“Engineered expression system in *K. phaffii* for expression of human alpha-glucosidases, their characterization and inhibition”

by

Nardo Esmeralda Nava Rodriguez

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Biology

Waterloo, Ontario, Canada, 2019

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	NAME: Dr. David McMillen Title: Associate Professor, Chemistry/Biophysics and Physical Chemistry University of Toronto Mississauga
Supervisor(s)	NAME: Dr. David Rose Title: Professor Department of Biology, Faculty of Sciences University of Waterloo
Internal Member	NAME: Dr. Todd Holyoak Title: Associate Professor, Department of Biology, Faculty of Science, University of Waterloo

Internal-external Member	NAME: Brian Ingalls Title: Associate professor, Department of Applied Mathematics, Faculty of Mathematics University of Waterloo
Other Member(s)	NAME: Dr. Andrew Doxey Title: Assistant Professor, Department of Biology, Faculty of Sciences University of Waterloo
	NAME: Dr. Roberto Quezada-Calvillo Title: Professor Biochemistry Department Faculty of Chemical Sciences Autonomous University of San Luis Potosi San Luis Potosi, Mx.

Authors declaration

“This thesis consists of material, all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.”

Statement of Contributions

Nardo E. Nava Rodriguez was the sole author for all the chapters in this thesis under the supervision of Dr. David Rose. The scientific contributions from the different collaborators are listed below:

Chapter 2

The work presented in this chapter was performed in its totality by the candidate. The cDNA for cloning Sucrase-isomaltase, as well as the pJJH1396, and the S. cerevisiae strain HOD208-c were obtained through collaboration with Dr. Hassan Naim and Dr. Mahdi Amiri at the Institut für Physiologische Chemie, University of Veterinary Medicine, Hannover, Germany. This work would not have been possible without the scientific support of Dr. Jin Duan and Dr. Mahdi Amidi who provided resources, protocols, training, and invaluable discussion.

Chapter 4

The work presented in this chapter was performed by the candidate. The synthesis of the castanospermine analogues was done by Dr. Sandrine Py's group at the Université Grenoble Alpes, France. Some results have been included in a manuscript in preparation in collaboration with Dr. Sandrine Py.

Abstract

This thesis describes a progression of experiments that lead from the design of a new cloning-expression vector for *K. phaffii*, to the successful expression of numerous human alpha-glucosidases, their characterization and inhibition with three novel compounds derived from castanospermine.

K. phaffii was chosen for its numerous advantages for the expression of foreign proteins. A convenient expression vector was engineered to provide extra benefits to the system, such as a simplified construct preparation, double selection marker, and efficient purification. All these improvements, together with an optimization of the expression protocols, resulted in a yield of expression for human recombinant alpha-glucosidases in a range of 0.79 mg/L to 3mg/L in batch production, along with a significant reduction in the cloning time and a suitable selection method to identify the cells with higher number of copies of the plasmid.

A collection of eight alpha-glucosidases was successfully cloned and expressed using this system. All of them showed enzymatic activity and the expected size according to their amino acid sequences and expected pattern of glycosylation. These products were used to test three castanospermine analogues, which addressed the effect of different orientation of the substituent groups and size of the side chain on selectivity among N- and C- terminal subunits of sucrase-isomaltase.

This research project opens the door to future investigation not only on the expression and structure/function analysis of alpha-glucosidases, but also as a good alternative for recombinant expression of other proteins.

Acknowledgments

I do not even know where to start thanking, should it be my parents? They raise me, they thought me to be curious, to do not give up, to be brave and never stop trying. But maybe, I should go deeper in my roots, maybe I should look at my ancestors. Those brave men and woman who invested their entire lives seeking for happiness and wellbeing. There was no obstacle that can stop them ones that they had made their minds (for better or for worse). I clearly have to thank my siblings, they are the corner stone of my life. They are my main inspiration, I have seen them work hard, smile in the middle of the worst storms, rebuild themselves after unimaginably struggles. They are example of strength, hope, and kindness; all of that with a nice pot of barbacoa spinning in the center of the table. Thanks to all my family, the ones that share my blood, but also the ones that share my soul, Oscarito, Birthe, Suheyl, Jin, Mahdi, and very specially to Mark for all his kindness, care, support; dear friends you not only made this possible, but also the process much nicer.

Thanks to Dr. Roberto Quezada, the main culprit! I just wanted to do an undergrad thesis and go get a job as engineer. He corrupted me and here I am, almost ten years later hopping to never leave the lab. Thanks to Dr. Naim Hassan, his generosity allowed me to discover a whole new word and changed my perspective in so many ways. Thanks to Dr. David Rose for giving me the opportunity to be here, for all his patience, and support over the last four years. It had not always been easy, but his unwavering temper had sat me back on the ground and helped me to keep realistic goals on mind.

Thanks to Brant for reminding me that I am still human, that there are more things that just the lab; for his company in the long nights of experiments, for keeping me safe and most of all for sharing with me his very valuable experience and wisdom.

Thanks to all the wonderful people who did not hesitate on sharing their knowledge with me over this journey: Dr. Jin Duan who patiently thought me the basic of molecular cloning, to Dr. Mahdi Amidi, who gave it a whole different meaning. To Dr. Birthe Gerike, the face of many of my favorite papers and memories. To Dr. Dragana Miskovic, who was the first on giving me the opportunity of TA a tutorial when I was still feeling unconfident about the language. To Dr. Christine Dupont, who has support me on my teaching development, to Dr. Bruce Wolf for trusting me his students a couple of times. To April Wettig and Lucy Satora for all their help and patience.

Special thanks to my sponsor Conacyt-Gobierno del estado de San Luis Potosi. Scholarship No. 410609 and scholarship tutor Dr. Elena Dibildox Alvarado

Table of contents

List of Figures

List of Tables

List of abbreviations

AOX: Alcohol Oxidase

BMGY: Buffered complex glycerol media

BMMY: Buffered complex methanol media

CAZy: Carbohydrate Active Enzymes

CSID: Congenital sucrase-isomaltase deficiency

CST: Castanospermine

CtMGAM: C-terminal subunit of maltase-glucoamylase (Glam)

CtSI: C-terminal subunit of sucrase-isomaltase (sucrase)

$CtSI_{cat}$: Truncated form of C-terminal subunit of sucrase-isomatase (sucrase)

DAS: Dihydroxyacetone synthase

EI: Enzyme-inhibitor complex

ER: Endoplasmic Reticulum

ES: Enzyme-substrate complex

ESI: Enzyme-substrate-inhibitor complex

FDH: Formate dehydrogenase

FLD: Formaldehyde dehydrogenase

GAP: Glyceraldehide 3-phosphate

GH31: Glucoside hydrolase family 31

LPH III, IV: Regions III and IV of lactase phlorizin hydrolase

LPH III: Region III of lactase phlorizin hydrolase

LPH: Lactase phlorizin hydrolase

MGAM:Maltase-glucoamylase

NHEJ: Non homologous end-joining

NJ: Nojirimycin

NMR: Nuclear Magnetic Resonance

NtMGAM: N-terminal subunit of maltase-glucoamylase (Mal)

NtSI: N-terminal subunit of sucrase-isomaltase (isomaltase)

PAD: Pichia adenine dropout

PAGE: Polyacrylamyde gel electrophoresis

PBS: Phosphate-buffered Saline

PGO: Peroxidase glucose-oxidase method

SI: sucrase-isomaltase

SI_{FL}: Sucrase-isomaltase full length

Ss: Signal sequence

TEV: Tobacco etch virus cleavage site

YPD: Yeast-peptone-glucose media

“If I have seen a little further it is by standing on the shoulders of Giants."

Sir Isaac Newton

Chapter I. Introduction

1.1 Expression of recombinant enzymes.

1.1.1 Recombinant enzymes

The production of recombinant enzymes reached the market in the early 1980s' when the advances in biotechnology allowed the development of tools for genetically engineering organisms (Vieira-Gomes et al., 2018). Nowadays, according to Markets and Markets Research, the "protein expression market" was evaluated at $1.6 billion in 2017 and is projected to reach $2.8 billion by 2022 (marketsandmarkets, 2018). Drug discovery is one of the leading forces driving this growth; between 2011 and 2016 alone 62 new recombinant proteins were approved by the regulatory system in the United States (Lagassé et al., 2017). However, recombinant proteins also play a critical role in the food industry, materials, bioremediation, and other fields (Vieira-Gomes et al., 2018).

1.1.2 Prokaryotic vs. Eukaryotic systems

The most common approach to expressing recombinant proteins is to use prokaryotic systems. They were the first to be developed; therefore, they are well known and offer multiple benefits such as yield/time efficiency (Vieira-Gomes, Souza-Carmo, Silva-Carvalho, Mendonça-Bahia, & Parachin, 2018). One of the main disadvantages of the prokaryotic expression systems is their limited capability of performing posttranslational modifications such as glycosylation. This disadvantage causes several problems in the expression of recombinant eukaryotic proteins as misfolding, formation of inclusion bodies, malfunction, or even complete lack of expression

(Juturu & Wu, 2018). Consequently, different eukaryotic systems have been explored and developed.

1.1.3 Eukaryotic systems.

There are a broad variety of eukaryotic expression systems with different features which make them suiTable for specific purposes; the selection relies on many factors that include but are not limited to the requirements of the target protein and more practical elements such as budget, required equipment and safety.

Furthermore, aspects such as downstream applications for the expressed protein are fundamental in the protein expression system selection. For example, for structural analysis, one of the most common techniques is X-ray crystallography. It is a powerful technique that provides valuable information about the structure of a protein in high detail. However, it requires a large amount of pure and highly concentrated protein ~10mg/mL. Therefore, it is necessary that the expression system offers a suitable protein purification option as well as high yield.

Yeast is a single cell organism; therefore, it is easy to modify and handle in comparison with higher eukaryotes systems. Also, yeast-based systems do not require specialized equipment, are less vulnerable to contamination, and require less maintenance.

1.1.4 *Saccharomyces cerevisiae* vs *K. phaffii*.

Among yeast expression systems, *Saccharomyces cerevisiae* is the most studied organism of this class and offers many options for products and protocols. However, like all other expression systems, this organism has limitations, and it is only suitable for successfully expressing certain proteins.

In the last 30 years, another yeast that has gained popularity for protein expression is *Komagataella phaffii* (also known as *Pichia pastoris*). One of the major advantages of *K. phaffii* is its capability to utilize methanol as a sole source of carbon. Also, genetic modifications have been made to utilize the same promoter for expressing both alcohol oxidase, a key enzyme for methanol metabolism, and the protein of interest. These modifications allow induction of protein expression by methanol addition.

Although both *Saccharomyces cerevisiae* and *K. phaffii* are yeast, there are important differences between the organisms. For example, *S. cerevisiae* has shown high efficiency at using homologous recombination as a DNA repair mechanism. Meanwhile, *K. phaffii* has a preference for non-homologous end-joining (NHEJ) mechanisms (Amen & Kaganovich, 2017). This characteristic allows the successful integration of linear vectors and linear DNA sequences in *S. cerevisiae*, unlike *K. phaffii*.

In addition, there are differences in protein production and modification. A clear example of this is N-glycosylation. *Komagataella phaffii*, unlike *Saccharomyces cerevisiae*, does not add alpha 1,3-terminal mannoses to oligosaccharides (Juturu & Wu, 2018). Also, *Saccharomyces*

cerevisiae is more likely than *K. phaffii* to produce hyperglycosylated proteins (Juturu & Wu, 2018).

Other noTable differences between *Saccharomyces cerevisiae* and *Komagataella phaffii* is that *K. phaffii* utilizes aerobic respiration, which allows high cell density cultures, and therefore improves the protein yield (Cohen, Chang, Boyer, & Helling, 1973). In addition, some *K. phaffii* strains have been modified to mimic the glycosylation patterns produced in mammalian systems (Laukens, Wachter, & Callewaert, 2015).

1.1.5 *Komagataella phaffii* expression systems

Initially, *K. phaffii* was thought of as a cheap alternative to produce high protein animal feed (Lin Cereghino et al., 2001). However, the oil crisis in the 1970s, as well as an important drop in soybean price, made this option uncompetitive (Lin Cereghino et al., 2001). Years letter, Phillips Petroleum Company, in collaboration with The Salk Institute of Biotechnology Industrial Associates (SIBIA) developed *Komagataella phaffii* as a heterologous gene expression system (L. Cereghino et al., 2000). They isolated the AOX1 gene and promoter and developed vectors, strains, and methods for genetic manipulation of *K. phaffii* (Cregg & R. Madden, 1987). Finally, in 1993, Phillips Petroleum Company sold its patent position to RCT and licensed Invitrogen^{MT} to sell components of the system for research purposes (L. Cereghino et al., 2000). As the yeast system, *K. phaffii* main benefits are its cost, high efficiency, and suitability for producing mammalian heterologous proteins (Mattanovich et al., 2012).

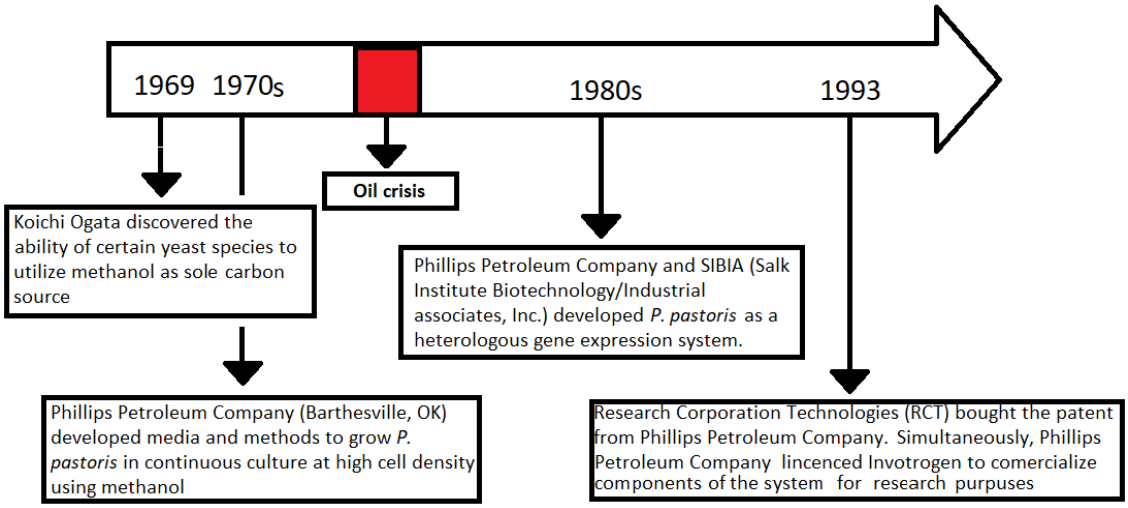


Figure 1.1 Timeline of the development of *Komagataella phaffii* as an expression system.

In summary, the remarkable features that make *K. phaffii* particularly interesting for recombinant expression is the induction of large-scale protein production by methanol addition (Raymond, Pownder, & Sexson, 1999), and the presence of posttranslational modifications as proteolytic processing, folding, disulfide bond formation, and glycosylation (both O- and N-)(L. Cereghino et al., 2000).

1.1.5.1 Methanol metabolism and its relationship with the induction of protein expression.

Komagataella phaffii can utilize methanol as a sole carbon source; this is due to its capability of metabolizing methanol to formaldehyde, and formaldehyde into formate and glyceraldehyde 3-phosphate (GAP) (Prielhofer et al., 2015). The first part of the process occurs in the peroxisomes (Figure 1.2) where alcohol oxidase (AOX) in the presence of O_2 hydrolyzes methanol to formaldehyde, releasing hydrogen peroxide as a secondary product. After this first reaction, a

portion of the formaldehyde is released to the cytosol and is oxidized by two dehydrogenases: formaldehyde dehydrogenase (FLD) and formate dehydrogenase (FDH) to produce energy (NADH) and CO_2. In parallel, the remaining formaldehyde enters a cyclic pathway catalyzed by dihydroxyacetone synthase (DAS) which produces GAP. The cell is protected from the toxic effects of hydrogen peroxide by retaining it in the peroxisomes, where it is metabolized to molecular oxygen (O_2) and water (J. L. Cereghino & Cregg, 2000).

The production of foreign protein requires the use of a specific promoter. Although several promoters have been developed for *K. phaffii* systems, pAOX1 is the most common and successful, with about 300 recombinant proteins expressed successfully as of 2005 (Macauley-Patrick, Fazenda, McNeil, & Harvey, 2005). AOX1 encodes for AOX, the protein which catalyzes the oxidation of methanol to formaldehyde, the first and critical step in the methanol metabolism. Some of its main features have been listed in Table 1.1.

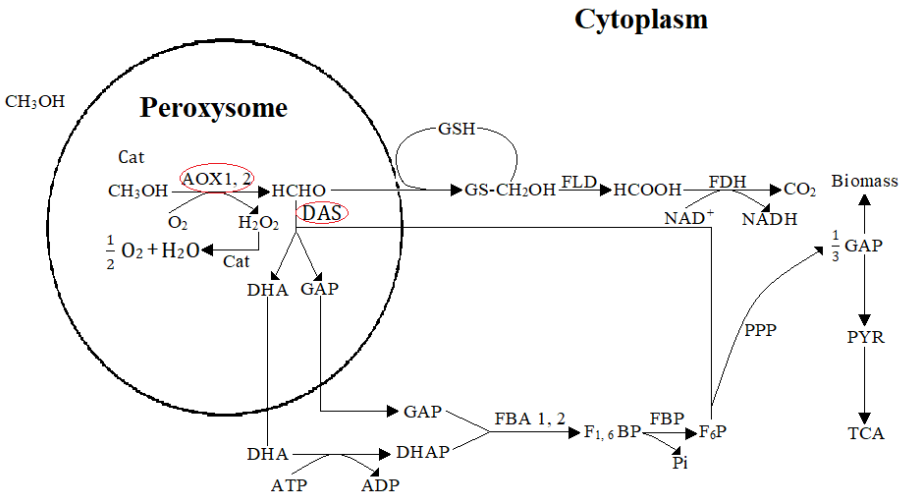


Figure 1.2 Methanol metabolism pathway in *K. phaffii*. AOX1, 2: Alcohol oxidase, DAS: dihydroxyacetone synthetase, Cat: catalase, FBA: fructose-1,6-biphosphate aldolase, BFP: fructose-1,6-biphosphatase, FLD: formaldehyde dehydrogenase, FLH: formaldehyde dehydrogenase (Juturu & Wu, 2018)**.**

pAOX1 is an inducible promoter, and its inducer molecule is methanol. The expression follows two steps (L. Cereghino et al., 2000). First, the biomass production is favored by a repression of the pAOX1 promoter. The accumulation of cell density is achieved, utilizing glycerol as a carbon source, blocking the methanol metabolism pathway (Vanz, Nimtz, & Rinas, 2014). Once the desired cell density is reached, the media is replaced by new media with the inducer, methanol. (Prielhofer et al., 2015). The presence of methanol in the system activates pAOX1, which regulates both AOX and the protein of interest.

Table 1.1 Advantages and disadvantages of pAOX as a promoter to regulate recombinant protein expression in *K. phaffii* expression systems (Juturu & Wu, 2018)**.**

Advantages	Disadvantages
Tightly regulated transcription controlled by repression/de-repression mechanism	Hard monitoring of methanol utilization
High yield of protein	Methanol flammability
Repression of the AOX1 gene in the absence of methanol ensures high cell growth before gene expression	Methanol utilization in food-grade applications is not preferable
Easy induction by addition of methanol	Uncertainty in the exact time when the carbon source changes

The utilization of methanol as inducer also requires a careful assessment of the methanol tolerance of the cells as at high concentrations it may be toxic, as well as affect the structural stability of the expressed protein (Santoso, Herawati, & Rubiana, 2012).

1.1.5.2 Posttranslational modifications.

Yeast can perform posttranslational modifications such as the processing of signal sequences, both pre- and pre- pro-type, the formation of disulfide bridges, ubiquitination, and glycosylations (O- and N-)(Vieira-Gomes et al., 2018). However, the pathways and characteristics of these posttranslational modifications vary even among the different yeast species (Vieira-Gomes et al., 2018). Therefore, some mammalian recombinant proteins successfully expressed in a specific yeast may not be successful in a different host.

Glycosylation is one of the most relevant posttranslational modifications in protein engineering (Lagassé et al., 2017). It may impact the folding, organization, localization, and even functionality of a protein. There are two types of glycosylation O- and N-.

In mammals, O-linked oligosaccharides include N-acetylgalactosamine, galactose, and sialic acid. Meanwhile, in organisms such as *Komagataella phaffii*, O-glycosylations are only composed of mannose residues (Juturu & Wu, 2018). Also, *K. phaffii* may or not glycosylate a protein that in the original host is glycosylated (Juturu & Wu, 2018). Similarly, the glycosylation sites may differ among hosts (Laron, 2001).

The general N-glycosylation process in eukaryotes starts in the endoplasmic reticulum (ER) with the transfer of a lipid-linked core unit ($Glc_3Man_9GlcNAc_2$) to asparagine at the recognition sequence Asp-X-Ser/Thr. This oligosaccharide core unit is trimmed to $Man_8GlcNAc_2$. After this step, the glycosylation pathways are different among Eukaryotes (Juturu & Wu, 2018).

In mammals, the N-glycosylation pathway leads to a series of trimming and additions that generated high-mannose oligosaccharides, a mixture of different complex sugars, as well as a combination of both, recognized as a hybrid type. In contrast, *K. phaffii* only has two patterns of glycosylation, carbohydrates structures similar to the core unit ($Man_{8\text{-}11}GlcNAc_2$) and hyper glycosylations (Trimble, Atkinson, Tschopp, Townsend, & Maley, 1991). An important consideration about expressing recombinant proteins in *Komagataella phaffii* is that for unknown reasons, it adds outer chains in some proteins (L. Cereghino et al., 2000). The addition of outer chains may impact protein folding and function.

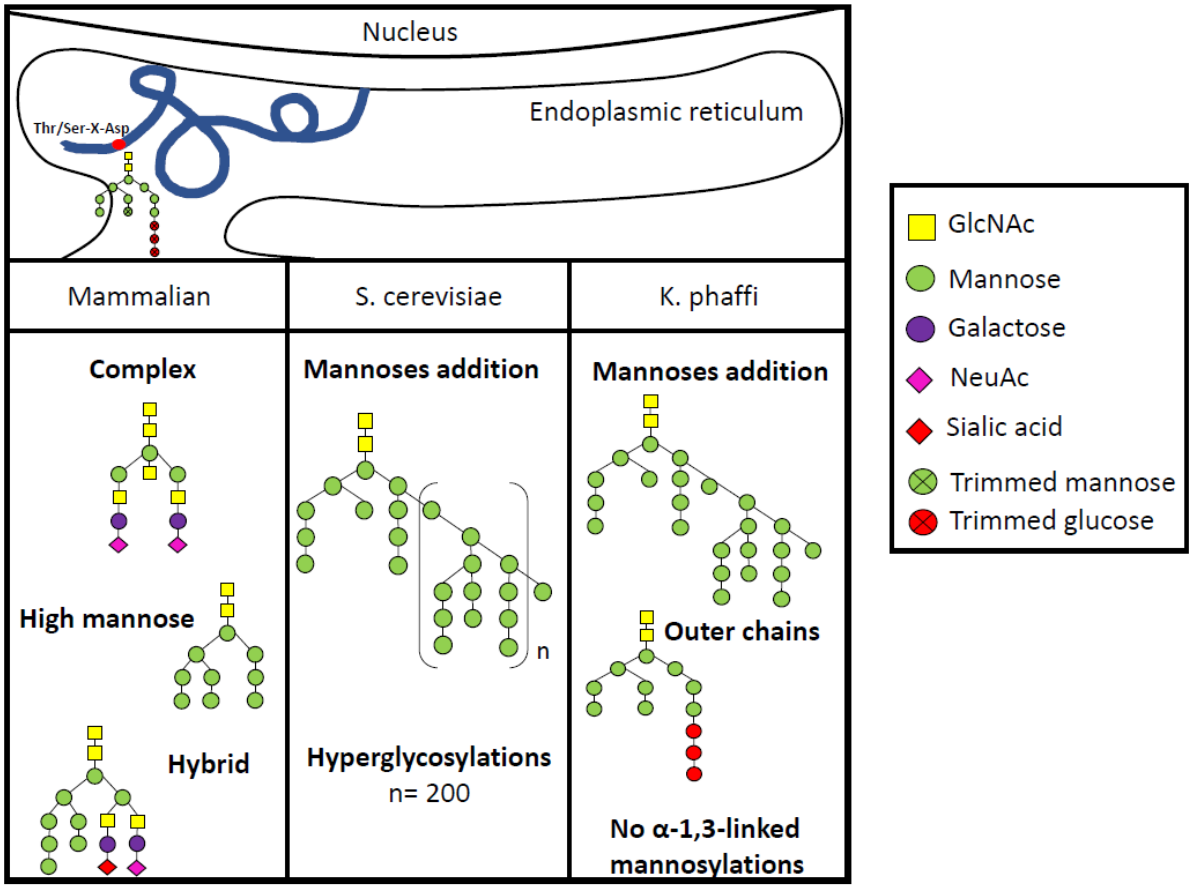


Figure 1.3. Generalized comparative Glycosylation pathways in mammals, S. cerevisiae and *K. phaffii*

Also, among species, there are small differences in the glycosylation patterns. For example, *S. cerevisiae* has a higher tendency to produce hyperglycosylated proteins (Conde, Cueva, Pablo, Polaina, & Larriba, 2004). *K. phaffii*, instead, typically adds outer chains of $Man_8GlcNAc_2$ (Bretthauer & Castellino, 1999). Moreover, proteins expressed in *K. phaffii* do not have alpha-1,3- linked mannosylations (Bretthauer & Castellino, 1999). However, there are a *K. phaffii* expression system that has been successfully engineered to mimic the mammalian glycosylation pattern (Hamilton & Gerngross, 2007).

1.1.5.3 Protein secretion and secretion signals

Secreted proteins have a recognition sequence that regulates their exit from the ER-Golgi complex. In native proteins, this sequence is already part of the gene, and therefore, it is characteristic of that organism. However, in the case of recombinant proteins, it is necessary to add this extra sequence, and the secretion efficiency depends on the secretion signal selected. There are a variety of secretion signals that can be used in yeast; some of them are compatible between species. Table 1.2 summarizes the main secretion signals utilized in *Komagataella phaffii.*

In 2011 the alpha-factor signal of *K. Phaffii* was sequenced, revealing that the pre-sequence is identical to the one from *S. cerevisiae* (Küberl et al., 2011). Meanwhile, the prosequence is larger (Küberl et al., 2011). However, in terms of secretion efficiency on *K. phaffii*, there is no difference (Küberl et al., 2011).

Table 1.2 Common signal sequences used in cloning for the secretion of recombinant proteins in *K. phaffii* expression systems: natural origin, and length (Juturu & Wu, 2018)**.**

Signal sequence	Organism	Length
Alpha- mating factor	Saccharomyces cerevisiae	89 aa & truncated forms
Alpha-amylase	Aspergillus niger	20 aa
STA-1	Aspergillus Nigger	18 aa
Inulinase	Saccharomyces diastaticus	16 aa
SUC2	Kluyveromyces Maximus	19 aa
Killer	Saccharomyces cerevisiae	26 aa
Lysozyme	Chicken lysozyme	26 aa
Albumin	Human serum albumin	18 aa

1.1.5.4 Expression vectors for *Komagataella phaffii*

A basic foreign expression cassette is composed of a DNA sequence containing a promoter, restriction sites for insertion of the foreign gene, terminator, and a selectable marker (L. Cereghino et al., 2000). More sophisticated expression vectors may also add a signal sequence (ss) for the secretion of the protein of interest, antibiotics resistance (in addition to the selection marker), as well as other sequences as tags for protein purification (L. Cereghino et al., 2000).

Promoters are classified into two main groups, constitutive and inducible (Kim, Yoo, & Kang, 2015). Constitutive promoters have the advantage of producing almost constant levels of recombinant protein during the entire expression; however, this may also cause a low secretion due to the accumulation of misfolded protein (Liu, Tyo, Martínez, Petranovic, & Nielsen, 2012).

On the other hand, the utilization of inducible promoters involves two stages in the expression protocol. First, in the biomass production stage, the promoter is repressed. Therefore the rate of growth is greater, producing higher cell density (Mumberg, Mulier, & Funk, 2012). The second stage is the induction. Here the cell density remains almost constant as the promoter is activated by the addition of the inductor molecule. During this second stage is when the protein of interest is expressed. The level of expression increases with the concentration of inductor (Prielhofer et al., 2015). Therefore, the addition of the inductor molecule is critical and must be periodically monitored according to the rate of utilization (Prielhofer et al., 2015).

In *K. phaffii* expressions, the most common inducible promoters are pAOX1, FLD1, and PEX8 (Juturu & Wu, 2018). pAOX1 has multiple advantages as its transcription is tightly regulated, and it is easily induced (Macauley-Patrick et al., 2005). It also has important disadvantages, mostly related to the use of methanol as an inducer molecule. Firstly, during expression, it is hard to monitor methanol concentration accurately (Juturu & Wu, 2018). Secondly, methanol is very flammable. Furthermore, the utilization of methanol in food grade applications is not recommended (Juturu & Wu, 2018). Therefore, continuous efforts are made to develop new and more suiTable promoters to improve this expression system; some of them are summarized in Table 1.3.

