Improving the protein secretion capacity of *Saccharomyces cerevisiae* with strain engineering

Heinrich Kroukamp

Dissertation presented for the degree of Doctor of Philosophy at Stellenbosch University

Promotor: W.H. van Zyl

Supervisor: R. den Haan

March 2015

DECLARATION

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

This dissertation includes one original paper published in peer-reviewed journals or books and two unpublished publications. The development and writing of the papers (published and unpublished) were the principal responsibility of myself and, for each of the cases where this is not the case, a declaration is included in the dissertation indicating the nature and extent of the contributions of co-authors. Selections of Chapter 2 were also published in two peer -reviewed review articles.

Signature: Date: 11 Feb. 2015

Declaration by the candidate:

With regard to Chapters 1 to 6, the nature and scope of my contribution were as follows:

Nature of contribution	Extent of contribution
Planning and execution of experimental work, data analysis and preparation of draft manuscript	90%

The following co-authors have contributed to Chapters 2 to 5:

Name	e-mail address	Nature of contribution	Extent of contribution
Dr. R den Haan	rdenhaan@uwc.ac.za	Did the SDS-PAGE and zymograms for Chapter 3. Constructed parental strains for Chapter 5.	20% of Chapter 3 5% of Chapter 5
Dr. N van Wyk	nvw@sun.ac.za	Constructed Cel6A producing strains	10% of Chapter 3
Prof WH van Zyl	whvz@sun.ac.za	Promoter – intellectual input	Input from promoter not quantifiable

Signature of candidate:

Date: 11 Feb. 2015

Declaration by co-authors:

The undersigned hereby confirm that

1. the declaration above accurately reflects the nature and extent of the contributions of the candidate and the co-authors to Chapters 2 to 5,
2. no other authors contributed to Chapters 2 to 5, besides those specified above, and
3. potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapters 2 to 5 of this dissertation.

Signature	Institutional affiliation	Date
R den Haan	University of the Western Cape	12 Dec 2014
N van Wyk	Stellenbosch University	12 Dec 2014
WH van Zyl	Stellenbosch University	12 Dec 2014

SUMMARY

The yeast *Saccharomyces cerevisiae* is frequently chosen for the production of industrial and pharmaceutical proteins, due to its rapid growth, microbial safety, eukaryotic post-translational processing and high-density fermentation capability. With the development of recombinant DNA technologies and efficient expression systems, this yeast also gained a prominent new role as a protein production host, due to its ease of genetic manipulation. Improving the production and secretion of recombinant proteins, whether for pharmaceutical, agricultural or industrial application, has the benefit of reducing the production costs and promoting accessibility to these technologies. This also holds true for the second generation biofuel production, where high levels of hydrolytic enzymes are required to break down the complex carbohydrates in lignocellulose. While high copy number expression vector systems, strong promoters and efficient secretion signals resulted in significant enhancement of protein production yields, these strategies are often limited by bottlenecks in the yeast secretion pathway. Although protein characteristics and host restrictions are likely to contribute to these bottlenecks, these factors are still poorly understood. Nonetheless, strain engineering approaches have shown great potential as a means to relieve these protein secretion bottlenecks and advancing recombinant protein production to new levels.

In this study, the potential of strain engineering, with regards to enhancing cellulase secretion for second generation bioethanol production, was evaluated. Further work involved the identification of novel genetic elements enhancing protein secretion and the elucidation of possible mechanisms involved in high cellulase secretion by *S. cerevisiae*.

When the native *PSEI* gene was expressed under the transcriptional control of the *PGK1* promoter in *S. cerevisiae*, the secreted yields of recombinantly produced *Neocallimastix patriciarum* Cel6A, *Trichoderma reesei* Cel7B and *Saccharomycopsis fibuligera* Cel3A were increased 1.15, 1.25 and 3.70 fold, respectively. The overexpression of *SOD1* did not increase any of the above mentioned cellulases. When *SOD1* was overexpressed in combination with *PSE1*, a synergistic enhancement in secreted Cel3A was obtained. To our knowledge, this is the first reported case where *SOD1* overexpression in *S. cerevisiae* resulted in higher heterologous protein secretion.

The effect of disrupting protein *N*-glycosylation elongation at various steps, on the secretion of heterologously produced *Neosartorya fischeri* Cel12A (lacking *N*-glycosylation sites) and the *S. fibuligera* Cel3A, was investigated. The deletion of the *MNN2* gene was shown to increase the extracellular Cel12A by 1.30 fold, while the deletion of *MNN11* resulted in a 1.26 fold increase in extracellular Cel3A. These results implicate the cell wall as a possible barrier to protein secretion. It was also found that Cel3A with shorter *N*-glycosylation chains had reduced cell wall retention, compared to enzymes resembling the native glycosylation pattern. The removal of the *PMR1* gene product (predicted to result in a general decrease in Golgi mannosyltranferase activities) was the only modification which enhanced the cell specific activities of both reporter cellulases, although this mutant's poor growth makes it an unlikely candidate for industrial application.

The *S. cerevisiae* M0341 strain that secretes high levels of recombinant *Talaromyces emersonii* Cel7A, was mated to the Y294 control strain to produce the H3 hybrid strain. Several high secreting progeny were selected after sporulating the H3 strain. Through genome shuffling and single nucleotide polymorphism (SNP) analysis of pooled segregants, five genomic regions of the M0341 strain were identified that contain putative alleles that are beneficial to Cel7A secretion. Identifying these alleles proved problematic due to strain instability. When the *T. emersonii CEL7A*, *N. fischeri CEL12A* and *S. fibuligera CEL3A* were expressed in selected H3 progeny on episomal plasmids, these strains had up to ~3.5 fold increased Cel7A secretion compared to the M0341 parental strain, but no increase were observed for the other cellulase. It was also shown that cell flocculation could improve secretion.

The strains constructed in this study represent a step toward efficient cellulase secreting yeasts for second generation biofuel production and presents a novel strategy to identify secretion enhancing elements for the protein production industry.

OPSOMMING

Die gis, *Saccharomyces cerevisiae*, word dikwels vir die produksie van industriële en farmaseutiese proteïene gekies, te danke aan sy vinnige groei tempo, veiligheid van gebruik, eukariotiese na-translasie modifisering van proteïene en hoë-digtheid fermentasie vermoë. Met die ontwikkeling van rekombinante DNS-tegnologieë en effektiewe geenuitdrukking sisteme, het die gis 'n prominente nuwe rol verwerf as 'n proteïen produksie gasheer vanweë die eenvoud waarmee dit geneties gemanulipeer kan word. Die verbetering in produksie en sekresie van rekombinante proteïene, hetsy vir farmaseutiese, landbou of industriële gebruik, bevoordeel laer produksie kostes en meer algemene toegang tot hierdie produkte. Hierdie voordele is ook belangrik vir die “tweede generasie” bio-brandstof bedryf, waar hoë vlakke hidrolitiese ensieme benodig word om die komplekse koolhidrate in lignosellulose af te breek. Hoë kopie getal uitdrukkings vektore, sterk promoters en effektiewe sekresie seine het aansienlike verbeteringe in proteïen produksie vlakke te weeg gebring, maar hierdie strategieë kan soms verstoppings in die sekresie weg veroorsaak. Alhoewel dit bekend is dat proteïen eienskappe en beperkinge van die uitdrukkings gasheer moontlik tot hierdie sekresie weg verstoppings kan bydrae, word hierdie faktore nog sleg verstaan. Nietemin, genetiese ingenieurswese van uitdrukkings rasse het die poteinsiaal om die verstoppings in die sekresie weg te verlig en sodoende die sekresie van rekombinante proteïene te verbeter.

In hierdie studie was die potensiaal van genetiese ingenieurswese, met betrekking tot die verbetering van sellulase sekresie vir tweede generasie bio-etanol produksie, geëvalueer. Verder was daar gepoog om nuwe genetiese elemente te identifiseer wat proteïen sekresie kan verhoog en om insig oor die meganismes wat betrokke is by hoë sellulase sekresie in *S. cerevisiae* te bekom.

Die ooruitdrukking van die natuurlike *PSE1* geen, onder die transkriptionele beheer van die *PGK1* promoter in *S. cerevisiae*, het die sekresie vlakke van rekombinant geproduseerde *Neocallimastix patriciarum* Cel6A, *Trichoderma reesei* Cel7B en *Saccharomycopsis fibuligera* Cel3A respektiwelik met 1.15, 1.25 en 3.70-voud verbeter. Die ooruitdrukking van die natuurlike *SOD1* geen kon nie die sekresie van die bogenoemde sellulases verhoog nie, maar die gesamentlike ooruitdruking van *SOD1* en *PSE1* was wel in staat om 'n sinergistiese verhoging in Cel3A sekresie vlakke teweeg te bring. Tot ons kennis, is dit die eerste geval waar verhoogde

heteroloë proteïen sekresie a.g.v. die ooruitdrukking van *SOD1* in die gis *S. cerevisiae*, wat aangemeld is.

Vervolgens het ons die effek wat die ontwrigting van proteïen *N*-glikosilering-ketting verlenging, gedurende verskillende stappe van die proses, op die heteroloë produksie van *Neosartorya fischeri* Cel12A en die *S. fibuligera* Cel3A ondersoek. Delesie van die *MNN2* geen het 'n 1.3-voudige toename in Cel12A sekresie teweeg gebring, terwyl die delesie van *MNN11* ekstrasellulêre Cel3A 1.26-voudig laat toeneem het. Hierdie resultate impliseer dat die selwand as 'n moontlike hindernis optree gedurende proteïen sekresie. Dit is gevind dat rekombinante Cel3A met korter *N*-glikosilerings kettings 'n laer affiniteit vir selwand assosiasie het, teenoor die ensiem met die natuurlike glikosilerings patroon. Die delesie van die *PMR1* geen (wat 'n algeneme verlaging in Golgi mannosieltransferase aktiwiteit veroorsaak) was die enigste verandering wat die sel-spesifieke aktiviteite van beide Cel3A en Cel12A kon verhoog. Ten spyte van dié mutante ras se verbeterde sekresie, is dit ongeskik vir industriële toepassing weens vertraagde groei.

Die *S. cerevisiae* M0341-ras wat hoë vlakke van die rekombinante *Talaromyces emersonii* Cel7A produseer, is met die Y294 kontrole ras gepaar om die H3 hibried ras te genereer. Nadat die H3-ras gesporuleer is, is menigte afstammelinge met ‘n hoë sekresie fenotipe geselekteer. Deur genoom skommeling en die analiese van enkel nukleotied-polimorfismes (ENPs) van die geselekteede afstammelinge, kon vyf genoom areas van die M0341 ras geïdentifiseer word wat moontlike allele bevat wat voordelig is vir die sekresie van Cel7A. Die identifikasie van die spesifieke allele was egter nie moontlik nie a.g.v. die onstabiliteit van die afstammeling genome. Dit kon wel gedemonstreer word dat sel flokkulasie die sekresie van Cel7A bevoordeel. Die uitdrukking van die *T. emersonii* CEL7A geen met hoë kopie getal episomale plasmiede in geseleteerde H3 afstammelinge, het 'n verhoging van tot ~3.5-voudig in vergelyking met die M0341 ras, tot gevolg gehad. Soortgelyke uitdrukking van die *S. fibuligera* CEL3A en *N. fischeri* CEL12A in hierdie afstammelinge het geen sekresie verbetering tot gevolg gehad nie.

Die rasse wat tydens hierdie studie gegenereer is, verteenwoordig beduidende vordering tot effektiewe sellulase sekresie in gis vir tweede generasie bio-brandstof produksie en demonstreer 'n nuwe strategie om genetiese elemente op te spoor wat die sekresie van verskeie industriële en farmaseutiese proteïen produkte kan verhoog.

To my parents, for all the sacrifices you made to further my education.

ACKNOWLEDGMENTS

I wish to express my sincere gratitude and appreciation to the following persons and institutions for their invaluable contributions to the successful completion of this study:

- My parents, **Marius** and **Anna-Rita Kroukamp**, who supported me (financially and emotionally) through all my years at university.
- My brother, **Divan Kroukamp**, for being my financial buffer and understanding housemate during this study.
- **Prof. Emile van Zyl**, for granting me the opportunity to grow as a molecular biologist and for giving me the freedom to pursue my, sometimes crazy, ideas in the laboratory.
- **Dr. Riaan den Haan**, for the academic guidance and proofreading over these past five years. His funny stories always lightened the mood in the laboratory.
- **Dr. Niël van Wyk**, who inspired me to do molecular microbiology and his continued interest in my work.
- **Prof. Marinda Bloom**, for all the helpful discussions, advice and assistance with the bursary applications.
- **Dr. Shaunita Rose** and **Kim Bester**, for proof reading and sharing your scientific experience.
- **Prof. Johan Thevelein** for the opportunity to work in his laboratory for six months.
- My friends from Worcester, Stellenbosch, Leuven and Birmingham, for your emotional support and encouragement.
- All the people that made Lab 335 feel like home over the past six years: Lisa Warburg, John-Henry van Zyl, Steffi Davison, Adri Prins, Marlin Mert, Annatjie Hugo, Arista Kloppers, Helba Bredell, Dr. Danie la Grange and Dr. Jacques Smith.
- Everyone in the Microbiology Department.
- The SANERI chair, the Erasmus Mundus Action 2, the National Research Foundation and the University of Stellenbosch for financial support.
- The **Lord Almighty** for the strength and guidance.

PREFACE

The dissertation is presented as a compilation of separately introduced scientific manuscripts. Chapter 3 is written in a style of a journal to which the manuscript was submitted. Chapter 4 and 5 are in preparation and is written in a style of a journal to which the manuscript will be submitted. Extracts of this dissertation was published in two review articles.

Chapter 3: Overexpression of native *PSE1* and *SOD1* in *Saccharomyces cerevisiae* improved heterologous cellulase secretion

Published in Applied Energy (2013) Vol 102: 150-156.

Chapter 4: Protein *N*-glycosylation chain length alters the protein secretion capacity and cell wall retention of *Saccharomyces cerevisiae*

In preparation for publication

Chapter 5: Identifying genes conferring high cellobiohydrolase secretion in *Saccharomyces cerevisiae*

In preparation for publication

Review articles contributed to:

den Haan R, Kroukamp H, Mert M, Bloom M, Görgens JF, van Zyl WH. (2013) Engineering *Saccharomyces cerevisiae* for next generation ethanol production. J Chem Technol & Biotechnol. 88:983-991.

den Haan R, Kroukamp H, van Zyl JD, van Zyl WH. (2013) Cellobiohydrolase secretion by yeast: Current state and prospects for improvement. Proc Biochem. 48:1-12.

INDEX

Addendum

LIST OF ABBREVIATIONS

ADP	adenosine diphosphate
Asn	asparagine
ATP	adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
BSA	bovine serum albumin
CBH	cellobiohydrolase
CBP	consolidated bioprocessing
cDNA	complementary deoxyribonucleic acid
Chr	chromosome
CMC	carboxymethyl cellulose
COP	coat protein complex
CPY	carboxypeptidase Y
DCW	dry cell weight
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMS	ethyl methanesulfonate
ER	endoplasmic reticulum
ERAD	endoplasmic reticulum associated degradation
ERGIC	ER-Golgi intermediate compartment
Fab	fragment antigen-binding
FAD	flavin adenine dinucleotide
FOA	5-Fluoroorotic acid
gDNA	genomic deoxyribonucleic acid
GH3	glycoside hydrolase family 3
Glc	glucose
GlcNAc	*N*-Acetylglucosamine
GPI	glycosylphosphatidylinositol
GPIT	glycosylphosphatidylinositol transamidase
GTP	guanosine-5'-triphosphate
HIV	human immunodeficiency virus
HSE	heat shock element
HSP	heat shock protein
IC	intermediate compartment
IGV	Integrative Genomic Viewer
Man	mannose
mRNA	messenger ribonucleic acid
MULac	methylumbelliferyl-β-D-lactoside
MWCO	molecular weight cut-off
NSC	no significant change
NSF	*N*-ethylmaleimide-sensitive factor
OD	optical density
ORF	open reading frame
OT	oligosaccharyltransferase
PASC	phosphoric acid swollen cellulose
PCR	polymerase chain reaction
PDIs	protein disulfide isomerases
PI	propidium iodide
PM	plasma membrane
***p*NP**	*p*-nitrophenol
***p*NPC**	*p*-nitrophenyl-β-D-cellobioside
***p*NPG**	*p*-nitrophenyl-β-D-glucopyranoside
Pro	proline
PSWGSA	pooled-segregant whole-genome sequencing analysis
qPCR	quantitative polymerase chain reaction
QTL	quantitative trait locus
ROS	reactive oxygen species
SAM	Sequence Alignment/Map
SC	synthetic complete
scFv	single-chain variable fragment
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel electrophoresis
Ser	serine
SM	Sec1/Munc18-like
SMM	sulfometuron methyl

SNAP	soluble NSF Attachment Protein
SNARE	soluble NSF attachment protein receptor
SNP	single nucleotide polymorphism
SNV	single nucleotide variance
SRP	signal-recognition peptide
TCP	total cell protein
tER	transitional endoplasmic reticulum
TGN	trans-Golgi network
Thr	threonine
TMD	trans-membrane domain
UDP	uridine diphosphate
UGGT	UDP-glucose:glycoprotein glucosyltransferase
UPR	unfolded protein response
UPRE	unfolded protein response element
UV	ultra-violet
YNB	yeast nitrogen base
YPD	yeast extract, peptone, glucose
YPG	yeast extract, peptone, galactose

LIST OF FIGURES

LIST OF TABLES

Chapter 1

General introduction and project aims

1.1. General introduction

The yeast *Saccharomyces cerevisiae* has been exploited by humans for hundreds of years in the baking, brewing and winemaking industries (Borneman et al. 2013). With the advent of molecular biology and genetic engineering, this yeast gained a prominent new role as an expression host for native and recombinant proteins for the biopharmaceutical, agricultural and enzyme industries (Demain & Vaishnav 2009). The products of these industries advanced numerous fields, including medicine, diagnostics, nutrition, detergents, textiles, paper and energy production. The bio-energy field, especially that of liquid fuel production, has received much attention in recent years due to greater interest in energy security and to mitigate the negative environmental impact of fossil fuel combustion. Ethanol, a prime candidate for petrol substitution in motor engines, is currently primarily produced from sugar and starch-based resources, however these resources directly compete with human food and animal feed supply. These shortcomings could be addressed by utilizing the lignocellulose within non-food plant biomass, to produce ethanol or other relevant biofuels. Cellulases and hemicellulases are required to hydrolyze lignocellulosic material into fermentable sugars, but these enzyme additions surmounts to one of the largest expenses of such industries (Stephanopoulos 2007).

