellular functions of N-linked glycans. *Science*, *291*(5512), 2364-2369. doi:10.1126/science.291.5512.2364

Helenius, A., & Aebi, M. (2004). Roles of N-linked glycans in the endoplasmic reticulum. *Annual Review of Biochemistry*, *73*, 1019–49. http://doi.org/10.1146/annurev.biochem.73.011303.073752

Henrissat, B., Vegetales, M., & Grenoble, F. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities, *Biochem. J, 280*, 309–316.

Herscovics, A. (1999a). Importance of glycosidases in mammalian glycoprotein biosynthesis. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1473*(1), 96–107. http://doi.org/10.1016/S0304-4165(99)00171-3

Herscovics, A. (1999b). Processing glycosidases of *Saccharomyces cerevisiae*. *Biochimica et Biophysica Acta - General Subjects*, *1426*(2), 275–285. http://doi.org/10.1016/S0304-4165(98)00129-9

Hettkamp, H., Legler, G., & Bause, E. (1984). Purification by affinity chromatography of glucosidase I, an endoplasmic reticulum hydrolase involved in the processing of asparagine-linked oligosaccharides. *European Journal of Biochemistry / FEBS*, *142*(1), 85–90. http://doi.org/10.1111/j.1432-1033.1984.tb08253.x

Higgins, D. R., & Cregg, J. M. (1995). Production of foreign proteins in the yeast *pichia pastoris*. *Canadian Journal of Botany, 73*(S1), 891-897. doi:10.1139/b95-336

Hitt, R., & Wolf, D. H. (2004). DER7, encoding α-glucosidase I is essential for degradation of malfolded glycoproteins of the endoplasmic reticulum. *FEMS Yeast Research*, *4*(8), 815–820. http://doi.org/10.1016/j.femsyr.2004.04.002

Ho, J.G.S., Middelberg, A.P.J., Ramage P., Kocher, H.P. (2003). The likelihood of aggregation during protein renaturation can be assessed using the second virial coefficient. *Protein Science, 12*(4), 708-716. 10.1007/s11095-006-0018-y

Hong, Y., Sundaram, S., Shin, D. J., & Stanley, P. (2004). The Lec23 Chinese hamster ovary mutant is a sensitive host for detecting mutations in α-glucosidase I that give rise to congenital disorder of glycosylation IIb (CDG IIb). *Journal of Biological Chemistry, 279*(48), 49894–49901. http://doi.org/10.1074/jbc.M410121200

Huang, Y., Zhen, B., Lin, Y., Cai, Y., Lin, Z., Deng, C., & Zhang, Y. (2014). Expression of codon optimized human bone morphogenetic protein 4 in *pichia pastoris*. *Biotechnology and Applied Biochemistry, 61*(2), 175-183. doi:10.1002/bab.1146

Invitrogen, L. T. (2013). User manual - EasySelect™ Pichia Expression Kit. *Cat. No. K1740-01*, (25). http://doi.org/Manual part no. 25-0172

Jaeken, J. (2003). Congenital disorders of glycosylation (CDG): It's all in it. *Journal of Inherited Metabolic Disease, 26(2)*, 99-118. doi:10.1023/A:1024431131208

Jaeken, J., Hennet, T., Freeze, H. H., & Matthijs, G. (2008). On the nomenclature of congenital disorders of glycosylation (CDG). *Journal of Inherited Metabolic Disease, 31*(6), 669-672. doi:10.1007/s10545-008-0983-x

Jiang, B., Sheraton, J., Ram, A. F., Dijkgraaf, G. J., Klis, F. M., & Bussey, H. (1996). *CWH41* encodes a novel endoplasmic reticulum membrane N-glycoprotein involved in beta 1,6-glucan assembly. *Journal of Bacteriology, 178*(4), 1162-1171.

Kalz-Fuller, B., Bieberich, E., & Bause, E. (1995). Cloning and expression of glucosidase I from human hippocampus. *European Journal of Biochemistry, 231*(2), 344–351. http://doi.org/10.1111/j.1432-1033.1995.tb20706.x

Kilker, J., R D, Saunier, B., Tkacz, J. S., & Herscovics, A. (1981). Partial purification from *Saccharomyces cerevisiae* of a soluble glucosidase which removes the terminal glucose from the oligosaccharide $Glc_3Man_9GlcNAc_2$. *Journal of Biological Chemistry, 256*(10), 5299.