Despite the important role of terminators not only for transcription, but also as mRNA stabilizers, in *K. phaffii* the terminator options are more limited and their role in protein output underappreciated (Vieira-Gomes et al., 2018). The most commonly used are CYC1 and ADH1t, both from *S. cerevisiae*, displaying high compatibility between the two systems.

Selectable markers can be classified into two groups: auxotrophy and antibiotic resistance (Juturu & Wu, 2018). The *K. phaffii* expression systems come from mutation derivatives of the original strain NRRL-Y11430 (Juturu & Wu, 2018). One of those mutations was on the auxotrophic genes to produce different mutants capable of growing in minimal media. Those mutations allow the use of specific media to screen for positive transformants.

Table 1.3 List of alternative promoters to control the induction of expression of recombinant proteins in *K. phaffii* expression systems (Juturu & Wu, 2018)**.**

Promoter	Type	Main features
GAP	Constitutive	Expresses proteins in all carbon sources
FLD1	Inducible	The same yield as pAOX1 but utilizes different carbon sources as methylamine and choline
ICL1	Inducible	Repressed by glucose presence and induced by ethanol
PEX8	Inducible	It is expressed in low amounts in glucose and is induced by methanol
YPT1	Constitutive	Low but continuous protein expression using as carbon source glucose, methanol, or mannitol.
NPS	Constitutive	Autoinducible sodium-phosphate symporter gene promoter. Phosphorus depletion required for expression.

There are two main methods for integrating the vector into the genome (L. Cereghino et al., 2000). Their efficiency varies with species, strain, and other factors. For example, in *S. cerevisiae* linear vectors can generate sTable transformants via homologous recombination; even using small terminal flanking sequences (~50bp) it is possible to get a 100% efficiency (Dudich et al., 2012), as *S. cerevisiae* uses homologous recombination as a major DNA repair pathway. Meanwhile, *K. phaffii* the recombination efficiency rates from only 1 to 30% even using long flanking sequences (1Kb) (Juturu & Wu, 2018).

1.1.5.5 *Komagataella phaffii* expression strains.

All *Komagataella phaffii* strains are derivatives of NRRL-Y11430 (Northern Regional Research Laboratories, Peoria, IL). Most of them have a HIS4 mutation for selection of expression vector. Nevertheless, other gene/auxotrophic combinations are available (Juturu & Wu, 2018).

An important adaptation for recombinant expression is the protease knock out. Especially during the fermentation, yeast vacuoles produce proteases that may degrade the protein of interest (L. Cereghino et al., 2000). As a consequence, *K. phaffii* has been mutated to prevent the production of the main proteases that cause protein degradation.

Two genes encode for the key proteases in *K. phaffii*: PEP4 and prb1. PEP4 encodes for protease A, which activates other main proteases such as carboxyl peptidase Y and proteinase B. The activity of the latter, for example, is doubled when activated by proteinase A; prb1 controls the production of proteinase B, itself (L. Cereghino et al., 2000).

Within the last decade, a series of *K. phaffii* strains have been developed with different mutations that mimic the N-glycosylation patterns obtained in mammalian systems, opening more options for expression of recombinant proteins (Hamilton & Gerngross, 2007).

These mutations, combined with the auxotrophy gene, give rise to a broad variety of *K. phaffii* strains. Some of them are protected by the RCT and Invitrogen patent, but also many are patent-free derivatives from the *K. phaffii* CBS7435. Indeed, CBS7435 strains have also been mutated

to enhance their homologous recombination efficiency, obtaining up to 90% (Näätsaari et al., 2012).

1.1.5.6 Protein expression in *K. phaffii* for structural studies

The most popular technique for structural studies is X-ray crystallography. It is a powerful technique that consists of crystallizing the protein and recording an X-ray diffraction pattern. This can be processed computationally to derive a three-dimensional representation of the atomic structure of the protein. Solving the structure of a protein is extremely useful for understanding its function, mechanism, substrate affinity, specificity, etc.

However, to crystallize a protein generally requires high concentrations (>10mg/mL) of pure protein (>95%). Many eukaryotic expression systems cannot meet this requirement. The *K. phaffii* system is one of the few exceptions, and for this reason, has gain popularity for structural biology purposes.

In comparison with other suitable organisms for expressing mammalian recombinant enzymes, *K. Phaffii* has advantages for nuclear magnetic resonance (NMR) as well, as it can grow in minimal media that can be modified for production of ^{15}N-labeled proteins or even triple labeled (^{15}N, ^{13}C, ^{2}H) proteins (L. Cereghino et al., 2000). Also, *K. Phaffii* glycosylations can be removed by endo H, as they are mostly high mannose, thus increasing the chances of crystallizing the protein (L. Cereghino et al., 2000).

1.1.5.7 *K. phaffii* commercial expression systems

Komagataella phaffii was first developed as a protein expression system in early 1980, and since then it had been the object of different modifications according to the necessities of the scientific community. Nowadays, Invitrogen offers a broad variety of vectors, modified *K. phaffii* strains, secretion signals, as well as protein expression kits. However, even with those improvements, it is not always possible to fulfill the requirements for expressing the different proteins. For the proteins of interest in this work, at least, it has been necessary to develop personalized versions of the system. Table 1.4 displays some of the features of the main *K. phaffii* expression systems available on the market.

Table 1.4 Comparison of the main K. phaffii expression systems available in the market (Invitrogen).

***K. phaffii* Expression System**	**Secretion sequence**	**HisTag**	**Antibiotics resistance in Yeast**	**Auxotrophy**	**E. coli compatibility**	**S. cerevisiae compatibility (Homologous recombination)**
Multi-Copy Pichia Expression Kit	•		•	His	•	
PichiaPink™ Secreted Protein Kit	•			Ade	•	
PichiaPink™ Secretion Optimization Kit	•			Ade	•	
EasySelect™ Pichia Expression Kit	•	Six his	•	His	•	

1.2 Alpha glucosidases.

1.2.1 Carbohydrate digestion

Carbohydrates have played an important role in human development and evolution (Quezada-Calvillo et al., 2008a). Since agriculture started, crop consumption increased dramatically, becoming in most of the civilizations, the base of human diets. Utilizing crops as the main source of calories caused a series of adaptative evolutionary processes in humans and domesticated animals (Karasov, Martínez del Rio, & Caviedes-Vidal, 2011). For example, to increase the efficiency of carbohydrate metabolism, the AMY2 gene, responsible for encoding pancreatic amylase, underwent a duplication, yielding AMY1, encoding salivary amylase (Perry et al., 2007).

Simple sugars such as glucose, galactose, and fructose are absorbed in the intestine and enter the different metabolism pathways without any extra transformation (C. C. Robayo-Torres, Quezada-Calvillo, & Nichols, 2006). However, more complex sugars, from disaccharides to oligosaccharides, require the action of specific enzymes, which hydrolyze the different linkages and release absorbable monosaccharides. Once the monosaccharides are absorbed into the bloodstream, they can be utilized to generate carbon and energy (Lee et al., 2016).

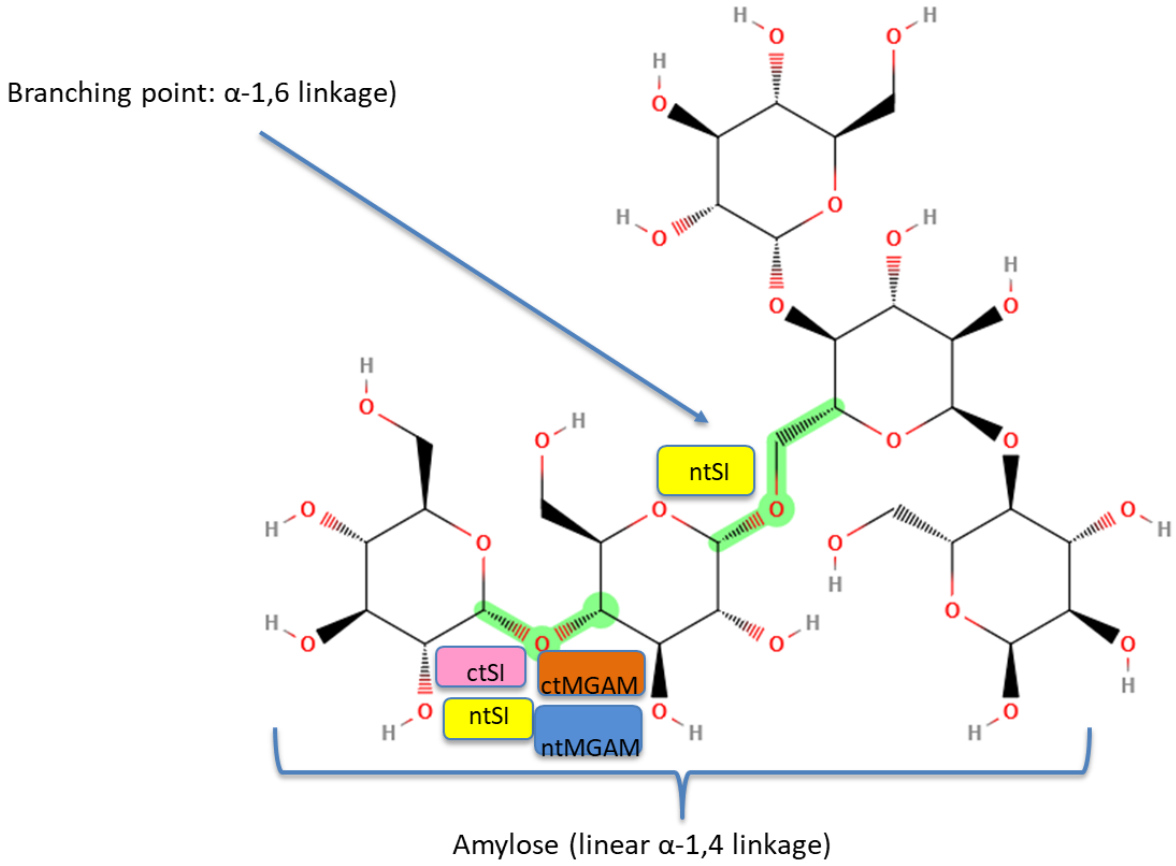


Figure 1.4 Starch linkages alpha-1,4 and alpha-1,6 highlighted in green and the intestinal alpha-glucosidases which hydrolyze them during the starch digestion.

Among carbohydrates, starch is responsible for a high portion of the caloric intake in the human diet worldwide. The digestion of starch starts in the mouth with salivary amylase hydrolyzing alpha-1,4 glycosidic linkages. However, together, saliva and pancreatic amylases only release less than 4% of the free glucose from starch digestion (Yook & Robyt, 2002). The next step in starch digestion takes place in the small intestine, where two main alpha glucosidases: Maltase-glucoamylase (MGAM) and sucrase-isomaltase (SI) hydrolyze the remaining alpha 1,4 linkages,

plus the alpha 1,2 and alpha 1,6, releasing the major portion of free glucose (Gericke, Amiri, & Naim, 2016) Figure 1.4.

1.2.2. Family GH 31

Sucrase- isomaltase and maltase- glucoamylase both belong to the Glucoside hydrolase family 31 (GH31) according to the Carbohydrate Active Enzymes (CAZy) data base (http://www.cazy.org/). Enzymes in this family have the catalytic $(\beta/\alpha)_8$ barrel structure and, with the exception of the alpha-glucan lyases, hydrolyze alpha-glucosidic linkages following a double displacement mechanism (Koshland retaining mechanism) Figure 1.5 (CAZypedia/Glycosil hydrolases). Two characteristic consensus motifs include the catalytic residues; the aspartic acid in WI**D**MNE acts as the catalytic nucleophile while the aspartic acid within the motif HWLG**D**N functions as the proton donor.

1.2.3. Maltase-glucoamylase

Located in the apical membrane of the small intestine, MGAM is another member of the glycoside hydrolase family 31 (GH31). The human MGAM gene is located in chromosome 7 (Nichols et al., 2003). The resultant protein includes 1857 amino acid residues and is highly N- and O- glycosylated, increasing its molecular weight to 336 KDa (Naim, Sterchi, & Lentze, 1988). MGAM is embedded in the plasma membrane by its N-terminus (Naim, Sterchi, & Lentze, 1988). MGAM (EC 3.2.148 and 3.2.1.3) contains two subunits: N-terminal (maltase) and C-terminal (glucoamylase). Both MGAM subunits have shown hydrolytic activity against alpha-1,4 glucosidic linkages; however, the N-terminal subunit (Nt-MGAM) has a preference for short glucose polymers. Meanwhile, the C-terminal subunit (CtMGAM) displayed a higher

affinity for longer chains (Sim, Quezada-Calvillo, Sterchi, Nichols, & Rose, 2008). As MGAM hydrolyzes the alpha-1,4 linked glucoses from the ends of the starch oligomers, it determines the overall rate of starch digestion (C. C. Robayo-Torres et al., 2006).

According to prior work, this enzyme is synthesized as a single polypeptide chain (H Y Naim, Sterchi, & Lentze, 1988). Then, the precursor molecule $MGAM_h$ undergoes posttranslational modifications in the ER adding N-linked high mannose residues. This first $MGAM_h$ molecule has a molecular weight of 285 KDa, of which about 20 KDa corresponds to glycans. Then the precursor molecule is transported to the Golgi apparatus where glycan maturation takes place. As a result, a final version of the protein $MGAM_m$ displays an increased molecular weight from 285KDa of the precursor to 335KDa. The O-glycosylations alone are estimated at 35KDa and the N-glycans (mostly complex sugars) add about 45 KDa.

Unlike sucrase-isomaltase, maltase- glucoamylase is matured as an intact complex and sorted in the apical membrane without cleavage of the two enzyme units.

The maltase and glucoamylase subunits have been successfully expressed individually in mammalian systems as COS-1 and CHO cells, as well as S2 cells and yeast, showing comparable enzymatic activity parameters according to BRENDA enzymes data base (EC. 3.2.1.20/ 3.2.1.3). Therefore, there is no indication of co-dependence of the subunits.

Maltase- glucoamylase has hydrolytic activity against alpha-1,4 linkages. However, the N-terminal subunit (maltase) has a preference for short glucosidic chains while the C-terminal (glucoamylase) has higher activity against long chains (Quezada-Calvillo et al., 2008b).

1.2.4. Sucrase-isomaltase

Sucrase- isomaltase (SI) is a heterodimer formed by two subunits which share more than 35% amino acid sequence identity and 41% conservation (Hunziker, Spiess, Semenza, & Lodish, 1986). SI is synthesized as a single polypeptide chain and N- glycosylated in the endoplasmic reticulum with mostly high mannose glycans. This precursor, known as $proSI_h$ (Hassan Y Naim, Sterchi, Lentze, & Bern, 1988), when immunoprecipitated from a human small intestinal biopsy, has a size of 210KDa, and 185KDa after deglycosylation with endo H, which specifically cleaves the N-glucans. $proSI_h$ is transported to the Golgi apparatus where, following trimming by alpha-mannosidases I and II, complex sugars are added at the N- and O-glycosylation sites. The modified Sucrase- isomaltase precursor called $proSI_c$ has been identified in Western blot as a 250KDa band, with about 20KDa corresponding to O-glycosylation. Once in the intestinal lumen, the pancreatic protease trypsin (in humans) cleaves the precursor separating it into the two subunits: sucrase (130KDa) and isomaltase (145KDa). Both subunits are N- and O-glycosylated; the deglycosylated versions are 94KDa for sucrase and 111KDa for isomaltase. The biosynthesis of sucrase- isomaltase had been schematized in Figure 1.5

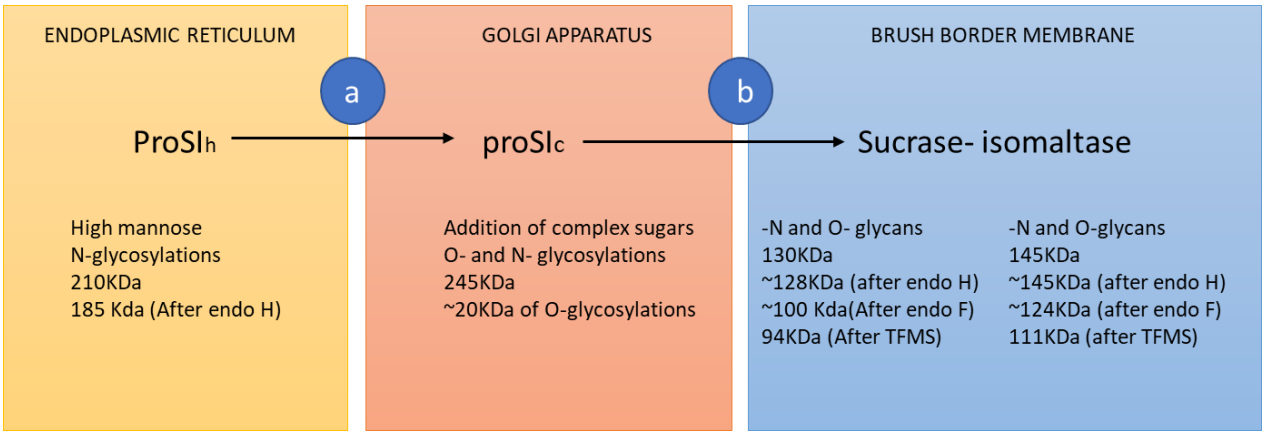


Figure 1.5 Scheme of the biosynthesis of Sucrase- isomaltase. In a) Once in Golgi, mannosidases I and II trim the high mannose residues and are replaced by complex sugars. In b) Once that the protein is assorted in the brush border membrane, trypsin cleaves the two subunits which reman associated via ionic interactions.

As the two subunits (sucrase and isomaltase) are cleaved after the single polypeptide chain has been completely folded and sorted into the brush border membrane (H Y Naim et al., 1988), it is possible that the subunits affect each other's folding. Indeed, in 2002 Naim et al. proposed sucrase as intramolecular chaperone crucial for proper transport of isomaltase from the ER to the Golgi apparatus and, therefore, for its correct maturation and localization, leading to a complete inactivation of the isomaltase subunit (Jacob, Pürschel, & Naim, 2002). However, *in vitro*, the isomaltase subunit was individually expressed in *Drosophila melanogaster* cells and shown to have maltase and isomaltase activity (K_M 7.1 and 11.1 mM, respectively); the crystal structure was solved (3.2 Å) in both apo form and in complex with kotalanol; a thiosugar highly potent as an alpha-glucosidase inhibitor (Sim et al., 2010b).

The assembly and structure of the isomaltase subunit is highly similar to the maltase subunit of MGAM (nt-MGAM), with a root mean square deviation of 0.5Å over 776 C_α residues in the crystal structures. Furthermore, most of those deviations are localized in the surface loops, as

well as in the trefoil region. Considering that the percentage of sequence identity between maltase (nt-MGAM) and isomaltase (nt-SI) is about 60%, this high similarity between the structures of the two proteins was expected.

A possible explanation for the contrasting isomaltase activity results described above may reside in the process for recombinant expression, particularly the host organism. Jacob et al. (2002) used protein expressed in African Green Monkey Cos-1 cells, while in 2009 Sim et al., used *Drosophila* S2 Cells (Sim et al., 2010b). These two organisms have important differences in terms of posttranslational modifications. Mammalian cells have more sophisticated pathways, and tighter quality control; therefore, a defective protein in one system might be detected and targeted for degradation or be retained in the ER, while in the other system, the same protein might be transported, modified in the Golgi, and secreted.

The sucrase subunit is the only intestinal alpha-glucosidase with alpha 1,2 hydrolytic activity (Treem, 2012); meanwhile, isomaltase hydrolyzes alpha 1,6 linkages (C. C. Robayo-Torres et al., 2006). Both sucrase-isomaltase subunits are also capable of breaking alpha 1,4 glucosidic linkages. The alpha 1,6 linkage generates the branches in the amylopectin. Consequently, its hydrolysis releases alpha-1,4 linked maltodextrins, which can be digested by all the alpha-glucosidases in the small intestine. Sucrase is responsible for most of the sucrase activity in humans, and together the sucrase and isomaltase subunits respond for 60 to 80% of the glucose released from maltose in the intestinal lumen (Treem, 2012).

Deficiencies in Sucrase-isomaltase cause carbohydrate malabsorption and consequently, different symptoms, including osmotic-fermentative diarrhea, vomiting, abdominal distention,

and flatulence; but most importantly, it also affects the absorption of the nutrients (Treem, 2012)(Belmont et al., 2002). The sucrase-isomaltase deficiencies are classified into two groups: induced and congenital (Gericke et al., 2016).

Induced SI deficiency is the result of intestinal injuries or malfunction derived from other environmental factors. Such factors include certain diseases such as infections and autoimmune disorders, as well as consumption of synthetic compounds, for example, some of those used in obesity and diabetes treatments (Gericke et al., 2016). Furthermore, chronic psychological stress and deficiency in iron and vitamin A have also been associated with a decreased sucrase-isomaltase activity (Boudry, Jury, Yang, & Perdue, 2007) (West & Oates, 2005)(Reifen, Zaiger, & Uni, 1998). In contrast, Congenital Sucrase-Isomaltase Deficiency (CSID) is genetically determined (Treem, 2012). This condition is due to mutations in the SI gene coding region (Sander et al., 2006). Such mutations may affect the SI intracellular trafficking, lead to functional deficits, or cause mis-sorting of sucrase-isomaltase in the lumen of the intestine (Sander et al., 2006).

1.2.5 Congenital Sucrase-Isomaltase deficiency (CSID)

Each mutation in the SI gene that causes CSID is elicited by a single nucleotide, which leads to a single amino acid exchange (Alfalah, Keiser, Leeb, Zimmer, & Naim, 2009). Depending on the mutation, one or both subunits may be affected (Gericke et al., 2016); furthermore, the enzymatic activity may be affected.

The sucrase-isomaltase function relies on three major factors: proper glycosylation and folding, trafficking, and sorting in the apical membrane (Alfalah et al., 1999). However, all these factors are intertwined, and their correlation might be direct or indirect. Improper glycosylation may cause misfolding (Kornfeld & Kornfeld, 1985); misfolding may affect protein trafficking or even cause degradation in the ER (Bernasconi & Molinari, 2011). Defective trafficking leads to mislocalization disabling SI from fulfilling its physiological function (Gericke et al., 2016).

In 1988 biopsies from CSID patients were analyzed leading to the classification of the SI gene mutants into seven phenotypes (H. Y. Naim et al., 1988), defined by trafficking and function (Table 1.5). The phenotype I is the most prevalent among European descendants, contributing about 83% of the CSID patients (Uhrich, Wu, Huang, & Scott, 2012). However, one of the challenges associated with SID is the lack of an appropriate diagnosis method, which leads to misdiagnosis and consequently to inaccurate statistics.

Table 1.5. Congenital sucrase-isomaltase deficiency phenotypes.

Phenotype	Enzymatic activity	Defect
I	Completely inactive	Blocked in the ER
II	Completely inactive	Blocked in the ER, ER-Golgi compartment, and cis-Golgi
III	Completely inactive	Apparent normal trafficking and folding unclear catalytic site disruption
IV	Active sucrase and isomaltase	Sorting defects
V	Active isomaltase, inactive sucrase	Intracellularly cleaved, sucrase degradation
VI	Active sucrase and isomaltase	Intracellularly cleaved. L340P mutation.
VII	Decreased sucrase activity, isomaltase completely inactive	C635R mutation causing misfolding, defective trafficking, altered sorting, and increased turnover rate.

The most accurate diagnosis requires testing duodenal or jejunal mucosa, obtained through a biopsy (Treem, 2012). Due to the invasiveness of this method, another technique was developed: the sucrose hydrogen breath test. This test is based on the analysis of exhaled gases post-sucrose/^{13}C-sucrose intake (Robayo-Torres et al., 2009).

The common treatment for sucrase-isomaltase deficiency is a combination of diet adaptation and enzyme replacement therapy (Gericke et al., 2016). Currently, Sucraid, from QOL Medical, LLC, is the only product in the market that offers an alternative for enzyme replacement. QOL uses sacrosidase (invertase) purified from *Saccharomyces cerevisiae* to compensate for the lack of alpha 1,2 hydrolytic activity of the sucrase subunit of SI (Treem, 2012).

1.2.6. Other diseases related to carbohydrate digestion

Type II diabetes is the most common of metabolic diseases (Ghani, 2015). The World Heath Organization has predicted that by 2030, 4.4% of the world population might suffer from this disease (Ghani, 2015). In addition to its symptoms and health consequences, diabetes is also a challenge for health services due to its high cost. The American Diabetes Association in the United States has reported an estimated cost of $245 billion of dollars in diabetes diagnosis during 2012, almost double that in 2007 (Ghani, 2015).

As the intestinal alpha-glucosidases are responsible for releasing most of the free glucose absorbed in the luminal membrane; they are the target of many treatments for type II diabetes (C. C. Robayo-Torres et al., 2006). Those treatments aim to selectively inhibit key subunits of

MGAM and SI, to decrease both the rate of carbohydrates digestion and the amount of free glucose released. Acarbose is one of the most used inhibitors worldwide (Li et al., 2005). This inhibitor has shown a preferential affinity for ct-MGAM with a Ki < 28 nM (Jones et al., 2011). However, the acarbose inhibition of MGAM causes such side-effects as abdominal cramps, bloating, and occasional diarrhea (Li et al., 2005). Therefore, there is an urgency for developing better and more specific treatments for diabetes (Ghani, 2015).

Other metabolic diseases related to carbohydrates digestion are obesity and metabolic syndrome, maltase-glucoamylase deficiency, pan disaccharidase deficiency, as well as lactose, fructose and sorbitol intolerance (Pontremoli et al., 2015; Quezada-Calvillo et al., 2008a).

1.2.7. Inhibition of intestinal alpha-glucosidases: MGAM and SI

Retarding or inhibiting the alpha-glucosidase activity is a good approach to control the postprandial hyperglycemia and blood glucose levels in diabetes type 2 patients (Mali et al., 2007). To achieve this, different inhibition compounds have been studied and developed. They reduce the rate of digestion or even completely prevent it, decreasing the glucose release and therefore, hyperglycemia(Krentz & Bailey, 2005). However, as the digestion is disrupted, the patients experience symptoms similar to the ones caused by sucrase-isomaltase deficiency: bloating, diarrhea, flatulence and other abdominal discomforts associated to maldigestion (Chiasson et al., 2004). Such problems cause that many patients drop their treatments threatening their wellbeing and safety.

The type of inhibition can be simplified into three main models: competitive, mixed, and uncompetitive; however, there are also partial inhibitions. Nevertheless, partial models of inhibition are harder to determine as they cannot be assessed through Lineweaver-Burk plots, but require secondary plots.

Figure 1.6 illustrates the main characteristics of the different types of inhibition and their kinetics. The competitive inhibition model refers to the compounds that compete with the substrate for the active site of the enzyme and once that the enzyme-inhibitor (EI) is formed they block the site preventing the substrate binding and further hydrolysis. A mixed model of inhibition often reflects allostery, as the inhibitor binds in a different binding site that is not the active site and therefore changes the conformation of the enzyme impeding the formation of the enzyme-substrate (ES) complex. A compound is identified as an uncompetitive inhibitor when it binds the ES complex avoiding the hydrolysis and release of products.

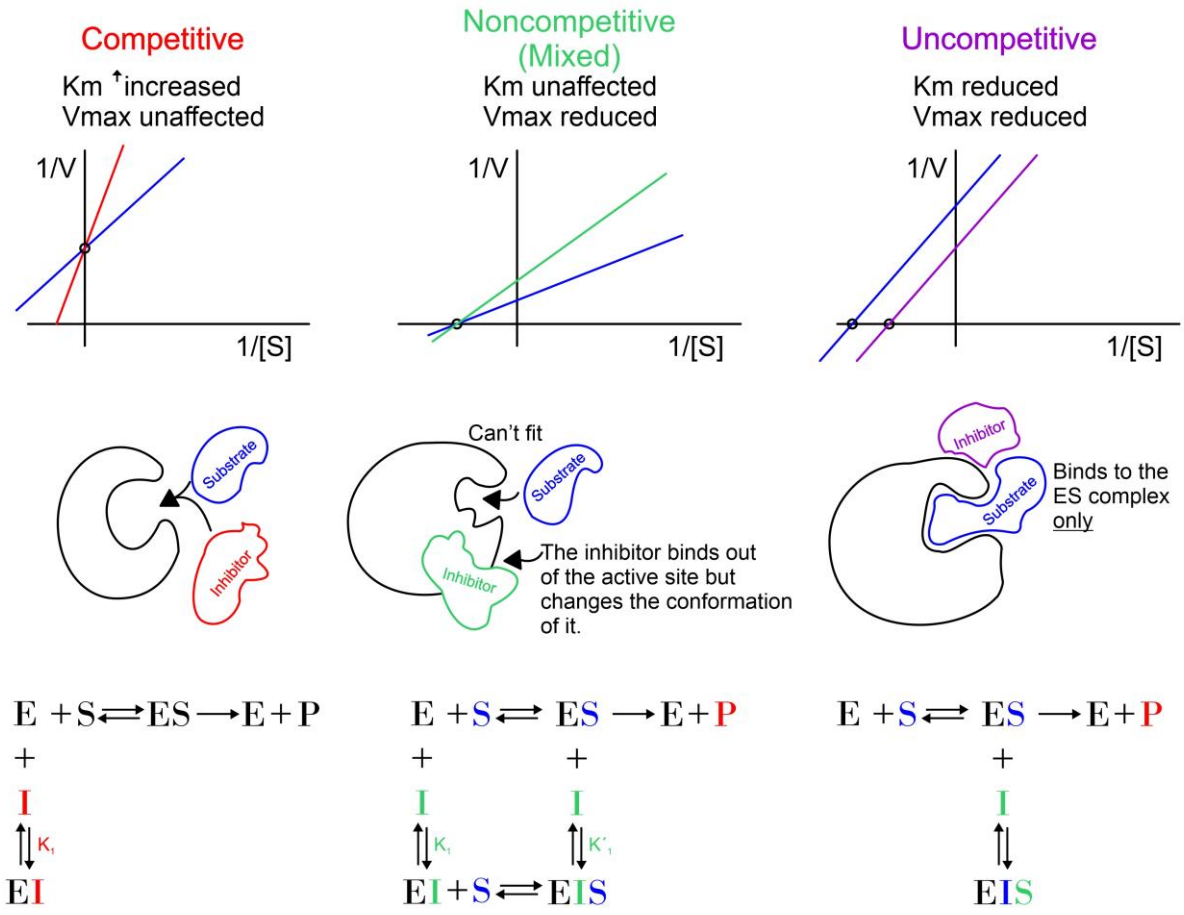


Figure 1.6. Types of inhibition and its main characteristics.