To realize cost effective second generation biofuel production, a consolidated bioprocessing (CBP) approach could reduce process cost by combining the substrate saccharification and subsequent fermentation (van Zyl et al. 2007). This approach entails the use of ethanologenic yeasts, able to hydrolyze and ferment the complex carbohydrates in lignocellulose. Although no natural yeast has been identified to date with all the desired qualities for such a process, many advances have been made in producing heterologous cellulases and hemicellulases in *S. cerevisiae* (van Rooyen et al. 2005; Den Haan et al. 2007; La Grange et al. 2001; Ilmén et al. 2011). Even with efficient expression vectors, codon optimization and secretion signals, the secreted levels of these enzymes in yeast are still inadequate to enable efficient lignocellulose conversion to ethanol (du Plessis et al. 2010).

Strain engineering is emerging as a promising technique to introduce beneficial qualities, such as high protein secretion, into expression hosts. Although numerous studies have demonstrated the use of strain engineering approaches, mostly by overexpressing and deleting native genes to enhance secreted protein yields of medically relevant proteins, few examples exist for enhancing

cellulase secretion in yeast. These semi-rational design strategies usually aim to relieve secretion bottlenecks, using the secretion pathway as a guide and relying on transcriptomic data to identify the cellular responses to secretion stress (Gasser et al. 2007). Although these approaches have been successfully applied to enhance secreted protein yields of a variety of proteins, a wide range of secreted protein titers were observed (Idiris et al. 2010). The secretion enhancing ability of many strain engineering strategies vary depending on the reporter protein characteristics and the properties of the protein products that influence their transit through the secretion pathway and the pathways indirectly interacting with secretion machinery are still poorly understood (Robinson et al. 1994; Wentz & Shusta 2008; Kroukamp et al. 2013).

1.2. Aims of the study

The aims of this study were (1) to evaluate the potential use of strain engineering approaches to improve the secreted yields of industrially relevant cellulases in *S. cerevisiae*, (2) to identify possible new genetic targets to use as strain engineering tools and (3) to elucidate some of the underlying mechanisms of these gene modifications.

The objectives identified to satisfy these aims were:

1. To investigate the effect of native gene overexpression and deletion on the production and secretion of heterologously expressed cellulases in *S. cerevisiae*.
2. To assess the value of combining different overexpessed native genes and its effect on the secretion of cellulases.
3. To investigate how the degree of protein *N*-glycosylation effects cellulase secretion.
4. To establish a bioinformatic workflow to evaluate genomic differences between high and low *T. emersonii* Cel7A secreting *S. cerevisiae* strains.
5. To identify native genes or allele variations that could enhance the secretion of *T. emersonii* Cel7A.

1.3. Outline of dissertation

This dissertation is presented as a number of chapters consisting of a review of the relevant scientific literature (Chapter 2) and the research conducted to meet the aims of the study (Chapters 3 to 5). A general discussion and conclusions are presented in Chapter 6.

Strain engineering strategies are dependent on prior knowledge and a thorough understanding of the molecular mechanisms and pathway interdependencies involved, in order to rationally design and interpret the outcomes of such endeavors. For this reason, an overview of the yeast secretion pathway is presented as part of Chapter 2. This chapter also includes examples of previously reported protein secretion pathway engineering strategies.

The first three objectives were addressed in Chapters 3 and 4. Chapter 3 details the benefits of strain engineering strategies for cellulase secretion and how these beneficial genetic elements can be combined. Chapter 4 describes a possible mechanism through which *N*-glycosylation might benefit protein secretion. Chapter 5 describes the identification of possible novel enhancing elements from high secreting strains, to be applied independently or in combination, to improve secreted protein yields (last two objectives).

Chapter 6 addressed the outcomes of the study, the limitations and future prospects regarding strain engineering strategies as a means to enhance protein secretion.

1.4. Literature cited

Borneman AR, Schmidt SA & Pretorius IS (2013) At the cutting-edge of grape and wine biotechnology. *Trends Genet.* 29: 263–271.

Demain AL & Vaishnav P (2009) Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* 27: 297–306.

Den Haan R, Rose SH, Lynd LR & van Zyl WH (2007) Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. *Metab. Eng.* 9: 87–94.

Du Plessis L, Rose SH & van Zyl WH (2010) Exploring improved endoglucanase expression in *Saccharomyces cerevisiae* strains. *Appl. Microbiol. Biotechnol.* 86: 1503–1511.

Gasser B, Sauer M, Maurer M, Stadlmayr G & Mattanovich D (2007) Transcriptomics-based identification of novel factors enhancing heterologous protein secretion in yeasts. *Appl. Environ. Microbiol.* 73: 6499–6507.

La Grange DC, Pretorius IS, Claeyssens M & van Zyl WH (2001) Degradation of xylan to D-xylose by recombinant *Saccharomyces cerevisiae* coexpressing the *Aspergillus niger* beta-xylosidase (*XLND*) and the *Trichoderma reesei xylanase* II (*XYN2*) genes. *Appl. Environ. Microbiol.* 67: 5512–5519.

Idiris A, Tohda H, Kumagai H & Takegawa K (2010) Engineering of protein secretion in yeast: strategies and impact on protein production. *Appl. Microbiol. Biotechnol.* 86: 403–417.

Ilmén M, den Haan R, Brevnova E, McBride J, Wiswall E, Froehlich A, Koivula A, Voutilainen SP, Siika-Aho M, la Grange DC, Thorngren N, Ahlgren S, Mellon M, Deleault K, Rajgarhia V, van Zyl WH, Penttilä M (2011) High level secretion of cellobiohydrolases by *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 4: 30.

Kroukamp H, den Haan R, van Wyk N & van Zyl WH (2013) Overexpression of native *PSE1* and *SOD1* in *Saccharomyces cerevisiae* improved heterologous cellulase secretion. *Appl. Energy* 102: 150–156.

Robinson AS, Hines V & Wittrup KD (1994) Protein disulfide isomerase overexpression increases secretion of foreign proteins in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 12: 381–384.

Stephanopoulos G (2007) Challenges in engineering microbes for biofuels production. *Science* 315: 801–804.

Wentz AE & Shusta EV (2008) Enhanced secretion of heterologous proteins from yeast by overexpression of ribosomal subunit RRP0. *Biotechnol. Prog.* 24: 748–756.

Van Rooyen R, Hahn-Hägerdal B, La Grange DC & van Zyl WH (2005) Construction of cellobiose-growing and fermenting *Saccharomyces cerevisiae* strains. *J. Biotechnol.* 120: 284–295.

Van Zyl WH, Lynd LR, den Haan R & McBride JE (2007) Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae*. *Adv. Biochem. Eng. Biotechnol.* 108: 205–235.

Chapter 2

Protein secretion by *Saccharomyces cerevisiae*: fundamentals and biotechnology

2.1. Introduction

With the advent of molecular biology and genetic engineering, budding yeast gained a prominent new role as an expression host for native and recombinant proteins for the biopharmaceutical, agricultural and enzyme industries (Demain & Vaishnav 2009). For a protein, native or heterologously expressed, to be secreted as a mature product, it has to progressively move through various intracellular compartments before it's release from the cells. This route by which newly synthesised proteins reach the cell membrane, the endosomes or vacuole, is generally referred to as the secretion pathway (Schekman 1982). Through genetic engineering, the transit of protein cargo can be made more efficient, by relieving bottlenecks at key points of the pathway.

To rationally engineer the secretion pathway or to understand the changes brought about by gene alterations, it is imperative to grasp its complexities, understand its limitations and elucidate its driving mechanisms. In this chapter, we will review the yeast secretion pathway, with emphasis on common genetic engineering target areas, including vesicle fusion events, protein folding and quality control. We also review previous genetic engineering of this pathway in order to enhance proteins production and secretion.

2.2. The protein secretion pathway of *Saccharomyces cerevisiae*

To facilitate cellular growth, the cell employs its polarized actin cytoskeleton and bilayer lipid vesicles for the transport of macromolecules to organelles and sites of cellular expansion (Heider & Munson 2012). This multi-step route by which proteins are transported across cellular membranes towards specific organelles and/or the cell membrane is generally referred to as the secretion pathway. In eukaryotes, membrane fission and fusion processes occur at numerous transport steps to move cargo between intracellular organelles, to transmit information between cells and organs, to respond to external stimuli and for releasing substrate hydrolyzing enzymes into the environment (Malsam et al. 2008). The precise order of events eluded scientist for many years, until the pioneering work of Palade in 1975, demonstrating that secretory proteins must cross the endoplasmic reticulum (ER) before being transported to the plasma membrane. The involvement of the endoplasmic reticulum was also confirmed *in vitro* by Blobel and Dobberstein (1975) later that same year. These early studies were hampered by low levels of

intracellular precursors, few putative secretion organelles and the short time newly synthesised proteins takes to reach the plasma membrane; with less than 5 minutes required for Suc2p, a secreted invertase of the yeast *S. cerevisiae*, to be detected in the extracellular environment after synthesis (Novick et al. 1981).

In the following years, the secretion pathway was slowly pieced together with a major breakthrough made by Novick et al. (1981), where the ordering the intracellular events through which a secretory protein destined for the extracellular environment was elucidated. Using conditionally lethal temperature-sensitive secretory (*sec*) mutants which trapped secretory proteins at distinct steps of phenotypically distinguishable stages and double-mutant analysis, they could show which mutant phenotype was overshadowed by upstream processes (Novick et al. 1980, 1981). They determined that proteins destined to be secreted, first enter the ER lumen and are then transported by 40-60 nm vesicles to the Golgi apparatus, where further glycosylation occurs, and are finally transported to the plasma membrane in 80-100 nm vesicles to the tip of the emerging bud.

Although this study will focus on the forward steps by which secretory proteins reach the extracellular environment (as depicted in **Fig. 2.1**), it should be kept in mind that there are many branching nodes to other organelles and that other pathways can directly interact and influence the secretion of proteins. These branching pathways include the retro-translocation of proteins from the ER lumen to the cytoplasm (reviewed by Tsai et al. 2002), the retrograde transport of molecules in COP I (coat protein complex one) coated vesicles from the Golgi apparatus to the endoplasmic reticulum (reviewed by Duden 2003; Spang 2013), the endocytic pathway (Harsay & Schekman 2002) and the transport of vesicles to (and from) the vacuole, endosomes and lysosomes (reviewed by Bryant & Stevens 1998; Bowers & Stevens 2005).

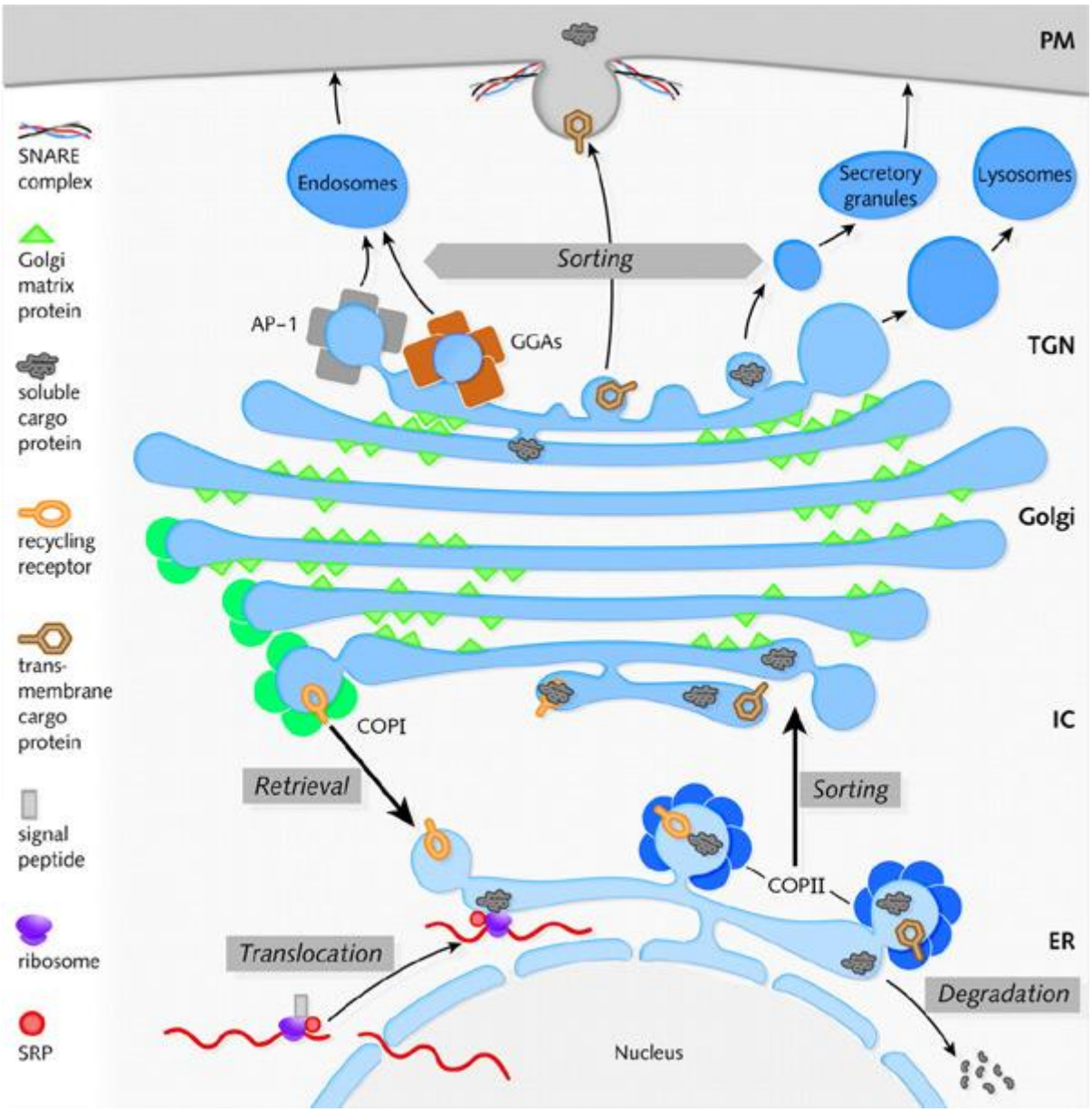


Fig. 2.1. A schematic representation of the secretion pathway of an eukaryotic cell. Newly synthesized polypeptides are delivered to the *cis* Golgi, mature as they transverse through intermediate compartments (IC) of the Golgi and get sorted by the trans-Golgi network (TGN) into secretion vesicles, destined for specific cell organelles or the plasma membrane (PM) (Figure obtained from http://embor.embopress.org/content/3/9/828).

2.2.1. Nascent protein targeting and translocation into the ER

Soluble proteins destined to be secreted from the cell or localized to internal compartments of cell organelles and membrane proteins are transported across the endoplasmic reticulum membrane. Soluble proteins usually have a cleavable amino acid signal sequence with a

characteristic 7-12 hydrophobic amino acid containing region which allows it to completely enter the endoplasmic reticulum lumen (Blobel & Sabatini 1971; Milstein et al. 1972). One or more segments of 20 hydrophobic amino acids are present in membrane proteins, with the hydrophilic regions residing on either side of the endoplasmic membrane when the protein is released laterally into the lipid bilayer.

Both soluble and membrane proteins are transported through the endoplasmic reticulum membrane resident Sec61 protein-conducting channel, either by co-translational translocation (SRP-dependent pathway) or posttranslational translocation (SRP-independent pathway) of the peptide (Blobel & Dobberstein 1975; Osborne et al. 2005). These pathways will be discussed in more detail below.

2.2.1.1. SRP-dependent pathway

In the SRP-dependent pathway, a signal sequence within proteins destined to be secreted into the extracellular environment, inserted into cell plasma-membrane or included in cellular organelles is recognised early during protein translation by a ribonucleoprotein called the signal recognition particle (SRP) (Walter & Johnson 1994).This SRP binds to the ribosome and cause temporary elongation arrest. The SRP directs this newly formed ribosome-SRP complex to the ER surface where it, with the addition of GTP, binds to the SRP-receptor protein located on the ER surface, bringing it into close proximity of a protein conducting channel with a hydrophilic interior. This evolutionarily conserved heterotrimetic membrane protein complex is termed the Sec61-complex in eukaryotes (also referred to as the translocon in older literature) (Walter & Lingappa 1986; Osborne et al. 2005).

The *S. cerevisiae* Sec61-complex consists of the Sec61p, Sbh1p and Sss1p. Sec61p, the largest of the subunits consist of ten membrane spanning domains, while Sbh1p and Sss1p each contribute one transmembrane alpha helix to the complex (Rapoport et al. 2004). Unlike the other two subunits, neither the Sbh1p, nor its closely related paralog Sbh2p are essential for growth; however temperature sensitivity and protein translocation defects are present when both are absent (Toikkanen et al. 1996). In *S. cerevisiae* a Sec61-like complex, the Ssh1 complex has been described which also facilitates cotranslational translocation, but unlike Sec61p, is not essential for cell viability (Finke et al. 1996).

The binding of the SRP-ribosome leads to the gating of the Sec61-channel and the nascent peptide chain is inserted into the channel. The SRP and its receptor protein are detached and the elongation arrest is lifted. As the peptide is being synthesised, the emerging peptide is pushed through the Sec61-channel into the ER-lumen and so co-translational translocation is achieved (Halic & Beckmann 2005). The SRP-dependent pathway is depicted in **Fig. 2.2**.