Kornfeld, R., & Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. *Annual Review of Biochemistry, 54*(1), 631-664. doi:10.1146/annurev.bi.54.070185.003215

Kurakata, Y., Uechi, A., Yoshida, H., Kamitori, S., Sakano, Y., Nishikawa, A., & Tonozuka, T. (2008). Structural insights into the substrate specificity and function of *Escherichia coli* K12 YgjK, a glucosidase belonging to the glycoside hydrolase family 63. *Journal of Molecular Biology*, *381*(1), 116-128. doi:10.1016/j.jmb.2008.05.061

Laemmli, U. K. (1970). Cleavage of structural proteins during assembly of head of bacteriophage-T4. *Nature*, *227*. http://doi.org/10.1038/227680a0

Leclerc, G., Leclerc, G., & Barredo, J. (2002). Real-time RT-PCR analysis of mRNA decay: Half-life of beta-actin mRNA in human leukemia CCRF-CEM and nalm-6 cell lines. *Cancer Cell International, 2*(1), 1-1-5. doi:10.1186/1475-2867-2-1

Lei, J., Guan, B., Li, B., Duan, Z., Chen, Y., Li, H., & Jin, J. (2012). Expression, purification and characterization of recombinant human interleukin-2-serum albumin (rhIL-2-HSA) fusion protein in *pichia pastoris*. *Protein Expression and Purification, 84*(1), 154-160

Liu, W., Cellmer, T., Keerl, D., Prausnitz, J.M., Blanch, H.W. (2005). Interactions of lysozyme in guanidinium chloride solutions from static and dynamic light-scattering measurements. *Biotechnology and Bioengineering, 90*(4), 482-490

Lobstein, J., Emrich, C. A., Jeans, C., Faulkner, M., Riggs, P., & Berkmen, M. (2012). SHuffle, a novel escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories, 11*(1), 56-56. doi:10.1186/1475-2859-11-56

Lubas, W. A., & Spiro, R. G. (1987). Golgi endo-alpha-D-mannosidase from rat liver, a novel N-linked carbohydrate unit processing enzyme. *Journal of Biological Chemistry, 262*(8), 3775.

Lubas, W. A., & Spiro, R. G. (1988). Evaluation of the role of rat liver golgi endo-alpha-D-mannosidase in processing N-linked oligosaccharides. *Journal of Biological Chemistry, 263*(8), 3990.

Marolda, C. L., Walter, P., Valvano, M. A., Helenius, J., Aebi, M., & Ng, D. T. W. (2002). Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature, 415*(6870), 447-450. doi:10.1038/415447a

Mehta, A., Zitzmann, N., Rudd, P. M., Block, T. M., & Dwek, R. A. (1998). α-glucosidase inhibitors as potential broad based anti-viral agents.*FEBS Letters, 430* (1): 17-22. doi:10.1016/S0014-5793(98)00525-0

Miyazaki, T., Matsumoto, Y., Matsuda, K., Kurakata, Y., Matsuo, I., Ito, Y., Tonozuka, T. (2011). Heterologous expression and characterization of processing α-glucosidase I from *Aspergillus brasiliensis* ATCC 9642. *Glycoconjugate Journal*, *28*(8-9), 563–571. http://doi.org/10.1007/s10719-011-9356-z

Neverova, I., Scaman, C. H., Srivastava, O. P., Szweda, R., Vijay, I. K., & Palcic, M. M. (1994). A spectrophotometric assay for glucosidase I. *Analytical Biochemistry, 222(1)*, 190-195. doi:10.1006/abio.1994.1472

Ou, W., Thomas, Cameron, P. H., D. Y., Bergeron & J. J. M. (1993). Association of folding intermediates of glycoproteins with calnexin during protein maturation. *Nature*, *364*(6440), 771-776. doi:10.1038/364771a0

Palcic, M. M., Scaman, C. H., Otter, A., Szpacenko, A., Romaniouk, A., Xia, Y., & Vijay, I. K. (1999). Processing α-glucosidase I is an inverting glycosidase, *Glycoconjugate Journal, 16*, 351–355.