1.2.8. Common inhibitors

There are two main groups of inhibitors: natural and synthetic (Ghani, 2015). Natural includes sugar mimics as iminosugars and thiosugars, but also a broad group of compounds such as chalcones, flavonoids, curcuminoids, terpenes, and phenols. The synthetic group is also very diverse containing some sugars mimics such as aminosugars, azasugars, and thiosugars, but also, an extensive variety of chalcones, cyclitol derivatives, thiadiazoles, among many others.

Currently, the market of alpha-glucosidases inhibitors for type 2 diabetes treatment is dominated by three compounds: acarbose, voglibose, and miglitol Figure 1.7.

Acarbose, commercially known as Glucobay (Europe and China), Precose (north America) and Prandase (Canada) and commercialized by Bayer AG is the first alpha-glucosidases inhibitor for treating Type 2 Diabetes. Acarbose is a pseudotetrasaccharide formed by four glucose units joined by three alpha 1-4 glucosidic linkages and one imino bridge. Due to the high structural similarity to oligosaccharides

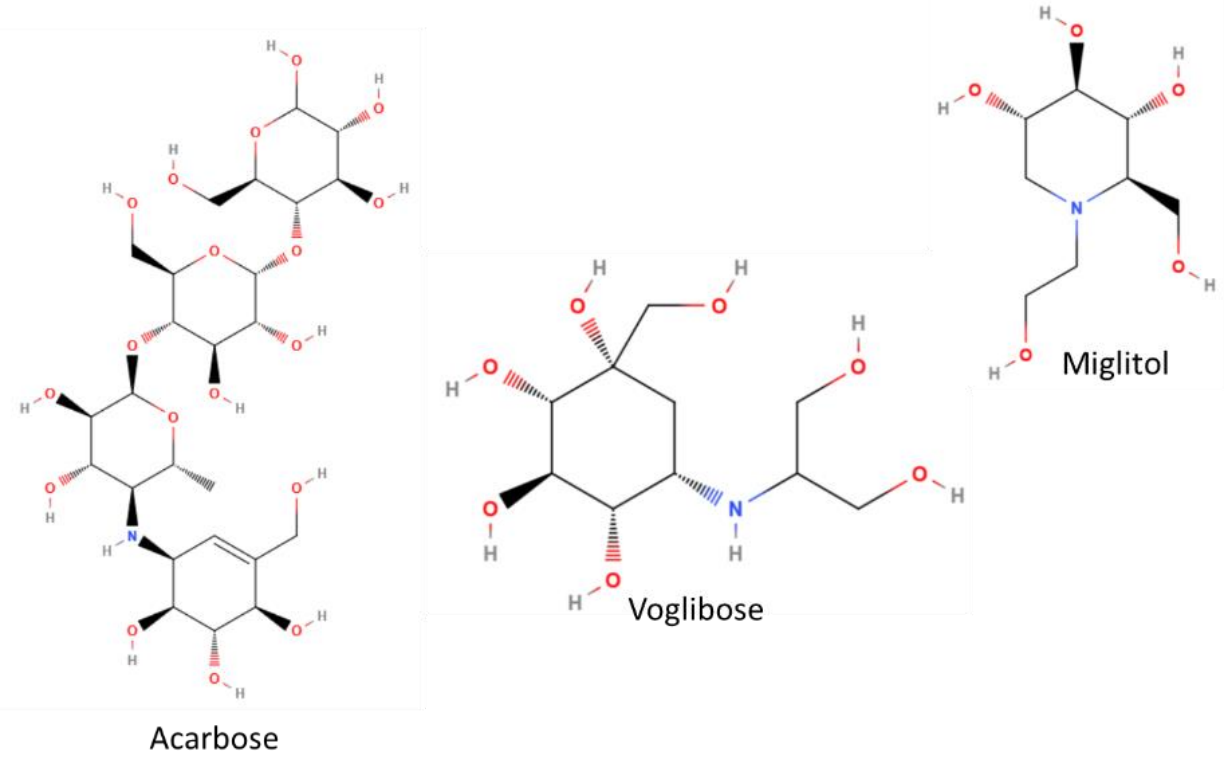


Figure 1.7 Common alpha-glucosidases inhibitors currently used to treat Diabetes Mellitus type II: Acarbose, voglibose, and miglitol.

Voglibose is a synthetic derivative of valiolamine, a potent inhibitor produced by Streptomyces hygroscopicus (Horii et al., 1986). Voglibose inhibits maltase and sucrase enzymes facilitating

the control of blood glucose with fewer side effects in comparison with acarbose. However, its efficiency is also lower (Vichayanrat, Ploybutr, & Tunlakit, 2002).

Miglitol is a derivative of Nojirimycin (NJ), a compound produced by Streptomyces roseochromogenes and has inhibitory activity against alpha and beta-glucosidases. Unlike the other two inhibitory compounds, miglitol is almost completely absorbed in the intestine, which may cause other systemic effects (Joubert, Foukaridis, & Bopape, 1987). However, there is no evidence of other effects in the organism (Sels, 1996).

Iminosugars

In imino sugars, a nitrogen atom replaces the ring oxygen of monosaccharides. These types of compounds can be monocyclic or bicyclic. Monocyclic iminosugars include pyrrolidines, piperidines, and zapanes. The bicyclic iminosugars include pyrrolizidines, indolizidines, and nortropanes (Asano, 2008). Figure 1.8.

Within the last thirty years, numerous iminosugars (natural and synthetic) have been studied as potential alpha-glucosidases inhibitors. As a result, numerous patents have been registered (Wadood et al., 2018).

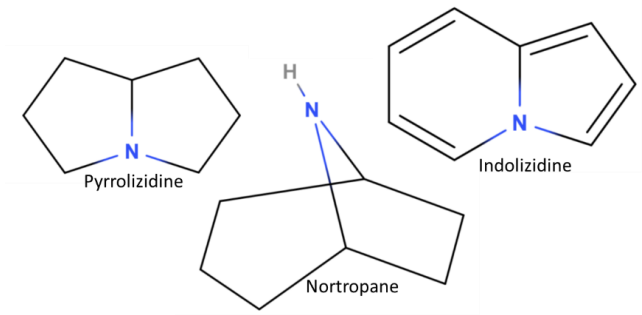


Figure 1.8 Example of bicyclic imino sugars relevant for inhibiton of alpha glusocidases.

Thiosugars

Like the iminosugars, in the thiosugars, the ring oxygen is replaced, but with sulfur rather than nitrogen. The most relevant group of thiosugars identified as alpha-glucosidases inhibitors have been isolated from *Salacia Reticulata*, a plant found in the submontane forest of southeast Asia (Yoshikawa, Murakami, Yashiro, & Matsuda, 1998). This group includes Kotalanol and Salacinol (Figure 1.9), both potent inhibitors with higher efficiency than Acarbose (Masayuki, Toshiyuki, Yoshikawa et al., 1997)(Yoshikawa et al., 1998).

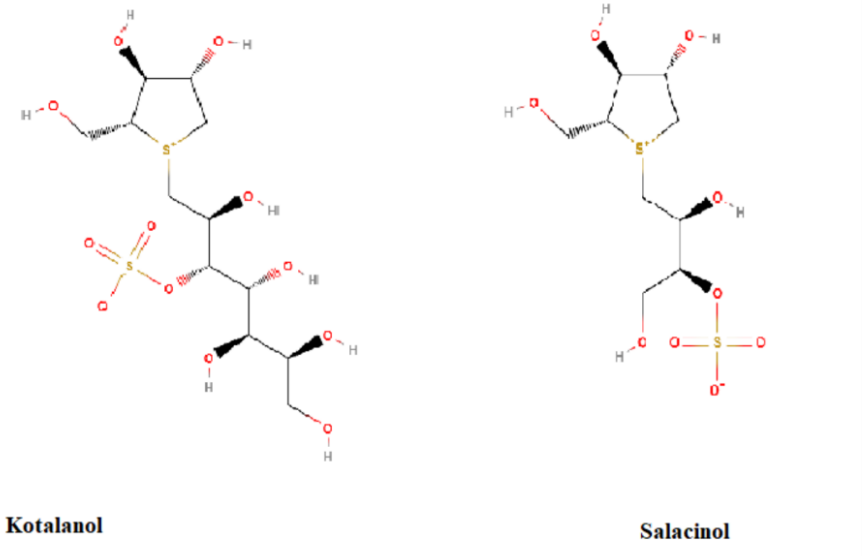


Figure 1.9 Structures of Kotalanol and Salacionol, the main thiosugars that have been proved to have inhibitory activity against alpha-glucosidases.

1.2.9. Lactase-phlorizin hydrolase (LPH).

At postnatal ages, milk is the main source of nutrients; it supplies about 70g/L of carbohydrates, from which close to 60 g are lactose (O-alpha-D-galactopyranosyl-[1-4]-alpha-D-glucopyranoside)(Foda, Kawashima, Nakamura, Kobayashi, & Oku, 2004). Therefore, lactose is the main caloric source. Lactose is hydrolyzed to galactose and glucose in the small intestine by Lactase-phlorizin hydrolase (LPH) EC 3.2.1.23 an enzyme anchored to the apical membrane by its C-terminus (H. Y. Naim, Sterchi, & Lentze, 1987).

Lactase-phlorizin hydrolase belongs to the GH family 1 (CAZy). it is a four-domain protein (Kruse et al., 1988). Domains I and II are recognized as intramolecular chaperones that assist in the proteins folding and are cleaved latter during the protein maturation (Mantei et al., 1988). The remaining two domains LPHIII and LPH IV each have one enzyme active site, both with phlorizin hydrolytic activity; however, only LPH IV is capable of digesting lactose (Jacob, Peters, & Naim, 2002). LPH is highly N- and O- glycosylated which seems to play an important role in both folding and assorting of the protein and therefore in its enzymatic activity (Jacob, Weiner, Stadge, & Naim, 2000).

LPH is synthesized as single polynucleotide chain. A a precursor form (215 KDa) is N-glycosylated with high mannose glycans. However once in the Golgi apparatus, the glycans are trimmed and replaced by complex sugars, and the precursor itself undergoes a series of proteolytic cleavages where all the domain I and about two thirds of the domain II (known as LPH_{α}) are removed. The remaining protein (LPH_{β}) is transported to the apical membrane for sorting.

Similar to SI and MGAM, the LPH domains III and IV share the (β/α)8 barrel structure and follow the double displacement mechanism; glutamic acid 1273 in domain III and 1749 in domain IV act as the nucleophiles, and glutamic acid in 1065 and 1538 as the proton donors, respectively. However, due to the lack of structural data for human LPH many questions regarding its hydrolytic mechanism, substrate affinity and folding remain unclear.

Due to its high level of complexity, the expression of a recombinant version of human LPH is challenging. Mammalian cell cultures rarely produce enough protein for structural studies. On the other hand, yeast or insect cells may not be able to perform the series of proteolytic trimmings required for the synthesis of the LPH mature form "LPH_{β} final".

LPH domain III (LPH III) can be expressed independently from the rest of the protein, and even be transported and conserve about the 40% of the phlorizin hydrolase activity. In the other hand, LPH IV, the only domain with lactose hydrolytic activity, is transport deficient and inactive by itself (Behrendt, Polaina, & Naim, 2010).

Chapter 2. *Komagataella phaffii* Expression System

2.1 Chapter overview.

This chapter focuses on the development of an expression system suitable for the expression of human alpha-glucosidases. These enzymes are relevant for the study of carbohydrate digestion and therefore related to numerous metabolic diseases such as type II diabetes, obesity and disaccharidase deficiencies.

The yeast organism *Komagataella phaffii* was selected as host for the expression due to the high yield of foreign protein that can be achieved during a successful expression via methanol induction, the simplicity of protein purification when it is secreted, the capability to perform posttranslational modifications, and the lower probability of hyperglycosylations in comparison with other yeast systems. The currently available expression systems showed different weaknesses such as a high vulnerability to contamination due to a lack of antibiotic resistance, inconvenient cloning methods for these types of proteins, difficulties in the protein purification, and low yield. Therefore, it was necessary to develop an expression system that overcomes these limitations.

The proposed system simplifies the obtention of the different constructs by replacing the classic cloning techniques with homologous recombination using *Saccharomyces cerevisiae* as an intermediary. To achieve this step a commercial expression vector for *K. phaffii* was modified by the addition of a *loxP* marker, a *Saccharomyces cerevisiae* promoter and an auxotrophy gene for selection in *S. cerevisiae*.

Another limitation in the expression of foreign proteins in *K. phaffii* is the number of copies of the plasmid; a higher number may increase the expression of protein during the induction period (Chen et al., 2005). To improve the selection of transformed cells, a double selection marker of adenine auxotrophy and zeocin resistance was introduced into the system. Positive transformants can grow in drop out media without adenine due to the presence of an *Ade* gene in the introduced plasmid. Furthermore, the cells with a higher number of plasmid copies can resist higher concentrations of antibiotic; therefore, in the developed system it is not only possible to differentiate among transformed and untransformed cells but also to filter them by number of plasmid copies using a gradient of antibiotic concentration.

The last modified feature on the expression vector was the successful addition of a 12-histidine chain to facilitate protein purification using nickel/cobalt affinity (Kimple, Brill, & Pasker, 2013). It is a common practice in yeast expression to use long (more than 6) histidine chains to overcome interference in the folded protein and enhance the affinity towards the nickel/cobalt resin selected for purification (Grisshammer & Tucker, 1997). In this case the addition of the extra 6-histidines improved the protein recovery by more than 50% in comparison with the conventional 6-histidine chain.

Finally, an incomplete factorial experimental design was used to generate a surface of response and determine the optimum conditions (methanol/time) for the highest yield of expression. The model predicted as a maximum point the addition of 13% methanol (daily) during an expression period of 120 hours (5 days). The predicted optimum concentration of methanol to be added exceeds the toxicity limit reported for other systems using *K. phaffii* (Santoso et al., 2012).

However, as the expression is done in batch, the cells are not intended to be reused once the expression is over. Therefore, the goal is to reach the highest possible yield of protein, regardless of the cell viability. The cell toxicity under these conditions was not measured directly and, therefore, it was not possible to compare with the value reported by Santoso. These predictions were tested using three of the expressed alpha-glucosidases, obtaining an average yield of 4.1 mg/L of expression.

2.2 Objectives

General.

Develop an expression system for production of human alpha-glucosidases in *Komagataella phaffii*.

Specific aims.

1. Design and clone a shuttle vector which enables cloning of alpha-glucosidases via homologous recombination in *S. cerevisiae* and transformation into *K. phaffii* for expression.
2. Implement a double selection marker for high-copy plasmid selection in transformed *K. phaffii* cells.
3. Achieve efficient protein purification using metal-affinity chromatography.
4. Create modified expression protocols and media recipes to maximize the protein production per expression using a batch method.

2.4 Methodology

2.4.1 Design of the pPinkNN_Z vector.

To assemble the shuttle vector the pPink HC expression vector from Invitrogen, Inc. was used as template Figure 2.1. This vector belongs to a commercial *K. phaffii* expression system and already contains certain desirable genes such as *Ade2*, for adenine auxotrophy, and the whole machinery for protein expression in *K. phaffii*. Therefore, no deletions from this vector were required, but only the addition necessary to make it suitable for *S. cerevisiae*. The elements added were two *loxP* sites, the 2-micron origin, and the *URA3* gene with promoter. First, forward and reverse primers were designed to amplify the sequence flanked by the *loxP1* and *loxP2* sequences in the pJJH1396 vector (Paululat & Heinisch, 2012). Those same primers were also used to add homology sites that match with the designated segments of the pPink HC sequence to perform the recombination (LoxP.fwd and LoxP2.rev) Appendix A. Then, the pPink HC vector was linearized with *EheI*, which has a unique restriction site located downstream of the *AmpR* promoter. The PCR products and linearized vector were then transformed into *S. cerevisiae* strain HOD228-3C (Kirchrath, Lorberg, Schmitz, Gengenbacher, & Heinisch, 2000) and plated in minimal media without uracil (-ura). After the plates were incubated at 30ºC for three days, random colonies were selected and inoculated into liquid -ura media. The plasmid DNA was extracted and confirmed via Sanger sequencing using AOX1.fwd and CYC1.rev primers.

Two other minimal modifications were done by classic cloning: 1) the addition of the pre-sequence alpha-mating factor from *S. cerevisiae*, as secretion signal, and 2) a TEV site followed

by a 6-histidine chain for protein purification. The resultant vector was called pPinkNN Figure 2.1.

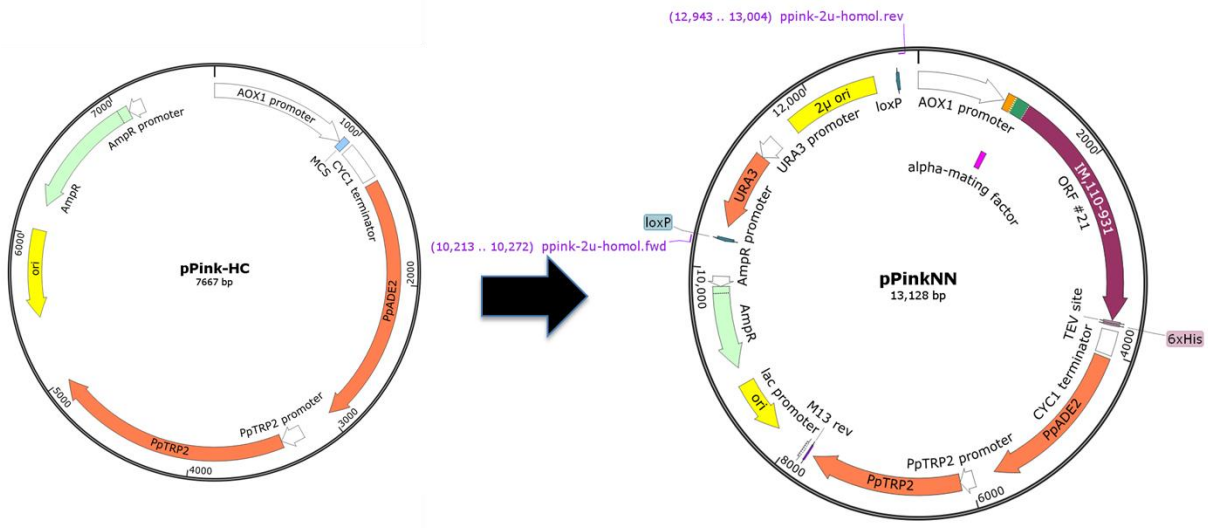


Figure 2.1 Progression pPinkHC (top) to pPinkNN (bottom) maps made in SnapGene version 4.3.10 GSL Biotech LLC.

The second major modification was the addition of the *S. hindustanus*-derived *ble* gene and its promoters to provide zeocin resistance as second selection marker (Benko & Zhao, 2011). A construct from Dr. Megan Barker, a former student in the Rose lab, was used as DNA template to amplify the EM7, TEF1, and Sh *ble* sequences. The forward and reverse primers were designed overlapping 20 bp from the desired sequence and adding compatible restriction sites: KpnI for the forward and SmiI for the reverse primer (Zeo_Kpn1.fwd and Zeo_SmiI.rev) Appendix A. Then a PCR amplification was performed, and in parallel, the pPinkNN vector was double digested with KpnI and SmiI restriction enzymes (Thermo scientific FastDigest). Once the PCR products were confirmed by Sanger sequencing and the vector linearized, they were ligated using T4 ligase and after incubation were transformed into *E. coli* TOP10 competent cells Invitrogen Cat. No. C404003) via heat shock and plated in LB media with ampicillin. After

screening, the successful colonies were used to inoculate liquid LB media with ampicillin and the plasmid DNA was extracted by alkaline lysis. The Plasmid DNA was confirmed by Sanger sequencing using AOX1.fwd and CYC1.rev primers Appendix A. The resultant vector was re-named as pPinkNN_Z Figure 2.2. Additionally, the histidine chain was extended by six residues to a total length of 12 histidine residues, after facing challenges in purifying proteins from preliminary expressions.

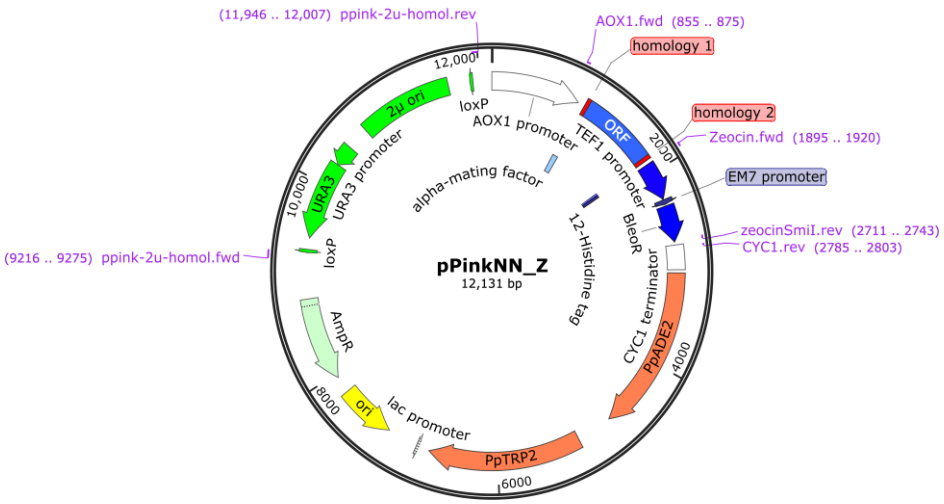


Figure 2.2 Map of the pPinkNN_Z vector after the addition the zeocin-resistance mechanisms and the extension of the 6-histidine chain.

2.4.2 Alpha-glucosidases construct generation.

The alpha-glucosidases selected to test the new expression system were Sucrase-isomaltase full length (SI_{FL}), its C-terminal subunit sucrase (CtSI), and the N-terminal subunit isomaltase (NtSI). The human cDNA for the cloning was obtained from Dr. Naim's group in Hannover and is shown in the Table 2.1.

Table 2.1 Proteins expressed in *K. Phaffii* using the pPinkNN_Z expression vector developed and their correspondent gene.

Protein	Abreviation	NCBI accession number	EC number
Sucrase- isomaltase (Full length)	SI_{FL}	NM_001041.4 From 392 to 5548	3.2.1.10/48
Isomaltase subunit of SI	Nt-SI	NM_001041.4 From 392 to 3085	3.2.1.10
Sucrase subunit of SI	Ct-SI	NM_001041.4 From 3086 to 5545	3.2.1.48

For homologous recombination, it was necessary to clone the selected homology sites from the pPinkNN_Z vector into the DNA sequences for each protein. For this purpose, specific forward and reverse primers with overhanging homology sites were designed for each protein (Appendix A, xSI_homol.fwd, xSI_homol.rev; where x represents the protein of interest). Then the DNA sequences with homology sites (H_1 for forward and H_2 for reverse) were amplified via PCR. The conditions for the amplification were optimized according to the size of each gene following the recommendations for the polymerase provider. Simultaneously, the pPinkNN_Z vector was linearized in the unique restriction site Age I located between the two homology sites. Then, both PCR products and linearized vector were transformed together into *S. cerevisiae* strain HOD228-c competent cells using a heat shock protocol. The transformed cells were plated in dropout minimal media without uracil and incubated for 3 days. After incubation well-isolated colonies were selected, grown in liquid media, and the plasmid DNA extracted using an alkaline lysis miniprep technique. The sequence of the isolated plasmids was confirmed via Sanger sequencing (BioBasic Inc.) using the AOX1.fwd and CYC1.rev primers (Appendix A) for each of the three proteins of interest.

Finally, the confirmed plasmids were transformed into freshly made electrocompetent *K. phaffii* cells strain P4 from Invitrogen Inc. via electroporation. The transformants were first plated in dropout media without adenine (auxotrophy selection marker) and further re-plated into YPD plates with 150 mg/mL of zeocin. The well-isolated colonies were then grown in liquid media and the cultures used for genomic DNA purification. The genomic DNA was used for a PCR using AOX1.fwd and CYC1.rev primers. The sequence of the amplified DNA was confirmed by Sanger sequencing. The positive colonies were then used for glycerol stocks and stored at -80ºC.

2.4.3 Cell growth curve after antibiotic exposure

To test zeocin resistance a growth curve was made using *K. phaffii* untransformed cells as control and NtSI_pPinkNN_Z transformed in the same strain of *K. phaffii*. These cells were first grown from glycerol stocks in BMGY solid media. Then, healthy, well-isolated colonies were selected to inoculate 12mL of liquid BMGY media at 0, 50, 100, and 150 mg/mL of zeocin. The liquid cultures were incubated at 28ºC in constant shaking at 200rpm in 125mL flasks with ventilated caps. The OD_{600} were measured and recorded for 72 hours (Figure 2.6).

2.4.4 Expression of human alpha-glucosidases in *K. phaffii*.

The expression of protein in *K. phaffii* can be divided into three stages: 1) Starter culture, 2) High cell density production, and 3) Induction of expression. During the first stage, glycerol stocks were used to inoculate 25mL of buffered complex glycerol media (BMGY) into 120mL flask and incubated at 28ºC with constant shaking at 200 rpm to an OD_{600} between 1.2 and 1.6. To produce the high cell density required for expression, start cultures were scaled up to 1000 mL of BMGY media, split equally among four 1L flasks with ventilated caps and incubated under

the same conditions until they reached OD_{600} of 2 to 6. For protein induction, the media containing glycerol must be replaced with media with methanol. The cultures were centrifuged at 3000xg for 5 minutes to collect the cells (pellet) and were further resuspend into 200 mL of buffered media with methanol (BMMY). To optimize expression conditions, an experimental design was performed in Design Expert version 12 (Stat-Easy Inc). This generated a matrix of experiments where the length of the expression and methanol addition were considered as main factors, and yield of expression as response. Expressions were performed with multiple methanol concentrations between 0.5% and 20% with incubation times from 1 to 5 days. The experiments were performed one enzyme at a time. Protein was harvested and incubated, with shaking, with the nickel resin for 12 hrs; protein purification, concentration, buffer exchange, and measurement were performed right after incubation.

2.4.5 Purification

After expression, cultures were centrifuged at 5000rpm for 15 min to separate the pellet and supernatant. As the pPinkNN_Z expression vector contains the pre-sequence alpha-mating factor, the protein of interest was recovered from the supernatant. The supernatant was mixed with the nickel/cobalt resin (Fisher scientific Cat. No. PI88222) according to the manufacturer instructions, and the proteins eluted with 100mM imidazole buffer for NtSI and CtSI, and 250mM for SI_{FL}. After elution the proteins were concentrated using 50KDa Amicon tubes (Millipore Cat. No. UFC905024 and UFC505024) to a volume between 1 and 0.5mL. Then, the imidazole buffer was replaced with PBS at pH 6.4. Protein concentration as measured in a nanodrop (Nanodrop 2000c, Spectrophotometer Thermo Scientific) as absorbance at 280nm. In

addition, 8% polyacrylamide gels were run at 80V through the staking gel (6%) and at 100V in the running gel to test for purity and correct protein size.

2.5 Results and Discussion

2.5.1 pPinkNN: shuttle vector *S. cerevisiae- K. phaffii* and construct preparation through homologous recombination.