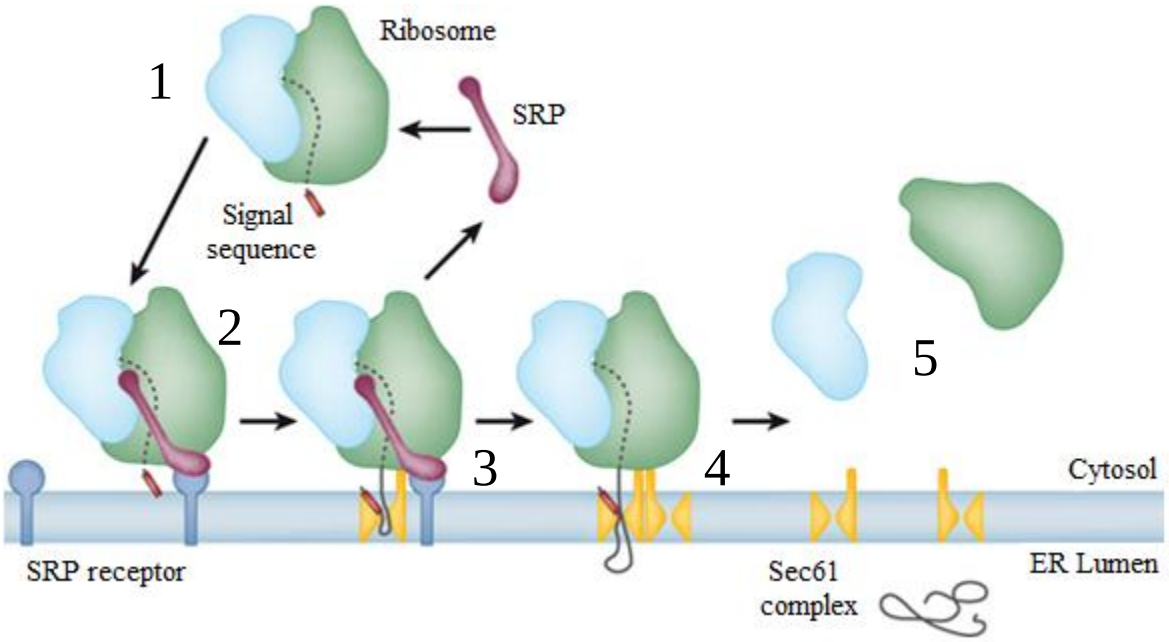


Fig. 2.2. The sequential steps of the co-translational translocation process of a peptide in the eukaryotic cell. (1) After the signal sequence of a peptide emerges, the signal recognition particle (SRP) binds to ribosome and translation is paused. (2) The SRP bound ribosome is targeted to the ER membrane, were the SRP attaches to the SRP-receptor on the ER membrane. (3) The signal sequence is inbedded in the membrane and the nascent peptide is subsequently inserted into the channel of the Sec61-complex (translocon). (4) The SRP/SRP-receptor detaches, translation resumes and the synthesised peptide is pushed into the ER-lumen. (5) In this example, the signal peptide gets cleaved off by signal peptidases, releasing the polypeptide into the ER lumen, concomitant with ribosome disassembly. This schematic was adapted from Rapoport 2007.

2.2.1.2. SRP-independent pathway

In many organisms the SRP-dependent pathway is essential for growth; however, *S. cerevisiae* mutant cells lacking the SRP or SRP receptor are viable, but display slow growth and impaired translocation of certain proteins across the ER membrane (Walter & Johnson 1994). This observation led to the elucidation of the SRP-independent pathway; the posttranslational mechanism by which unfolded polypeptides enter the ER-lumen without the assistance of the

SRP (**Fig. 2.3**). These peptides typically have a less hydrophobic signal sequence than those associated with the SRP-dependent pathway (Ng et al. 1996).

A different set of proteins is associated with the Sec61-channel for the SRP-independent translocation of proteins, namely the luminal ATPase Kar2p (member of the Hsp70 family and BiP homolog) and the tetrameric Sec62/63 membrane protein complex (Deshaies et al. 1991; Panzner et al. 1995). Also part of the Sec62/63-complex are two non-essential subunits, Sec71p and Sec72p, which are absent in the mammalian version of this complex (Meyer et al. 2000).

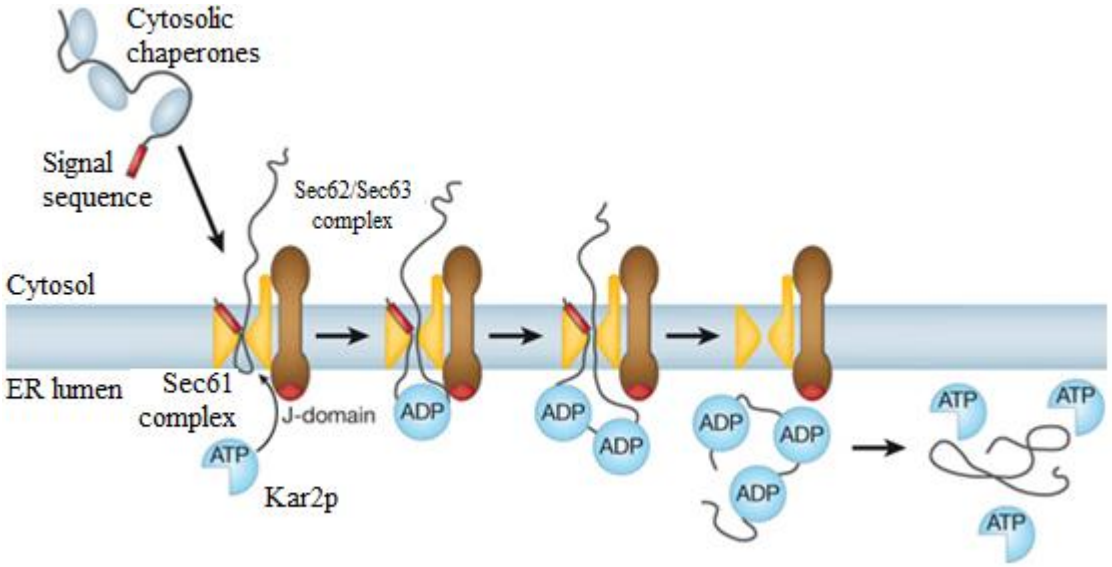


Fig. 2.3. Model of eukaryotic post-translational translocation. The interaction of Kar2p with the J-domain of the Sec63 complex results in hydrolysis of ATP and the closure of the Kar2p peptide-binding pocket around the emerging peptide. The binding of Kar2p prevents the translocated region of a polypeptide from slipping back into the cytosol. This schematic was adapted from Rapoport 2007.

Cytosolic chaperones and other putative targeting factors guide the unfolded polypeptides with a specific signal sequence towards the Sec61-complex, where the nascent peptide is inserted into the channel (Deshaies & Scheckman 1989). Posttranslational translocation of a peptide is proposed to occur by a ratcheting mechanism (Brownian ratchet), since the polypeptide can slide in any direction, but the binding to the ADP-Kar2p inside the ER-lumen prevents backward movement into the cytosol, resulting in a net forward (toward the ER-lumen) translocation (Matlack et al. 1999).

2.2.2. Protein quality control and folding in the endoplasmic reticulum

Proper post-translational modification and folding is essential for cell viability. Accumulation of aberrant proteins in the ER elicits a rapid cellular response to relieve this stress or succumb to programmed cell death (Shen et al. 2004). In addition to maintaining conditions conducive to proper protein folding through high chaperone concentrations and the unfolded protein response, the ER lumen also provides the environment for several covalent protein modifications to occur, including signal peptide removal, protein glycosylation and disulfide bond formation (Braakman & Hebert 2013). The interdependent process of protein folding and glycosylation occurs simultaneously and will be discussed in more detail below.

2.2.2.1. Protein glycosylation

Sugar-amino acid linkages are found in all living organisms and display great diversity in sugar type, oligosaccharide complexity, linked amino acid and position of linkage (Spiro 2002). Glycosylation is the biosynthesis of glycans, linked to proteins, and unlike DNA and protein synthesis, it is template independent and may be branched (Varki et al. 2009). At least thirteen different monosaccharides and eight amino acids have been implicated in this process. In eukaryotes, these co- or post-translational protein modifications play a role in cell physiology, cell polarity, protein targeting, protein recognition, regulation, cell-cycle progression, cell-cell interactions, cell wall properties, filamentation, protein folding, solubility, stability and protection against proteases (Varki 1993; Roth et al. 2010; Benton et al. 1996; Fiedler & Simons 1996; De Groot et al. 2005; Petkova et al. 2012; Yang et al. 2009; Loibl & Strahl 2013).

Based on the linkage type, protein glycosylation in yeast is divided into five groups, namely, glypiation, phosphoglycation, *C*-mannosylation, *N*- and *O*-glycosylation (Spiro 2002). The prominent types of protein glycosylation present in *S. cerevisiae* are glypiation, *N*- and *O*-glycosylation, and will be discussed in detail in the following sections.

2.2.2.1.1. N-linked protein glycosylation

The *β*-glycosylamine linkage of *N*-Acetylglucosamine (GlcNAc) to Asn (*GlcNAc*-β-*Asn*) was first described by Johansen and colleagues in 1961 for ovalbumin, and soon thereafter discovered in a vast array of other eukaryotic proteins. *N*-linked protein glycosylation has also been

described in archaea and eubacteria, although less prevalent than in eukaryotic systems (Yang & Haug 1979; Erickson & Herzberg 1993). These glycans are large, flexible hydrophilic modifications that can extend up to 3 nm from the glycoprotein surface (Hebert et al. 2005). *N*-linked glycosylation in eukaryotes occurs at certain Asn-Xaa-Ser/Thr sequons, where Xaa can be any amino acid except Pro (Weerapana & Imperiali 2006). It has also been shown that the amino acids directly adjacent to the Asn-Xaa-Ser/Thr sequon could influence the likelihood of oligoglycan transfer (Petrescu et al. 2004).

During the process of co-translational glycosylation, the signal peptidase complex containing Sec11p, cleaves the signal sequence and the ER-luminal oligosaccharyl transferase (OT) transfers a preassembled tetradecasaccharide "core unit" ($Glc_3Man_9GlcNAc_2$) to suitable asparagine amino acids as the nascent peptide emerges from the Sec61-channel (Bohni et al. 1988; Knauer & Lehle 1999). The transfer of the "core unit" is highly dependent on recognition of its three glucose units, since the transfer of $Glc_2Man_9GlcNAc_2$ occurs at greatly reduced rates in *alg10* deletion strains (Burda 1998).

The assembly of the core glycan requires the action of numerous Asparagine Linked Glycosylation (Alg) mannosyltransferase proteins and Rft1p, which is required for translocation of $Man_5GlcNAc_2$-PP-Dol from the cytoplasmic side to the luminal side of the endoplasmic membrane (**Fig. 2.4**). This assembly pathway is known as the dolichol pathway and is reviewed by (Weerapana & Imperiali 2006).

The addition of the core oligosaccharide to the peptide occurs concurrent with peptide synthesis by the ribosome, making the contribution of global folding and tertiary structure of the protein of minute consequence in determining if an Asn-Xaa-Ser/Thr sequon is suitable for oligosaccharide transfer (Weerapana & Imperiali 2006). More important however, is the local peptide secondary structure flanking the glycosylation site, since the structure should allow for a conformation switch from an Asx-turn to a β-turn after the addition of the oligosaccharide (Imperiali & O'Connor 1999; Imperiali & Hendrickson 1995; Imperiali & Rickert 1995). *N*-glycosylation is thus directly involved in protein folding and structure.

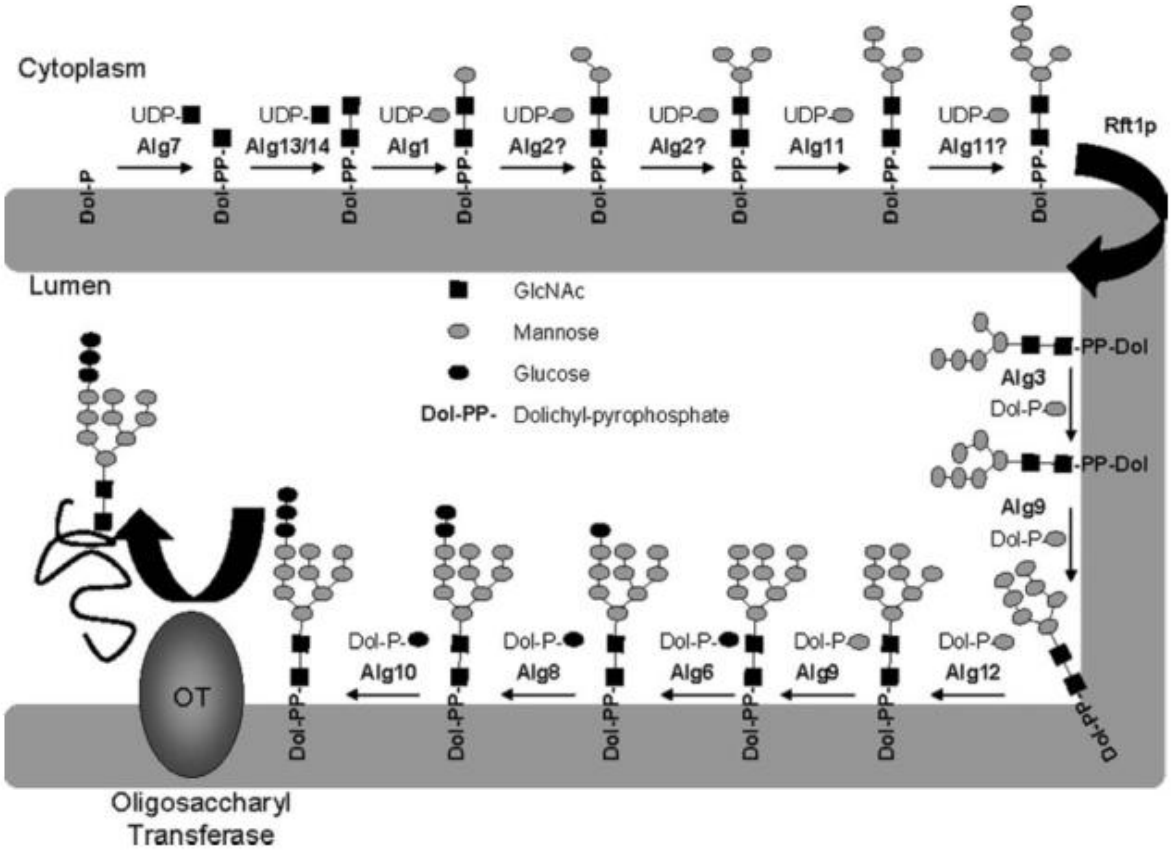


Fig. 2.4. The dolichol pathway of *S. cerevisiae*, illustrating the assembly and transfer of the conserved core oligosaccharide to newly synthesised proteins in the endoplasmic reticulum (Weerapana & Imperiali 2006).

After the addition of the core oligosaccharide to the polypeptide chain, the two terminal glucose units are removed by glucosidase I (Cwh41p) and II (Rot2p) respectively; allowing the protein to enter the lectin folding cycles (Brada & Dubach 1984; Hammond et al. 1994; Hammond & Helenius 1994). The lectin folding cycle will be discussed in more detail in **section 2.3.2.2.1.3**, along with other chaperones and protein folding mechanisms of the endoplasmic reticulum. The third glucose of the original $Glc_3Man_9GlcNAc_2$ core unit is subsequently removed by glucosidase II. Incorrectly folded proteins get reglucosidated and re-enter the folding cycles, while correctly folded proteins undergo α-1,2-mannose cleavage by the endoplasmic reticulum-resident mannosidase, Mns1p. The folded $Man_8GlcNAc_2$-containing glycoproteins are transported to the Golgi, where it will undergo further *N*-glycan processing (**section 2.2.3.2)**.

2.2.2.1.2. O-linked protein glycosylation

Another conserved type of glycosylation in eukaryotes is *O*-linked mannosylation, a prominent modification of many secreted and cell wall proteins (De Groot et al. 2005). Unlike animal cells, *O*-linked glycan synthesis in yeast begins in the endoplasmic reticulum instead of the Golgi (Gemmill & Trimble 1999). A single mannose is tranferred from Man-P-dolichol to the hydroxyl groups of specific serine and threonine amino acids in newly synthesised proteins, with the addition of additional mannose units after transfer to the Golgi.

These *O*-linked transferase reactions are catalysed by protein O-mannosyltransferase (PMT)-family enzymes, which comprise of several trans-membrane spanning regions and are grouped into 3 major clades, referred to as PMT1, PMT2 and PMT4 (Girrbach et al. 2000). Numerous proteins have been shown to serve as specific substrates for Pmt1p-Pmt2p and Pmt4p complexes, however in most cases these complexes share protein substrates; each complex acting on different regions of a given protein (Gentzsch & Tanner 1997).

Unlike the *N*-glycosylation Asn-Xaa-Ser/Thr sequon, no sequons or motifs have been identified for either of the *O*-mannosyltransferase complexes, suggesting a secondary or tertiary structure requirement for recognition inferring that it occurs after primary structure-based *N*-glycosylation (Loibl & Strahl 2013). This is however contradicted by Ecker et al. (2003), who showed that non-native *N*-linked glycosylation occurred for Ccw5p, in the absence of Pmt4p. They inferred from their study that *O*-mannosylation (at least that of the Pmt4p complex) takes place before *N*-glycosylation and sterically hinders the *N*-glycosylation machinery for site accessibility.

Although the recognition motifs of individual *O*-mannosylation sites are still in question, it has been observed that *O*-mannosylations are predominantly acquired by secreted proteins, cell wall proteins and proteins located in the secretion pathway (De Groot et al. 2005). Many of these proteins have one or more serine/threonine rich region. In *S. cerevisiae*, 45% of its signal peptide-containing proteins have a 20 amino acid region of at least 40% serine/threonine content (González et al. 2012).

As with *N*-glycosylation, *O*-glycosylation also plays a crucial role in protein folding and endoplasmic reticulum homeostasis, as defects in *O*-mannosylation constitutively induce the unfolded protein response (**section 2.2.2.2.2**), in addition to activating the cell wall integrity (CWI) pathway and the STE vegetative growth (pheromone response) pathway: another cell wall

affecting pathway (Arroyo et al. 2011). These extreme cellular responses to buffer the resulting cell wall aberrations emphasises the importance of the mannosylated protein fraction in cell wall integrity; being responsible for cell-cell interactions, cell wall permeability and biosynthesis (De Groot et al. 2005).