Pan, Y. T., Hori, H., Saul, R., Sanford, B. A., Molyneux, R. J., & Elbein, A. D. (1983). Castanospermine inhibits the processing of the oligosaccharide portion of the influenza viral hemagglutinin. *Biochemistry, 22*(16), 3975-3984. doi:10.1021/bi00285a038

Papandréou, M., Barbouche, R., Guieu, R., Kieny, M. P., & Fenouillet, E. (2002). The alpha-glucosidase inhibitor 1-deoxynojirimycin blocks human immunodeficiency virus envelope glycoprotein-mediated membrane fusion at the CXCR4 binding step. *Molecular Pharmacology, 61*(1), 186-193. doi:10.1124/mol.61.1.186

Parodi, A. J., Behrens, N. H., Leloir, L. F., & Carminatti, H. (1972). The role of polyprenol-bound saccharides as intermediates in glycoprotein synthesis in liver. *Proceedings of the National Academy of Sciences of the United States of America, 69*(11), 3268-3272. doi:10.1073/pnas.69.11.3268

Praeter, C. M. De, Gerwig, G. J., Bause, E., Nuytinck, L. K., Vliegenthart, J. F. G., Breuer, W., Coster, R. N. Van. (2000). A Novel Disorder Caused by Defective Biosynthesis of N-Linked Oligosaccharides Due to Glucosidase I Deficiency, *American Journal of Human genetics*, 1744–1756.

Perry, S. T., Buck, M. D., Plummer, E. M., Penmasta, R. A., Batra, H., Stavale, E. J., Shresta, S. (2013). An iminosugar with potent inhibition of dengue virus infection in vivo. *Antiviral Research, 98*(1), 35-43. doi:10.1016/j.antiviral.2013.01.004

Qi, X., Sun, Y., & Xiong, S. (2015). A single freeze-thawing cycle for highly efficient solubilization of inclusion body proteins and its refolding into bioactive form. *Microbial Cell Factories, 14*, 24. http://doi.org/10.1186/s12934-015-0208-6

Ray M.K., Yang, J., Sundaram, S., Stanley, P. (1991). A novel glycosylation phenotype expressed by Lec23, a Chinese hamster ovary mutant deficient in alpha-glucosidase I. *J Biol Chem*, 266, 22818–22825

Romaniouk, A., & Vijay, I. K. (1997). Structure-function relationships in glucosidase I: amino acids involved in binding the substrate to the enzyme. *Glycobiology, 7*(3), 399–404. http://doi.org/10.1093/glycob/7.3.399

Romero, P. A., Dijkgraaf, G. J. P., Shahinian, S., Herscovics, A., & Bussey, H. (1997). The yeast CWH41 gene encodes glucosidase I, *Glycobiology, 7*(7), 997–1004.

Pukazhenthi, B. S., Varma, G. M., & Vijay, I. K. (1993). Conserved structural features in glycoprotein processing glucosidase I from several tissues and species. *Indian Journal of Biochemistry & Biophysics, 30*(6), 333.

Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Engineering, 12*(2), 85

Rye, C. S., & Withers, S. G. (2000). Glycosidase mechanisms, *Current Opinion on Chemical Biology*, 573–580. doi:10.1016/S1367-5931(00)00135-6

Sadat, M. A, Moir, S., Chun, T.-W., Lusso, P., Kaplan, G., Wolfe, L., Rosenzweig, S. D. (2014). Glycosylation, hypogammaglobulinemia, and resistance to viral infections. *The New England Journal of Medicine*, *370*(17), 1615–25. http://doi.org/10.1056/NEJMoa1302846

Santos, C. A., Beloti, L. L., Toledo, M. A. S., Crucello, A., Favaro, M. T. P., Mendes, J. S.,Souza, A. P. (2012). A novel protein refolding protocol for the solubilization and purification of recombinant peptidoglycan-associated lipoprotein from Xylella fastidiosa overexpressed in Escherichia coli. *Protein Expression and Purification*, *82*(2), 284–289. http://doi.org/10.1016/j.pep.2012.01.010

Saunier, B., Kilker, J., R D, Tkacz, J. S., Quaroni, A., & Herscovics, A. (1982). Inhibition of N-linked complex oligosaccharide formation by 1-deoxynojirimycin, an inhibitor of processing glucosidases. *The Journal of Biological Chemistry*, *257*(23), 14155.

Schachter, H. (2001). The clinical relevance of glycobiology. *Journal of Clinical Investigation*, *108*(11), 1579–1582. http://doi.org/10.1172/JCI200114498

Shailubhai, K., Pratta, M. a, & Vijay, I. K. (1987). Purification and characterization of glucosidase I involved in N-linked glycoprotein processing in bovine mammary gland. *The Biochemical Journal*, *247*(3), 555–62.