The most commonly recommended classic cloning approach for *K. phaffii* expression systems includes a triple ligation (*PichiaPink* ^{TM} *Expression System*, 2010). This strategy, though convenient, is not suiTable for large proteins, such as those of interest in this chapter, and did not provide any useful results for the cloning. A second approach was then designed, using different vectors as intermediaries to clone the alpha-mating factor and histidine chain into the protein sequence and finally into the expression vector (unpublished data, Dr. Jin Duan). Despite the success achieved with this approach, factors such as the time consumed and multiple PCR amplification, digestions, transformation, etc., made it inconvenient for the preparation of multiple constructs (Table 2.2). The length of the coding sequences represents a challenge for cloning manipulation, as the probability of multiple restrictions sites or binding sites for primers increases with the sequence length, along with chances of random mutations over the cloning process. Therefore, reducing the number steps during the construct preparation, and consequently minimizing the amount of DNA manipulation, is highly desirable. To fulfill this purpose homologous recombination was chosen as a cloning strategy, instead of the two previous approaches.

The chosen host organism for the expression, *K. phaffii*, uses non-homologous end joining as the preferred DNA repair mechanism; therefore, its efficiency at performing homologous recombination is limited (1-30%) (Vieira-Gomes et al., 2018). On the other hand, *S. cerevisiae* uses homologous recombination with an efficiency between 97 and 100%. For this reason, it was more effective to assemble the DNA sequence of the protein of interest in the expression vector using homologous recombination in *S. cerevisiae* and then transform the isolated plasmid into *K. phaffii*, taking advantage of the compatibility between the two organisms. This strategy required the design of a shuttle vector to be used in *S. cerevisiae* for homologous recombination and in *K. phaffii* for protein expression.

Homologous recombination, as a cloning method, is convenient not only because minimizes the DNA manipulation to only one PCR and one transformation, but also because it is more time efficient than traditional cloning (Kunes, Schatz, & Botstein, 1987). For this reason it has been extensively used and adapted to numerous systems (Kuijpers et al., 2013; van Leeuwen, Andrews, Boone, & Tan, 2015). Although *S. cerevisiae* requires longer incubation times than *E. coli*, the reduction in the number of steps and intermediaries during the cloning decreased the cloning time from over a month to a week for CtSI and SI_{FL} (Table 2.2).

Table 2.2 Comparison between the classic cloning and homologous recombination approaches for the preparation of the constructs to express alpha-glucosidases in *K. phaffii*.

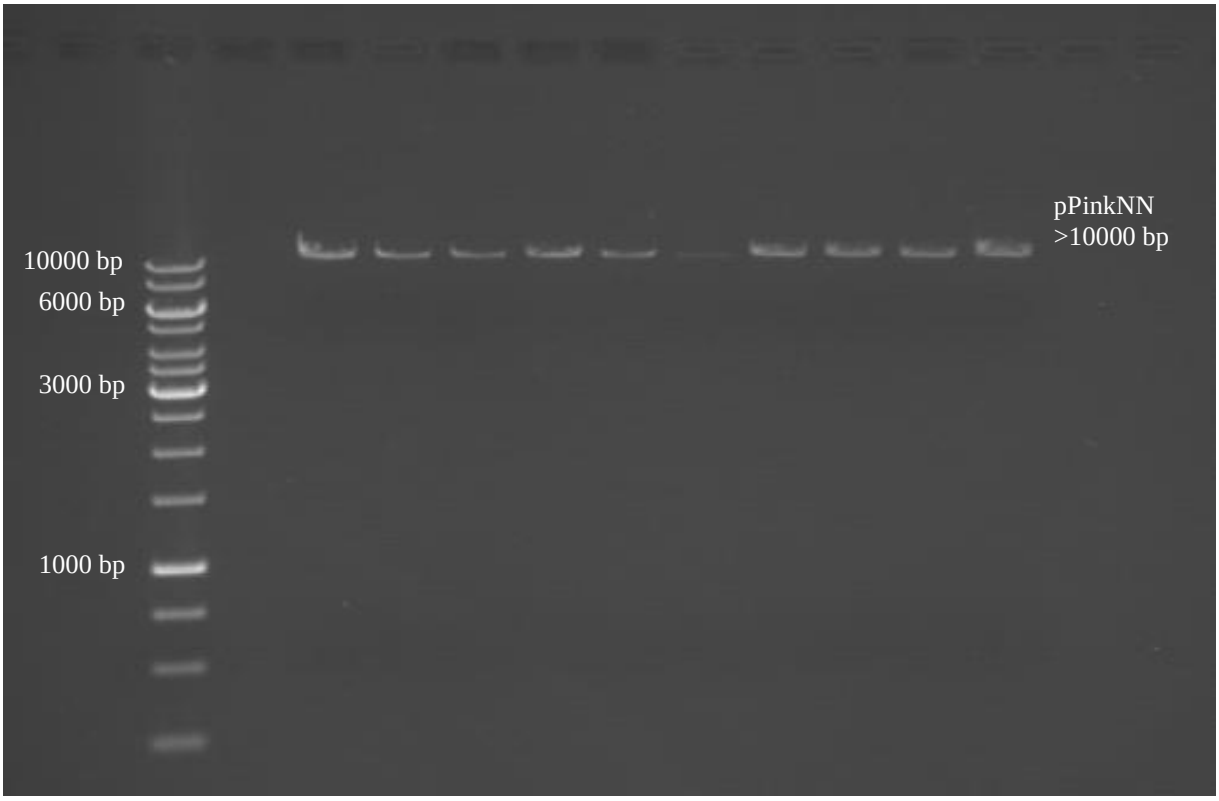
Classic cloning	**Homologous recombination**
Amplification of DNA sequence of interest	Amplification of DNA sequence of interest via PCR with homology primers and linearization of pPinkNN
Cloning into pBS via REN, ligation and transformation (addition of secretion signal)	Transformation in *S. cerevisiae* (Homologous recombination)
Cloning into pET30a via REN, ligation and transformation (addition of TEV site and polyhistidine chain)	Linearization of pPinkNN plasmid
Cloning into *K. phaffii* vector via REN, ligation and transformation	Transformation in *K. phaffii*
Linearization of *K. phaffii* plasmid	
Transformation in *K. phaffii*	
Estimated time: 1 month	Estimated time: 1 week

Therefore, it was necessary to design a vector that simplified the cloning process. To achieve this purpose, the alpha-mating factor (secretion signal), TEV-site, and histidine chain, were combined into a *K. phaffii* vector. In addition, this vector was made compatible with *S. cerevisiae* to be able to integrate the sequence of the protein of interest into the vector through homologous recombination.

The pPink HC expression vector from Invitrogen Inc. used as template for the shuttle vector has 7.7 Kb, ampicillin resistance in *E. coli*, and an *ADE2* gene with its promoter for adenine auxotrophy. To avoid disrupting these features, the *URA3* gene, its promoter, and the 2-micron origin were cloned via homologous recombination in *S. cerevisiae* at the end of the vector (7667-0). The transformants were plated in dropout solid media without uracil, well isolated colonies

were selected for liquid inoculation and further alkaline lysis (miniprep) to isolate the plasmid DNA (Figure 2.3). The agarose gel revealed a band slightly larger than 10kb, the expected size of the modified vector, which indicates that the insertion was successful. The plasmid was also confirmed via Sanger sequencing.

Figure 2.3 Agarose gel showing the alkaline lysis (miniprep) for plasmid isolation from *S. cerevisiae* cultures in dropout media without uracil. In the right the Gene ruler ladder from Thermo scientific, all the other wells were loaded with 5µL of isolated pPinkNN plasmid.



Additionally, using classic cloning techniques, the alpha-mating factor pre-sequence and a TEV site followed by a chain of six histidines were cloned into the MCS of pPink HC (941-991). The homology sites 1 and 2 were chosen (1012-1051 and 3662-3701) respectively. As a result, a final

vector was 13.1 kb, including the sequence for NtSI between the homology sites 1 and 2 (Figure 2.2).

This first version of the expression vector was named pPinkNN and was used to prepare the constructs to express the two subunits of sucrase-isomaltase (NtSI and CtSI) as well as the SI full length (SI_{FL}). The constructs were prepared according to the methodology described in section 2.4.2. After miniprep, the isolated plasmids were amplified via PCR using primers in *aox1* and *cyc1*. Gel electrophoresis (Figure 2.4) identified bands at the expected sizes of the corresponding DNA sequences, showing that cloning was successful. The PCR products were then confirmed via Sanger sequencing.

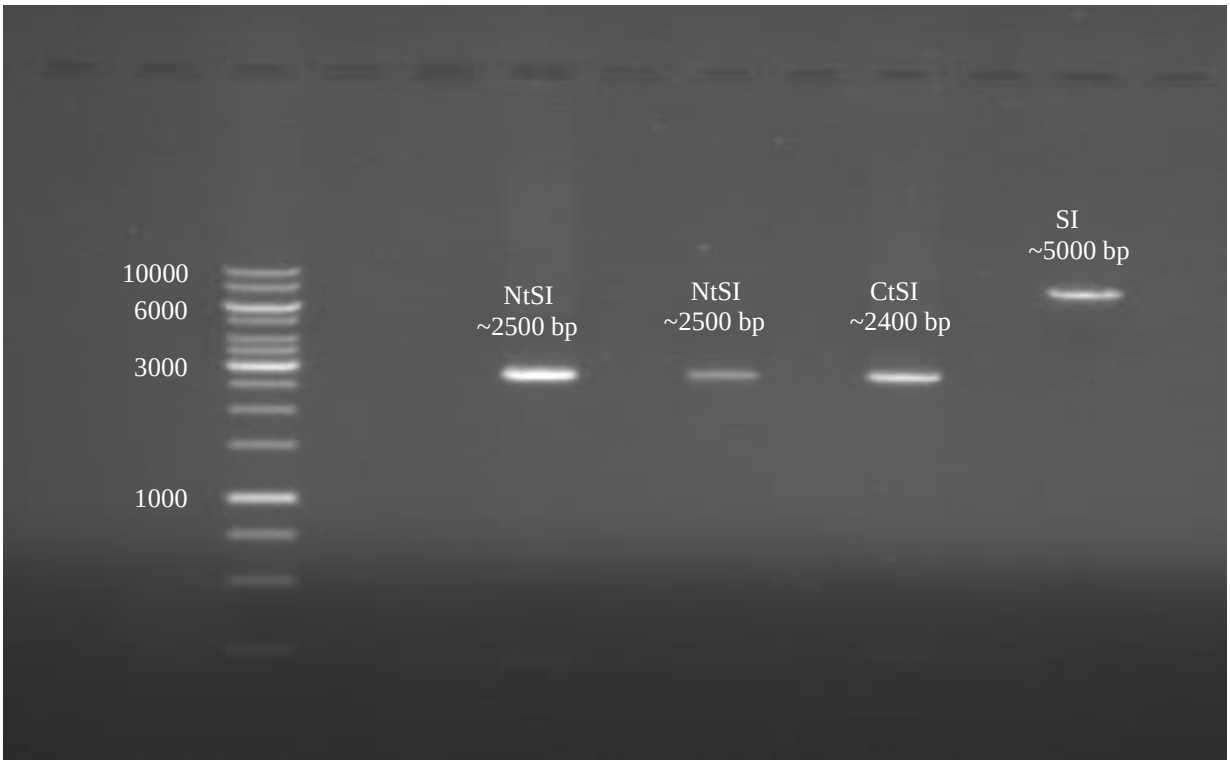


Figure 2.4 Agarose gel showing the PCR products resulted from the amplification of the three different plasmids (NtSI_pPinkNN, CtSI_pPinkNN, and SI_pPinkNN) with the AOX1.fwd and CYC1.rev primers.

After confirmation, the pPinkNN plasmids were linearized and electroporated into *K. phaffii* electrocompetent cells. The transformants were plated in solid dropout media without adenine, and only white colonies were selected for liquid inoculation and further genomic DNA extraction. The extracted DNA was confirmed performing the same PCR as that used to confirm the *S. cerevisiae* plasmids and the PCR products were sent for sequencing. The sequencing results confirm that the protein sequences were successfully inserted into the vector and transformed into *K. phaffii*.

The successful transformation of the three sucrase-isomaltase constructs proved:

1) That pPinkNN is indeed acting as a shuttle vector.
2) The homologous recombination approach worked and simplified the cloning process to only one PCR and one transformation to clone the desired DNA sequence into pPinkNN.
3) The auxotrophy in both *S. cerevisiae* (-*URA*) and *K. phaffii* (-*ADE*) worked as selection marker to identify transformed colonies.

2.5.2 Double selection marker: Adenine auxotrophy and zeocin resistance.

The expression of proteins in *K. phaffii* is induced by the addition of methanol. The metabolism of methanol depends on the expression on alcohol-oxidase, and the protein of interest is controlled by the same protomer (AOX1) (L. Cereghino et al., 2000); therefore, increasing the amount of methanol fed to the system increases yield of proteins of interest. However, there are

some challenges. If the cell count of *K. phaffii* is too low, the methanol feed may accumulate until it reaches toxic levels. For example, if the expression culture gets contaminated, especially in early stages (start culture/cell growth), methanol may accumulate even after the correct OD is reached, depending on the proportion of *K. phaffii* cells to contaminant cells. Another challenge is the selectivity between number of plasmid copies. The -*Ade* selection marker helps to identify the successfully transformed cells; however, determining the number of copies of the plasmid relies on a visual color identification where a stronger pink colour means lower number of copies (*PichiaPink* ^{TM} *Expression System*, 2010). However, the colour development takes an extra 3 to 5 days after the colonies grow and develops better at cold temperatures. Consequently, colonies with low copy number of the plasmid are commonly mistaken as promising for expression before developing colour.

These two challenges were simultaneously addressed by the addition of the Sh *ble* gene and promoter into the pPinkNN vector. This gene confers antibiotic resistance to zeocin (Higgins et al., 1998). The addition of zeocin to the start cultures and during the cell density production prevents the growth of contaminant organisms, keeping the OD_{600} measurement representative of the number of *K. phaffii* cells. Additionally, the presence of zeocin in the -*Ade* plates limits the growth of colonies with low copies of the vector, as a higher number of copies confers higher tolerance to the antibiotic (Higgins et al., 1998). It can also be used to re-plate the successful colonies using a gradient concentration of zeocin to select only the ones with the highest number of copies.

The modifications to the pPinkNN vector to add the antibiotic resistance are described in the section 2.4.1. After the new pPinkNN_Z vector was confirmed, it was used to clone the sucrase-isomaltase proteins (NtSI, CtSI, and SI_{FL}) via homologous recombination. The pPinkNN_Z constructs were prepared following the same methodology as for pPinkNN (section 2.5.1), and the transformants were plated in solid YPD media, adenine drop-out media, YPD with 50mg/mL of zeocin, and adenine dropout media with 50mg/mL of zeocin. An example of the results is illustrated in Figure 2.5.

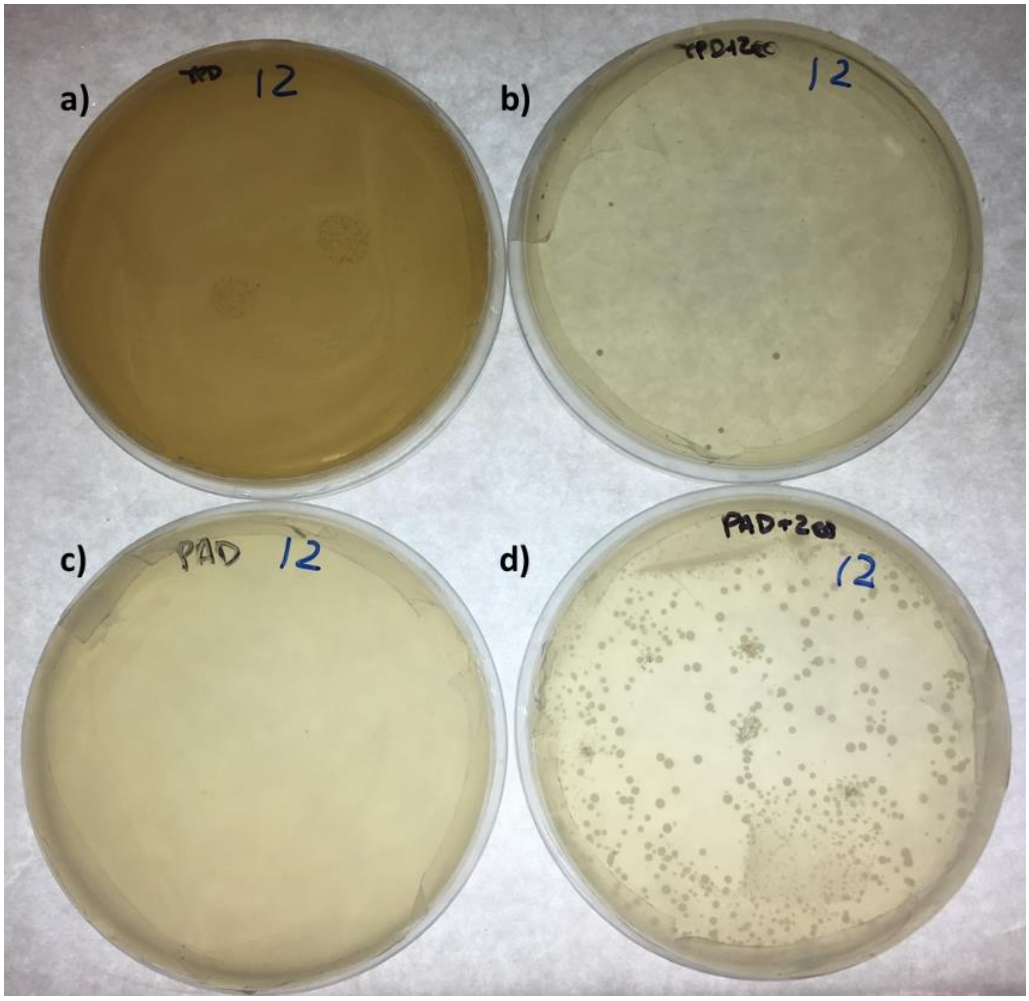


Figure 2.5 N-terminal sucrase-isomaltase construct (NtSI) plated in the different selection media. A) rich media Yeast-peptone-glucose (YPD), b) YPD with 50mg/mL of zeocin, c) adenine dropout media, and d) adenine dropout media with 50mg/mL of zeocin.

These results showed an extreme overgrowth in the media without antibiotics, as the surface of both YPD and dropout without adenine media (PAD) were completely depleted. In addition, the YPD plate showed two morphologies of colonies, one growing on top of the other, indicating possible contamination. In contrast, the plates containing zeocin showed a more limited growth, but better isolated colonies with a homogenous morphology. Notably, the plate with double selection marker had a significantly higher number of colonies than the single-marker YPD with zeocin. The double selection marker was expected to be the most challenging media for cell growth, as the cells must overcome the lack of adenine in addition to resist the presence of the antibiotic. However, this response was consistent among thirteen constructs plated in the same conditions. This phenomenon had been observed in other studies where the antibiotic does not work well in combination with the media for auxotrophy selection (Hentges, Van Driessche, Tafforeau, Vandenhaute, & Carr, 2005). Even though the two selection markers do not seem to be compatible together, they can be used sequentially, as the cells require a couple of hours to develop the antibiotic resistance. Right after electroporation, the transformed cells can be plated in adenine dropout media, and then the selected colonies re-plated in YPD using a gradient of zeocin. The colonies with higher resistance to zeocin are also the ones with the higher number of plasmid copies and therefore, the most promising for expression (Higgins et al., 1998).

The results for the growth curve for zeocin resistance are illustrated in Figure 2.6.

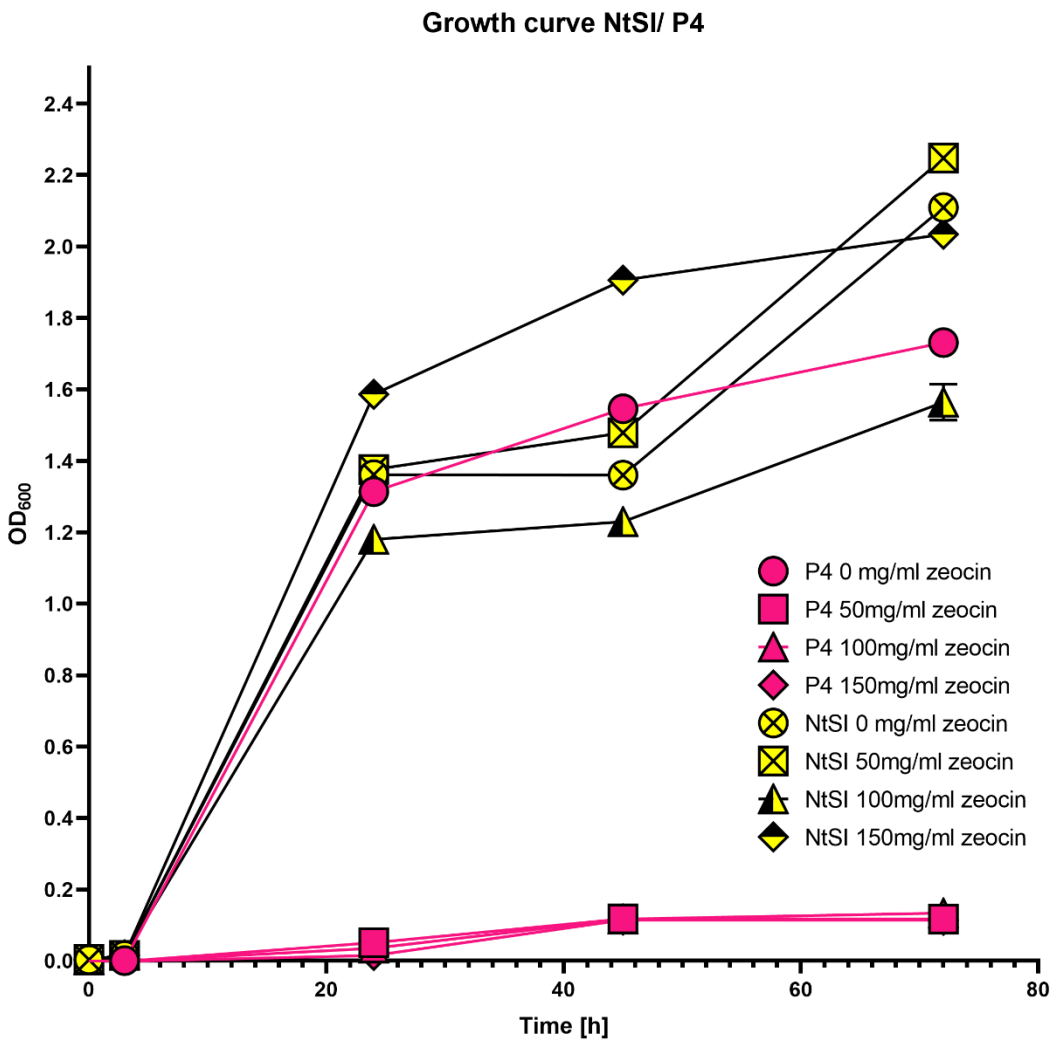


Figure 2.6 Growth curve for untransformed *Komagataella phaffii* cells (in pink) and *K. phaffii* transformed cells with NtSI_pPinkNN_Z (yellow) in BMGY media with four different concentrations of zeocin. Most of the error bars are masked by the symbols.

From these results several observations can be made:

1) Both untransformed and transformed *K. phaffii* cells can grow in BMGY media without antibiotics at a comparable rate for >44 hours.
2) The untransformed cells were unable to grow in any of the concentrations of zeocin during the first 24 hours. Between 24 and 72 hours minimal growth was observed, with

OD_{600} <0.2. This growth can be attributed to a possible decrement on the antibiotic activity over the time.

3) The transformed *K. phaffii* cells grew in all the concentrations of zeocin (50-150 mg/mL) indicating that they are resistant to the antibiotic.
4) The transformed cells showed different rates of growth in each of the concentrations of antibiotic.
5) The transformed cells with the highest concentration of antibiotic reached the highest OD_{600} after 48 hours.

The fact that the untransformed cells cannot grow in media with antibiotics proves that the antibiotic resistance is acquired from the pPinkNN_Z plasmid; therefore, only successfully transformed cells are able to grow in presence of zeocin. The difference in the growth rates may indicate that the cells need time to synthesize the protein to develop tolerance to the antibiotic which may increase in the metabolic load on the cells, slowing growth. For the transformed cells the first 24 hours seems to be crucial for the adaptation to the different concentration of antibiotics. A potential explanation for this is that within the first few hours many cells died and only the ones able to produce the high concentration of the protein necessary to bind the antibiotic and prevent the DNA damage managed to survive. This stress may have limited the growth rate initially. However, once the selection was done, the remaining cells produced offspring with more copies of the plasmid and consequently, BleorR protein. Therefore, more resources in the growth media were utilized by surviving cells allowing them a higher rate of growth than cultures with a slower rate of initial selection. These cells are also the ones with a higher number of copies of the plasmid and as consequence the most promising for protein expression.

2.5.3 Protein purification using a 12-Histidine chain as tag.

The pPinkNN vectors includes a secretion signal and consequently the protein of interest is expected to be secreted. Therefore, the first step in purification is to centrifuge cultures to separate the cells (pellet) and protein (supernatant). Recovery of the protein from the supernatant is performed by immobilized metal-affinity chromatography using the chain of six histidines cloned into the pPinkNN vector. The supernatant was combined with Nickel resin and after washing, buffers with different concentration of imidazole were tested to find the appropriate imidazole concentration for elution. All the eluted products were run in a polyacrylamide gel to identify the one with a band in the expected size of the protein or interest. Unfortunately, as can be seen in Figure 2.7, the amount of protein in the washes is comparable with the one in the main elution buffer (100mM imidazole). This indicates that the binding between protein and nickel resin was weak, reducing protein recovery from the supernatants.

To overcome this disadvantage, the length of the polyhistidine chain was increase to twelve residues. This approach has been shown to enhance the affinity histidine tag-transition metal ion and has been used successfully in other systems (Grisshammer & Tucker, 1997). This diminished affinity towards the transition metal ion is more commonly seen in non-*E. coli* systems, as they are more likely to have consecutive histidine residues within the protein sequence (Crowe et al., 1994).

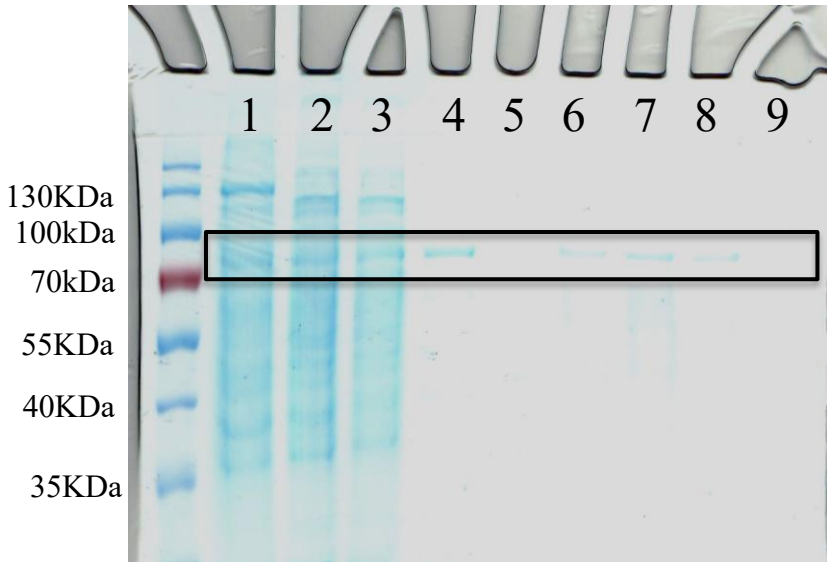


Figure 2.7 NtSI expressed in the vector with a chain of 6 histines, 1: supernatant; 2 and 3: washes (PBS+25mM Imidazole); 4: elution buffer 100mM imidazole; 5: elution buffer 150mM imidazole; 6: elution buffer 200mM Imidazole; 7: elution buffer 250mM imidazole; 8: elution buffer 300mM imidazole; 9: elution buffer 500mM imidazole.

After the addition of the extra six histidines to the expression vector, new constructs were prepared, and the purification process repeated. The polyacrylamide gel confirmed that the longer histidine chain improved the affinity between the protein and the nickel resin, as the washes showed no band at the size of the protein of interest (Figure 2.8).

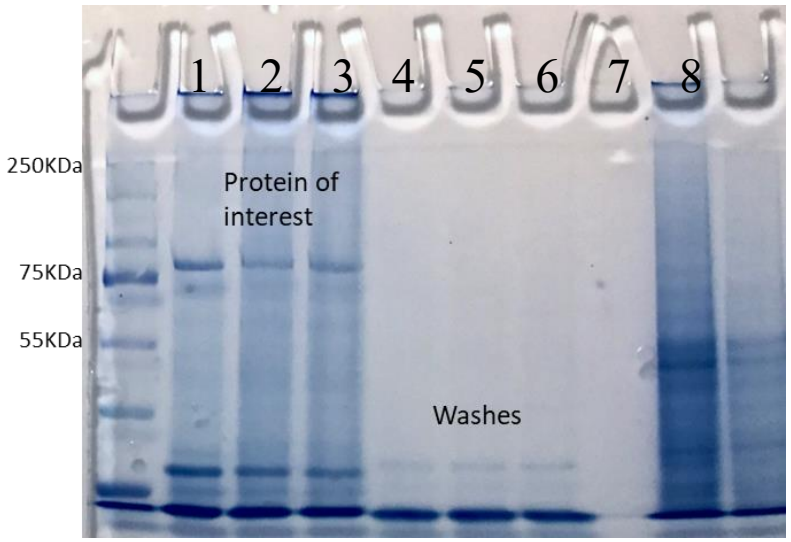


Figure 2.8. PAGE purification of NtSI expressed in the vector with a chain of 12, 1, 2, and 3: supernatant; 4, 5 and 6: washes (PBS+25mM Imidazole)

In yeast the secretion of proteins is not an efficient process; therefore, it is likely that protein remains in the pellet. However, as the protein recovered from the pellet may or not be completely folded or glycosylated, that aspect was not explored in this thesis.