2.2.2.1.3. GPI anchors

Glycosylphosphatidylinositol (GPI) anchoring is a posttranslational protein modification initiated in the endoplasmic reticulum where a complex glycophospholipid (GPI) is added to the C-terminal end of a protein, anchoring it to the cell surface (Orlean & Menon 2007). These GPI anchored proteins contains long stretches of serine and threonine amino acids and are functionally diverse; responsible for normal cell morphology, cell-cell interactions, acting as receptors for environmental signals while many also have hydrolytic activities (Caro et al. 1997; Hamada et al. 1998b). GPI synthesis is the most metabolically expensive lipid posttranslational modification that has been described so far, but is indispensible since blocking steps in this pathway are lethal in yeast cells (Leidich et al. 1994). The saturated fatty acyl chains found in the monolayer penetrating GPI lipid portion permits it to associate with sphingolipid-rich rafts in the membrane, allowing movement throughout the membrane for their role in trafficking and signalling (Simons & Toomre 2000; Helms & Zurzolo 2004).

For a protein to receive a GPI attachment, it must enter the ER lumen and contain a C-terminal GPI signal anchor sequence. This GPI attachment recognition sequence consists of several distinct regions up and downstream of the amino acid receiving the GPI-moiety (referred to as the ω-amino acid), as indicated by **Fig. 2.5**. Towards the N-terminal side of the ω-amino acid lies a region of approximately 10 polar amino acids, while the C-terminal region consists of a moderately polar region followed by hydrophobic amino acids (Eisenhaber et al. 1998). The endoplasmic reticulum resident GPI transamidase complex (GPIT) recognises the GPI signal sequence and catalyses the displacement of the C-terminal region of the signal with GPI (Ohishi et al. 2000). This GPI transfer by the five subunit transamidase complex competes with molecular water.

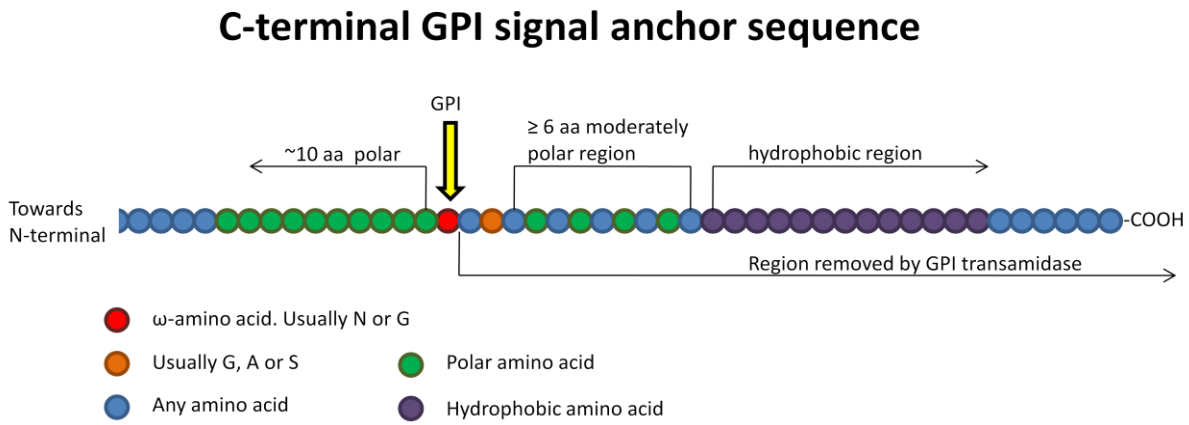


Fig. 2.5. A graphical illustration of the recognition signals present in the GPI anchor signal sequence.

In addition to modifications to the GPI moiety itself, another modification of the GPI occurs specifically in fungi by which some of the GPI proteins become cross-linked to the cell wall glycans, which takes place after passing through the secretion pathway (Lu et al. 1994, 1995) The GPI of a membrane-anchored protein is cleaved within the glycan region (losing the inositol-containing lipid), which is subsequently attached to cell wall β-1,6-glucans (Kapteyn et al. 1996). A study by Hamada et al. (1999) on the C-terminal signal suggests that these cell wall linkages could be GPI anchor signal dependent, since the presence of two basic amino acids in the polar region (N-terminal side of the ω-amino acid), in general, results in linkage to the cell wall glycans. The absence or removal of these basic amino acids, shifts the protein attachment to the cell membrane (Hamada et al. 1998a; 1999), although this might not be the sole determining factor for cell wall localization (Frieman & Cormack 2003).

2.2.2.2. Protein folding, recycling and ER homeostasis

The ER serves as the primary gateway for the secretion pathway, and to fulfill its role, it provides the catalytic machinery to stabilize newly synthesized polypeptides and assists them in obtaining their native conformations (Fewell et al. 2001). For many years it was believed that the ER quality control mechanisms used a fixed standard to determine if a protein obtained its native form, before it was allowed to continue its journey along the secretion pathway. This model, however, could not explain cell-specific secretion efficiencies of the same protein (Sekijima et al. 2003). Sekijima et al. (2005) proposed a cell type scoring system that took into account the energetics of the protein fold properties, quaternary structure assembly and chaperone enzyme

distributions, which competes with ER-associated degradation. Using this model they successfully predicted in which tissue a reporter protein will be secreted and in which it will be degraded before maturation.

Wiseman et al. (2007) refined this model by incorporating Michaelis-Menten kinetics of the associated pathways, providing an adaptable prediction model for protein homeostasis in the ER. This refined model evidently demonstrates that no single feature dictates folding and transport efficiency. It is not possible to cover all the network dynamics responsible for ER homeostasis in this dissertation; instead, key insights about the major processes involved in protein export efficiency, as described by Wiseman et al. (2007), will be discussed. These processes include folding and maturation of proteins, product export and ER-associated degradation. We will also discuss how the ER detects and responds to the accumulation of unfolded proteins in order to restore ER homeostasis.

2.2.2.2.1. Protein folding

To maintain protein homeostasis in the yeast cell, mechanisms within the cell that target proteins to their destined locations, assist peptides in obtaining their functional native form and prevent aggregation in cellular compartments are crucial processes for a cell's survival (Verghese et al. 2012). In eukaryotic cells, almost all membrane and secretory proteins enter the endoplasmic reticulum, where these newly synthesised peptides adopt an energetically favourable conformation, in which hydrophobic amino acids are buried on the inside of soluble proteins and hydrophilic residues are mostly found in solvent-accessible sites (Christis et al. 2008). The folded protein's conformation is kept in place by stabilizing hydrogen bonds, electrostatic, van der Waals' interactions and in some cases covalent bonds. The formation of native secondary and tertiary structure is called protein folding and the extension to this folding process which results in the quaternary structure is referred to as oligomerization.

According to the concept of funnel-like energy landscapes, peptides will generally adopt a conformation of minimum energy (Dobson 2004). Multiple routes and conformational intermediates exists for each protein falling towards this minimal energy state, but not all peptides are able to obtain their native functional form unassisted, especially in the crowded environment within the cell where the risk of aggregation is high (Leopold et al. 1992).

To reduce the formation of stuck intermediate folding conformations and unwanted protein aggregations, cells have a large number of different chaperones and folding factors to smoothen the energy landscape to increase the likelihood of proteins reaching their native form (Leopold et al. 1992). All cellular compartments of eukaryotic cells harbours protein chaperones, many of which are shared between the different organelles such as heat shock proteins (**section 2.2.2.2.1.1**), although the oxidative environment of the endoplasmic reticulum also exhibits folding mechanisms absent from the cytoplasm, such as lectin-associated folding (**section 2.2.2.2.1.3**) and disulfide bond formation (**section 2.2.2.2.1.2**) (Verghese et al. 2012). Although it is predicted that up to one-third of all *S. cerevisiae* proteins fold in the endoplasmic reticulum, even proteins destined for secretion might be assisted by cytosolic folding factors as seen for posttranslational translocated proteins (**section 2.2.1.2**) (Ghaemmaghami et al. 2003).

2.2.2.2.1.1. Heat shock proteins

In response to elevated environmental temperatures, yeasts elicit the highly conserved heat shock response, involving the repression of protein synthesis and the induction of cytoprotective genes, referred to as heat shock proteins (HSPs) (Verghese et al. 2012). Many of these HSPs are involved in the secretion pathway, at basal levels, assisting protein folding even under normal growth conditions. The HSPs involved in the secretory pathway of *S. cerevisiae* are listed in **Table 2.1**.

The sole endoplasmic reticulum resident Hsp70 family member present in the yeast ER is the mammalian BiP homolog, Kar2p. Kar2p is an essential protein with multiple and diverse roles, assisting with protein folding, peptide translocation, regulation of the unfolded protein response pathway and retrograde transport of aberrant peptides to the cytosol for proteasomal degradation (Marcinowski et al. 2011; Kimata et al. 2003; Matlack et al. 1999). Information on Kar2p's role in the detection and removal of non-native folded proteins from the endoplasmic reticulum lumen will be discussed in **sections 2.2.2.2.2** and **2.2.2.2.3**. Concurrent with its multiple functions, the Kar2p polypeptide has two targeting signals; an N-terminal secretory signal sequence for targeting to the endoplasmic reticulum and a C-terminal endoplasmic reticulum retention signal (HDEL). The mature protein consists of an N-terminal nucleotide-binding domain and a C-terminal substrate-binding domain (Tokunaga et al. 1998).

Table 2.1. Endoplasmic reticulum heat shock chaperones and co-chaperones.

Class	Protein	Function	References
Hsp70			
GRP170	Lhs1p	Kar2p nucleotide exchange, substrate binding	(Baxter et al. 1996)
Hsp70	Kar2p	Protein folding, translocation, unfolded protein response regulation, karyogamy	(Rose et al. 1989)
Hsp70 NEF	Sil1p	Kar2p nucleotide exchange	(Tyson & Stirling 2000)
Hsp40			
	Sec63p	Kar2p ATPase activator, translocation	(Young et al. 2001)
	Scj1p	Kar2p ATPase activator	(Schlenstedt et al. 1995)
	Jem1p	Kar2p ATPase activator, karyogamy	(Cyr et al. 1994)

The functionally diverse roles of Kar2p are matched by its transcriptional control mechanisms. The *KAR2* promoter has three *cis*-acting elements controlling the levels of transcription, depending on cellular needs and is induced by heat shock as well as accumulated unfolded proteins in the endoplasmic reticulum (Kohno et al. 1993). To facilitate peptide translocation and folding under non-stress conditions, the *KAR2* promoter allows basal constitutive expression through a GC-rich region that governs constitutive expression. This constitutive promoter region is flanked by a 20-bp heat shock element (HSE) (Amin et al. 1988; Xiao & Lis 1988) and a 22-bp unfolded protein response element (UPRE) (Mori et al. 1992); each element allowing several fold increases in transcription, in response to their specific stress condition and having an additive effect when both stress factors are present to allow maximal induction of the *KAR2* gene transcription (Kohno et al. 1993).

Kar2p has been implicated in both the early and late stages of protein translocation into the endoplasmic reticulum lumen for both co-translational and posttranslational translocation, although a more direct role was suggested for the latter process (Brownian ratchet theory; see **section 2.2.1.2**) (Brodsky et al. 1995). A temperature sensitive depletion of Kar2p reserves in yeast resulted in the accumulation of both prepro-α-factor and invertase precursors on the cytosolic side of the ER that prevented the maturation of posttranslational and co-translational translocated polypeptides, respectively (Sanders et al. 1992; Vogel 1990). Kar2p/BiP assistance

in protein folding is well documented (Normington et al. 1989) and a good example of its chaperone-function in yeast was documented by Simons et al. (1995). Using the well-characterized vacuolar glycoprotein carboxypeptidase Y (CPY) as reporter system and three temperature sensitive *KAR2* mutant strains, it was shown that Kar2p is necessary for protein maturation and it prevented the accumulation of CPY aggregates in the endoplasmic reticulum.

2.2.2.2.1.2. Thiol oxidoreductases

In eukaryotic cells, some intracellular proteins and the majority of secreted proteins require additional stabilization by disulfide bond formation to obtain their native functional forms (Holst et al. 1997). Disulfide bonds covalently link two regions of a polypeptide; thereby lowering its entropy and potentially generating a hydrophobic core of the folded protein.

Protein disulfide isomerases (PDIs) catalyse the formation or rearrangement of polypeptide disulfide bonds (Darby & Creighton 1995). PDI, a member of the thioredoxin family, contains the characteristic CXXC motif, with the two catalytically active cysteine residues. The PDI-catalysed reaction involves the reduction and oxidation of an internal disulfide bond between cysteine residues (Martin 1995). Although it is unsure whether the X-amino acids influence the activity of the thioredoxin motif, there is evidence that the position of this site within the protein secondary structure is important. Site directed mutagenesis studies have shown that the **a** thioredoxin-like domain of Pdi1p is catalytically more active than the **a'** domain (with identical amino acids to the **a** domain) of the same protein; yeasts harbouring a mutated Pdi1 with a deactivated **a'** catalytic site display no growth defects or reduction of carboxypeptidase Y folding rate in the ER (Nørgaard et al. 2001; Holst et al. 1997).

Eukaryotic PDI and PDI-like proteins reside in the oxidizing environment of the ER lumen. There are five ER resident thioredoxin family members in *S. cerevisiae* (**Fig. 2.6**), known as Mpd1p, Mpd2p, Eug1p, Eps1p and the well studied Pdi1p (Frand et al. 2000). Being responsible for the majority of the PDI activity in the ER, Pdi1p deletion is lethal to the cell (Farquhar et al. 1991; Nørgaard et al. 2001). The substrate specificity of the different yeast PDIs is unclear, but some functional overlap has been demonstrated. The overexpression of any of the other yeast PDIs suppresses the lethality of *PDI1* deletion, although efficient protein folding is only partially restored. Pdi1p and Mpd1p are the only yeast PDIs that, upon overexpression, are able to

suppress lethality in a yeast strain with deletions for all the other genes in the *PDI1* family (Nørgaard et al. 2001).

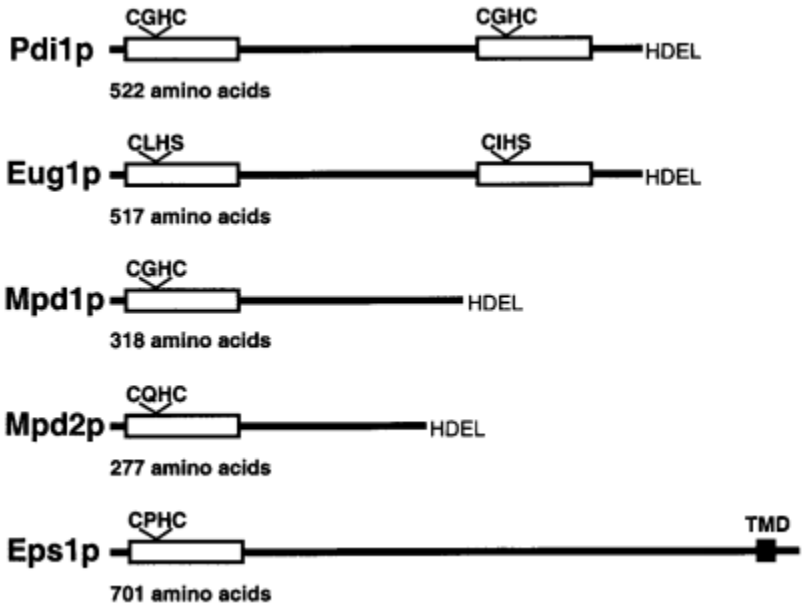


Fig. 2.6. *Saccharomyces cerevisiae* contains four soluble and one transmembrane domain (TMD) containing PDI-like proteins in the ER. Pdi1p and Eug1p have similar amino acid sequences, although the CLHS/CIHS motives of Eug1p are capable of disulfide isomerase activity, but unlike Pdi1p, Eug1p is unable to oxidize the substrate thiol groups (Holst et al. 1997; Nørgaard et al. 2001).

At least two (Pdi1p and Mpd2p) members of the yeast thioredoxin family members require the action of the ER thiol oxidase, Ero1p, to oxidize the thiol groups of their catalytic sites (Frand & Kaiser 1999). Ero1p is believed to partially maintain the oxidative conditions of the ER lumen. This was demonstrated by the enhanced tolerance of yeast strains overexpressing Ero1p to the reductant dithiothreitol (DTT) and by the observation that addition of the thiol oxidant diamide can restore carboxypeptidase Y folding and cell viability in *ero1* mutants (Cuozzo & Kaiser 1999). The flavoprotein Ero1p is able to reduce molecular oxygen and generating disulfide bonds to PDIs, generating hydrogen peroxide in the process (Gross et al. 2006; Benham et al. 2013).

All yeast PDI proteins, with the exception of Mpd1p, have been shown to possess additional chaperone abilities besides thio-group oxidation and disulfide bond reshuffling functions (Kimura et al. 2004; Hatahet & Ruddock 2009). Pdi1p has also been implied in the retrotranslocation of misfolded glycoproteins, by introducing a disulfide bond into the

mannosidase homology domain of Mnl1p (**section 2.2.2.2.1.3**) and directly interacting with it as a functional subunit (Tsai et al. 2002; Sakoh-Nakatogawa et al. 2009).

2.2.2.2.1.3. Lectin chaperone cycle

Sugar-chain processing is an important mechanism that eukaryotic cells employ to achieve proper folding and assembly of *N*-linked glycoproteins, which represent the majority of secretory proteins passing through the endoplasmic reticulum lumen (Lederkremer & Glickman 2005). After the transfer of the core oligosaccharide ($Glc_3Man_9GlcNAc_2$) to the protein, the two terminal glucose units are removed. Proteins with the monoglycosylated oligosaccharide ($Glc_1Man_9GlcNAc_2$) are retained in the ER, then recognised by Cne1p (a calnexin homologue), an ER lectin able to enhance the folding of glycoproteins and prevent their aggregation in a concentration dependent manner (**Fig. 2.7**) (Xu et al. 2004a; 2004b).

The third glucose is subsequently removed from the oligosaccharide by Glucosidase II; consequently allowing the glycoprotein to either (1) be marked as misfolded and proceed to the ERAD machinery for retrotranslocation and degradation (**section 2.2.2.2.3**), by removing a B-branch mannose by Mns1p (ER mannosidase 1) or (2) be allowed to exit the ER (Hebert et al. 2005; Knop et al. 1996; Jakob et al. 1998b).