Shailubhai, K., Pukazhenthi, B. S., Saxena, E. S., Varma, G. M., & Vijay, I. K. (1991). Glucosidase I, a transmembrane endoplasmic reticular glycoprotein with a luminal catalytic domain. *Journal of Biological Chemistry*, 266(25), 16587.

Singh, S. M., & Panda, A. K. (2005). Solubilization and refolding of bacterial inclusion body proteins. *Journal of Bioscience and Bioengineering*, *99*(4), 303-310. doi:10.1263/jbb.99.303

Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M., & Panda, A. K. (2015). Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microbial Cell Factories*, *14*(1), 41. http://doi.org/10.1186/s12934-015-0222-8

Tsumoto, K., Ejima, D., Kumagai I., Arakawa, T. (2003). Practical considerations in refolding proteins from inclusion bodies. *Protein Expression and Purification*, *28*, 1-8. 10.1016/S1046-5928(02)00641-1

Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler, M.E. (2009). Essentials of glycobiology. *Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press.* ISBN-13: 9780879697709

Verbert, A., Cacan, R., & Cecchelli, R. (1987). Membrane transport of sugar donors to the glycosylation sites. *Biochimie, 69*(2): 91-99. doi: 10.1016/0300-9084(87)90240-9

Völker, C., Praeter, C. M. De, Hardt, B., Breuer, W., Kalz-füller, B., Coster, R. N. Van, & Bause, E. (2002). Processing of N-linked carbohydrate chains in a patient with glucosidase I deficiency (CDG type IIb), *Glycobiology, 12*(8), 473–483.

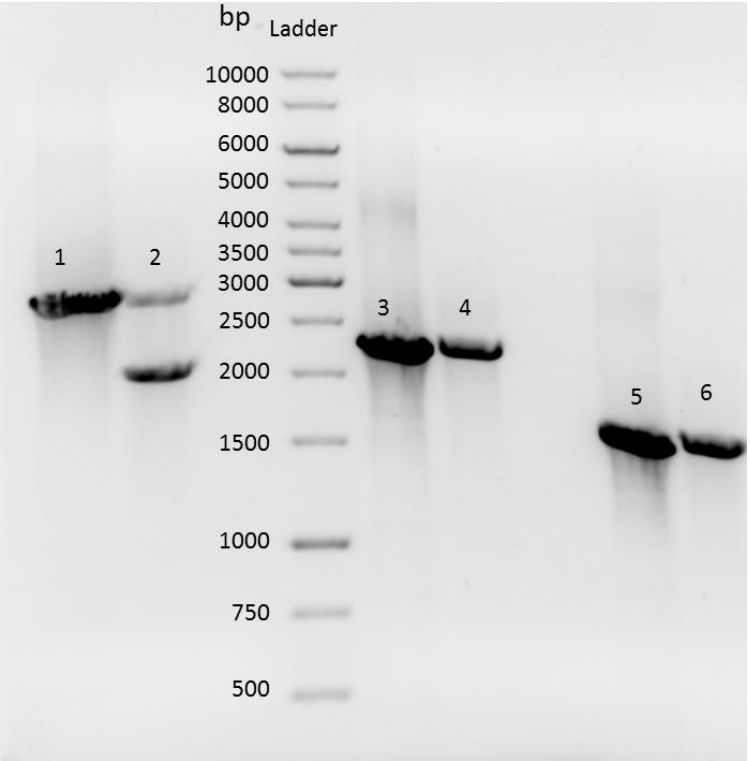
Wu, S., Lee, C., Liao, C., Dwek, R. A., Zitzmann, N., & Lin, Y. (2002). Antiviral effects of an iminosugar derivative on flavivirus infections. *Journal of Virology, 76*(8), 3596-3604. doi:10.1128/JVI.76.8.3596-3604.2002

Yin, J., Li, G., Ren, X., & Herrler, G. (2007). Select what you need: A comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *Journal of Biotechnology, 127*(3), 335-347. doi:10.1016/j.jbiotec.2006.07.012