2.5.4 Optimization of the conditions for protein expression using an experimental design.

Once the three constructs were confirmed by PCR and Sanger sequencing, the colonies were selected using a zeocin gradient. Those colonies were used to prepare glycerol stocks, and for a pilot expression to determine if they could produce the protein of interest. The pilot expression resulted in bands on polyacrylamide gels at the expected sizes for each of the proteins of interest. Then, those colonies were used to scale up and optimize the expression.

The expression of recombinant human alpha-glucosidases in *K. phaffii* was done in batch using flasks. For this protocol, there are two major factors to optimize to maximize yield: the length of the expression and the concentration of methanol used for the induction. As these factors may be affected by the protein of interest, they must be tested for each case. For this type of optimization, the most convenient method was to use an experimental design to generate a surface of response to predict the best combination of methanol and length of expression. The matrix of experiments, shown in Table 2.3, was constructed considering only those two factors and the recovered protein after purification as response. The same matrix was used for the three proteins SI, CtSI, and NtSI.

Table 2.3 Matrix of experiments with responses for the three enzymes.

	Factor 1	Factor 2	Response		
Run	A: methanol	B: time	SI	NtSI	CtSI
	%	hours	mg/L	mg/L	mg/L
1	0.5	120	0.25	0.553	0.228
2	25	24	0.978	0.026	0.301
3	10.3	72	1.26	0.1281	0.768
4	25	72	0.79	0.0672	0.196
5	13.8525	24	0.308	0.0325	0.504
6	20.835	120	0.7986	1.12	0.908
7	11.8925	120	4.8888	3.58475	4.0785
8	0.5	72	0.93225	0.110825	0.7024
9	19.4875	48	1.345	0.036	1.05
10	20.835	120	2.552	1.1193	2.88
11	25	72	2.94	0.2241	2.405
12	10.3	72	2.32	1.022	3.232
13	5.4	48	1.336	0.0575	0.686
14	0.5	24	1.295	0.026	0.903
15	5.4	96	3.9875	0.292	2.9
16	10.3	72	3.06065	0.3669	3.285

The surface of response obtained for all three proteins was very similar (Figure 2.9), indicating maximum yield with 13-15% of methanol and a length of expression of about 120 hours.

The surface of response corresponded to a polynomial model. This is not ideal as it indicates a poor correlation between the response and the two factors evaluated. However, this is not surprising considering that this is a complex system with many independent and interacting factors, and that the response comes from an indirect measurement.

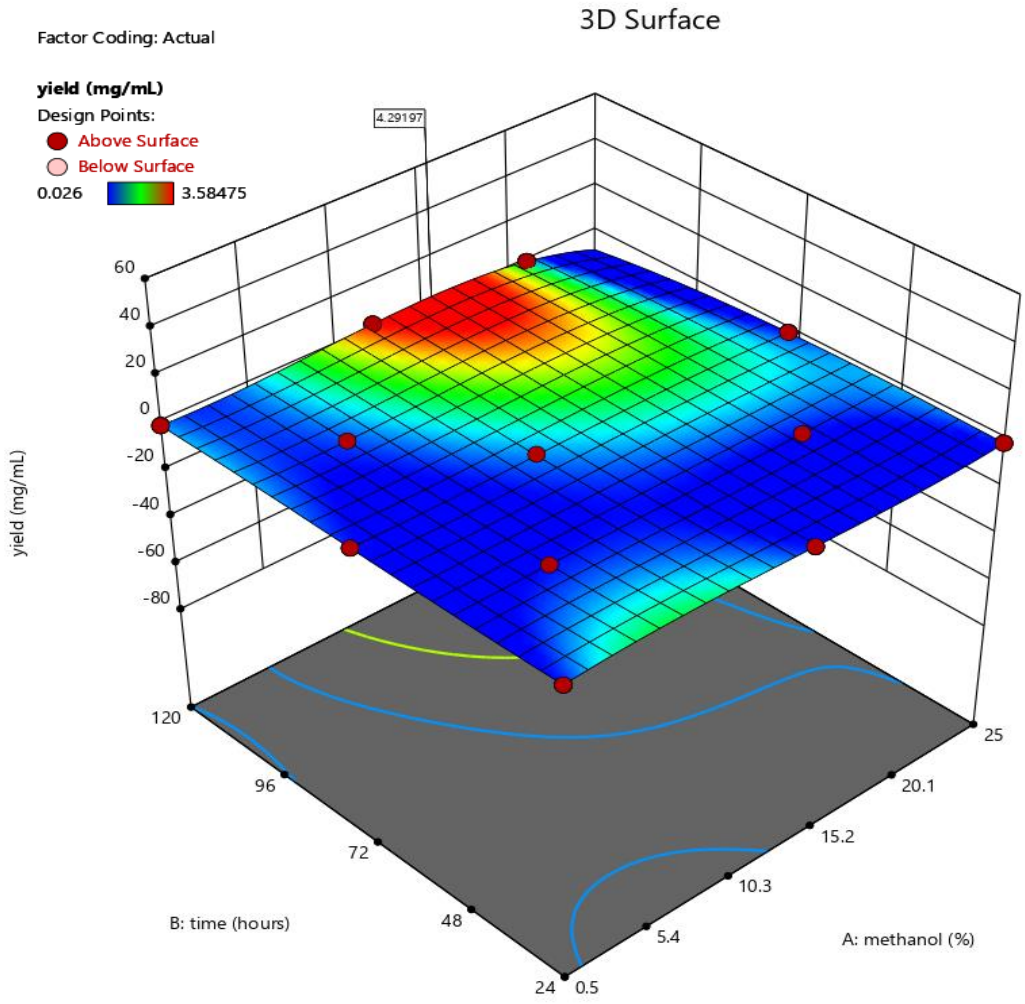


Figure 2.9 Surface of response obtained for NtSI. Design Xpert version 12, Stat-Easy, Inc.

Table 2.4 Predicted optimum conditions for NtSI, CtSI and SI_{FL} based in the surface of response generated with a matrix of 16 experiments.

Optimum conditions			
Enzyme	**Methanol (%)**	**Time (hours)**	**Yield mg/L**
CtSI	14.908	112.019	1.246
NtSI	15.217	120	4.292
SI	13.626	120	5.074

Using the model, conditions for the maximum yield of expression were predicted as shown in Table 2.4 for each of the proteins. Despite the ambiguity of the model, the conditions seem to be quite consistent among the three proteins. This may indicate that the multiple modifications made to the vector may have a bigger impact over the condition for the expression than the proteins, themselves (Chen et al., 2005; Karim, Curran, & Alper, 2013).

2.6. Conclusions

1. The developed expression system in *Komagataella phaffii* can successfully express human alpha-glucosidases: sucrase-isomaltase full length, its two individual subunits (NtSI and CtSI).
2. The utilization of the adenine auxotrophy and zeocin resistance double selection marker successfully discriminates between cells with and without the insert, preventing the growth of untransformed cells and diminishing the growth of positive transformed cells with low copies of the plasmid as the concentration of the antibiotic (zeocin) increased.
3. The implementation of a longer histidine chain improved the protein-resin affinity, helping to overcome the interferences present in the culture and allowing higher yields of protein to be recovered (~4mg/L).

4. Increased concentration of methanol in the culture, as a result of the optimization, enhanced the protein production, increasing the yield up to ~10% in comparison with the yield obtained adding 0.5% of methanol as suggested in previous protocols.

Chapter 3. Expression and characterization of intestinal alpha-glucosidases: Sucrase-isomaltase, maltase-glucoamylase, and lactose- phlorizin hydrolase.

Chapter 3. 3.1 Chapter overview

As a yeast, *K. phaffii* belongs to the simplest eukaryotic system. Therefore, its capability for expressing enzymatically active human proteins needs to be confirmed. In Chapter II, the engineered expression system was tested for successful cloning and expression of human alpha-glucosidases. In this chapter, the enzymatic activity and size of the expressed proteins were tested and compared with their human counterparts, as a preliminary indication of their correct assembly and post-translational modification.

For these tests, eight alpha-glucosidases were successfully expressed in *K. phaffii* using the developed system: SI_{FL}, NtSI, CtSI, NtMGAM, CtMGAM, LPH III, and LPH III, IV, introduced in Chapter 1. The size of each of those proteins was determined by electrophoresis running a polyacrylamide gel. The results showed that in all the cases the bands are at least 15KDa bigger that the predicted size based on their amino acid sequences. This is an indicator that the studied proteins are likely glycosylated. From the enzymatic activity test, the substrate used as reference for comparison was maltose, as it that activity is conserved across most of the alpha-glucosidases studied in this chapter. However, the substrate affinity was also tested for some enzymes and discussed in the corresponding results section. With maltose as substrate, the values of K_M fell

into the range of K_M previously reported in BRENDA data base. This supports the proposal that the human enzymes expressed in *K. phaffii* conserve their catalytic features.

Further examination of the recombinant enzymes would confirm the correct folding of the proteins, the type and level of glycosylation, as well as their behavior with different and more complex substrates.

3.2 Objectives

General:

Successfully express human alpha-glucosidases in concentrations that allowed for their study and characterization using the expression system developed for *K. phaffii*., in terms of enzymatic activity, protein size, and DNA sequence.

Specific:

1. Test the capability of the developed *K. phaffii* expression system to successfully express different recombinant alpha-glucosidases from human genes.
2. To express sufficient amounts of human recombinant proteins in *K. phaffii* to allow their investigation and characterisation.
3. Confirm the correct expression of the selected human alpha-glucosidases by size comparison between the expressed proteins, the predicted size according to their amino acid sequence, and previous observations in immunoprecipitated enzymes from human intestine biopsies.

4. Determine and compare the enzymatic kinetic parameters for the expressed enzymes using simple carbohydrates with the reported values for the same enzymes expressed in different host organisms.

3.4 Materials and methods

3.4.1 Cloning and Expression in *K. phaffii*

All the proteins studied in this section were cloned into the pPinkNN_Z expression vector via homologous recombination in *S. cerevisiae* as explained in Section 2.2.3 and further transformed into *K. phaffii* (strain P4, Invitrogen Inc.) by electroporation. All the sequences were confirmed through Sanger sequencing and further alignment with the corresponding sequences for each gene (Table 3.1).

3.4.2 Protein purification

The first step in the purification was to centrifuge the cultures at 3000xg for 15 min to separate the cells from the supernatant, which should be clear and free of turbidity. The pellets were then frozen at -80 C and kept for further lysis.

Table 3.1 Proteins expressed in *K. phaffii* using the pPinkNN_Z expression vector and the corresponding gene

Protein	Abbreviation	NCBI accession	EC number
Sucrase-isomaltase (Full length)	SI_{FL}	NM_001041.4 From 392 to 5548	3.2.1.10/48
Isomaltase subunit of SI	NtSI	NM_001041.4 From 392 to 3085	3.2.1.10
Sucrase subunit of SI	CtSI	NM_001041.4 From 3086 to 5548	3.2.1.48
Maltase subunit of maltase-glucoamylase (MGAM)	NtMGAM	NM_004668. From 157 to 3135	3.2.1.20
Glucoamylase subunit of MGAM	CtMGAM	NM_004668.2 From 3136 to 5625	3.2.1.20
Region III of Lactase Phlorizin Hydrolase (LPH)	LPH III	NM_002299.4 From 2620 to 4060	3.2.1.23 & 3.2.1.62
Regions III and IV of LPH associated	LPH III, IV	NM_002299.4 From 4061 to 5620	3.2.1.23 & 3.2.1.62

The pPinkNN_Z expression vector contains a 12 histidine chain for nickel affinity; therefore, the supernatants were mixed with NiNTA resin (Thermo Fisher Scientific, Cat. No. PI88222) and an equal volume of equilibration buffer (PBS with 10 at pH 7.4mM Imidazole) shaken in a platform rocker for 1 hour at 20rpm to facilitate the binding between protein and resin. Then the mixture was transferred to a gravity flow column where all the supernatant was eluted, the resin washed with 25mM imidazole buffer, and the protein eluted with the proper buffer as illustrated in the Table 3.2.

Once eluted, the proteins were concentrated using Amicon tubes in the appropriate pore size according to the provided instructions and the size of the protein of interest. In addition to protein concentration, a buffer exchange was also performed at the end of the concentration, rinsing the imidazole buffer three times with PBS pH 6.4 and recovering the protein with PBS.

Table 3.2 Protein purified via Nickel affinity in a gravity flow column and the imidazole buffer used for its elution.

Protein	Elution buffer
SI_{FL}	100mM Imidazole, 300mM NaCl and 20mM NaH_2PO_4; pH 7.4
NtSI	100mM Imidazole, 300mM NaCl and 20mM NaH_2PO_4; pH 7.4
CtSI	100mM Imidazole, 300mM NaCl and 20mM NaH_2PO_4; pH 7.4
NtMGAM	100mM Imidazole, 300mM NaCl and 20mM NaH_2PO_4; pH 7.4
CtMGAM	250mM Imidazole, 300mM NaCl and 20mM NaH_2PO_4; pH 7.4
LPH III	100mM Imidazole, 300mM NaCl and 20mM NaH_2PO_4; pH 7.4
LPH III,IV	100mM Imidazole, 300mM NaCl and 20mM NaH_2PO_4; pH 7.4

3.4.3 Polyacrylamide gel electrophoresis

Electrophoresis was used to confirm the presence and purity of the different proteins. The percentage of polyacrylamide gels (SDS-PAGE) was selected according to the expected size of each protein (Table 3.3). This technique was used throughout the process at critical points, for example to monitor the protein expression during the induction, at the end of the expression before and after centrifugation, during the purification to analyze the different fractions eluted with the imidazole buffers, and after concentration to confirm the purity and presence of the protein of interest.

All polyacrylamide gels were made using polyacrylamide 30 (BioRad, cat. No. 1610158) cast in 1mm glass trays, loaded with 24 µL: 20uL sample and 4 µL 6x SDS dye after heated at 95C for

5 min. The chamber was filled with SDS buffer. The gels were run at 80 V for the stacking and 120V for separation.

Table 3.3 Protein expressed in *K. phaffii* with the expected size calculated based in their amino acid sequence calculated by ExPASY, SIB Swiss Institute of Bioinformatics (https://web.expasy.org/compute_pi/).

Protein	Expected size (KDa)
SI_{FL}	197
NtSI	102
CtSI	95
NtMGAM	111
CtMGAM	95
LPH III	55
LPH III, IV	115

3.4.4 Determination of the enzymatic activity: PGO (Peroxidase Glucose- oxidase method).

All the proteins studied in this section have the ability to hydrolyse at least one substrate to glucose (Table 3.4). Therefore, the peroxidase glucose-oxidase method (Quezada-Calvillo, Robayo-Torres, Ao, et al., 2007) was suitable to test enzymatic activity. PGO is a colorimetric technique based in a series of redox equilibria among the glucose released and the different chemical species present in the reagents, leading to a change in the O-dianosidine from a reduced state to an oxidized one; this causes a change in the color from colorless to brown-orange. The change in the color can be related to the glucose released by measuring the change in the absorbance at 450nm and making a correlation with a glucose standard curve following the Beer-Lambert law.

First, a standard curve was build using glucose at 400, 200, 100, and 50 µg/ml in PBS (pH 6.4). The absorbance values were analyzed taking only the linear section and adjusting the valid range of measure. Then, a maltose stock of 50mM (maltose/PBS) were used to prepare serial dilutions. 10µL of each enzyme were tested adding 10µL of the different concentrations of maltose and 180µL of PGO reagent. The absorbance was measured every 2 min for one hour at 37 C and with a 2s shake before each read. Based on the absorbance readings, the optimal range of maltose concentration (abs >0.1 and < 1 at 450nm) were determined for each enzyme and used to evaluate K_M.

3.5 Results

After running each of the proteins in a polyacrylamide gel, the size of the bands was compared, first against the predicted size according to the amino acid sequence and second, against the reported values in the BRENDA protein data base (Tables 3.4 and 3.5). The results obtained from the gels (Tables 3.4 and 3.5) showed that in all the cases, the size of the protein was higher than the predicted value but comparable with the ones reported for human enzymes. These observations indicate that the expressed enzymes may be glycosylated.

Glycosylation is a desirable modification for the alpha-glucosidases, as it may be related to the substrate affinity, enzymatic activity, and proper folding; however, it is well known that *K. phaffii* does not possess the complex glycosylation machinery that mammalian systems does. Consequently, only high mannose residues can be attached to the proteins (Bretthauer & Castellino, 1999). The effect of the glycosylation pattern over the enzyme activity and folding are unknown. Furthermore, it has been observed that other yeast systems such as *S. cerevisiae*

may produce hyperglycosylated proteins. In *K. phaffii* this is in general an uncommon output, and for this specific set of proteins, very unlikely as the size of the bands only differs by about 15 to 30 KDa from the predicted from the amino acid sequence. Nevertheless, the possibility of changes in the glycosylation sites should also be considered.

Table 3.4 Summary of the results obtained from the characterisation of the recombinant human alpha-glucosidases expressed in *K. phaffii* in comparison with the values previously reported in the literature.

Enzyme expressed in *K. phaffii*	**Size exp. KDa**	**Size obs. KDa**	**K_M Maltose (mM) exp.**	**K_M Maltose (mM) obs.**
SI_{FL}	197	230	n.a.	15
NtSI (isomaltase)	102	120	6.83-7.1	6.5
CtSI (sucrase)	95	120	n.a.	12.1
NtMGAM (glam)	111	130	4.3- 7.71	3
CtMGAM (mal)	95	130	n.a.	2.7

Table 3.5 Summary of the results obtained from the characterisation of the recombinant human Lactase-phlorizin hydrolase expressed in *K. Phaffii* in comparison with the values previously reported in the literature.

Enzyme expressed in *K. phaffii*	**Size exp. KDa**	**Size obs. KDa**	**K_M lactose (mM) exp**	**K_M lactose (mM) obs.**	**K_M phlorizin (mM) exp.**	**K_M phlorizin (mM) obs.**
LPH III	55	75	21-30*	N/D	0.44*	15
LPH III, IV	115	130	21-30*	12	0.44*	6.9

From the seven enzymes expressed in *K. phaffii*, five of them are variants of sucrase-isomaltase and maltase-glucoamylase; all of them expected to have maltase activity. Therefore, maltose was used as the substrate of reference not only to compare each enzyme with their human-counterpart but also among them. Reference K_M values for human enzymes were taken from BRENDA

protein data base and reported in Table 3.4. These five proteins expressed in *K. phaffii* showed maltase activity (Table 3.4), and the observed K_M values are comparable with those previously observed in human enzymes. However, this comparison may or not be representative as the information available for human enzymes is limited. The specific results are more extensively discussed in the following sections.

3.5.1 Sucrase-isomaltase

Three forms of sucrase-isomaltase were expressed and studied in this thesis: Sucrase-isomaltase full length (SI_{FL}) which includes both subunits, sucrase and isomaltase, and the two independent subunits, CtSI and NtSI, respectively. Naturally, the enzyme is expressed and folded with the two subunits joined and interacting, but they are cleaved after sorting to the apical membrane; however, they maintain strong non-covalent interactions. Consequently, it is uncertain what to expect from the two joined subunits before cleavage. On the other hand, the expression of each subunit completely independent from the other raises questions about possible misfolding. In previous studies, the independent subunits have been successfully expressed and shown promising enzymatic activity and substrate affinity.

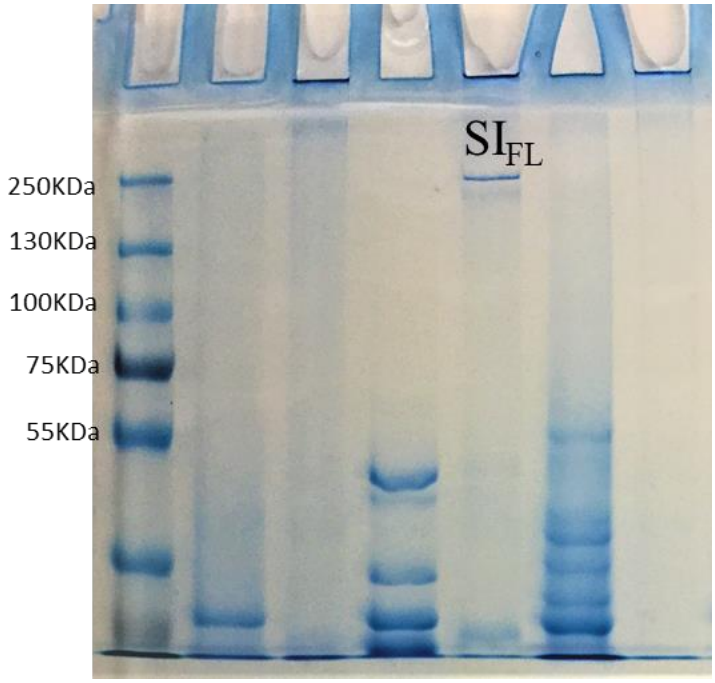


Figure 3.1. 8% SDS-PAGE showing sucrase-isomaltase full length in the fourth lane after purification by gravity flow chromatography using Nickel resin affinity, eluted with 100mM imidazole elution buffer.

The polyacrylamide gel (Figure 3.1) showed SI_{FL} as a well-defined band slightly below 250 KDa after the purification with nickel resin and elution. The lack of contaminant bands demonstrates a clear discrimination between proteins with and without histidine tag. In addition, this level of purity discards the possibility of other proteins or fragments interfering with the enzymatic activity assays.

Another interesting observation regarding SI_{FL} was its possible self-cleavage over time (Figure 3.2). The purified protein was left at 4ºC for about two weeks, then a sample was run in an SDS-PAGE under the same conditions as before. It was observed that the band at 230KDa were diminished and a strong band about 75KDa appears, suggesting that cleavage can occur in that timeframe.

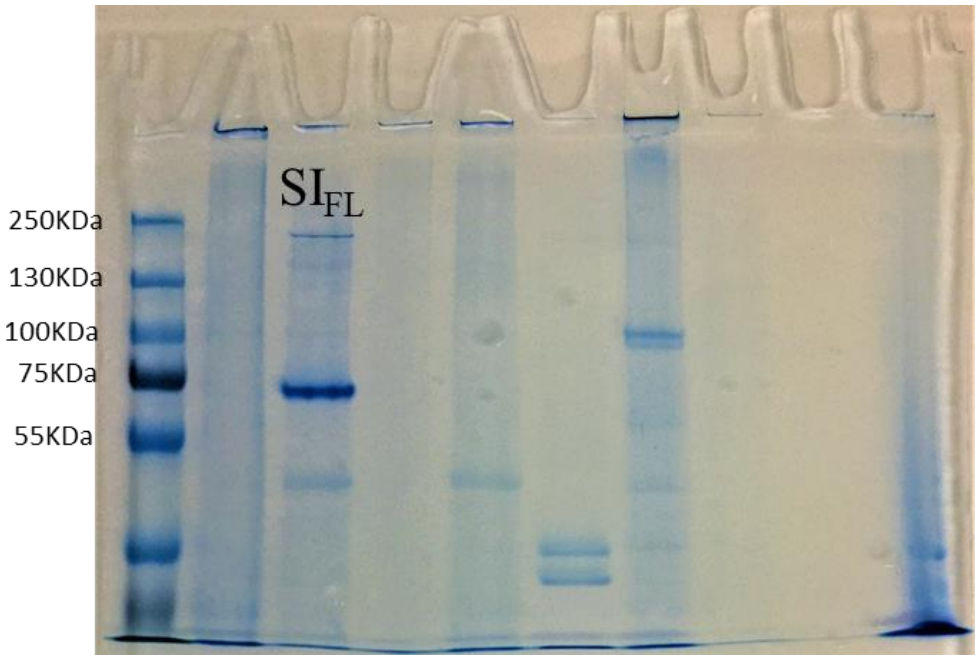


Figure 3.2. 8% SDS-PAGE showing purified sucrase-isomaltase full length in the second lane after incubation at 4C for two weeks.

The enzymatic activity assays reveal an interesting behaviour: the independent subunits showed higher maltase activity than the full length. This phenomenon may be circumstantial considering the size and complexity of sucrase-isomaltase full length. For example, it is not known whether the uncleaved protein is less active against maltose as result of interactions between the domains.

Regarding NtSI, the SDS-PAGE displayed a band above 100KDa and a smaller band under 55KDa (Figure 3.3); however, this second band was only observed occasionally in the elution buffer, but most commonly came off the during the washes. This suggest that the 55kDa band may belong to contaminant protein.

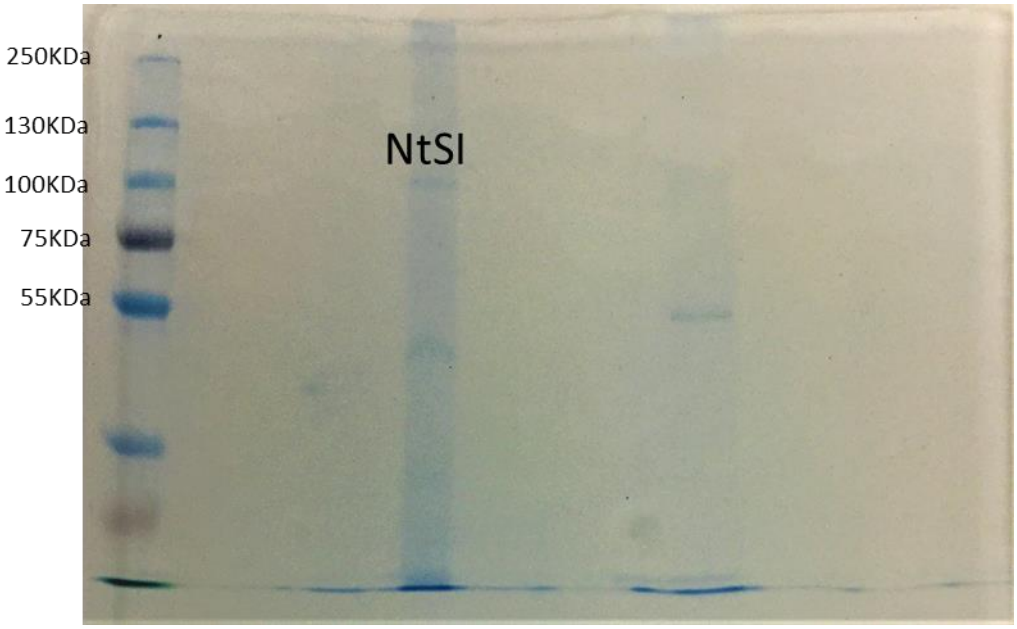


Figure 3.3. 8% SDS-PAGE gel with NtSI (isomaltase) in the third well and wash in the sixth well.

The maltose K_M value obtained for NtSI falls well within the range of values reported for the immunoprecipitated enzyme from human intestine (Table 3.4), as well as values observed in other recombinant systems. Nevertheless, it was expected from the literature that NtSI would have lower maltose activity than CtSI. However, the maltose K_M for CtSI was about double that for NtSI. The explanation for this may be in the polyacrylamide gel Figure 3.4. In addition to the CtSI band at 120kD, there is a contaminant band at a much higher molecular weight in the elution samples. If this band represents a protein with maltase activity, such as a CtSI dimer, SI_{FL} or NtSI from a contaminating construct, that could affect the activity measurement.

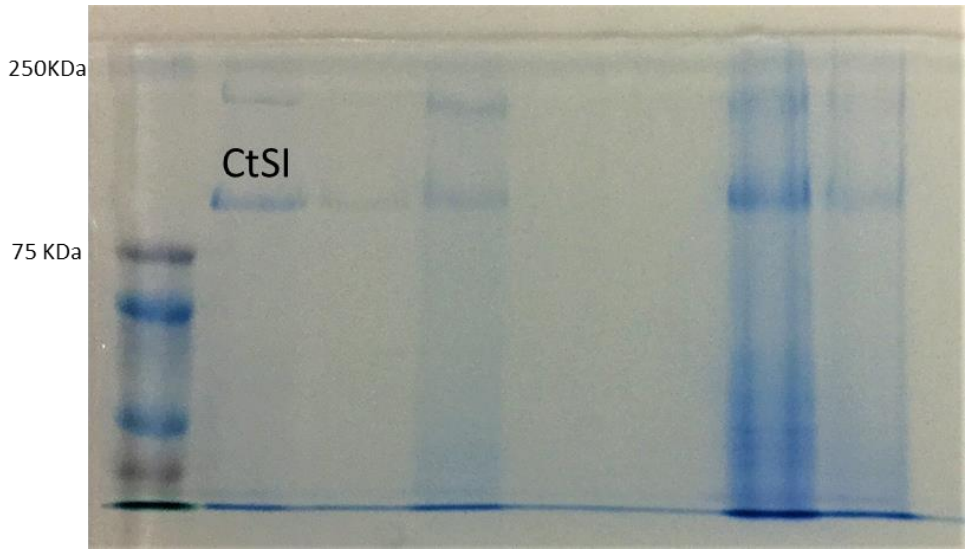


Figure 3.4. 8% SDS-PAGE showing the elutions of CtSI in all the wells after purification by gravity flow chromatography using Nickel resin affinity, the protein was eluted with 100mM imidazole elution buffer.