In eukaryotes studied up to date, with the exception of *S. cerevisiae*, UDP-glucose: glycoprotein glucosyltransferase (UGGT) recognises (Glc_0)Man9GlcNAc2 containing glycoproteins and reglucosylates those displaying hydrophobic regions; consequently redirecting it back into the calnexin/calreticulin-folding cycle (Fernández et al. 1994). In *S. cerevisiae*, the lack of a UGGT in glycoprotein quality control might result in a leaky system, and based on data obtained from their laboratory, Jakob and co-workers (1998b) proposed alternative mechanisms by which yeast ensures efficient folding of glycoproteins. They argued that since the calnexin-oligosaccharide interaction prevents glucose cleavage (which serves as a signal to enter the subsequent quality control steps), a similar mechanism might be in place for Cne1p, but it might have more stringent requirements with regards to the proteins folding state before it is released from this lectin-chaperone (Zapun et al. 1997). Cne1p interaction might therefore be the determining factor for designating a protein for export or glycoprotein ER assosiated degradation (ERAD) (Jakob et al. 1998b; Suzuki & Lennarz 2002).

Jakob et al. (1998a) predicted the existence of a lectin, similar to the EDEM (ER degradation-enhancing α-mannosidase-like protein) lectins found in mammalian cells, able to recognise incorrectly folded $Man_9GlcNAc_2$-containing proteins, trimming one specific α-1,2-mannose residue from the B-branch which serves as a signal for ERAD. They concluded that the ER-localized α-1,2-mannosidase, Mns1p fulfils this role in *S. cerevisiae*, due to its *in vitro* removal of the B-branch mannose of free $Man_9GlcNAc_2$ oligosaccharides and delaying the degradation of ERAD model proteins in Mns1p-deficient cells (Jakob et al. 1998a). $Man_8GlcNAc_2$ oligosaccharides present on misfolded polypeptides serve as a signal for degradation and is recognized by Mnl1p (mannosidase-like 1) which removes an α1,2-mannose unit from the *N*-linked oligosaccharide C-branch to expose an α-1,6-mannosyl residue of the $Man_7GlcNAc_2$ glycan (Quan et al. 2008; Clerc et al. 2009). Although this cleavage of one specific mannose is seen as the minimum requirement for progression toward the protein's destruction, additional mannosyl residues might also be trimmed ($Man_{5-7}GlcNAc_2$) and recognised; supported by an observation that Mnl1p processing can be bypassed in *Δalg3* producing $Man_5GlcNAc_2$-glycan structures with an already exposed α-1,6-mannosyl residue (Jakob et al. 1998a).

The misfolded $Man_{5-7}GlcNAc_2$ polypeptides subsequently associate with an ER-membrane lectin, Yos9p, that recognises exposed α-1,6-mannosyl residues of Mnl1p trimmed products (Quan et al. 2008). Yos9p is part of the HRD ubiquitin ligase complex, that includes luminal quality control proteins and lectins (Yos9p, Kar2p and Hrd3p), the ubiquitin ligase Hrd1p, and several other transmembrane (Ubx2p and Der1p) and cytosolic factors (Cue1p, Ubc7 and Cdc48) involved in extraction and ubiquitination of the translocating polypeptide (Deak & Wolf 2001; Gauss et al. 2006).

Recognition of misfolded glycoproteins by Hrd3p and Yos9p serves as a bipartite quality control mechanism that assesses both folding state and glycan signal, before committing a polypeptide for retrotranslocation through the Sec61p channel to the cytocol, ubiquitination and subsequent degradation by the 26S proteasome (Denic et al. 2006; Hirsch et al. 2009; Määttänen et al. 2010). Detail on the ERAD machinery is discussed in **section 2.2.2.2.3** and for the cytosolic events leading to proteosomal peptide hydrolysis of misfolded polypeptides, the reader is referred to reviews by Nakatsukasa & Brodsky (2008) and Hirsch et al. (2009).

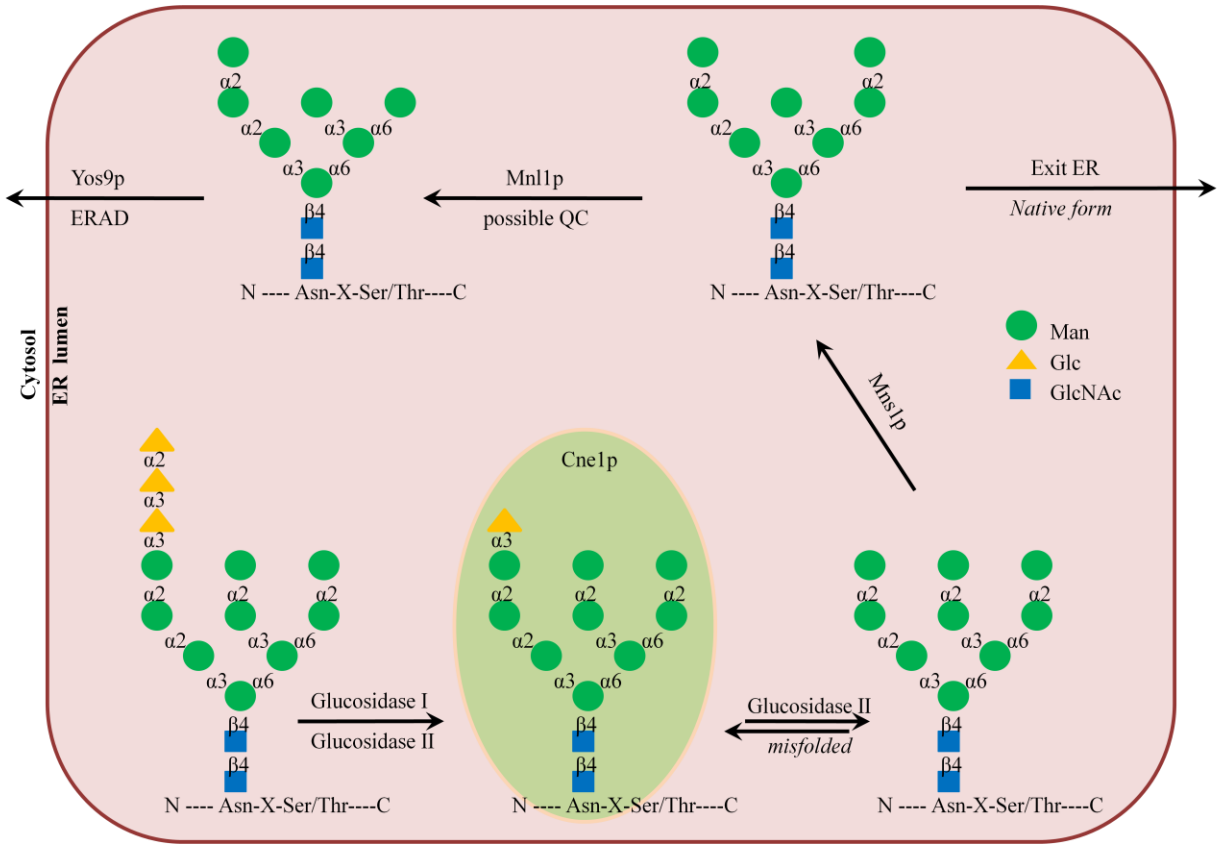


Fig. 2.7. The lectin-chaperone cycle in *S. cerevisiae.* After the transfer of the core oligosaccharide to the polypeptide, the outermost glucose units are cleaved. The resulting $Glc_1Man_9GlcNAc_2$ oligosaccharide is recognized by Cne1p, which binds to the glycoprotein and assist in its folding. Another quality control step determines if the protein is fit to continue through the secretion pathway, after removal of the final glucose unit. The subsequent removal of mannose units secures aberrant proteins' destruction by the ER-associated degradation pathway.

2.2.2.2.2. The unfolded protein response

For proteins to be secreted they must pass through the endoplasmic reticulum (ER) to be properly folded and modified (Chawla et al. 2011). However, when the folding demand of nascent polypeptides exceeds the capacity of the ER protein folding machinery, the cell elicits a stress response know as the unfolded protein response (UPR) through a signal transduction network from the ER to the nucleus (Schroder & Kaufman 2006).

The induction of the UPR is facilitated by the ER-transmembrane protein, Ire1p, which undergoes autophosphorylation and oligomerization when unfolded proteins are detected (**Fig. 2.8**). Oligomerization activates the endo-RNase activity of Ire1, which excises the intron from the HAC1 mRNA, making it available for translation.

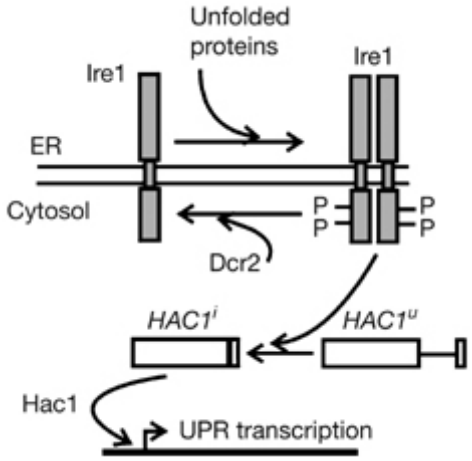


Fig. 2.8. A schematic model of the events leading to the induction of the UPR (Guo & Polymenis 2006).

Hac1 is a transcription factor that induces the expression of genes coding for ER-resident chaperones and other UPR-specific proteins (Travers et al. 2000). Not only does this increase the folding capacity of the ER, it also expedites the removal of incorrectly folded proteins by a process called ER-associated degradation (ERAD). The ERAD and UPR function in a dynamic way, such that an induced UPR increases the ERAD capacity, while constitutive UPR induction is observed with the loss of ERAD.

2.2.2.2.3. The ER associated protein degradation

In the ER, most misfolded proteins will be reintroduced into the ER quality control for several rounds in order to obtain their native conformation; however, when these subsequent rounds of folding fail to produce the desired product, the aberrant protein is set aside for proteosomal degradation (Lederkremer 2009). It is estimated that up to a third of all newly synthesized proteins fail to obtain their native folded state and are destroyed shortly after they have entered the endoplasmic reticulum (Schubert et al. 2000). Many factors are attributed to this high prevalence of aberrant proteins, including errors in translation and transcription, folding defects, subunit imbalances, radiation and chemical damaged mature proteins (Hirsch et al. 2009). The process by which misfolded polypeptides are detected in the ER, retrotranslocated to the cytosol and degraded by an ubiquitylation-dependent proteasome is referred to as the ER associated degradation (ERAD).

Aberrant proteins ultimately arrive at the ubiquitination machinery and are degraded by the cytosolic 26S proteasome, but the events that initiate the recognition of unglycosylated proteins are poorly understood. Yeasts have two ER membrane E3 ubiquitin ligase complexes involved in ERAD; the SSM4 ubiquitin ligase complex and the HRD ubiquitin ligase complex. These two complexes have overlapping substrate specificities, but the SSM4 ubiquitin ligase complex prefers cytosolic substrates, while the HRD complex is mostly involved in the ubiquitylation of misfolded ER luminal proteins and proteins with transmembrane domains (Swanson et al. 2001).

The HRD ubiquitin ligase complex subunits Hrd1p/Der3p consist of six transmembrane domains and a cytosolic RING-finger domain that can ubiquitylate target proteins (Deak & Wolf 2001). Ubiquitin, a highly conserved protein, is covalently added to lysine residues of the retrotranslocating polypeptide. Polyubiquitin chains serve as signals for degradation and are produced by the addition of ubiquitin units to the lysine 48 of the previous ubiquitin.

The luminal subunits of the HRD ubiquitin ligase complex consist of Hrd3p and Yos9p (**section 2.2.2.2.1.3**) that are required for substrate recognition. Although the mechanisms for recognition and remodelling of unfolded glycoproteins have been unravelled, it is still unclear how aberrant non-glycoproteins are identified. Yeast strains lacking the lectin subunit Yos9p exhibit an accumulation of unfolded, *N*-glycosylated proteins in the ER, but this does not prevent the interaction between the unfolded substrate and the other HRD ligase members (Gauss et al. 2006). Also, the removal of glycosylation sites from an inherently misfolded mutant version of CPY did not abolish its interaction with the ligation machinery. Considering these observations, it is tempting to speculate that a generic recognition mechanism, based on Kar2p and Hrd3p affinity to bind to exposed hydrophobic regions, might mark a protein for ubiquitin-proteasome degradation and that unfolded glycoproteins need to pass an additional quality control step before retrotranslocation (Hirsch et al. 2009).

Hrd1p, Ubc1p and the Cue1p recruited Ubc7p ubiquitinates polypeptides as they are retrotranslocated to the cytocol (Bays et al. 2001). The AAA^+ ATPase Cdc48p provides the driving force required to release ubiquitylated polypeptides from the ER membrane. This extraction from the ER membrane is proposed to occur through a similar 'molecular ratchet' mechanism as described for SRP-independent translocation (2.1.2.) (Amm et al. 2014; Zhang et al. 2000). Rad23p and Dsk2p binds to an ubiquitylated polypeptide through their C-terminal

UBA (ubiquitin associated) motif and direct it to the proteasome. Peptide:N-glycanase (PNGase), encoded by *PNG1*, discriminates between non-native and folded glycoproteins in the cytosol and deglycosylates the unbiquitylated substrate in transit (Suzuki et al. 2000; Hirsch et al. 2003).

2.2.3. Early secretion pathway

2.2.3.1. COPII vesicle sorting

Following protein synthesis, maturation and transport across the ER, transport vesicles bud from the surface at specialized ER domains, known as transitional ER (tER) sites, and ultimately fuse with the Golgi apparatus (Burgoyne 1988; Farquhar & Palade 1981). Export of vesicles from the ER and their transport to the Golgi is dependent on the generation of vesicles formed by the essential coat protein complex II (COPII), in turn formed by Sec proteins in a GTP-dependent manner (Barlowe 1994). The correct sorting of secretory proteins, that have been correctly folded and matured into their correct, secretion-competent form, requires that resident proteins be sufficiently segregated from cargo molecules as transport vesicles are being sculpted from the donor ER membrane (Elrod-erickson 1998). The latter is essential to ensure that the correct cargo is transported to the Golgi while resident proteins remain in the ER, which is made even more important as protein cargo is not a prerequisite for vesicle biogenesis (Lee & Miller 2007).

The COPII coat is largely responsible for the physical deformation of the ER membrane that drives vesicle formation, whilst also being accountable for the capture of cargo proteins within these newly formed vesicles (Sato 2004). The stepwise assembly of the COPII coat on the cytosolic face of the ER is able to couple cargo selection and membrane deformation, effectively generating functional transport vesicles (Copic et al. 2012).

Three components of the yeast COPII – the Sar1p GTPase, the Sec23/24 subcomplex and the Sec13/31 subcomplex (outer coat complex) – combine to regulate and organize this complex series of events leading to COP II vesicle formation (Matsuoka et al. 1998). Since more than a third of the entire proteome is directed to the ER, there are a large number of components that assist the core COPII constituents in functions related to specialized cargo sorting (Herrmann et al. 1999). A subset of accessory proteins, including Shr3p, Chs7p and Vma22p, are able to interact with secretory cargo proteins and assist in facilitating their incorporation into COPII

vesicles. Whilst all components of the COPII coat are widely considered to contribute to membrane curvature for vesicle formation, the exact mechanism of membrane deformation by the COPII coat remains unclear (Copic et al. 2012; Bickford et al. 2004). Additionally, the COPII budding machinery is organized into separate higher-order structures at a multitude of transitional zones on the ER surface. At these locations, the large multi-domain Sec16 protein performs an essential function (Miller & Barlowe 2010).

In eukaryotes, the formation of the COPII-coated vesicles at the ER membrane is triggered by the activation of the Ras-like small GTPase Sar1p, in turn activated through GDP/GTP exchange (Sato 2004). The activated Sar1p promotes COPII coat assembly whilst subsequent GTP hydrolysis by Sar1p leads to coat protein disassembly, after which these proteins are recycled for recurring cycles of vesicle formation. It is thus likely that the GTPase cycle of Sar1p is responsible for the regulation of COPII coat assembly and disassembly. Recent studies have illustrated a continual flux of GTPase cycles in Sar1p, which facilitates the concentration of cargo proteins into vesicles that ultimately bud from the cytosolic ER membrane (Miller & Barlowe 2010).

2.2.3.2. Protein sorting and maturation in Golgi

The Golgi is a dynamic structure, generally consisting of several stacks of disk-like membrane sacs, called cisternae and was first described by Golgi (1898). The widely accepted cisternal maturation model for cargo transport through the Golgi states that cisternae are able to form *de novo*, progressively mature, and ultimately dissipate (Glick & Malhotra 1998). Newly synthesized secretory proteins are delivered from the tER sites, via COPII vesicles, to the *cis* face of the stack. Several large protein complexes are involved in the tethering and fusion of COPII vesicles with either the ER-Golgi intermediate compartment (ERGIC) or *cis* Golgi, and are reviewed elsewere (Oka & Krieger 2005). As these proteins progress through the medial and *trans*-Golgi compartments as the cisternae mature, they are modified and processed by early- and late-acting enzymes, present during the different maturation steps (Griffiths & Simons 1986). The resident Golgi and ER proteins and mature secretory proteins are subsequently sorted by the trans-Golgi network (TGN), and dispatched to their destined cellular locations (Rothman 1981). Golgi and ER resident proteins containing retention signals are recognised and recycled through the retrograde transport of coat protein complex I (COPI) coated vesicles, while proteins destined

for the endosome or cell membrane are sorted into clathrin coated vesicles (Mellman & Warren 2000).

Unlike the closely related *Pichia pastoris*, *S. cerevisiae* has an unusual Golgi structure, consisting of individual cisternae that are scattered throughout the cytoplasm (Suda & Nakano 2012). Independent of its structure, the Golgi membranes are believed to function similarly, playing a crucial role in the maturation, transport of proteins and sorting into secretion vesicles. Of these maturation processes, including sulfation, phosphorylation and proteolytic cleaving, oligosaccharide modifications are the most prominent and best understood (Mellman & Warren 2000). In *S. cerevisiae*, this stepwise elongation of the core glycan-chain can proceed, to produce structures (with up to 200 mannose residues) containing α-1,6-linked residues, with α-1,2-linked branches which terminate in α-1,3-linked mannosyl residues as demonstrated in **Fig. 2.9**. The addition of this hyperglycosylated structure is highly specific, mostly limited to secreted enzymes and structural protein in the cell wall (Jungmann & Munro 1998).