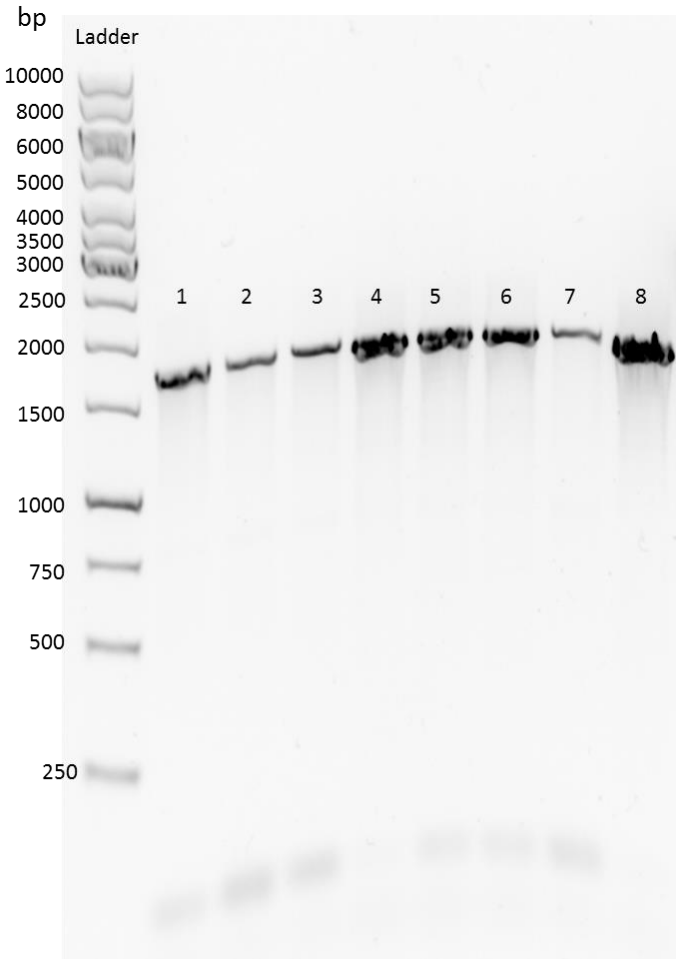
Zeng, Y. C., & Elbein, A. D. (1998). Purification to homogeneity and properties of plant glucosidase I. *Archives of Biochemistry and Biophysics*, *355*(1), 26–34. http://doi.org/10.1006/abbi.1998.0717

Zielinska, D. F., Gnad, F., Wiśniewski, J. R., & Mann, M. (2010). Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell, 141*(5), 897-907. doi:10.1016/j.cell.2010.04.012