These results were only seen in the CtSI expressed using the latest version of the expression vector (pPinkNN_Z), as previous expressions had shown a single band with higher maltase activity. Therefore, it is unlikely that these results are due to a flaw on the expression system but more likely a possible contamination of the constructs. Further testing of the glycerol stocks and a new expression may shed some light on this issue.

3.5.2 Maltase-glucoamylase

The two subunits of maltase-glucoamylase, NtMGAM and CtMGAM were cloned and expressed as independent constructs in one of the previous versions of the expression vector pPinkNN. This vector has a six-histidine tag, and this affected significantly the efficiency of the purification. The maltase activity of these proteins was evaluated and reported in Table 3.4. and the details about the expression can be find in Appendix C. Currently new constructs for NtMGAM and CtMAGA had been cloned into pPinkNN_Z and are been tested for expression.

3.5.3 Lactase phloridizine hydrolase (LPH)

This is perhaps the first recombinant version of human LPH expressed in a foreign organism with both phlorizin and lactose hydrolytic activity. Previous studies have been made in COS-1 cells expressing fragments and truncated versions of the enzyme, but it has never before been highly expressed and isolated.

LPH is a large protein with complex assembly and post-translational modifications that not only include N- and O-glycosylations, but also a series of trimmings and re-assortments. To achieve the expression of an active form of LPH III,IV , first the expression of LPH region III was tested. The result was a 75kDA protein after glycosylation and a deglycosylated form of about 55KDa (Figure 3.7).

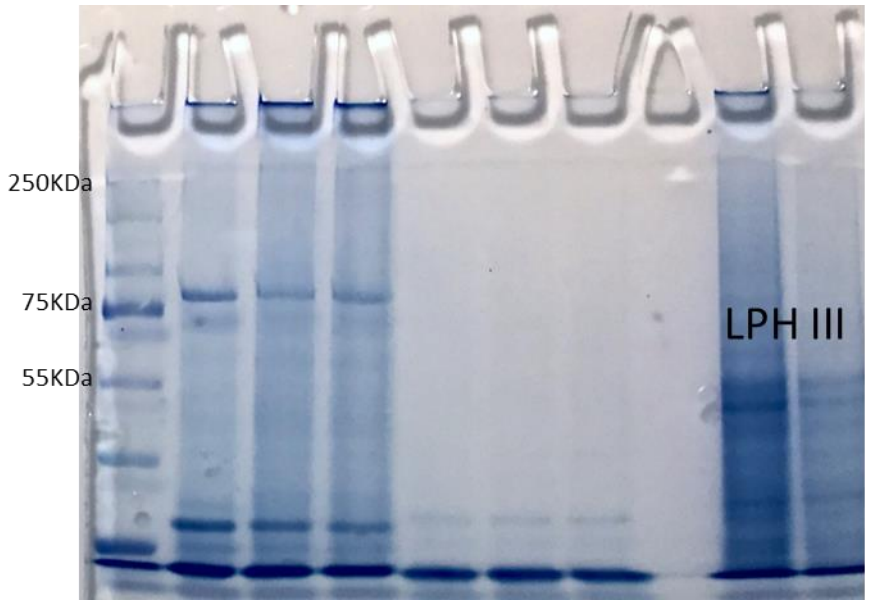


Figure 3.7. 8% SDS-PAGE showing the elutions of LPH III (last two lanes) after purification by gravity flow chromatography using Nickel resin affinity, eluted with 100mM imidazole elution buffer.

The enzymatic activity of LPH III was heavily diminished in comparison with the LPH_f immunoprecipitated from human biopsies (Appendix B); however, this result was consistent with the ones observed by Dr. Naim Hassan's group (Diekmann, Behrendt, Amiri, & Naim, 2017), as LPH III barely shows 4% of the phlorizin hydrolitic activity and no lactase activity.

The success at expressing LPH III allowed the second step, the expression of LPH regions III and IV. Region IV of LPH possess the only lactase active site of this protein, but it is transport incompetent and requires at least region III for proper folding (Gericke, Schecker, Amiri, & Naim, 2017); therefore, expressing regions III and IV together presents the opportunity of obtaining a form of the enzyme with the two active sites. Furthermore, while LPH has four regions, regions I and II are considered intramolecular chaperones and are trimmed before the protein leaves the ER; therefore, expressing regions III and IV may be the most representative version of LPH_f.

LPH III, IV was cloned and expressed in *K. phaffii* using the developed system, obtaining a band in a PAGE about 130 KDa, potentially glycosylated, as the predicted size was 115KDa (Figure 3.8). The expressed LPH III, IV showed enzymatic activity against the two substrates phlorin and lactase. LPH has two active sites for the hydrolysis of phlorizin, one in the region III and the other in region IV. This is one of the potential reasons why the LPH III protein showed reduced phlorizin activity. On the other hand, the phlorizin K_M of LPH III, IV were about half of the K_M for LPH III; however this is still more than 10-fold higher than that observed in the immunoprecipitated protein from human biopsies. For lactose, LPH III did not show any detectable actvity, as expected. For LPH III, IV, the K_M for lactose was approximately half of

that previously reported on BRENDA (Table 3.5). This indicates that the expressed form of LPH III, IV was even more active than the one immunoprecipitated from human intestine. This can be attributed to different reasons from the characteristics of the biopsy sample, its management, immunoprecipitation, and so on, to the differences between the recombinant LPH expressed in *K. phaffii* and the native form.

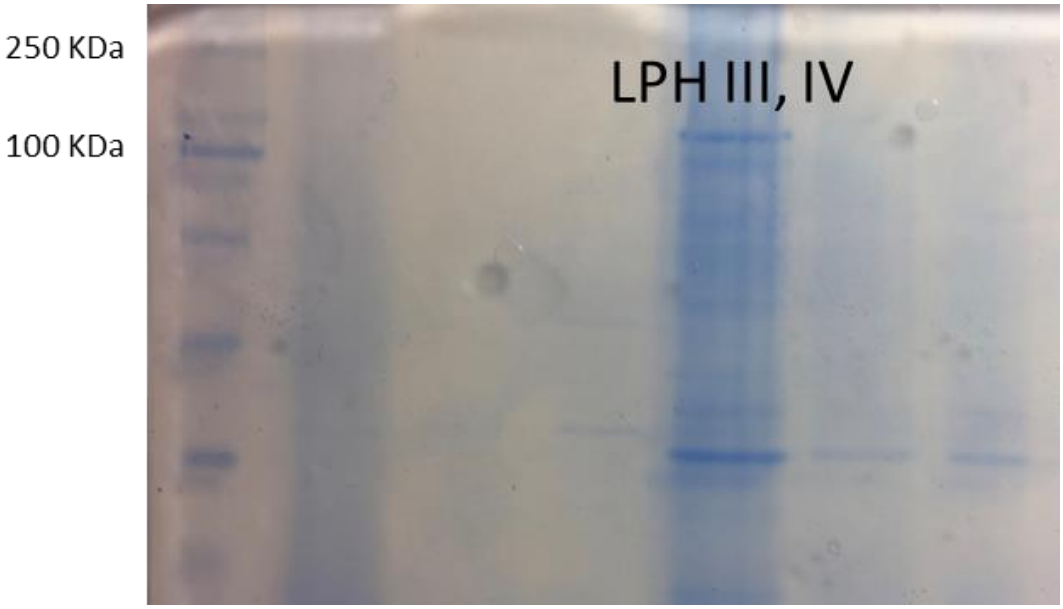


Figure 3.8. 8% SDS-PAGE showing the eluted LPH III, IV after purification.

3.6 Conclusions

From the expression of eight human intestinal alpha-glucosidases in the developed *K. phaffii* expression system I conclude:

1. The developed system is suitable for a quick cloning (~ one week) and proper expression of different recombinant alpha-glucosidases from human. Eight out of eight attempted

human enzymes were successfully expressed, enzymatically active, glycosylated, and in the expected sizes.

2. All the expressed proteins showed bands equal to, or larger than, the expected sizes, consistent with glycosylation of the proteins and comparable with the sizes observed with immunoprecipitated human enzymes during SDS-PAGE.
3. The K_M values observed for the enzymes expressed in *K. phaffii* fell into the range reported for the correspondent enzyme in different systems. In comparison with human enzymes isolated via immunoprecipitation, the obtained K_M values were slightly different, but these differences can be attributed to the differences in the glycosylation mechanisms between yeast and mammalian cells; however, further research is required to fully understand the differences.

Chapter 4. Inhibition studies

4.1 Chapter overview

In this chapter the analysis of three iminosugars is described. The compounds were synthesized by Dr. Sandrine Py's group at the Université Grenoble Alpes, France. The three compounds tested are castanospermine (CST) analogues. Compounds 1 and 3 are indolizines (D-gluco- and L-ido-, respectively) and compound 2 is a D-gluco-quinolizidine Figure 4.1.

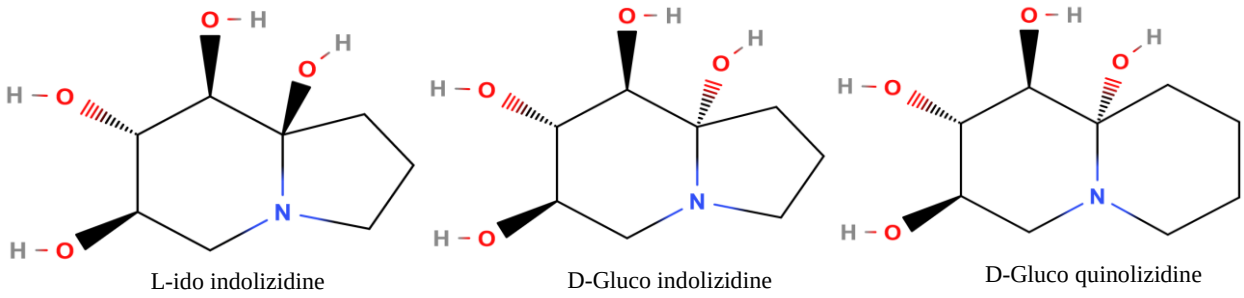


Figure 4.1 Chemical structure of the castanospermine analogues synthesized by Dr. Sandrine Py's group.

Two enzymes were used to test the inhibitory activity of these compounds: human C and N terminal units of Sucrase-isomaltase. The inhibitory activity was tested using the Peroxidase, glucose-oxidase method (Quezada-Calvillo, Robayo-Torres, Opekun, et al., 2007) using maltose as substrate. To determine the type of inhibition and kinetic parameters the results were analyzed using Sigma Plot 14 (Systat Software Inc. San Jose, CA), enzyme kinetics module.

For the D-gluco quinolizidine, the results pointed towards either a mixed or competitive type of inhibition with K_i < 1µM for both C and N terminal subunits. For the L-ido indolizidine, the subunit N-terminal showed competitive inhibition, however, the low R^2 value revealed an important lack of fit (R^2= 0.735). The C-terminal, inhibition displayed a mixed or competitive

model of inhibition with a similar lack of fit (R^2= 0.740). Therefore, the results were interpreted as inconclusive. The K_i values obtained with L-ido indolizidine are over 10 times higher for all the enzymes tested in comparison with the results with the D-gluco indolizidine. This indicates that the orientation of the hydroxylic group plays a relevant role in the enzyme-inhibitor affinity. This last compound, D-gluco indolizidine clearly showed mixed inhibition.

4.2 Chapter objectives

General

The aim of this section was to evaluate the effectiveness and mechanism of inhibition of modified versions of castanospermine (CST) in the inhibition of the two subunits of human sucrase-isomaltase expressed in *K. phaffii*.

Specific Aims

Perform a series of experiments to test the castanospermine analogues for inhibitory activity against the two different subunits of human sucrase- isomaltase expressed in *Komagataella phaffii*.

Determine the type of inhibition and inhibitory parameters for each of the compounds tested with the two different enzymes.

Identify the possible effect of the specific modifications (size of the imino ring and orientation of the hydroxylic group) in the castanospermine analogues in the inhibitory parameters, enzyme selectivity and type of inhibition.

4.3 Methodology

4.3.1. Sample preparation

Inhibitors

The inhibitory compounds were received as powder in glass vials. All of them were solubilized in 100 µL of MilliQ water to a stock concentration. For each compound a series of preliminary assays were performed to determine the appropriate range of working concentrations. The main stocks were preserved frozen at -20ºC. Meanwhile the working dilutions were kept at 4ºC to avoid multiple freezing- thawing cycles.

Enzymes

The human enzymes were expressed in *K. phaffii* according to the method reported in Chapter 2. The proteins were purified from the supernatants using Nickel resin (Thermo Fisher Cat. No. IP88222 and elution buffer containing 150mM imidazole. The eluted proteins were concentrated using 15mL Amicon concentrator tubes (Millipore cat no. UFC905024 and UFC505024). After concentration the imidazole buffer was exchanged for phosphate buffer, pH 6.4. The enzymes were aliquoted and conserved at -20ºC. Each aliquot only had one cycle of freezing/ thaw.

Maltose (Sigma Aldrich Cat. No. N9171-100G) was used as substrate for all assays. A stock solution of 1mM was prepared using PBS pH 6.4 as solvent, filter sterilized using a 0.22 µm vacuum filter and handled under sterile conditions. The main stock was used to prepare serial dilutions (Maltose/PBS).

4.3.2. Inhibition assays: Peroxidase-Glucose Oxidase method (PGO).

The inhibition activity of each compound was evaluated using the peroxidase, glucose-oxidase method (PGO) as specified in (Quezada-Calvillo, Robayo-Torres, Opekun, et al., 2007). 10µL of maltose were loaded followed by 180 µL of PGO reagent: 1mg/ml Peroxidase (Sigma cat. No. P-8250), 10mg/ml units of glucose-oxidase (Sigma, cat. No. G2133), 10mg/ml O-dianosidine (Sigma cat. No. D-3252) onto 96-well microplates (Greiner Bio-one Cat. No. 655101). Plates were mixed for 20s at low speed in the spectrophotometer (Spectramax PLUS 384, Marshall Scientific) and incubated at 37ºC for 10 min. After incubation 10 µL of the appropriate inhibitor dilution as well as 10 µL of maltose (at different concentrations) were added. The plates were read for 60 minutes at 2 minutes intervals at 450nm in the spectrophotometer with temperature control at 37ºC and mixing before every read for 5 seconds.

Each condition was run in triplicate. A blank of each enzyme, substrate and inhibitor were run in the same plate as well a standard glucose curve. The first read was considered as blank and subtracted from each absorbance value.

4.3.3. Data analysis

The data were collected as a .pda file and exported as a text file. The data preparation was done in Excel (Microsoft, Redmond, WA). This preparation consisted of sorting the results and subtracting the correspondent blanks.

The absorbance values from the glucose standard curve were analyzed and the linear section was used as a model to correlate absorbance values with glucose concentration. The correlation

equation was then used to transform the absorbance data collected from the PGO assays. The transformed data were exported to Sigma Plot 14 (Systat Software Inc. San Jose, CA) and analyzed under the Enzyme kinetics module using a non-linear correlation analysis.

4.4. Results and discussion

4.3.2 Inhibition of human Nt-SI and Ct-SI with castanospermine analogues.

As was discussed in the introduction, castanospermine has a natural inhibitory activity against alpha- and beta- glucosidases; however, it is unable to differentiate among them. This lack of selectiveness makes castanospermine unsuitable for therapeutic applications as it may interfere with other mechanisms in the organism (Boisson et al., 2015). To overcome this disadvantage, different castanospermine analogues have been synthesized.

The compounds in this study were generated, firstly, by the addition of a second ring to the piperidine cycle of castanospermine; this results in a reduction in the conformational mobility of the structure and facilitates the adjustment of the substituent orientation (Borges de Melo, da Silveira Gomes, & Carvalho, 2006). Secondly, a quaternary chirality center was built to allow another substituent in the iminosugar. The presence of this substituent was expected to improve the affinity of the castanospermine analogues to the active site of specific glucosidases (H Furneaux et al., 1997). Following this rational, three compounds were synthesized: L-ido indolizidine, D-gluco indolizidine and D-gluco quinolizidine (Figure 4.1)

The K_i values and type of inhibition for maltase activity for each enzyme/compound combination are reported in Table 4.1. For both enzymes, maltase activity was strongly inhibited by D-gluco indolizidine, with a mixed mode of inhibition. On the other hand, L-ido indolizidine, with high Ki values, showed low affinity to the two subunits of sucrase-isomaltase and the type of inhibition remains unclear. The most interesting results were obtained with D-gluco quinolizidine, which not only showed low Ki values for both NtSI and CtSI, but also selectivity between them. However, the results for the type of inhibition are not conclusive due to the similar R^2 for more than one model.

Table 4.1 Castanospermine analogues inhibitory effects on human sucrase-isomaltase C and N terminal subunits.

		NtSI	**CtSI**
L-ido indolizidine	Ki (μM)	2.6±0.9	9.4±5.4
	Type of inhibition	Mixed/Competitive R^2= 0.735	Mixed/Competitive 0.740
D-Gluco indolizidine	Ki (μM)	0.1±0.01 μM	0.1±0.01
	Type of inhibition	Mixed R^2= 0.997	Mixed R^2= 0.997
D-Gluco quinolizidine	Ki (μM)	0.7±0.07 (mixed) 0.7±0.04 (competitive)	0.004±0.004 (Mixed) 0.004±0.001 (competitive)
	Type of inhibition	Mixed/Competitive R^2= 0.987	Mixed/Competitive R^2=0.999

According to the results obtained D-gluco compounds seem to have high affinity towards Sucrase-isomaltase displaying a K_i <0.7μM for NtSI and as low as 0.004 for CtSI. These results are comparable with the ones observed with Kotalanol (thiosugar), a compound that also shares the same chirality of the substituent groups in the pyranose ring, in the case of kotalanol, a thiopyranose ring. In addition, the structure of kotalanol also has a hydroxylic group in the same orientation as the D-gluco castanospermine analogues in the quaternary chirality center. Most interestingly, kotalanol and d-Gluco quinolizidine have higher affinity to CtSI than to NtSI (10

and 100 fold respectively). This may indicate that the size of the secondary chain plays a role on the selectivity between subunits.

In comparison with other common commercial inhibitors, the D-gluco indolizidine and quinolizidine showed lower K_i than acarbose for both N- and C- terminal subunits. D-gluco quinolizidine seems to be a potent inhibitor for CtSI with a K_i value at least 10 times lower than acarbose, salacinol, blintol, miglitol, and kotalanol (Table 4.2). D-gluco indolizidine showed higher affinity than the same set of common inhibitors against NtSI.

Table 4.2 Commercially available alpha-glucosidases inhibitors taken from BRENDA data base.

Inhibitor/Enzyme	NtSI (μM)	CtSI(μM)	Reference
Acarbose	14	0.246	(Sim et al., 2010a)
Blintol	0.16	0.029	(Jones et al., 2011)
Kotalanol	0.6	0.042	(Sim et al., 2010a)
Salacinol	0.000277	0.047	(Jones et al., 2011)

Values reported for E.C Number 3.2.1.48 and 3.2.1.10.

For L-ido indolizidine, despite showing inhibitory activity against both CtSI and NtSI, the K_i values are more than 10 times higher than the ones observed with commercial inhibitors. In addition, the results showed an important lack of fit to the three typical models of inhibition and a deviation of about 50%. Therefore, no more than general conclusions can be made.

L-ido indolizidine:

The L-ido indolizidine a low inhibitory activity against sucrase-isomaltase, with K_i values more than 30 times higher than the D-gluco counterpart for NtSI and almost 100 times for CtSI. High K_i values indicate low affinity; a possible interpretation for this is that the system may be quite dynamic, with the ligand intermittently binding and being released from the different binding sites of sucrase-isomaltase.

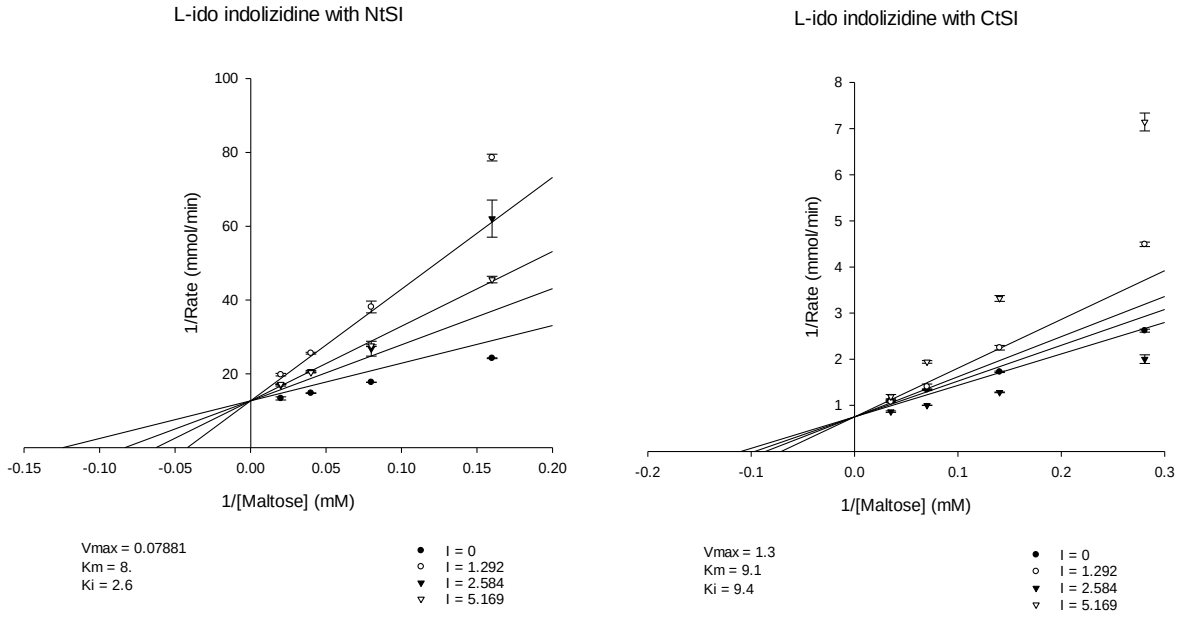


Figure 4.2 Lineweaver-Burk plots for NtSI and CtSI with L-ido indolizidine. The inhibitor concentrations are expressed in µM and the rate of reaction measured as mmol of glucose released per min.

Using Lineweaver-Burk, the inhibition results did not fit any complete model of inhibition. The closest complete model indicates competitive inhibition, however, with a R^2 of 0.735 for NtSI and 0.740 for CtSI. The lack of fit does not support this type of inhibition as an explanation for the behaviour of this compound as inhibitor.

One of the main problems when using Lineweaver-Burk plots to determine the type of inhibition is the inequality of the weights given to the different points. As Lineweaver-Burk models inhibition with reciprocals, it undervalues the points obtained at low concentrations of substrate. An alternative method to determine the type of inhibition is Eadie-Hofstee; however, in this case, due to the variability of the results obtained, this method did not give better results (Figure 4.3).

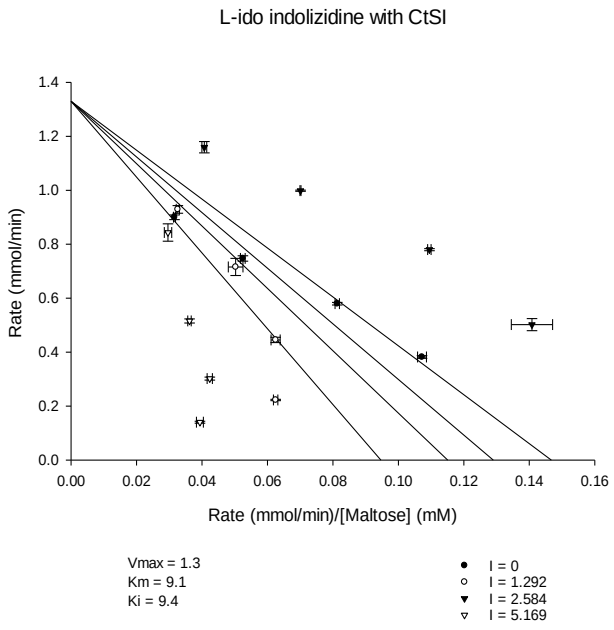


Figure 4.3 Eadie-Hofstee plot for L-ido indolizidine with CtSI. The inhibitor concentrations are expressed in µM and the rate of reaction measured as mmol of glucose released per min.

For inhibition of NtSI, the best R^2 values obtained point to a partial mixed model of inhibition (R^2 =0.937) closely followed for a partial competitive inhibition (R^2 =0.935) as indicated in Tables 4.3 and 4.4. These results were obtained using the Enzymatic kinetics module of SigmaPlot, which uses a series of secondary plots and analyse them as hyperbolic functions (http://sigmaplot.co.uk/products/sigmaplot/enzyme-mod.php).

Table 4.3 Enzyme kinetics model comparison for L-ido indolizidine with NtSI. Single substrate-single inhibitor with 3 replicates.

Equation	R^2	Test	Convergence
Mixed (Partial)	0.93769	pass	Yes
Competitive (Partial)	0.93513	pass	Yes
Noncompetitive (Partial)	0.89895	fail	Yes
Competitive (Full)	0.73538	pass	Yes
Mixed (Full)	0.73538	pass	Yes
Noncompetitive (Full)	0.71295	fail	Yes
Uncompetitive (Full)	0.69581	pass	Yes
Uncompetitive (Partial)	0.69581	pass	Yes

"Partial" inhibition means that catalysis is not completely blocked but instead diminished; therefore, the EI or ESI complex, according to type of inhibition, will still release some product (Grant, 2018).

For CtSI, the results did not show good fit to any of the models of inhibition (Table 4.3) and the results showed an important deviation in the K_i calculation (~50%). Therefore, it is dangerous to draw any conclusions from this analysis. To improve the accuracy of the determination of the type of inhibition, it will be necessary to test more concentrations of inhibitor and substrate. This will provide more points to use for the fit and open the range of measurement.

Table 4.4 Enzyme kinetics model comparison for L-ido indolizidine with CtSI. Single substrate- single inhibitor with 3 replicates.

Equation	R^2	Test	Convergence
Mixed (Partial)	0.74059	pass	Yes
Competitive (Full)	0.74039	pass	Yes
Mixed (Full)	0.74039	pass	Yes
Competitive (Partial)	0.74039	pass	Yes
Noncompetitive (Partial)	0.72968	pass	Yes
Noncompetitive (Full)	0.72968	pass	Yes
Uncompetitive (Full)	0.72057	pass	Yes
Uncompetitive (Partial)	0.72057	pass	Yes

D-gluco indolizidine:

The D-orientation of the hydroxylic group in the pyrrolidine seems to increase the affinity of this indolizidine with both subunits of sucrase-isomaltase in comparison with L-ido indolizidine. As a result, the K_i values decreased significantly, 30- and 100-fold for NtSI and CtSI, respectively. Also, the fit to the type of 4.4 shows, the intercept between the inhibition curves occurs close to the origin, crossing simultaneously the Y and X axis. Therefore, it is still hard to differentiate between a mixed or competitive type of inhibition, and the difference in the R^2 between the two models is < 0.001. This same result was observed with both the N- and C-terminal subunits of sucrase-isomaltase. Nevertheless, the high value of alpha (42 and 330.3) may indicate a lack of fit, pointing towards a competitive kind of inhibition. In order to confidently determine the type of inhibition, further analysis would be required. Some options for that further analysis could be trying an independent fitting of the data to a non-linear model, an independent fitting of the inhibition results choosing ranges of inhibitor concentration, or adjusting a set of equations that describes the mechanism of this inhibition.

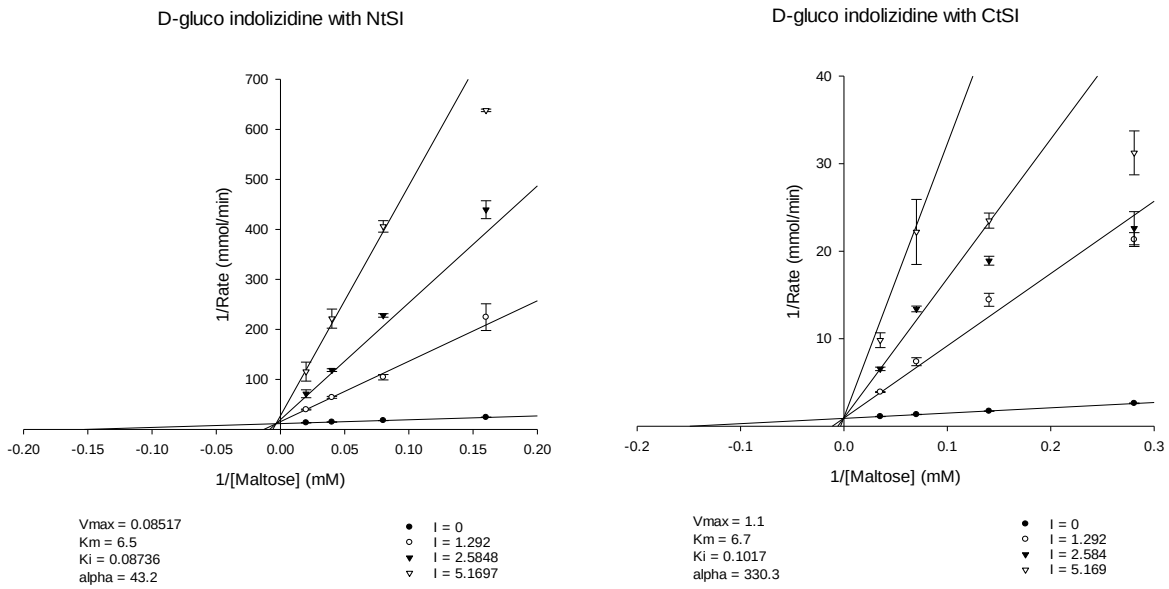


Figure 4.4. Lineweaver-Burk plot for D-Gluco indolizidine with NtSI (left) and CtSI (right). The inhibitor concentrations are expressed in µM and the rate of reaction measured as mmol of glucose released per min.