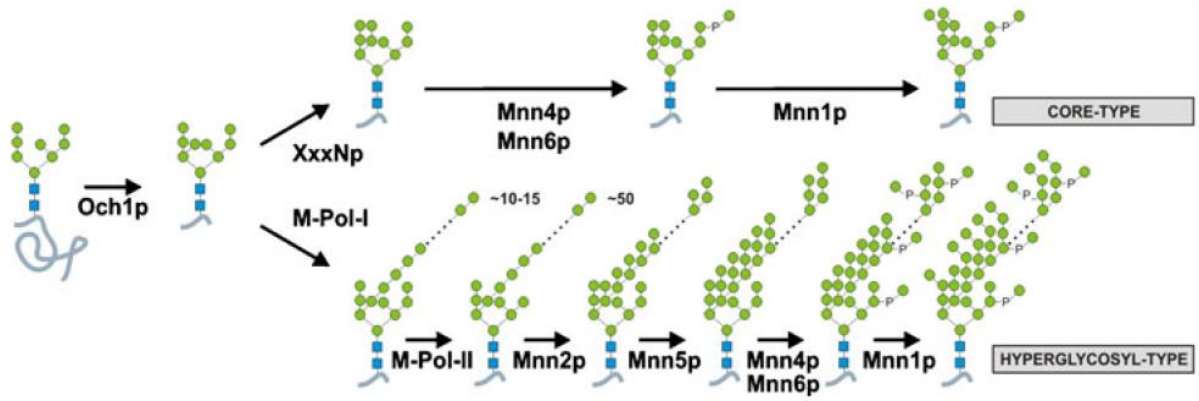


Fig. 2.9. Protein *N*-linked glycan elongation in the yeast Golgi-apparatus. Elongation of the $Man_8GlcNAc_2$ glycan outer chain is initiated by Och1p, adding a single mannose through a α-1,6-glycosidic bond. Mannan polymerase I (consisting of Van1p and Mnn9p) is then able to extend the mannose outer chain with up to 15 mannosyl residues. Mannan polymerase II (Hoc1p, Mnn8p, Mnn9p, Mnn10p and Mnn11p) further extend this outer chain up to 50 mannosyl residues. Mnn2p and Mnn6p are responsible for branching of the outer chain, adding single mannosyl residues through α-1,2-glycosidic and phosphate bonds, respectively. Mnn1p terminates the elongation by adding α-1,3-linked mannosyl residues to chain ends. Adapted from De Pourcq et al. (2010).

2.2.4. Late secretion pathway

The polarized transit of clathrin-coated secretion vesicles is initiated when the Myo2p tail recognises the vesicular Myo2p-binding sites on a vesicle destined for the cell membrane This

tail is believed to interact directly with Sec4p (Schott et al. 1999; Loubéry & Coudrier 2008). The Myo2p-vesicle complex consists of Smy1p, Sec2p and Sec4p; all three resulting in lethality when absent. This Myo2p-vesicle moves along the polarized actin bundles to the sites of secretion, but for tethering and vesicle fusion to take place it must first interact with the plasma membrane localized exocyst complex through its Sec4p subunit, a Rab GTPase, that brings the vesicle-SNARES (Soluble *N*-ethylmaleimide-sensitive factor (NSF) attachment protein receptor) on the vesicle in close proximity to the target-SNARES on the plasma membrane to allow a mature SNARE complex to form (Carr et al. 1999).

The assembly of the exocyst complex starts with the localization of the Sec3p to the tip of the emerging bud and after a sufficient bud size is reached, shifts to the mother-daughter neck region, attaching itself to the cap actin bundles as a landmark for the other components to assemble (Finger & Novick 1998). Unlike the rest of the exocyst components, Sec3p does not depend on active vesicle transport for its localization at the plasma membrane. The exocyst is a large complex consisting of at least eight proteins, Sec3p, Sec5p, Sec6p, Sec8p, Sec10p, Sec15p, Exo70p and Exo84p (TerBush et al. 2001). **Fig. 2.10** depicts the interaction of exocyst subunits to each other, GTPases and the exocytic SNARE complex, needed for vesicle fusion (**Fig. 2.11**). The exocyst tethers the secretion vesicle to the plasma membrane, but its direct interaction of the Sec15p and Exo70p subunits with different GTPases (Sec4p, Rho3p and Cdc42p) links it with the global cellular signalling pathways, implying a secretion regulating mechanism in addition to its tethering role (Robinson et al. 1999).

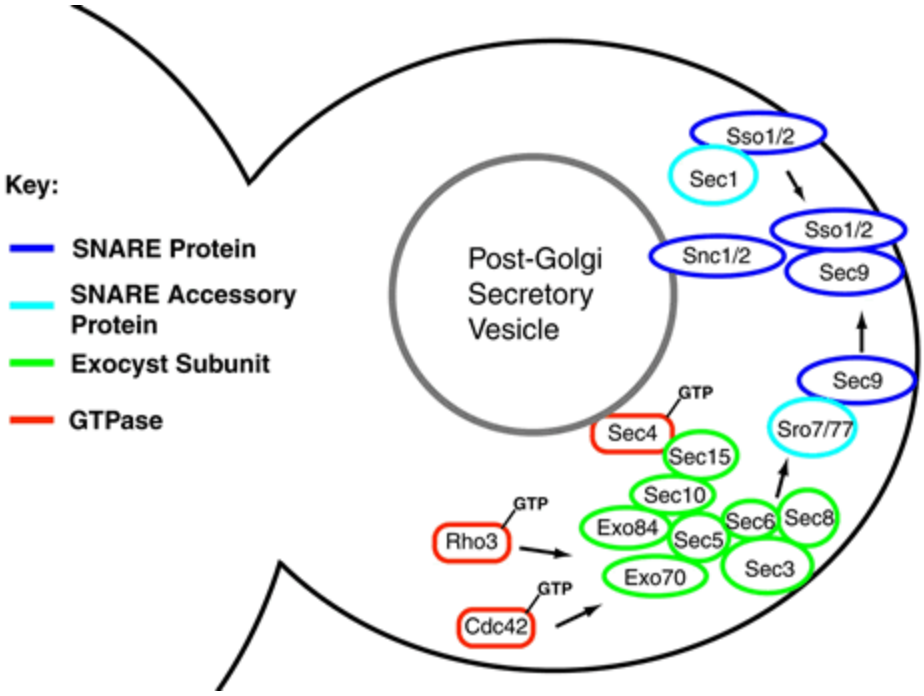


Fig. 2.10. The yeast bud depicting a late secretory vesicle, interacting with the exocyst complex and the exocytic SNARE complex (http://www.med.unc.edu/cellbiophysio/faculty/brennwald).

Fusion of the late secretory vesicles depends on the formation of a trans-SNARE complex or SNAREpin, by complementary SNARES present on the donor (secretion vesicle) and acceptor (plasma membrane) membranes (Weber et al. 1998; Burri & Lithgow 2004). The individual SNARE proteins consist of a region consisting of heptad repeat SNARE-motifs that can participate in coiled-coil formation and are typically C-terminally anchored in their resident membrane (Pelham 1999; Bonifacino & Glick 2004; Bock et al. 2001). For the trans-SNARE complex to form, and to allow membrane fusion, SNARE-motifs of the donor membrane SNAREs must bundle in parallel with the helixes of the acceptor membrane SNARES, resulting in a four helix parallel bundle and bringing the membranes in close proximity, an energetically favourable state for fusion (Chen & Scheller 2001). By studying reconstituted SNAREs in lipid bilayers, it was determined that membrane fusion occurs in the absence of ATP and that the protein folding during pin formation provides enough energy to overcome the repulsive forces between the two membranes (Weber et al. 1998; Hu et al. 2003).

The SNARE-motif of Sncp, the yeast donor membrane exocytic SNARE (v-SNARE), interacts with the acceptor membrane SNARES (syntaxin family, t-SNARES), Ssop and Sec9p, contributing one and two α-helixes respectively (Burri & Lithgow 2004). Both Sncp and Ssop

have two functionally redundant homologues, namely Snc1/2p and Sso1/2p, originating from the whole-genome duplication event in *S. cerevisiae*'s early history (Grote et al. 2000). Sso1/2p is distributed isotropically over the cell membrane, suggesting that the localization of SNAREpin formation is dependent on the regulation by Sec4p and the exocyst complex through Sec1p, a SM-like protein that binds to the SNAREpin to allow fusion and that interacts directly with the Sec6p subunit of the exocyst as regulatory component (Carr et al. 1999; Brennwald et al. 1994; Grote et al. 2000). This function of Sec1p in yeast is in contrast with its mammalian homologue where it only has a regulatory role, since it only binds to syntaxin, but not the assembled SNARE complex (Yang 2000).

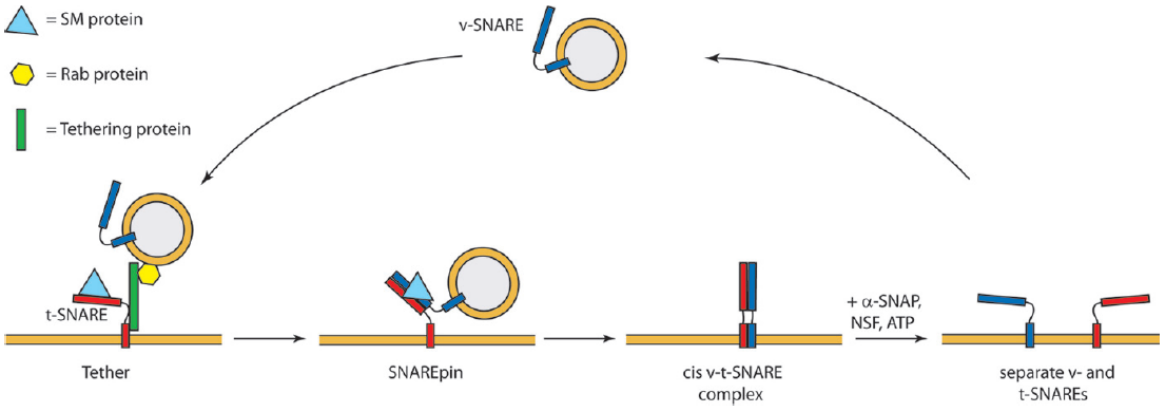


Fig. 2.11. The basic model for SNARE mediated vesicle fusion and recycling of the v-SNAREs. The late secretion vesicle is tethered to the plasma membrane by the exocyst complex, interacting directly with the Sec4p of the vesicle and the t-SNARE, Sec9p. After the SNAREpin is formed, vesicle fusion is initiated by the SM-protein Sec1p, expelling the cargo from the cell. The resulting cis-SNARE complex is dissembled by Sec18p and Sec17p (NSF and α-SNAP), allowing recycling of the v-SNARE through endocytosis (Malsam et al. 2008).

After the SNARE-mediated membrane fusion allows the expulsion of its content into the environment, the SNARE bundle remains as a cis-SNARE complex on the plasma membrane (Malsam et al. 2008; Hong & Lev 2014). The first steps in recycling these SNARE proteins, involves the ATP-dependent action of Sec18p (yeast NSF homologue), which binds to the cis-SNARE complex in the presence of its partner Sec17p (yeast α-SNAP homologue) (Söllner et al. 1993; Grote et al. 2000). The ability of Sec18p to disassemble SNARE complexes is not limited only to plasma membrane SNAREpins, but causes blocks at multiple steps of the secretion pathway when deactivated (Graham & Emr 1991; Jahn et al. 2003). These released SNARE

proteins are recycled by endocytosis and transferred to the endosomes before being cycled back to the ER, but these retrograde pathways are beyond the scope of the current study (reviewed by Duden 2003).

2.3. Tailoring the secretion pathway for high secretion

Yeast is a frequent choice for the production of industrial and pharmaceutical proteins, due to their rapid growth, microbial safety, eukaryotic post-translational processing, high-density fermentation capability and their ability to express and secrete heterologous proteins (Sodoyer 2004; De Pourcq et al. 2010). Improving the growth rates of host strains or increasing the transcription levels of the product protein, does not always lead to higher levels of secreted protein (Lodish 1988; Biemans et al. 1991; Inan et al. 2006). Overproducing proteins bestow a burden on the molecular pathways responsible for folding, protein modifications, quality control and intracellular transport through the cell. This may lead to secretion bottlenecks and cellular stress that impairs the secretion rate or induce protein degradation pathways (Mattanovich et al. 2004). In order to relieve these secretion bottlenecks, numerous studies aimed to increase either the volumetric or cell specific production, using strain engineering approaches (Inan et al. 2006; Idiris et al. 2010b).

Through systematic screening and semi-rational design of recombinant yeast strains, many genetic targets were identified whose genetic modification enhance the secretion of native and heterologous proteins. Examples of such gene alterations are listed in **Table 2.2**. Many of the published secretion enhancing gene modifications are proposed to affect the protein production and secretion pathway in a direct manner, by assisting in product folding or secretion vesicle fusion events (Aalto et al. 1993; Määttänen et al. 2010). Less obvious mechanisms by which modifications could modulate secretion, include the general maintenance of cellular health under environmental and general secretion pathway stress conditions (Smith et al. 1985; Raimondi et al. 2008).

Table 2.2. List of strain engineering examples by altering single genes. The following shorthand is used: **NSC** - *no significant change in secretion;* **small caps** - *gene product inactivation;* ↑ - *gene overexpressed;* ↓ - *gene expression repressed;* **underlined values** - *reduced secretion.*

Gene and location	Expression host	Reporter protein of interest	Fold increase*	Reference**
Nucleus				
↑***BFR2***	*S. cerevisiae*	Anti-HIV1 2F5-Fab (fed-batch)	1.23	(Gasser et al. 2007a)
		Anti-HIV1 2F5-Fab (shake flasks)	1.40	
↑***BMH2***	*S. cerevisiae*	Anti-HIV1 2F5-Fab (fed-batch)	1.32	(Gasser et al. 2007a)
		Anti-HIV1 2F5-Fab (shake flasks)	1.50	
↑***HAC1***	*Pichia pastoris*	Anti-HIV1 2F5-antibody binding fragment (Fab)	~1.29	(Gasser et al. 2006)
		Anti-HIV1 2F5-Fab (shake flasks)	1.50	(Gasser et al. 2007a)
		Surface displayed mouse interferon-γ	1.8	(Guerfal et al. 2010)
		Surface displayed human interferon-β	<1	
		Surface displayed human thrombomodulin	1.9	
		Surface displayed human erythropoietin	1.3	
		Secreted mouse interleukin-10	2.2	
		Secreted *Trypanosoma cruzi* trans-sialidase	2.1	
	S. cerevisiae	*Bacillus amyloliquefaciens* α-amylase	~2.0	(Valkonen et al. 2003)
		Trichoderma reesei Cel7B	NSC	
		S. cerevisiae invertase	~1.6-2.0	
		Talaromyces emersonii Cel7A	NSC	Addendum A
		Human albumin	~1.40	(Payne et al. 2008)

Gene and location	Expression host	Reporter protein of interest	Fold increase*	Reference**
		Surface displayed *Burkholderia gladioli* carboxyesterase	1.25	(Breinig et al. 2006)
hac1	*S. cerevisiae*	Green fluorescent protein	NSC	(Huang et al. 2008)
		Bacillus amyloliquefaciens α-amylase	0.25-0.30	(Valkonen et al. 2003)
		Trichoderma reesei Cel7B	0.5-0.6	
		S. cerevisiae invertase	NSC	
		Surface displayed *Burkholderia gladioli* carboxyesterase	NSC	(Breinig et al. 2006)
		Talaromyces emersonii Cel7A	NSC	Addendum A
		Neosartorya fischeri Cel12A	NSC	Chapter 4
		Saccharomycopsis fibuligera Cel3A	NSC	
↑PSE1	*S. cerevisiae*	*Saccharomycopsis fibuligera* Cel3A	3.70	(Kroukamp et al. 2013)
		Trichoderma reesei Cel7B	1.15	
		Neocallimastix patriciarum Cel6A	1.25	
		S. cerevisiae killer toxin	>1	(Chow et al. 1992)
		S. cerevisiae α-factor	>1	
		S. cerevisiae acid phosphatase	>1	
		Talaromyces emersonii Cel7A	NSC	Addendum A
rgr1	*S. cerevisiae*	Mouse α-amylase	17.2	(Sakai, A. et al. 1988)
		S. cerevisiae invertase	1.27	
		Human β-endorphin	11.9	
↑RPP0	*S. cerevisiae*	Surface displayed 7/15-single-chain T-cell receptor	~2.1	(Wentz & Shusta 2007a)
		7/15-Single-chain T-cell receptor (20°C)	~1.9	

Gene and location	Expression host	Reporter protein of interest	Fold increase*	Reference**
		LWHI-single-chain T-cell receptor (20°C)	NSC	
		4-4-20-single-chain variable fragment (20°C)	~1.8	
		OX26-single-chain variable fragment (20°C)	~1.9	
		Bovine pancreatic trypsin inhibitor	NSC	(Huang et al. 2008)
		Green fluorescent protein	~2.2	
Cytoplasm				
↑***CCS1***	*Kluyveromyces lactis*	*Arxula adeninivorans* glucoamylase	NSC	(Raimondi et al. 2008)
↑***PSA1***	*Hansenula polymorpha*	Human urokinase	>1.0	(Agaphonov et al. 2001)
	K. lactis	Human serum albumin	~3.0	(Uccelletti et al. 2005)
		Arxula adeninivorans glucoamylase	~4.0	
	S. cerevisiae	*Talaromyces emersonii* Cel7A	NSC	Addendum A
↓***RPN5***	*S. cerevisiae*	*S. cerevisiae* Hsp150p	5.3	(Davydenko et al. 2004)
ski5	*S .cerevisiae*	*S. cerevisiae* K1 killer toxin protein	>1	(Bussey et al. 1983)
↑***SOD1***	*S. cerevisiae*	*Saccharomycopsis fibuligera* Cel3A	NSC	(Kroukamp et al. 2013)
		Trichoderma reesei Cel7B	1.14	
		Neocallimastix patriciarum Cel6A	NSC	
		Talaromyces emersonii Cel7A	1.14	Addendum A
	K. lactis	Human serum albumin	~4.3	(Raimondi et al. 2008)
		Arxula adeninivorans glucoamylase	~2.2	