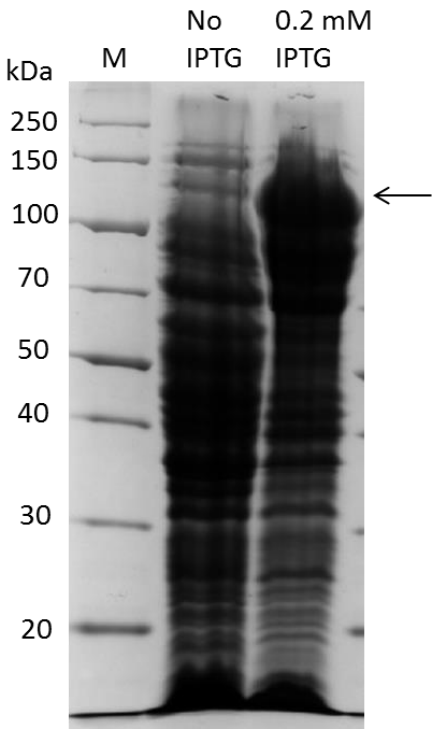
Appendices



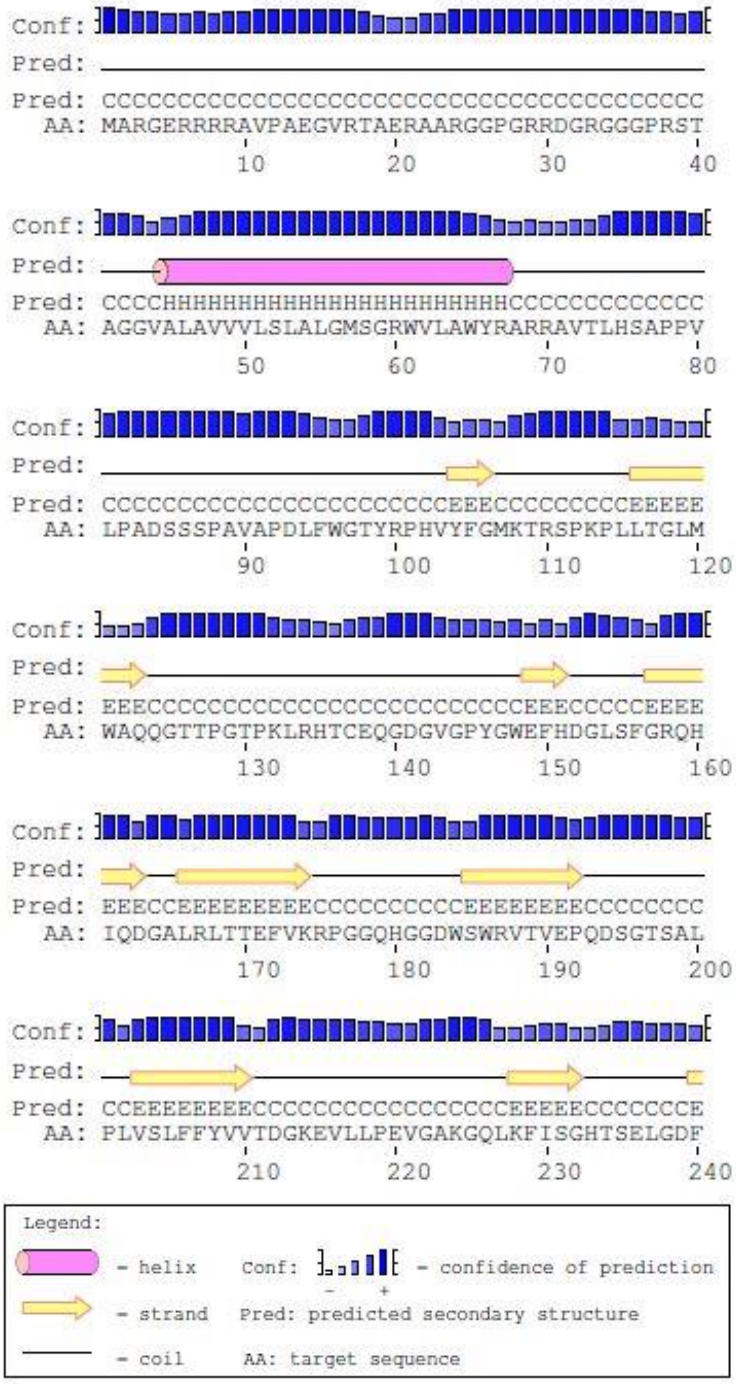
Appendix Figure 1. Colony PCR analysis for the positive transformant of pPICZαB-*MOGS*$_{\Delta 1\text{-}64}$ on agarose gel. 1, 3 &5, plasmid was the PCR positive template; 2, 4 &6, the yeast genomic DNA from the transformant was the PCR template . 1&2, 5' AOX /3' AOX primer was used; 3&4, *MOGS* Inter 1/3'AOX primer was used; 5&6, *MOGS* Inter 2/3'AOX primer was used.



Appendix Figure 2. Colony PCR analysis for the positive transformants of pPICZαB-*MOGS*$_{\Delta 1\text{-}59}$ /pPICZαB-*MOGS*$_{\Delta 1\text{-}71}$ on agarose gel. Yeast genomic DNA was the PCR template, *MOGS* Inter 2/3'AOX primer was used. 1&2, pPICZαB-*MOGS*$_{\Delta 1\text{-}59}$ transformed into GS115; 3&4, pPICZαB-*MOGS*$_{\Delta 1\text{-}71}$ transformed into GS115; 5&6, pPICZαB-*MOGS*$_{\Delta 1\text{-}59}$ transformed into X33; 7&8, pPICZαB-*MOGS*$_{\Delta 1\text{-}71}$ transformed into X33. The expected size was around 1500bp, the curve caused the bigger size.



Appendix Figure 3. SDS-PAGE (12%) analysis of the induction of MBP-MOGS$_{\Delta 1\text{-}63}$ in the presence of isopropyl-D-thiogalactopyranoside (IPTG). The SDS-PAGE gel was stained by Coomassie blue. The arrow indicates the expected protein band.



Appendix Figure 4. Secondary structure prediction of MOGS (amino acids 1-240) by using PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/), the residues between Ala45 to Arg67 are predicted to be the first N-terminal α-helix.