The model of a competitive type of inhibition implies that the substrate and inhibitor are competing for the active site of the enzyme. In the other hand, a mixed model of inhibition indicates that the inhibitor binds in a different site of the enzyme and distorts the conformation of the enzyme, decreasing the affinity between the substrate and the active site of the enzyme. In this case, as castanospermine and its analogues are small molecules, it is possible that the inhibitor can bind in more than one site of the enzyme. This will also explain the difficulty of finding a good fit to the different models of inhibition for L-ido indolizidine with the C- and N-terminal subunits of sucrase-isomaltase.

D-gluco quinolizidine

Like D-gluco indolizidine, the mode of D-gluco quinolizidine inhibition of sucrase-isomaltase is also unclear. Figure 4.5 shows the Lineweaver-Burk plots for D-gluco quinolizidine with NtSI and CtSI. These plots show convergence close to the origin.

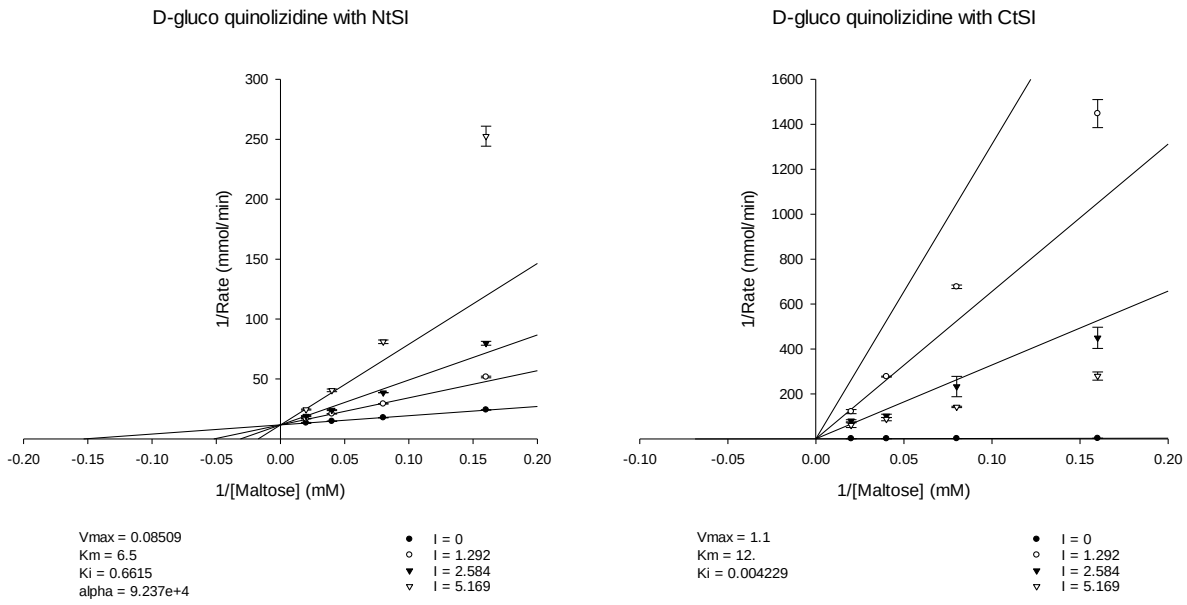


Figure 4.5 Lineweaver-Burk plot for D-guco quinolizidine with NtSI (left) and CtSI (right). The inhibitor concentrations are expressed in µM and the rate of reaction measured as mmol of glucose released per min.

The most remarkable result found with the castanospermine analogues, was the high selectivity between N- and C- terminal subunits displayed by D-Gluco quinolizidine. With a difference in the K_i of more than 150 fold between NtSI (K_i=0.66µM) and CtSI (0.004 µM), this compound is more efficient in the inhibition of CtSI than the most popular inhibitors currently found on the market including acarbose (K_i=0.246µM), salacinol (0.047µM), blintol (0.029µM), and miglitol (0.130µM). The ability to differentiate among the sucrase-isomaltase subunits is a highly desirable feature in inhibitors for treating Type II Diabetes, and there are currently few inhibitors with this feature. Blintol has a K_i for CtSI 5 times lower than for NtSI, and for Kotalanol it is

10-fold lower; however, none of the common inhibitors have shown nearly as high selectivity as this castanospermine analogue. Furthermore, no other inhibitor so far has shown a K_i as low as D-Gluco indolizidine for CtSI.

4.3.3. Changes in the inhibition pattern and affinity due to changes in the orientation of the hydroxylic group.

The difference between L-ido indolizidine and D-gluco indolizidine is the orientation of the hydroxylic group in the quaternary chiral center. This orientation seems to play an important role in the binding process with the catalytic site of the alpha-glucosidase enzymes (H Furneaux et al., 1997). According to the results observed by Winchester et al. (1990) and Boisson et al. (2015) the hydroxylic groups located in the positions C3,C4, and C5 in the pyrano ring are enough to allow the binding of compounds to alpha-glucosidases (Winchester et al., 1990). However, the addition of another chiral center in the pyrrolidine increases the affinity to the active site of specific alpha-glucosidases (Boisson et al., 2015). The results observed in this study showed that the L-ido orientation was less favorable, displaying less affinity for the subunits of sucrase-isomaltase in comparison with the D-gluco orientation (Table 4.2). This result was already hypothesized by Winchester et al. (1990) who considered the R orientation as "prohibited."

4.3.4 Effect of an extra carbon in the imino ring.

The K_i of D-gluco indolizidine with NtSI is about eight-fold lower than the K_i obtained with D-gluco quinolizidine; therefore, the affinity of D-gluco indolizidine could be interpreted as higher. This would indicate that the smaller furanose ring found indolizidine is more a favorable structure to bind NtSI. In contrast, the results with CtSI showed the opposite behaviour, though to a lesser extent. While D-gluco indolizidine showed a similar affinity for NtSI and CtSI, D-gluco quinolizidine displayed high selectivity between the two subunits, with CtSI favoured (Table 4.2). D-gluco indolizidine and D-gluco quinolizidine are only differentiated by the presence of a single carbon (pyrano vs furano rings) attached to the piperidine. This difference seems to be a key for the selectivity between N- and C- terminal subunits of sucrase-isomaltase.

The phenomenon observed with D-gluco quinolizidine is desirable, as compounds with specific affinity means fewer side effects and more specific treatments. However, it is not common and the specifics of how this selectivity works remain unclear. A few other inhibitory compounds have shown this behaviour, in lower degree. D-gluco quinolizidine can be compared with Kotalanol, a natural compound isolated from *Salacia reticulata* used in Ayurvedic traditional medicine as a diabetes treatment (Jones et al., 2011). Despite the substitution of a nitrogen for a sulfur atom, and the size of the ring, the structures of kotalanol and D-gluco quinolizidine are similar (Figure 4.6). They both share the same orientation in the substituents in the chiral centers 3, 4, and 5 predicted to be key for binding alpha-glucosidases (Winchester et al., 1990), and a side chain with an additional chiral center (in kotalanol case, more than one).

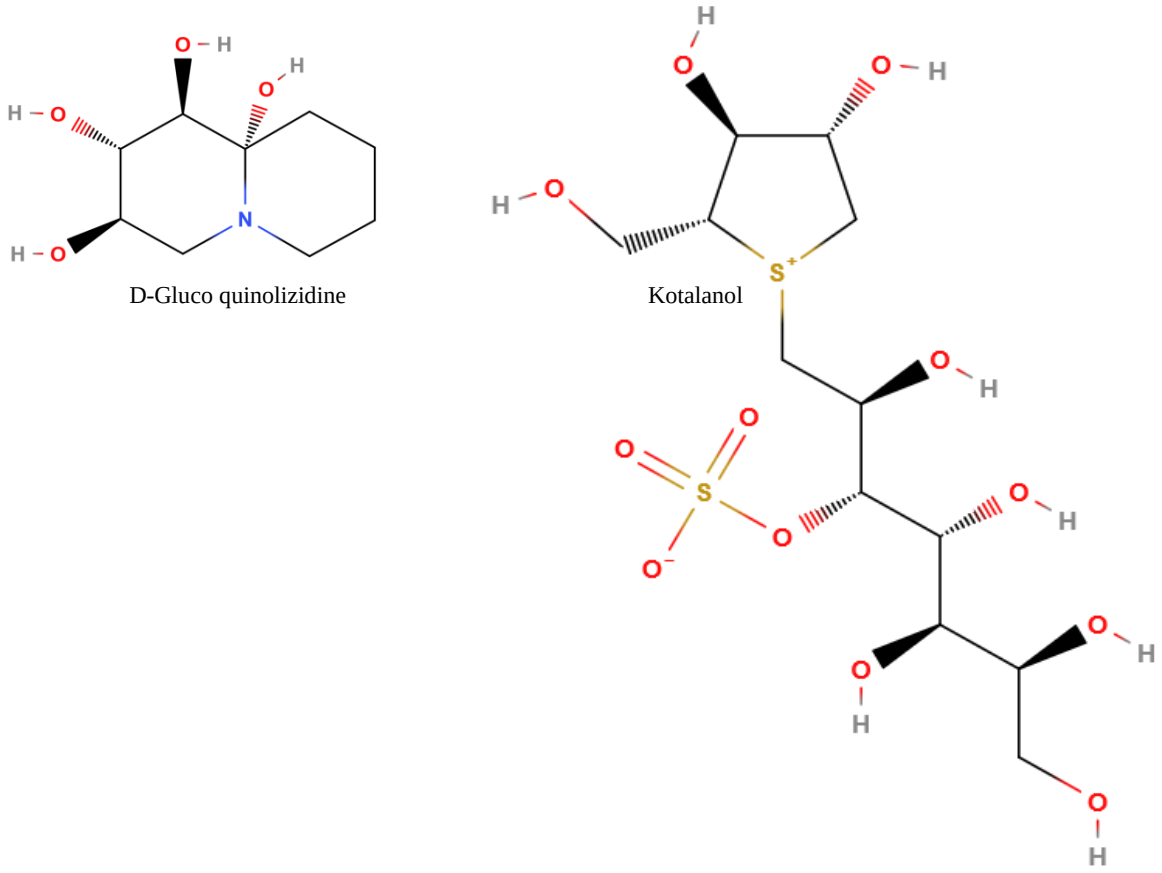


Figure 4.6. Chemical structures of D-gluco quinolizidine (left) and Kotalanol (right).

In previous studies the crystal structure of NtSI with kotalanol was solved at 2.15Å resolution, indicating that the inhibitor binds in the active site (Sim et al., 2010b). However, as the D-gluco quinolizidine is a more compact and less substituted molecule than kotalanol, it is still possible for it to bind additional sites in sucrase-isomaltase, which is suggested by the similar fit for both mixed and competitive modes of inhibition (Table 4.1). However, no crystal structure of CtSI is available, which impedes studying the interaction between D-gluco quinolizidine and CtSI.

4.4. Conclusions

From this study four main points can be highlighted:

1. All three castanospermine analogues shown inhibitory activity against both subunits of human sucrase-isomaltase.

2. The compounds with D- orientation displayed K_i values under 0.66µM, which make them comparable to commercial inhibitors such as miglitol and salacinol. The L-ido indolizidine showed a significantly reduced affinity to both subunits of sucrase-isomaltase with K_i values greater than 10 µM for both subunits.

3. The patterns of inhibition observed with the castanospermine analogues were mixed or competitive. For D-gluco quinolizidine, the two types of inhibition showed the same or similar fit, which indicates that the compounds may bind in the catalytic pocket or potentially in other binding sites of the protein.

4. The extra carbon in the imino ring seems to improve the affinity of the castanospermine analogue with Ct-SI, as the observed K_i was decreased 25-fold. In addition, the increase of the size of the ring seems to be key for selectivity between CtSI and NtSI, as the K_i is more than 100 times lower for CtSI.

Chapter 5. Conclusions and Future Directions

From this research work the following conclusions can be made:

- The engineered vector for protein cloning and expression in *K. phaffii* pPinkNN_Z offers

  1) A simplified cloning method for obtaining the constructs

  2) an improved method for colony selection that not only differentiates among transformed and untransformed cells, but also number of plasmid copies.

  3) Increased affinity of the histidine tag-transition metal for a more efficient purification.

- The optimized expression system allowed the cloning and expression of eight human alpha-glucosidases with sizes ranging from 936 to over 5000 bp. The typical protein yield ranged between 0.79mg/mL and 4.88 mg/mL which is above the average commonly reported for batch expression in *K. phaffii.* All the proteins expressed in *K. phaffii* were enzymatically active and showed bands in polyacrylamide gels consistent with the sizes expected from their amino acid sequences.

- The inhibition of sucrase-isomaltase and its subunits by castanospermine compounds demonstrated that the substituent groups in the chiral centers C3, C4, and C5 are crucial for the enzyme-inhibitor binding. The addition of a quaternary chiral center in the D-gluco quinolizidine proved to play a role in the selectivity among C- and N- terminal subunits. These findings can play a key role for the design of highly specific inhibitory

compounds, and in the investigation of the differences of substrate affinity among NtSI, CtSI, NtMGAM, and CtMGAM.

The expression system developed in this research shows promise for facilitating the expression and purification of heterologous proteins from Eukaryotic organisms. Therefore, it would be interesting to test its the feasibility for expressing other human proteins as well as proteins from different sources that cannot be successfully expressed in Prokaryotic systems.

The expression results for the different alpha-glucosidases is the first step in characterizing the products. Subsequently, a more detailed characterization including their glycosylation patterns and sites is required. The amino-acid sequences can be confirmed by Mass spectrometry. Ultimately, the proteins expressed in *K. phaffii* would facilitate structural analysis, as they are mostly N-glycosylated with high mannose residues that are easily removable, and the typical yields are sufficient.

In terms of enzymatic activity, the expressed proteins could be tested against a broad set of substrates including different alpha-glucosidic linkages, lengths, and distribution to distinguish better between their substrate affinity and catalytic activity. Specifically, for sucrase-isomaltase, the investigation of the differences between the full length and its individual subunits may help to elucidate if the interaction between C- and N- terminal domains influence folding or activity. Additional studies regarding the protein- protein interactions as well as coupled assays using different combinations of enzymes and substrates may shed some light over their interaction in

the intestine and how their catalytic activities complement each other for a more efficient digestion of carbohydrates.

For lactase phlorizin hydrolase, structural studies would be a priority as protein plays such a crucial role in lactose digestion and there is almost nothing known regarding its overall conformation or its catalytic site.

Finally, the high selectivity of D-gluco quinolizidine among the C- and N- terminal subunits of SI represents an opportunity for a better understanding of the differences in affinity among the subunits, and the further development of selective inhibitors.

References

Alfalah, M., Jacob, R., Preuss, U., Zimmer, K. P., Naim, H., & Naim, H. Y. (1999). O-linked glycans mediate apical sorting of human intestinal sucrase-isomaltase through association with lipid rafts. *Current Biology*, *9*(11), 593–596. https://doi.org/10.1016/S0960-9822(99)80263-2

Alfalah, M., Keiser, M., Leeb, T., Zimmer, K. P., & Naim, H. Y. (2009). Compound Heterozygous Mutations Affect Protein Folding and Function in Patients With Congenital Sucrase-Isomaltase Deficiency. *Gastroenterology*, *136*(3), 883–892. https://doi.org/10.1053/j.gastro.2008.11.038

Amen, T., & Kaganovich, D. (2017). Integrative modules for efficient genome engineering in yeast. *Microbial Cell*, *4*(6), 182–190. https://doi.org/10.15698/mic2017.06.576

Asano, N. (2008). Sugar-mimicking glycosidase inhibitors: bioactivity and application. *Cell. Mol. Life. Sci*, *66*, 1479–1492. https://doi.org/10.1007/s00018-008-8522-3

Beau, I., Cotte-Laffitte, J., Geniteau-Legendre, M., Estes, M. K., & Servin, A. L. (2007). An NSP4-dependant mechanism by which rotavirus impairs lactase enzymatic activity in brush border of human enterocyte-like Caco-2 cells. *Cellular Microbiology*, *9*(9), 2254–2266. https://doi.org/10.1111/j.1462-5822.2007.00956.x

Beaulieu, J. F., Weiser, M. M., Herrera, L., & Quaroni, A. (1990). Detection and characterization of sucrase-isomaltase in adult human colon and in colonic polyps. *Gastroenterology*, *98*(6), 1467–1477. https://doi.org/10.1016/0016-5085(90)91077-J

Behrendt, M., Polaina, J., & Naim, H. Y. (2010). *Structural Hierarchy of Regulatory Elements in the Folding and Transport of an Intestinal Multidomain Protein *. 285*(6), 4143–4152.

https://doi.org/10.1074/jbc.M109.060780

Belmont, J. W., Reid, B., Taylor, W., Baker, S. S., Moore, W. H., Morriss, M. C., … Schwartz, I. D. (2002). Congenital sucrase-isomaltase deficiency presenting with failure to thrive , hypercalcemia, and nephrocalcinosis. *BMC Pediatrics*, *7*, 1–7.

Benko, Z., & Zhao, R. Y. (2011). Zeocin for selection of bleMX6 resistance in fission yeast. *BioTechniques*, *51*(1), 57–60. https://doi.org/10.2144/000113706

Bernasconi, R., & Molinari, M. (2011). ERAD and ERAD tuning: Disposal of cargo and of ERAD regulators from the mammalian ER. *Current Opinion in Cell Biology*, *23*(2), 176–183. https://doi.org/10.1016/j.ceb.2010.10.002

Boisson, J., Thomasset, A., Racine, E., Cividino, P., Banchelin Sainte-Luce, T., Poisson, J. F., … Py, S. (2015). Hydroxymethyl-Branched Polyhydroxylated Indolizidines: Novel Selective α-Glucosidase Inhibitors. *Organic Letters*, *17*(15), 3662–3665. https://doi.org/10.1021/acs.orglett.5b01505

Borges de Melo, E., da Silveira Gomes, A., & Carvalho, I. (2006). α- and β-Glucosidase inhibitors: chemical structure and biological activity. *Tetrahedron*, *62*(44), 10277–10302. https://doi.org/10.1016/J.TET.2006.08.055

Boudry, G., Jury, J., Yang, P. C., & Perdue, M. H. (2007). Chronic psychological stress alters epithelial cell turn-over in rat ileum. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *292*(5), G1228–G1232. https://doi.org/10.1152/ajpgi.00358.2006

Bretthauer, R. K., & Castellino, F. J. (1999). Glycosylation of Pichia pastoris -derived proteins. *Biotechnology and Applied Biochemistry*, *30*(3), 193–200. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1470-8744.1999.tb00770.x/full%5Cnpapers3://publication/doi/10.1111/j.1470-

8744.1999.tb00770.x

Cereghino, J. L., & Cregg, J. M. (2000). Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiology Reviews*, *24*(1), 45–66. https://doi.org/10.1016/S0168-6445(99)00029-7

Cereghino, L., Geoffrey, P., Godfrey, L., de la Cruz, B. J., Johnson, S., Khuongsathiene, S., … Cregg, J. M. (2000). Recombinant protein expression in Pichia pastoris. *Applied Biochemistry and Biotechnology - Part B Molecular Biotechnology*, *16*(1), 23–52. https://doi.org/10.1385/MB:16:1:23

Chen, X., Yuan, H., He, W., Hu, X., Lu, H., & Li, Y. (2005). Construction of a novel kind of expression plasmid by homologous recombination in Saccharomyces cerevisiae. *Science in China. Series C, Life Sciences*, *48*(4), 330–336.

Chiasson, J.-L., Josse, R. G., Gomis, R., Hanefeld, M., Karasik, a, & Laakso, M. (2004). Acarbose for the prevention of Type 2 diabetes, hypertension and cardiovascular disease in subjects with impaired glucose tolerance: facts and interpretations concerning the critical analysis of the STOP-NIDDM Trial data. *Diabetologia*, *47*(6), 969–975; discussion 976-977. https://doi.org/10.1007/s00125-004-1409-4

Cohen, S. N., Chang, A. C. Y., Boyer, H. W., & Helling, R. B. (1973). Construction of Biologically Functional Bacterial Plasmids In Vitro (R factor/restriction enzyme/transformation/endonuclease/antibiotic resistance). *Pnas*, *70*(11), 3240–3244.

Conde, R., Cueva, R., Pablo, G., Polaina, J., & Larriba, G. (2004). A search for hyperglycosylation signals in yeast glycoproteins. *Journal of Biological Chemistry*, *279*(42), 43789–43798. https://doi.org/10.1074/jbc.M406678200

Conklin, K A, Yamashiro, K. M., & Gray, G. M. (1975). Human intestinal sucrase-isomaltase.

Identification of free sucrase and isomaltase and cleavage of the hybrid into active distinct subunits. *Journal of Biological Chemistry* , *250*(15), 5735–5741. Retrieved from http://www.jbc.org/content/250/15/5735.abstract

Conklin, Kenneth A., Yamashiro, K. M., & Gray, G. M. (1975). Human Intestinal Sucrase-Isomaltase. *The Journal of Biological Chemistry*, (15), 5735–5742.

Cregg, J., & R. Madden, K. (1987). Biological research on Yeast Vol II. In G. G. Stewart, I. Russell, R. D. Klein, & R. R. Hiebsch (Eds.), *DEVELOPMENT OF YEAST TRANSFORMATION SYSTEMS AND CONSTRUCTION OF METHANOL-UTILIZATION-DEFECTIVE MUTANTS OF PICHIA PASTORI BY GENE DISRUPTION*. Boca Raton, Florida, US: CRC Press.

Crowe, J., Dobeli, H., Gentz, R., Hochuli, E., Stiiber, D., & Henco, K. (1994). 6xffis-Ni-NTA Chromatography as a Superior Technique in Recombinant Protein Expression/Purification. In A. J. Harwood (Ed.), *Protocols for Gene Analysis* (pp. 371–387). https://doi.org/10.1385/0-89603-258-2:371

Diekmann, L., Behrendt, M., Amiri, M., & Naim, H. Y. (2017). Structural determinants for transport of lactase phlorizin-hydrolase in the early secretory pathway as a multi-domain membrane glycoprotein. *Biochimica et Biophysica Acta - General Subjects*. https://doi.org/10.1016/j.bbagen.2016.10.016

Dudich, E., Dudich, I., Semenkova, L., Benevolensky, S., Morozkina, E., Marchenko, A., … Tatulov, E. (2012). Engineering of the Saccharomyces cerevisiae yeast strain with multiple chromosome-integrated genes of human alpha-fetoprotein and its high-yield secretory production, purification, structural and functional characterization. *Protein Expression and Purification*, *84*(1), 94–107. https://doi.org/10.1016/j.pep.2012.04.008

Foda, M. I., Kawashima, T., Nakamura, S., Kobayashi, M., & Oku, T. (2004). Composition of Milk Obtained From Unmassaged Versus Massaged Breasts of Lactating Mothers. *Journal of Pediatric Gastroenterology and Nutrition*, *38*(5). Retrieved from https://journals.lww.com/jpgn/Fulltext/2004/05000/Composition_of_Milk_Obtained_From_Unmassaged.5.aspx

Fransen, J. a M., Hauri, H. P., Ginsel, L. a., & Naim, H. Y. (1991). Naturally occurring mutations in intestinal sucrase-isomaltase provide evidence for the existence of an intracellular sorting signal in the isomaltase subunit. *Journal of Cell Biology*, *115*(1), 45–57. https://doi.org/10.1083/jcb.115.1.45

Gericke, B., Amiri, M., & Naim, H. Y. (2016). The multiple roles of sucrase-isomaltase in the intestinal physiology. *Molecular and Cellular Pediatrics*, *3*(1), 2. https://doi.org/10.1186/s40348-016-0033-y

Gericke, B., Schecker, N., Amiri, M., & Naim, H. Y. (2017). Structure-function analysis of human sucrase-isomaltase identifies key residues required for catalytic activity. *Journal of Biological Chemistry*, *292*(26), 11070–11078. https://doi.org/10.1074/jbc.M117.791939

Ghani, U. (2015). Re-exploring promising a -glucosidase inhibitors for potential development into oral anti-diabetic drugs : Finding needle in the haystack. *European Journal of Medicinal Chemistry*, *103*, 133–162. https://doi.org/10.1016/j.ejmech.2015.08.043

Grant, G. A. (2018). The many faces of partial inhibition: Revealing imposters with graphical analysis. *Archives of Biochemistry and Biophysics*, *653*(May), 10–23. https://doi.org/10.1016/j.abb.2018.06.009

Grisshammer, R., & Tucker, J. (1997). Quantitative evaluation of neurotensin receptor purification by immobilized metal affinity chromatography. *Protein Expression and*

Purification, *11*(1), 53–60. https://doi.org/10.1006/prep.1997.0766

H Furneaux, R., J Gainsford, G., M Mason, J., C Tyler, P., Hartley, O., & G Winchester, B. (1997). The chemistry of castanospermine, part V: synthetic modifications at C-1 and C-7. *Tetrahedron*, *53*(1), 245–268. https://doi.org/10.1016/S0040-4020(96)00967-2

Hamilton, S. R., & Gerngross, T. U. (2007). Glycosylation engineering in yeast: the advent of fully humanized yeast. *Current Opinion in Biotechnology*, *18*(5), 387–392. https://doi.org/10.1016/j.copbio.2007.09.001

Hentges, P., Van Driessche, B., Tafforeau, L., Vandenhaute, J., & Carr, A. M. (2005). Three novel antibiotic marker cassettes for gene disruption and marker switching in Schizosaccharomyces pombe. *Yeast*, *22*(13), 1013–1019. https://doi.org/10.1002/yea.1291

Higgins, D. R., Busser, K., Comiskey, J., Whittier, P. S., Purcell, T. J., & Hoeffler, J. P. (1998). Small Vectors for Expression Based on Dominant Drug Resistance with Direct Multicopy Selection. In D. R. Higgins & J. M. Cregg (Eds.), *Pichia Protocols* (pp. 41–53). https://doi.org/10.1385/0-89603-421-6:41

Horii, S., Fukase, H., Matsuo, T., Kameda, Y., Asano, N., & Matsui, K. (1986). Synthesis and -D-Glucosidase Inhibitory Activity of N-Substituted Valiolamine Derivatives as Potential Oral Antidiabetic Agents. *J. Med. Chem.*, *56194*(1982), 1038–1046. https://doi.org/10.1021/jm00156a023

Hunziker, W., Spiess, M., Semenza, G., & Lodish, H. F. (1986). The sucrase-isomaltase complex: Primary structure, membrane-orientation, and evolution of a stalked, intrinsic brush border protein. *Cell*, *46*(2), 227–234. https://doi.org/10.1016/0092-8674(86)90739-7

Jacob, R., Peters, K., & Naim, H. Y. (2002). The prosequence of human lactase-phlorizin hydrolase modulates the folding of the mature enzyme. *Journal of Biological Chemistry*,

277(10), 8217–8225. https://doi.org/10.1074/jbc.M111500200

Jacob, R., Pürschel, B., & Naim, H. Y. (2002). Sucrase is an intramolecular chaperone located at the C-terminal end of the sucrase-isomaltase enzyme complex. *Journal of Biological Chemistry*, *277*(35), 32141–32148. https://doi.org/10.1074/jbc.M204116200

Jacob, R., Weiner, J. R., Stadge, S., & Naim, H. Y. (2000). Additional N-glycosylation and its impact on the folding of intestinal lactase-phlorizin hydrolase. *Journal of Biological Chemistry*, *275*(14), 10630–10637. https://doi.org/10.1074/jbc.275.14.10630

Jones, K., Sim, L., Mohan, S., Kumarasamy, J., Liu, H., Avery, S., … Rose, D. R. (2011). Mapping the intestinal alpha-glucogenic enzyme specificities of starch digesting maltase-glucoamylase and sucrase-isomaltase. *Bioorganic and Medicinal Chemistry*, *19*(13), 3929–3934. https://doi.org/10.1016/j.bmc.2011.05.033

Joubert, P. H., Foukaridis, G. N., & Bopape, M. L. (1987). *Miglitol May Have a Blood Glucose Lowering Effect Unrelated to Inhibition of Alpha-Glucosidase*. 723–724.