Gene and location	Expression host	Reporter protein of interest	Fold increase*	Reference**
↑***SSA4***	*P. pastoris*	Anti-HIV1 2F5-Fab (fed-batch)	1.73	(Gasser et al. 2007a)
		Anti-HIV1 2F5-Fab (shake flasks)	1.60	
↑***SSE1***	*P. pastoris*	Anti-HIV1 2F5-Fab (fed-batch)	2.29	(Gasser et al. 2007a)
		Anti-HIV1 2F5-Fab (shake flasks)	1.40	
↑***UBI4***	*K. lactis*	Human serumalbumin (duplicated *UBI4* gene)	>8.80	(Bao & Fukuhara 2001)
		Human serumalbumin (PGK1 promoter-*UBI4*)	>5.70	
		Human interleukin 1β	NSC	
	S. cerevisiae	Human leucocyte elastase inhibitor	~7.0	(Chen et al. 1994)
		Saccharomyces cerevisiae α-factor	NSC	
ubi4	*S. cerevisiae*	Human leucocyte elastase inhibitor	<u>≤0.33</u>	(Chen et al. 1994)
ubx2	*K. lactis*	*Bacillus amyloliquefaciens* α-amylase	>1	(Bartkeviciute & Sasnauskas
		Human growth hormone	~5-7	2003)
		K. lactis invertase	1.20	
	S. cerevisiae	*Bacillus amyloliquefaciens* α-amylase	>1	(Bartkeviciute & Sasnauskas
		S. cerevisiae invertase	1.40	2003)
vps34	*Schizosaccharomyces pombe*	Human growth hormone	<u>0.58</u>	(Idiris et al. 2010a)
Endoplasmic reticulum				
↑***CNE1***	*H. polymorpha*	Human interferon-γ	1.4 - 2.5	(Klabunde et al. 2007)
		Human serum albumin	1.5 - 2.0	

Gene and location	Expression host	Reporter protein of interest	Fold increase*	Reference**
		Azotobacter vinelandii AlgE1	1.4 - 2.5	
		Synthetic fungal phytase	1.1 - 1.6	
		Bovine follicle-stimulating hormone α-subunit	1.5	(Qian et al. 2009)
		Bovine follicle-stimulating hormone β-subunit	6.0	
	S. cerevisiae	Human A_2a adenosine receptor	~0.80	(Butz et al. 2003)
		Human Substance P receptor	~0.75	
		Single-chain T-cell receptor	1.9	(Shusta et al. 2000)
		Measles virus hemagglutinin	NSC	(Ciplys et al. 2011)
cne1	*S. cerevisiae*	*Saccharomyces cerevisiae* α-factor	NSC	(Parlati et al. 1995)
		Saccharomyces cerevisiae acid phosphatase	NSC	
		Mammalian α-antitrypsin	2-2.6	
		Hydrophobic domain-lysozyme fusion	2.58	(Arima et al. 1998)
		Hen egg white lysozyme	NSC	(Song et al. 2001)
		Hen egg white lysozyme (unstable mutant)	<1	
↑ERO1	*S. cerevisiae*	Surface displayed 7/15-single-chain T-cell receptor (20°C)	~1.35	(Wentz & Shusta 2007a)
		7/15-Single-chain T-cell receptor (20°C)	~3.2	
		LWHI-single-chain T-cell receptor (20°C)	~1.6	
		4-4-20-single-chain variable fragment (20°C)	~1.7	
		OX26-single-chain variable fragment (20°C)	~7.3	

Gene and location	Expression host	Reporter protein of interest	Fold increase*	Reference**
	K. lactis	Human serum albumin	~15.0	(Lodi et al. 2005)
		Human interleukin 1β	NSC	
	P. pastoris	Anti-HIV1 2F5-Fab (shake flasks)	1.40	(Gasser et al. 2007a)
↑***EUG1***	*S. cerevisiae*	Surface displayed single-chain T-cell receptor	NSC	(Shusta et al. 2000)
↑***JEM1***	*S. cerevisiae*	Human albumin	~1.90	(Payne et al. 2008)
↑***KAR2***	*H. polymorpha*	*Aspergillus niger* glucose oxidase	~0.1	(van der Heide et al. 2002)
	P. pastoris	HSA-interleukin-1 receptor antagonist fusion	~0.67	(Shen et al. 2012)
		HSA-human growth hormone fusion	~0.33	
		A33 single-chain antibody fragment	3.0	(Damasceno et al. 2007)
		Anti-HIV1 2F5-Fab (shake flasks)	NSC	(Gasser et al. 2007a)
	S. cerevisiae	Bovine prochymosin	1.50	(Harmsen et al. 1996)
		Plant thaumatin	0.78	
		Antithrombotic hirudin	2.50	(Kim et al. 2003)
		Human erythropoietin	~5.00	(Robinson & Wittrup 1995)
		Bovine pancreatic trypsin inhibitor	NSC	(Robinson et al. 1996)
		Schizosaccharomyces pombe acid phosphatase	NSC	
		Human granulocyte colony-stimulating factor	NSC	
		Pyrococcus furiosus β-glucosidase	1.20	(Smith et al. 2004)
		Human A_2a adenosine receptor	NSC	(Butz et al. 2003)

Gene and location	Expression host	Reporter protein of interest	Fold increase*	Reference**
		Single-chain antibody fragment	~1.20	(Xu et al. 2005)
		McPC603 Single-chain antibody fragment (scFv)	~2.4	(Shusta et al. 1998)
		McPC603 Single-chain antibody fragment (3 mutants) 4-4-20	~3.9-4.8	
		Single-chain antibody fragment	~4.5	
		Single-chain T-cell receptor	2.0	(Shusta et al. 2000)
		OX26-single-chain variable fragment (20°C)	~3.8	(Hackel et al. 2006)
↓***KAR2***	*S. cerevisiae*	Bovine pancreatic trypsin inhibitor	~ 0.1	(Robinson et al. 1996)
		Schizosaccharomyces pombe acid phosphatase	~ 0.1	
		Human granulocyte colony-stimulating factor	~ 0.1	
↑***LHS1***	*S. cerevisiae*	Human albumin	~1.15	(Payne et al. 2008)
↑***PDI1***	*K. lactis*	Human serum albumin (duplicated *PDI1* gene)	>15.0	(Bao & Fukuhara 2001)
		Human serum albumin (PGK1 promoter-*PDI1*)	>16.5	
		Human interleukin 1β	NSC	
		Human serum albumin (HSA)	~15.0	(Lodi et al. 2005)
	Pichia pastoris	Human parathyroid hormone	2.75	(Vad et al. 2005)
		HSA-interleukin-1 receptor antagonist fusion	2.40	(Shen et al. 2012)
		HSA-human growth hormone fusion	~2.67	
		A33 single-chain antibody fragment	NSC	(Damasceno et al. 2007)
		Necator americanus ASP1	3.9-7.9	(Inan et al. 2006)
		Plasmodium falciparum trasmission-blocking peptide	~3.0	(Tsai et al. 2006)

Gene and location	Expression host	Reporter protein of interest	Fold increase*	Reference**
		Anti-HIV1 2F5-antibody binding fragment (Fab)	1.90	(Gasser et al. 2006)
		Bovine follicle-stimulating hormone	6.0	(Huo et al. 2007)
	S. cerevisiae	*Schizosaccharomyces pombe* acid phosphatase	~4.0	(Robinson et al. 1994)
		Human growth factor B homodimer	~10.0	
		Human granulocyte colony-stimulating factor	NSC	
		McPC603 Single-chain antibody fragment (scFv)	~2.3	(Shusta et al. 1998)
		McPC603 Single-chain antibody fragment (3 mutants) 4-4-20	~3.3-4.7	
		Single-chain antibody fragment	~1.9	
		Pyrococcus furiosus β-glucosidase	1.25	(Smith et al. 2004)
		Human A_2a adenosine receptor	~0.75	(Butz et al. 2003)
		Substance P receptor	~0.60	
		Human lysozyme (Pdi1p active sites disrupted)	1.57	(Hayano et al. 1995)
		Single-chain antibody fragment	~1.70	(Xu et al. 2005)
		Single-chain T-cell receptor	NSC	(Shusta et al. 2000)
		OX26-single-chain variable fragment (20°C)	~4.2	(Hackel et al. 2006)
		Trichoderma reesei Cel7A	1.62	(Xu et al. 2014a)
		S. cerevisiae invertase	NSC	
		4m5.3 human IgG	15.0	(Rakestraw et al. 2009)
pmt5	*S. cerevisiae*	*Phanerochaete chrysoporium* exocellulase	4.3	(Wang et al. 2013)

Gene and location	Expression host	Reporter protein of interest	Fold increase*	Reference**
↓***RER2***	*S. cerevisiae*	*S. cerevisiae* Hsp150p	1.29	(Davydenko et al. 2004)
↑***SBH1***	*S. cerevisiae*	*Bacillus* α-amylase	2.1	(Toikkanen et al. 2004)
↑***SCJ1***	*S. cerevisiae*	Human albumin	NSC	(Payne et al. 2008)
↑***SEC61***	*S. cerevisiae*	*Bacillus* α-amylase	NSC	(Toikkanen et al. 2003)
↑***SIL1***	*S. cerevisiae*	Human albumin	NSC	(Payne et al. 2008)
↑***SSS1***	*S. cerevisiae*	*Bacillus* α-amylase	NSC	(Toikkanen et al. 2003)
Early secretion pathway				
↑***SEC31***	*P. pastoris*	Anti-HIV1 2F5-Fab (shake flasks)	NSC	(Gasser et al. 2007a)
↑***SLY1***	*S. cerevisiae*	Human insulin precursor	NSC	(Hou et al. 2012)
		Aspergillus oryzae α-amylase	1.43	
		S. cerevisiae invertase	NSC	
Golgi apparatus				
↑***COG6***	*P. pastoris*	Anti-HIV1 2F5-Fab (shake flasks)	NSC	(Gasser et al. 2007a)
↑***COY1***	*P. pastoris*	Anti-HIV1 2F5-Fab (shake flasks)	NSC	(Gasser et al. 2007a)
↑***IMH1***	*P. pastoris*	Anti-HIV1 2F5-Fab (shake flasks)	NSC	(Gasser et al. 2007a)
kex2	*S. cerevisiae*	Insulin-containing fusion protein	>1.0	(Zhang et al. 2001)
mnn1	*S. cerevisiae*	*Neosartorya fischeri* Cel12A	NSC	Chapter 4
		Saccharomycopsis fibuligera Cel3A	NSC	

Gene and location	Expression host	Reporter protein of interest	Fold increase*	Reference**
mnn2	*S. cerevisiae*	*Neosartorya fischeri* Cel12A	1.30	Chapter 4
		Saccharomycopsis fibuligera Cel3A	NSC	
mnn6	*S. cerevisiae*	*Neosartorya fischeri* Cel12A	NSC	Chapter 4
		Saccharomycopsis fibuligera Cel3A	NSC	
mnn9	*S. cerevisiae*	*Neosartorya fischeri* Cel12A	NSC	Chapter 4
		Saccharomycopsis fibuligera Cel3A	NSC	
mnn10	*K. lactis*	α-amylase	>1.0	(Bartkeviciūte & Sasnauskas 2004)
		Human growth hormone	~3-4	
		K. lactis invertase	~1.5	
	S. cerevisiae	*S. cerevisiae* invertase	~1.8	(Bartkeviciūte & Sasnauskas 2004)
		Phanerochaete chrysoporium exocellulase	6.0	(Wang et al. 2013)
		Neosartorya fischeri Cel12A	NSC	Chapter 4
		Saccharomycopsis fibuligera Cel3A	NSC	
mnn11	*S. cerevisiae*	*S. cerevisiae* invertase	~1.5	(Bartkeviciūte & Sasnauskas 2004)
		Neosartorya fischeri Cel12A	NSC	Chapter 4
		Saccharomycopsis fibuligera Cel3A	1.26	
och1	*K. lactis*	Human serum albumin	~2.5	(Uccelletti et al. 2006)
		Arxula adeninivorans glucoamylase	~4.0	

Gene and location	Expression host	Reporter protein of interest	Fold increase*	Reference**
	S. cerevisiae	*Neosartorya fischeri* Cel12A	NSC	Chapter 4
		Saccharomycopsis fibuligera Cel3A	0.68	
pmr1	*H. polymorpha*	Human urokinase	>1.0	(Agaphonov et al. 2007)
	K. lactis	*Arxula adeninivorans* glucoamylase	~1.67	(Uccelletti et al. 1999)
	P. pastoris	Recombinant hirudin	1.55	(Ni et al. 2008)
	S. cerevisiae	Bovine prochymosin	3.50	(Harmsen et al. 1996)
		Plant thaumatin	1.12	
		Calf prochymosin	11.48	(Smith et al. 1985)
		Bovine growth hormone	15.0	
		Trichoderma reesei Cel7A	2.62	(Xu et al. 2014a)
		S. cerevisiae invertase	~0.93	
		Human urinary plasminogen	<7.0	(Rudolph et al. 1989)
		S. cerevisiae invertase	NSC	
		Neosartorya fischeri Cel12A	1.69	Chapter 4
		Saccharomycopsis fibuligera Cel3A	1.36	
↑***YND1***	*K. lactis*	Human serum albumin	~3.5	(Uccelletti et al. 2007)
		Arxula adeninivorans glucoamylase	~2.0	
Vacuole				
↑***CUP5***	*P. pastoris*	Anti-HIV1 2F5-Fab (fed-batch)	1.83	(Gasser et al. 2007a)
		Anti-HIV1 2F5-Fab (shake flasks)	1.70	

Gene and location	Expression host	Reporter protein of interest	Fold increase*	Reference**
pep4	*P. pastoris*	Human serum albumin - parathyroid hormone fusion	~1.6	(Wu et al. 2013)
	S. cerevisiae	Human interferon-*β*	~3.4	(Tomimoto et al. 2013)
prb1	*S. cerevisiae*	Human interferon-*β*	~5.0	(Tomimoto et al. 2013)
vps4	*S. cerevisiae*	Insulin-containing fusion protein	<1.0	(Zhang et al. 2001)
vps8	*S. cerevisiae*	Insulin-containing fusion protein	<1.0	(Zhang et al. 2001)
vps10	*S. cerevisiae*	*S. cerevisiae* invertase-λ repressor protein fusion	~3.4	(Hong et al. 1996)
		Trichoderma reesei Cel7A	1.62	(Xu et al. 2014a)
		S. cerevisiae invertase	~0.87	
	Sz. pombe	Human growth hormone	1.62	(Idiris et al. 2010a)
vps13	*S. cerevisiae*	Insulin-containing fusion protein	<1.0	(Zhang et al. 2001)
vps22	*Sz. pombe*	Human growth hormone	0.63	(Idiris et al. 2010a)
vps35	*S. cerevisiae*	Insulin-containing fusion protein	<1.0	(Zhang et al. 2001)
vps36	*S. cerevisiae*	Insulin-containing fusion protein	<1.0	(Zhang et al. 2001)
Mitochondria				
↑GEM1	*S. cerevisiae*	*S. cerevisiae* Pep4p	>1	(Wolff et al. 1999)
ssq1	*S. cerevisiae*	Calf prochymosin	11.9	(Smith et al. 1985)
		Bovine growth hormone	1.33	
Late secretion pathway				

Gene and location	Expression host	Reporter protein of interest	Fold increase*	Reference**
ddi1	*S. cerevisiae*	*Talaromyces emersonii* Cel7A	1.23	Addendum A
↑***KIN2***	*P. pastoris*	Anti-HIV1 2F5-Fab (fed-batch)	2.24	(Gasser et al. 2007a)
		Anti-HIV1 2F5-Fab (shake flasks)	1.50	
mon2	*S. cerevisiae*	*Cypridina noctiluca* luciferase	2.5	(Kanjou et al. 2007)
↑***SEC1***	*S. cerevisiae*	Human insulin precursor	1.30	(Hou et al. 2012)
		Aspergillus oryzae α-amylase	1.62	
		S. cerevisiae invertase	~1.45	
		Talaromyces emersonii Cel7A	NSC	Addendum A
↑***SEC4***	*S. cerevisiae*	*Bacillus* α-amylase	~3.0	(Toikkanen et al. 2003)
↑***SNC1***	*S. cerevisiae*	*Talaromyces emersonii Cel7A*	~1.74	(Van Zyl et al. 2014)
		Saccharomycopsis fibuligera Cel3A	~1.20	
↑***SNC2***	*S. cerevisiae*	*Talaromyces emersonii Cel7A*	~1.48	(Van Zyl et al. 2014)
		Saccharomycopsis fibuligera Cel3A	~0.77	
↑***SNC9***	*S. cerevisiae*	*Talaromyces emersonii Cel7A*	~1.36	(Van Zyl et al. 2014)
		Saccharomycopsis fibuligera Cel3A	~1.20	
↑***SSO1***	*S. cerevisiae*	*Bacillus* α-amylase	1.8	(Ruohonen et al. 1997)
		Talaromyces emersonii Cel7A	~1.36	(Van Zyl et al. 2014)
		Saccharomycopsis fibuligera Cel3A	~1.40	
		Bacillus α-amylase	2.2	(Toikkanen et al. 2004)
		Trichoderma reesei Cel7A	1.10	(Xu et al. 2014a)

Gene and location	Expression host	Reporter protein of interest	Fold increase*	Reference**
		S. cerevisiae invertase	1.17	
		Talaromyces emersonii Cel7A	1.28	Addendum A
↑***SSO2***	*P. pastoris*	Anti-HIV1 2F5-Fab (shake flasks)	NSC	(Gasser et al. 2007a)
	S. cerevisiae	*Trametes versicolor* laccase	~2.0	(Larsson et al. 2001)
		Human parathyroid hormone	NSC	(Vad et al. 1998)
		S. cerevisiae invertase	1.5	(Ruohonen et al. 1997)
		Bacillus α-amylase	1.5	
		Trichoderma reesei Cel7B	NSC	
		Talaromyces emersonii Cel7A	NSC	Addendum A
		Talaromyces emersonii Cel7A	~1.23	(Van Zyl et al. 2014)
		Saccharomycopsis fibuligera Cel3A	NSC	
vsm1	*S. cerevisiae*	Total secreted protein	~3.4	(White et al. 2011)
vta1	*S. cerevisiae*	*α-amylase*	1.35	(Liu et al. 2014)
Cell wall				
↑***CCW12***	*S. cerevisiae*	Surface displayed 7/15-single-chain T-cell receptor	~2.1	(Wentz & Shusta 2007a)
		7/15-Single-chain T-cell receptor (20°C)	~2.5	
		LWHI-single-chain T-cell receptor (20°C)	NSC	
		4-4-20-single-chain variable fragment (20°C)	~4.0	
		OX26-single-chain variable fragment (20°C)	~4.1	
		Bovine pancreatic trypsin inhibitor	NSC	(Huang et al. 2008)

Gene and location	Expression host	Reporter protein of interest	Fold increase*	Reference**
		Green fluorescent protein	NSC	
↑***CWP2***	*S. cerevisiae*	Surface displayed 7/15-single-chain T-cell receptor	~1.3	(Wentz & Shusta 2007a)
		7/15-Single-chain T-cell receptor (20°C)	~1.8	
		LWHI-single-chain T-cell receptor (20°C)	NSC	
		4-4-20-single-chain variable fragment (20°C)	NSC	
		OX26-single-chain variable fragment (20°C)	~1.5	
mkc7	*P. pastoris*	Human serum albumin - parathyroid hormone fusion	NSC	(Wu et al. 2013)
	S. cerevisiae	*Manduca sexta* diuretic hormone	~1.27	(Copley et al. 1998)
↑***SED1***	*S. cerevisiae*	Surface displayed 7/15- single-chain T-cell receptor	~1.2	(Wentz & Shusta 2007a)
		7/15-Single-chain T-cell receptor (20°C)	~1.6	
		LWHI-single-chain T-cell receptor (20°C)	~1.5	
		4-4-20-single-chain variable fragment (20°C)	~1.6	
		OX26-single-chain variable fragment (20°C)	NSC	
yps1	*P. pastoris*	Human serum albumin - parathyroid hormone fusion	~1.4	(Wu et al. 2013)
	S. cerevisiae	Human albumin	>1.18	(Kerry-Williams et al. 1998)
		Human glucagon	4.9	(Egel-Mitani et al. 2000)
		Glucagon-like-peptide 1	1.6	
		Glucagon-like-peptide 2	7.4	
		Ampphetamine regulated transcript peptide	2.1	
		Human parathyroid hormone	~7.0	(Kang et al. 1998)
		Manduca sexta diuretic hormone	~1.09	(Copley et al. 1998)

Gene and location	Expression host	Reporter protein of interest	Fold increase*	Reference**
yps3	*P. pastoris*	Human serum albumin - parathyroid hormone fusion	NSC	(Wu et al. 2013)
yps7	*P. pastoris*	Human serum albumin - parathyroid hormone fusion	NSC	(Wu et al. 2013)

* Fold increase in secreted protein of interest, compared to its wild-type counterpart; unless otherwise indicated

** Reference excluded due to limited space:

Idiris et al. (2006a) evaluated 52-protease deficient strains with secretions ranges between 0.55 -1.44.