Juturu, V., & Wu, J. C. (2018). Heterologous Protein Expression in Pichia pastoris: Latest Research Progress and Applications. *ChemBioChem*, *19*(1), 7–21. https://doi.org/10.1002/cbic.201700460

KANO, T., USAMI, Y., ADACHI, T., TATEMATSU, M., & HIRANO, K. (1996). Inhibition of Purified Human Sucrase and Isomaltase by Ethanolamine Derivatives. *Biological & Pharmaceutical Bulletin*, *19*(3), 341–344. https://doi.org/10.1248/bpb.19.341

Karasov, W. H., Martínez del Rio, C., & Caviedes-Vidal, E. (2011). Ecological Physiology of Diet and Digestive Systems. *Annual Review of Physiology*, *73*(1), 69–93. https://doi.org/10.1146/annurev-physiol-012110-142152

Karim, A. S., Curran, K. A., & Alper, H. S. (2013). Characterization of plasmid burden and copy

number in Saccharomyces cerevisiae for optimization of metabolic engineering applications. *FEMS Yeast Research*, *13*(1), 107–116. https://doi.org/10.1111/1567-1364.12016

Kim, H., Yoo, S. J., & Kang, H. A. (2015). Yeast synthetic biology for the production of recombinant therapeutic proteins. *FEMS Yeast Research*, *15*(1), 1–16. https://doi.org/10.1111/1567-1364.12195

Kimple, M. E., Brill, A. L., & Pasker, R. L. (2013). Overview of affinity tags for protein purification. *Current Protocols in Protein Science*, (SUPPL.73), 608–616. https://doi.org/10.1002/0471140864.ps0909s73

Kirchrath, L., Lorberg, A., Schmitz, H. P., Gengenbacher, U., & Heinisch, J. J. (2000). Comparative genetic and physiological studies of the MAP kinase Mpk1p from Kluyveromyces lactis and Saccharomyces cerevisiae. *Journal of Molecular Biology*, *300*(4), 743–758. https://doi.org/10.1006/jmbi.2000.3916

Kornfeld, R., & Kornfeld, S. (1985). Assembly of asparangine-linked oligosaccharides. *Annual Review of Biochemistry*, *54*(1), 631–664.

Krentz, A. J., & Bailey, C. J. (2005). Oral Antidiabetic Agents. *Drugs*, *65*(3), 385–411. https://doi.org/10.2165/00003495-200565030-00005

Kruse, T. A., Bolund, L., Grzeschik, K. H., Ropers, H. H., Sjöström, H., Norén, O., … Semenza, G. (1988). The human lactase-phlorizin hydrolase gene is located on chromosome 2. *FEBS Letters*, *240*(1–2), 123–126. https://doi.org/10.1016/0014-5793(88)80352-1

Küberl, A., Schneider, J., Thallinger, G. G., Anderl, I., Wibberg, D., Hajek, T., … Pichler, H. (2011). High-quality genome sequence of Pichia pastoris CBS7435. *Journal of Biotechnology*, *154*(4), 312–320. https://doi.org/10.1016/j.jbiotec.2011.04.014

Kuijpers, N. G. A., Solis-Escalante, D., Bosman, L., van den Broek, M., Pronk, J. T., Daran, J. M., & Daran-Lapujade, P. (2013). A versatile, efficient strategy for assembly of multi-fragment expression vectors in Saccharomyces cerevisiae using 60 bp synthetic recombination sequences. *Microbial Cell Factories*, *12*(1), 1–13. https://doi.org/10.1186/1475-2859-12-47

Kunes, S., Schatz, P. J., & Botstein, D. (1987). Plasmid construction by homologous recombination in yeast (Saccharomyces. *Science*, *58*, 201–216.

Lagassé, H. A. D., Alexaki, A., Simhadri, V. L., Katagiri, N. H., Jankowski, W., Sauna, Z. E., & Kimchi-Sarfaty, C. (2017). Recent advances in (therapeutic protein) drug development. *F1000Research*, *6*, 113. https://doi.org/10.12688/f1000research.9970.1

Laron, Z. (2001). Insulin-like growth factor 1 (IGF-1): A growth hormone. *Journal of Clinical Pathology - Molecular Pathology*, *54*(5), 311–316. https://doi.org/10.1136/mp.54.5.311

Lau, H. K. (1987). Physicochemical characterization of human intestinal lactase. *The Biochemical Journal*, *241*(2), 567–572. https://doi.org/10.1042/bj2410567

Laukens, B., Wachter, C. De, & Callewaert, N. (2015). Engineering the Pichia pastoris N-Glycosylation Pathway Using the GlycoSwitch Technology. In A. Castilho, Alexandra (University of Natural and Life Science, Viena (Ed.), *Glyco-Engineering: Methods and Protocols, Methods in Molecular Biology, vol. 1321* (pp. 103–122). https://doi.org/10.1007/978-1-4939-2760-9

Lee, B. H., Rose, D. R., Lin, A. H. M., Quezada-Calvillo, R., Nichols, B. L., & Hamaker, B. R. (2016). Contribution of the Individual Small Intestinal α-Glucosidases to Digestion of Unusual α-Linked Glycemic Disaccharides. *Journal of Agricultural and Food Chemistry*. https://doi.org/10.1021/acs.jafc.6b01816

Li, C., Begum, A., Numao, S., Kwan, H. P., Withers, S. G., & Brayer, G. D. (2005). Acarbose rearrangement mechanism implied by the kinetic and structural analysis of human pancreatic ??-amylase in complex with analogues and their elongated counterparts. *Biochemistry*, *44*(9), 3347–3357. https://doi.org/10.1021/bi048334e

Lin Cereghino, G. P., Lin Cereghino, J., Jay Sunga, A., Johnson, M. A., Lim, M., Gleeson, M. A. G., & Cregg, J. M. (2001). New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of Pichia pastoris. *Gene*, *263*(1–2), 159–169. https://doi.org/10.1016/S0378-1119(00)00576-X

Liu, Z., Tyo, K. E. J., Martínez, J. L., Petranovic, D., & Nielsen, J. (2012). Different expression systems for production of recombinant proteins in Saccharomyces cerevisiae. *Biotechnology and Bioengineering*, *109*(5), 1259–1268. https://doi.org/10.1002/bit.24409

Macauley-Patrick, S., Fazenda, M. L., McNeil, B., & Harvey, L. M. (2005). Heterologous protein production using the Pichia pastoris expression system. *Yeast*, *22*(4), 249–270. https://doi.org/10.1002/yea.1208

Mali, K. C., Carson, M. L., Miller, M. P., Turowski, M., Bell, M., Wilder, D. M., & Reeves, M. S. (2007). High-Viscosity Hydroxypropylmethyl- cellulose Blunts Postprandial Glucose and Insulin Responses. *Diabetes Care*, *30*(5). https://doi.org/10.2337/dc06-2344.Abbreviations

Mantei, N., Villa, M., Enzler, T., Wacker, H., Boll, W., James, P., … Semenza, G. (1988). Complete primary structure of human and rabbit lactase-phlorizin hydrolase: implications for biosynthesis, membrane anchoring and evolution of the enzyme. *The EMBO Journal*, *7*(9), 2705–2713. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2460343

Masayuki, Toshiyuki, Yoshikawa, M., Shimada, H., Matsuda, H., Yamahara, J., Tanabe, G., &

Muraoka, O. (1997). Salacinol, potent antidiabetic principle with unique thiosugar sulfonium sulfate structure from the Ayurvedic traditional mecine Salacia reticulata in Sri Lanka and India. *Tetrahedron Lett.*, *38*(48), 8367–8370.

Mattanovich, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M., & Porro, D. (2012). Recombinant protein production in Yeasts. In Argelia Lorence (Arkansas Biosciences Institute) (Ed.), *Recombinant Gene Expression Third Edition* (Third, pp. 329–347). https://doi.org/10.1007/978-1-61779-433-9

Mumberg, D., Mulier, R., & Funk, M. (2012). *Nar00049-0251*. *22*(25), 1–2. Retrieved from papers2://publication/uuid/77ED3ADF-55EC-49A0-884D-3AEEFB53A5C1

Näätsaari, L., Mistlberger, B., Ruth, C., Hajek, T., Hartner, F. S., & Glieder, A. (2012). Deletion of the pichia pastoris ku70 homologue facilitates platform strain generation for gene expression and synthetic biology. *PLoS ONE*, *7*(6). https://doi.org/10.1371/journal.pone.0039720

Naim, H. Y., Roth, J., Sterchi, E. E., Lentze, M., Milla, P., Schmitz, J., & Hauri, H. P. (1988). Sucrase-isomaltase deficiency in humans. Different mutations disrupt intracellular transport, processing, and function of an intestinal brush border enzyme. *Journal of Clinical Investigation*, *82*(2), 667–679. https://doi.org/10.1172/JCI113646

Naim, H. Y., Sterchi, E. E., & Lentze, M. J. (1987). Biosynthesis and maturation of lactase-phlorizin hydrolase in the human small intestinal epithelial cells. *Biochemical Journal*, *241*(2), 427–434. https://doi.org/10.1042/bj2410427

Naim, H Y, Sterchi, E. E., & Lentze, M. J. (1988). Structure, biosynthesis, and glycosylation of human small intestine maltase-glucoamylase. *Journal of Biological Chemistry*, *263*(36), 19709–19717.

Naim, Hassan Y, Sterchi, E. E., Lentze, M. J., & Bern, C.-. (1988). *Biosynthesis of the Human Sucrase-Isomaltase Complex WITHIN THE ENZYME COMPLEX *. 263*(15), 7242–7253.

Paululat, A., & Heinisch, J. J. (2012). New yeast/E. coli/Drosophila triple shuttle vectors for efficient generation of Drosophila P element transformation constructs. *Gene*. https://doi.org/10.1016/j.gene.2012.09.058

Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., … Stone, A. C. (2007). Diet and the evolution of human amylase gene copy number variation. *Nature Genetics*, *39*(10), 1256–1260. https://doi.org/10.1038/ng2123

PichiaPink ™ Expression System. (2010).

Pontremoli, C., Mozzi, A., Forni, D., Cagliani, R., Pozzoli, U., Menozzi, G., … Sironi, M. (2015). Natural selection at the brush-border: Adaptations to carbohydrate diets in humans and other mammals. *Genome Biology and Evolution*, *7*(9). https://doi.org/10.1093/gbe/evv166

Prielhofer, R., Cartwright, S. P., Graf, A. B., Valli, M., Bill, R. M., Mattanovich, D., & Gasser, B. (2015). Pichia pastoris regulates its gene-specific response to different carbon sources at the transcriptional, rather than the translational, level. *BMC Genomics*, *16*(1), 1–17. https://doi.org/10.1186/s12864-015-1393-8

Quezada-Calvillo, R., Robayo-Torres, C. C., Ao, Z., Hamaker, B. R., Quaroni, A., Brayer, G. D., … Nichols, B. L. (2007). Luminal substrate “brake” on mucosal maltase-glucoamylase activity regulates total rate of starch digestion to glucose. *Journal of Pediatric Gastroenterology and Nutrition*, *45*(1), 32–43. https://doi.org/10.1097/MPG.0b013e31804216fc

Quezada-Calvillo, R., Robayo-Torres, C. C., Opekun, A. R., Sen, P., Ao, Z., Hamaker, B. R., …

Nichols, B. L. (2007). Contribution of mucosal maltase-glucoamylase activities to mouse small intestinal starch alpha-glucogenesis. *The Journal of Nutrition*, *137*(7), 1725–1733. https://doi.org/137/7/1725 [pii]

Quezada-Calvillo, R., Sim, L., Ao, Z., Hamaker, B. R., Quaroni, A., Brayer, G. D., … Nichols, B. L. (2008a). Luminal starch substrate "brake" on maltase-glucoamylase activity is located within the glucoamylase subunit. *The Journal of Nutrition*, *138*(4), 685–692. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18356321

Quezada-Calvillo, R., Sim, L., Ao, Z., Hamaker, B. R., Quaroni, A., Brayer, G. D., … Nichols, B. L. (2008b). Luminal starch substrate "brake" on maltase-glucoamylase activity is located within the glucoamylase subunit. *The Journal of Nutrition*, *138*(4), 685–692.

Raymond, C. K., Pownder, T. A., & Sexson, S. L. (1999). General Method for Plasmid Construction Using. *BioTechniques*, *26*(January), 134–141. https://doi.org/10.1073/PNAS.0901477106

Reifen, R., Zaiger, G., & Uni, Z. (1998). Effect of vitamin A on small intestinal brush border enzymes in a rat. In *International journal for vitamin and nutrition research. Internationale Zeitschrift für Vitamin- und Ernährungsforschung. Journal international de vitaminologie et de nutrition* (Vol. 68).

Robayo-Torres, C. C., Quezada-Calvillo, R., & Nichols, B. L. (2006). Disaccharide digestion: Clinical and molecular aspects. *Clinical Gastroenterology and Hepatology*, *4*(3), 276–287. https://doi.org/10.1016/j.cgh.2005.12.023

Robayo-Torres, C., Opekun, A. R., Quezada-Calvillo, R., Villa, X., Smith, E. O., Navarrete, M., & Baker, S. S. (2009). 13C-breath tests for sucrose digestion in congenital sucrase isomaltase-deficient and sacrosidase-supplemented patients. *Journal of Pediatric*

Gastroenterology and Nutrition, *48*(4), 412–418. https://doi.org/http://dx.doi.org/10.1097/MPG.0b013e318180cd09

Rossi, E. J., Sim, L., Kuntz, D. a., Hahn, D., Johnston, B. D., Ghavami, A., … Rose, D. R. (2006). Inhibition of recombinant human maltase glucoamylase by salacinol and derivatives. *FEBS Journal*, *273*(12), 2673–2683. https://doi.org/10.1111/j.1742-4658.2006.05283.x

Sander, P., Alfalah, M., Keiser, M., Korponay-Szabo, I., Kovács, J. B., Leeb, T., & Naim, H. Y. (2006). Novel mutations in the human sucrase-isomaltase gene (SI) that cause congenital carbohydrate malabsorption. *Human Mutation*, *27*(1), 119. https://doi.org/10.1002/humu.9392

Santoso, A., Herawati, N., & Rubiana, Y. (2012). Effect of methanol induction and incubation time on expression of human Erythropoietin in methylotropic yeast Pichia pastoris. *MAKARA of Technology Series*, *16*(1), 29–34. https://doi.org/10.7454/mst.v16i1.1041

Sels, J. P. J. E. (1996). *Miglitol (Bay m 1099) has no extraintestinal effects on glucose control in healthy volunteers*. 503–506.

Sim, L., Willemsma, C., Mohan, S., Naim, H. Y., Pinto, B. M., & Rose, D. R. (2010a). *Structural Basis for Substrate Selectivity in Human Maltase-Glucoamylase and Sucrase-Isomaltase*. *285*(23), 17763–17770. https://doi.org/10.1074/jbc.M109.078980

Sim, L., Willemsma, C., Mohan, S., Naim, H. Y., Pinto, B. M., & Rose, D. R. (2010b). Structural basis for substrate selectivity in human maltase-glucoamylase and sucrase-isomaltase N-terminal domains. *Journal of Biological Chemistry*, *285*(23), 17763–17770. https://doi.org/10.1074/jbc.M109.078980

Skovbjerg, H., Sjostrom, H., & Noren, O. (1981). Purification and characterisation of amphiphilic lactase/phlorizin hydrolase from human small intestine. *European Journal of*

Biochemistry, *114*(3), 653–661. https://doi.org/10.1111/j.1432-1033.1981.tb05193.x

Treem, W. R. (2012). *Clinical aspects and treatment of congenital sucrose-isomaltase deficiency*. *55*(November), 7–13. https://doi.org/10.1097/01.mpg.0000421401.57633.90

Trimble, R. B., Atkinson, P. H., Tschopp, J. F., Townsend, R. R., & Maley, F. (1991). Structure of oligosaccharides on Saccharomyces SUC2 invertase secreted by the methylotrophic yeast Pichia pastoris. *Journal of Biological Chemistry*, *266*(34), 22807–22817.

Uhrich, S., Wu, Z., Huang, J.-Y., & Scott, C. R. (2012). Four Mutations in the SI Gene Are Responsible for the Majority of Clinical Symptoms of CSID. *Journal of Pediatric Gastroenterology and Nutrition*, *55*, S34–S35. https://doi.org/10.1097/01.mpg.0000421408.65257.b5

van Leeuwen, J., Andrews, B., Boone, C., & Tan, G. (2015). Rapid and Efficient Plasmid Construction by Homologous Recombination in Yeast. *Cold Spring Harbor Protocols*, *2015*(9), 853–862. https://doi.org/10.1101/pdb.prot085100

Vanz, A. L., Nimtz, M., & Rinas, U. (2014). Decrease of UPR- and ERAD-related proteins in Pichia pastoris during methanol-induced secretory insulin precursor production in controlled fed-batch cultures. *Microbial Cell Factories*, *13*(1), 1–10. https://doi.org/10.1186/1475-2859-13-23

Vichayanrat, A., Ploybutr, S., & Tunlakit, M. (2002). *Efficacy and safety of voglibose in comparison with acarbose in type 2 diabetic patients*. *55*, 99–103.

Vieira-Gomes, A., Souza-Carmo, T., Silva-Carvalho, L., Mendonça-Bahia, F., & Parachin, N. (2018). Comparison of Yeasts as Hosts for Recombinant Protein Production. *Microorganisms*, *6*(2), 38. https://doi.org/10.3390/microorganisms6020038

Wadood, A., Ghufran, M., Khan, A., Sikander, S., Jelani, M., & Uddin, R. (2018). Glycosidase

inhibitors : A patent review (2012 – present). *International Journal of Biological Macromolecules*, *111*, 82–91. https://doi.org/10.1016/j.ijbiomac.2017.12.148

West, A. R., & Oates, P. S. (2005). Decreased sucrase and lactase activity in iron deficiency is accompanied by reduced gene expression and upregulation of the transcriptional repressor PDX-1. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *289*(6), G1108–G1114. https://doi.org/10.1152/ajpgi.00195.2005

Winchester, B. G., Cenci di Bello, I., Richardson, A. C., Nash, R. J., Fellows, L. E., Ramsden, N. G., & Fleet, G. (1990). The structural basis of the inhibition of human glycosidases by castanospermine analogues. *Biochemical Journal*, *269*(1), 227–231. https://doi.org/10.1042/bj2690227

Wuthrich, M., & Sterchi, E. E. (1997). Human lactase-phlorizin hydrolase expressed in COS-1 cells is proteolytically processed by the lysosomal pathway. *FEBS Letters*, *405*(3), 321–327. https://doi.org/10.1016/s0014-5793(97)00206-8

Yook, C., & Robyt, J. F. (2002). Reactions of alpha amylases with starch granules in aqueous suspension giving products in solution and in a minimum amount of water giving products inside the granule. *Carbohydrate Research*, *337*(12), 1113–1117. https://doi.org/10.1016/S0008-6215(02)00107-6

Yoshikawa, M., Murakami, T., Yashiro, K., & Matsuda, H. (1998). Kotalanol, a potent alpha-glucosidase inhibitor with thiosugar sulfonium sulfate structure, from antidiabetic ayurvedic medicine Salacia reticulata. *Chem. Pharm. Bull*, *46*, 1339–1340.

Appendices

Appendix A. Primers

Primer	Sequence (5'-3')
+6His_BshTI.twd	ATGaccggtGCGGCCGCAGAAAACCTGTATTTTCAGGGCCTCGAGCACC ACCACCACCAC
+6His_KpnI.rev	TTAGGTACCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG
alphaFactor.fwd	gaattcATGAGATTTCCTTCAATTTTTACTGCAGTTTTATTCGCAGCATCC TCCGCATTAGCT
AlphaFactor.rev	actagtAGCTAATGCGGAGGATGCTGCGAATAAAACTGCAGTAAAAATT GAAGGAAATCTCAT
AOX1.fwd	GACTGGTTCCAATTGACAAGC
CYC1.rev	GCGTGAATGTAAGCGTGAC
CtSI_homol.fwd	ATTCGCAGCATCCTCCGCATTAGCTGGGGGATCCACTAGTATAAAGT TACCTTCTGACCC
CtSI_homol.rev	GGTGCTCGAGGCCCTGAAAATACAGGTTTTCTGCGGCCGCtaaggtgtatcta atatttag
LoxP1.fwd	ATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCTTCGTAC GCTGCAGGTCGAC
LoxP2.rev	CGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCATAGG CCACTAGTGGATCTG
ntSI_homol.fwd	ATTCGCAGCATCCTCCGCATTAGCTGGGGGATCCACTAGTactaagacacct gctgttgatg
ntSI_homol.rev	GGTGCTCGAGGCCCTGAAAATACAGGTTTTCTGCGGCCGCccattgaacact aaagtttc
SI_homol.fwd	ATTCGCAGCATCCTCCGCATTAGCTGGGGGATCCACTAGTactaagacacct gctgttgatg
SI_homol.rev	GGTGCTCGAGGCCCTGAAAATACAGGTTTTCTGCGGCCGCtaaggtgtatcta atatttag
Zeo_KpnI.fwd	GGTACCAATGTTTCTACTCCTTTTTT
Zeo_SmiI.rev	ATTTAAATTCAGTCCTGCTCCTCGGCCACGAAG

Appendix B. K_M and size of the selected alpha-glucosidases according to BRENDA data base.

Human Protein	Size (KDa)	K_M Maltose [mM]	Host organism	reference
SI_{FL}	130	/	Human intestinal mucosa	(Kano, Usami, Adachi, Tatematsu, & Hirano, 1996)
	220	/	Adult jejunum	(Beaulieu, Weiser, Herrera, & Quaroni, 1990)
	245	/	Human intestine biopsies	(Fransen, Hauri, Ginsel, & Naim, 1991)
	280	/	Human intestinal mucosa	(K A Conklin, Yamashiro, & Gray, 1975)
NtSI	100	7.1	Drosophilla S2 cells	(Sim et al., 2010a)
	/	6.83	Duodenal biopsies homogenate	(Quezada-Calvillo et al., 2008b)
	148	/	Adult jejunum	(Beaulieu et al., 1990)
	150	/	Adult colon	(Beaulieu et al., 1990)
	151	/	Human intestine biopsies	(Fransen et al., 1991)
CtSI	130	/	Human intestinal mucosa	(Kenneth A. Conklin, Yamashiro, & Gray, 1975)
	145	/	Human intestine biopsies	(Fransen et al., 1991)
MGAM	/	0.58	Human intestine biopsies	(Quezada-Calvillo et al., 2008b)
NtMGAM (mal)	/	4.3	Drosophilla S2 cells	(Jones et al., 2011)
		7.71	Human intestine biopsies	(Quezada-Calvillo et al., 2008b)
	105.4	/	Drosophilla S2 cells	(Rossi et al., 2006)
CtMGAM (glam)	/	/	/	/
LPH (mature form)	170	/	Human Brush-border membranes	(Lau, 1987)
	160	28.9 (lactose)	Caco-2 cells	(Beau, Cotte-Laffitte, Geniteau-Legendre, Estes, & Servin, 2007)
	160	30 (lactose)	Caco-2 cells	(Beau et al., 2007)
	160	0.44 (phlorizin)/ 21 (lactose)	Autopsy intestine	(Skovbjerg, Sjostrom, & Noren, 1981)
	145	/	COS-1 cells	(Wuthrich & Sterchi, 1997)
LPH III	/	/	/	/
LPH III, IV	/	/	/	/

This enzyme also comprises two subunits, but unlike SI_{FL} the subunits are folded and transported completely independent from each other, therefore, there is no need for expressing a MGAM full length. From the expression of the individual subunits, two enzymes similar in size were obtained (Figures.3.5 and 3.6). This was expected as the predicted size based on amino acid sequence only show a difference <10KDa. Interestingly, based on the amino acid sequence, similar sizes were predicted between the MGAM subunits and their SI homologous, but the MGAM enzymes displayed slightly larger bands. A possible explanation for this observation may be part of the glycosylation pattern, as glycosylations are not considered in the size prediction.

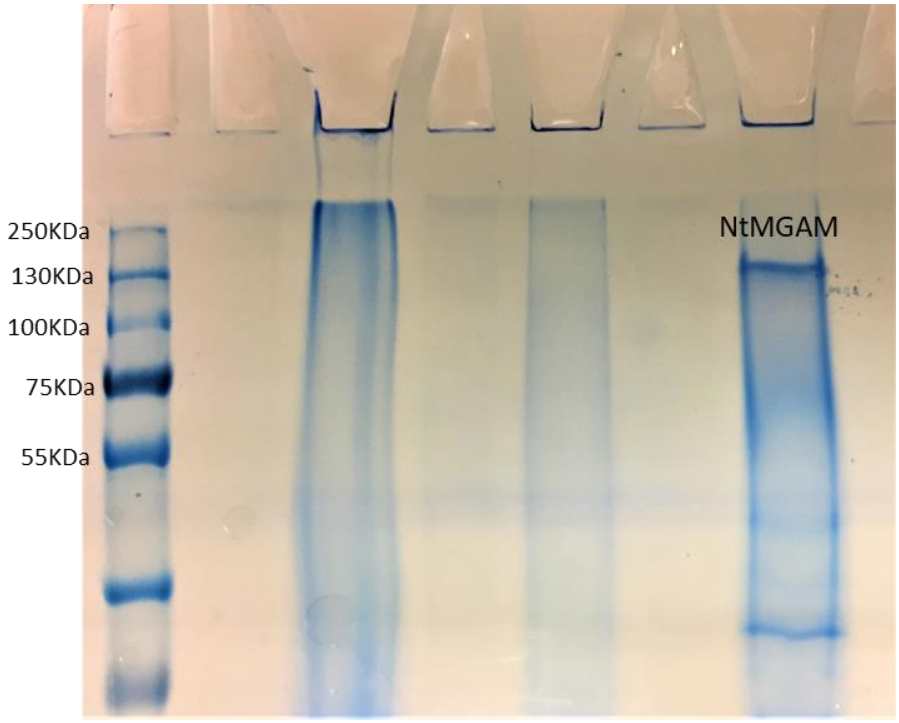


Figure 3.5. 8% SDS-PAGE showing the elutions of NtMGAM. The protein eluted with 100mM imidazole elution buffer is shown in the last lane.

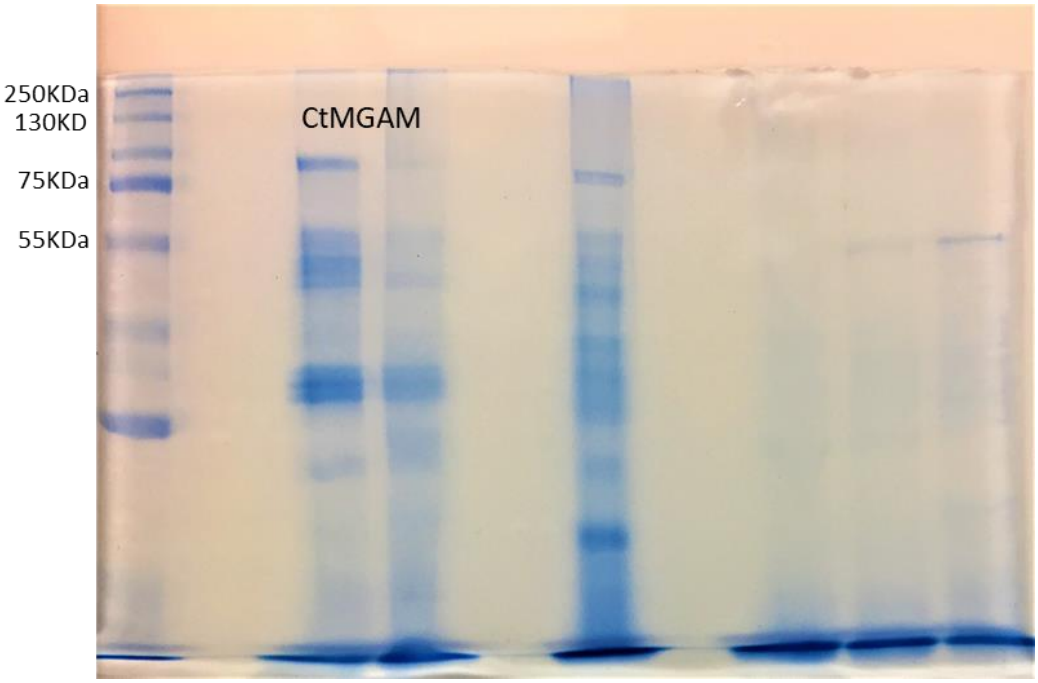


Figure 3.6. 8% SDS-PAGE showing the elutions of CtMGAM after purification by gravity flow chromatography using Nickel resin affinity, eluted with 100mM imidazole elution buffer.

The MGAM proteins here presented were expressed using one of the previous versions of the vector (pPinkNN), with a shorter histidine chain; therefore, the purification was less efficient, and as consequence, the K_M values obtained are less confident. Despite that, all the controls were run in the same plate and none of the endogenous proteins in *K. phaffii* showed any maltase activity.

The K_M values obtained with this version of the proteins were similar between NtMGAM and CtMGAM (Table 3.4). However, they follow expected trend of CtMGAM showing higher affinity towards maltose than NtMGAM (Jones et al., 2011).