Wang et al. (2013) evaluated 10-late secretory pathway deletions and 47-glycosylation related deletions.

Hoshida et al. (2013) evaluated 44 knockout mutants of six different functional catagoties; 13 mutants had more than 2-fold increased secretion.

The secretion enhancing ability of many gene alterations vary depending on the reporter protein characteristics, although these properties are still poorly understood (Robinson et al. 1994; Wentz & Shusta 2008; Kroukamp et al. 2013). This disparity in secretion enhancement was demonstrated by Robinson et al. (1994). Overexpression of *PDI1* in *S. cerevisiae*, enhanced the secretion of human growth factor B 10-fold, but failed to enhance in human granulocyte colony-stimulating factor secretion.

Depending on their beneficial mechanism, secretion enhancing gene alterations may act synergistically (Smith et al. 2004; Idiris et al. 2006b; Shusta et al. 1998; Kroukamp et al. 2013; Xu et al. 2014b). By overexpressing Kar2p and Pdi1p in *S. cerevisiae*, the secretion of single-chain antibody fragments were improved 2.4- and 2.3-fold, respectively, and in combination up to 8.3-fold more secreted product was obtained (Shusta et al. 1998). The synergistic effect of Kar2p and Pdi1p, as observed with single gene alterations, was reporter protein specific (Lodi et al. 2005). Lodi and co-workers were able to obtain significant increases in secreted human serum albumin by overexpressing either Kar2p or Pdi1p, but found no further enhancement with the co-overexpression of these proteins. It should be noted that similar high secretion values were obtained for the individual overexpressions; implying that there is a maximum, potentially protein specific, secretion capacity that exist for each protein secretion host.

The reporter protein specificity observed in numerous secretion engineering studies, is still a poorly understood topic, although protein size, hydrophobicity, presence of disulfide bonds and degree of glycosylation have all been implicated as obstacles to high level secretion (Shuster 1991; Shusta et al. 1998; Agaphonov et al. 2007; Idiris et al. 2010b). In addition to these reporter protein characteristics, studies such as Wentz & Shusta (2007a) and Kroukamp et al. (2013) would suggest that some of these enhancing elements' effects are highly dependent on the physiological condition of the cells. The negative effect of elevated cultivation temperature on single-chain T-cell receptor secretion, was shown to be partly relieved by overexpressing the genes, *CCW12, SED1, CWP2*, *ERO1* and *RPP0* (Wentz & Shusta 2007a). With the exception of *ERO1*, no significant differences were observed in the native-gene overexpressing strains' secretion at 20°C. However, up to 2.1-fold enhancement in secretion was observed for all engineered strains at 30°C and 37°C. Superoxide dismutase overexpression has been previously shown to increase heterologous protein secretion, although no effect was observed by Kroukamp et al. (2013) for recombinant *β*-glucosidase secretion in *S. cerevisiae* (Raimondi et al. 2008). Only after *SOD1* was overexpressed in the

high secreting strain (overexpressing *PSE1*) did its secretion enhancing ability became apparent. From these and other synergy examples, it was clear that many secretion enhancing alterations might only have phenotypic effects under specific protein secretion stress conditions.

Secretion pathway engineering is a powerful tool for increasing protein yields, but many of these high secreting recombinant strains display growth defects and increased sensitivity to inhibitors and elevated temperatures (Idiris et al. 2006b; Agaphonov et al. 2007; Payne et al. 2008). This problem can be partly mitigated through a novel mutagenesis technique, as demonstrated by Abe et al. (2009). By introducing an error-prone DNA polymerase into a triple deletion mutant of *S. cerevisiae* (*och1Δ*, *mnn1Δ* and *mnn4Δ*), they were able to significantly increase the growth rate, temperature tolerance and the resistance to cell wall affecting inhibitors of the strain.

2.3.1 Identifying secretion enhancing elements

In order to fully exploit strain engineering as a means to enhance the secretion of relevant proteins, numerous studies have focussed on generating 'supersecreting' strains and identifying secretion enhancing genetic elements (Smith et al. 1985; Sakai, A. et al. 1988; Wentz & Shusta 2007b; Gasser et al. 2007a; Liu et al. 2014). In one of the earlier studies, Smith et al. (1985) identified 39 mutant strains with high prochymosin secretion from a collection of ~120 000 mutagenized colonies. These 'supersecreting' mutants were selected through plate assays, and the unknown putative mutant genes responsible for the phenotype referred to as *ssc* (for supersecretion). The recessive *ssc1* allele behaved as a simple Mendelian trait and depending on the reporter protein, improved secretion up to 10-fold. *SSC1* was later shown to be identical to the Golgi Ca^{2+}/Mn^{2+}-ATPase *PMR1* (Rudolph et al. 1989) and is a frequent target to improve heterologous protein yields (Harmsen et al. 1996; Agaphonov et al. 2007; Xu et al. 2014a).

Sakai et al. (1988) were also able to isolate high secretion mutants, using plate assays and ethyl methanesulfonate treated cells. Although they were able to link the *rgr1* allele to chromosome XII, they were unable to accurately identify the other mutated genes. Sakai et al. (1988) also attempted to classify the mechanisms by which mutations may alter the secretion enhancing allele in addition to changing its coding sequence, namely; (I) on transcriptional level (mutations in promoter, *cis*-acting), (II) mutations in *trans*-acting gene elements and

(III) mutations in other genes influencing the product of the secretion enhancing allele (e.g. post-transcriptional modification).

A means to accurately identify alleles responsible for the high secretion phenotype was elusive for many years. The elucidation of the secretion pathway and the cellular responses to heterologous protein expression, gradually allowed researchers to move towards rational design strategies (Idiris et al. 2010b). By analysing the transcriptomes of *P. pastoris* overexpressing human trypsinogen, compared to non-expressing strains, Gasser et al. (2007a) evaluated 13 upregulated genes (**Table 2.2**) involved in the secretion pathway and stress response (Gasser et al. 2007b). Twelve of the evaluated genes, when overexpressed, increased the secretion of the human antibody Fab fragment, with up to 2.5-fold increase in production rates. From this study it was clear that a depletion of protein folding and secretion factors in protein production hosts, was met with a cellular response aimed to restore the chaperonin equilibrium. However, the native regulatory networks might not be able to completely restore folding capacity to the unstressed state. In an earlier example of this, Robinson and Wittrup (1995) determined that the prolonged production of heterologous proteins in *S. cerevisiae* reduced the extractable levels of Kar2p and Pdi1p from these cells. By replenishing these ER-chaperones, they were able to enhance the production of human erythropoietin up to 5-fold.

Recent advances in high throughput DNA sequencing technologies allowed researchers to overcome the limitations of SNP identification in supersecreting strains generated by random mutagenesis (Idiris et al. 2010b). Liu et al. (2014) obtained an *S. cerevisiae* strain, through UV-random mutagenesis, that was capable of producing 5-fold more heterologous secreted amylase than the original strain. Unlike the studies of Smith et al. (1985) and Sakai et al. (1988), they were able to discover several point mutations in their strain through whole-genome sequencing; allowing them to identify an allele that is responsible for 35% of the observed secretion enhancement, namely the *VTA1* allele.

The *S. cerevisiae* genome sequence was completed in 1996, unveiling a plethora of unidentified genes; and potential new genomic targets to improve heterologous secretion (Goffeau et al. 1996). Initiatives to analyse and assign functions to each of the ~6000 identified open reading frames, like the EUROFAN projects (Winzeler et al. 1999), produced valuable overexpression libraries and gene deletion strains to interrogate for secretion enhancing factors (Davydenko et al. 2004; Kanjou et al. 2007).

Davydenko et al. (2004) explored 63 strains obtained from the EUROSCARF collection that each had one essential gene under the 'tuneable' doxycycline *tetO*$_7$-CYC1 promoter (Garí et al. 1997). Using the native secreted Hsp150p as a reporter, they were able to identify *RPN5* as a novel secretion enhancing gene when its expression was downregulated in the presence of doxycycline (Davydenko et al. 2004).

An intuitive high throughput approach to screen yeasts for high secretion was demonstrated by Wentz & Shusta (2007b). By displaying the reporter protein on the cell surface and the subsequent immunolabeling thereof, they were able to measure the fluorescence intensity of each cell with flow cytometry. The yeasts displaying the reporter protein, in this case single-chain T-cell receptor, were transformed with a *S. cerevisiae* cDNA overexpression library (Wentz & Shusta 2007b). This method allowed them to rapidly identify yeasts with high secretion and subsequently, the open reading frames (ORFs) they were overexpressing. These identified ORFs were able to increase the secreted yields of non-cell wall bound single-chain T-cell receptor and other antibody fragments, in addition to their surface display enhancement. The identified ORFs included the previously reported secretion enhancer, *ERO1* and four novel genes.

Depending on the properties of the reporter protein, it is evident that effectiveness of many previously identified secretion genes vary (Wentz & Shusta 2007b; Idiris et al. 2010b; Kroukamp et al. 2013). Allele alterations responsible for improved secretion may imply a range of up- and downregulation or even deletion of the allele in question. The synergistic effect of these genetic changes (Lodi et al. 2005; Damasceno et al. 2007; Kroukamp et al. 2013) and the partial secretion enhancing effect of *VTI1* allele compared to the original mutant strain (Liu et al. 2014), does not represent simple Mendelian traits. The phenotype of high secretion was indeed shown to be a polygenetic phenomenon (**Chapter 5**) as may have been implied by the vast array of gene products involved in the protein secretion process.

In order to decipher the genetic basis of the high secretion trait, we set out to identify these genetic elements in a high secretion *S. cerevisiae* strain M0341 (described in **Chapter 5**). To identify multiple genetic elements, all contributing towards one phenotypic trait, we employed chromosome shuffling and pooled-segregant genome sequencing (Parts et al. 2011) to determine the chromosomal regions of the M0341 strain, associated with the high secretion phenotype. Although a promising technique to simultaneously identify the genetic elements responsible for the high secretion of a specific reporter protein, the authors were unable to

pinpoint these elements due to chromosomal instability in the recombinant strains (**Chapter 5**).

2.4. References

Aalto MK, Ronne H & Keranen S (1993) Yeast syntaxins Sso1p and Sso2p belong to a family of related membrane proteins that function in vesicular transport. *EMBO J.* 12: 4095–4104.

Abe H, Takaoka Y, Chiba Y, Sato N, Ohgiya S, Itadani A, Hirashima M, Shimoda C, Jigami Y, Nakayama K (2009) Development of valuable yeast strains using a novel mutagenesis technique for the effective production of therapeutic glycoproteins. *Glycobiology* 19: 428–436.

Agaphonov M, Plotnikova T, Fokina V, Romanova N, Packeiser N, Kang H & Ter-Avanesyan MD (2007) Inactivation of the *Hansenula polymorpha PMR1* gene affects cell viability and functioning of the secretory pathway. *FEMS Yeast Res.* 7: 1145–1152.

Agaphonov M, Packeiser N, Chechenova MB, Choi ES & Ter-Avanesyan MD (2001) Mutation of the homologue of GDP-mannose pyrophosphorylase alters cell wall structure, protein glycosylation and secretion in *Hansenula polymorpha*. *Yeast* 18: 391–402.

Amberg DC (1998) Three-dimensional imaging of the yeast actin cytoskeleton through the budding cell cycle. *Mol. Biol. Cell* 9: 3259–3262.

Amin J, Ananthan J & Voellmy R (1988) Key features of heat shock regulatory elements. *Mol. Cell. Biol.* 8: 3761–3769.

Amm I, Sommer T & Wolf DH (2014) Protein quality control and elimination of protein waste: the role of the ubiquitin-proteasome system. *Biochim. Biophys. Acta* 1843: 182–196.

Arima H, Kinoshita T, Ibrahim HR, Azakami H & Kato a (1998) Enhanced secretion of hydrophobic peptide fused lysozyme by the introduction of *N*-glycosylation signal and the disruption of calnexin gene in *Saccharomyces cerevisiae*. *FEBS Lett.* 440: 89–92.

Arroyo J et al. (2011) Functional and genomic analyses of blocked protein *O*-mannosylation in baker's yeast. *Mol. Microbiol.* 79: 1529–1546.

Bao WG & Fukuhara H (2001) Secretion of human proteins from yeast: stimulation by duplication of polyubiquitin and protein disulfide isomerase genes in *Kluyveromyces lactis*. *Gene* 272: 103–110.

Barlowe C (1994) COPII: A membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. *Cell* 77: 895–907.

Bartkeviciute D & Sasnauskas K (2003) Studies of yeast *Kluyveromyces lactis* mutations conferring super-secretion of recombinant proteins. *Yeast* 20: 1–11.

Bartkeviciūte D & Sasnauskas K (2004) Disruption of the *MNN10* gene enhances protein secretion in *Kluyveromyces lactis* and *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 4: 833–840.

Baxter BK, James P, Evans T & Craig EA (1996) *SSI1* encodes a novel Hsp70 of the *Saccharomyces cerevisiae* endoplasmic reticulum. *Mol. Cell. Biol.* 16: 6444–6456.

Bays NW, Gardner RG, Seelig LP, Joazeiro CA & Hampton RY (2001) Hrd1p/Der3p is a membrane-anchored ubiquitin ligase required for ER-associated degradation. *Nat. Cell Biol.* 3: 24–29.

Benham AM, van Lith M, Sitia R & Braakman I (2013) Ero1-Pdi interactions, the response to redox flux and the implications for disulfide bond formation in the mammalian endoplasmic reticulum. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 368: 1-8.

Benton BK, Plump SD, Roos J, Lennarz WJ & Cross FR (1996) Over-expression of *S. cerevisiae* G 1 cyclins restores the viability of *alg1 N*-glycosylation mutants. *Curr. Genet.* 29: 106–113.

Bickford LC, Mossessova E & Goldberg J (2004) A structural view of the COPII vesicle coat. *Curr. Opin. Struct. Biol.* 14: 147–153.

Biemans R, Thines D, Rutgers T, De Wilde M & Cabezon T (1991) The large surface protein of hepatitis B virus is retained in the yeast endoplasmic reticulum and provokes its unique enlargement. *DNA Cell Biol.* 10: 191–200.

Blobel G & Dobberstein B (1975) Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J. Cell Biol.* 67: 835–851.

Blobel G & Sabatini DD (1971) Ribosome-membrane interaction in eukaryotic cells. Manson, LA, ed. *Biomembranes* 193–195.

Bock JB, Matern HT, Peden AA & Scheller RH (2001) A genomic perspective on membrane compartment organization. *Nature* 409: 839–841.

Bohni PC, Deshaies RJ & Schekman RW (1988) *SEC11* is required for signal peptide processing and yeast cell growth. *J. Cell Biol.* 106: 1035–1042.

Bonifacino J & Glick B (2004) The mechanisms of vesicle budding and fusion. *Cell* 116: 153–166.

Bowers K & Stevens TH (2005) Protein transport from the late Golgi to the vacuole in the yeast *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* 1744: 438–454.

Braakman I & Hebert DN (2013) Protein folding in the endoplasmic reticulum. *Cold Spring Harb. Perspect. Biol.* 5: 1-17.

Brada D & Dubach UC (1984) Isolation of a homogeneous glucosidase II from pig kidney microsomes. *Eur. J. Biochem.* 141: 149–156.

Breinig F, Diehl B, Rau S, Zimmer C, Schwab H & Schmitt MJ (2006) Cell surface expression of bacterial esterase A by *Saccharomyces cerevisiae* and its enhancement by constitutive activation of the cellular unfolded protein response. *Appl. Environ. Microbiol.* 72: 7140–7147.

Brennwald P, Kearns B, Champion K, Keranen S, Bankaitis V & Novick P (1994) Sec9 is a SNAP-25-like component of a yeast SNARE complex that may be the effector of Sec4 function in exocytosis. *Cell* 79: 245–258.

Bretscher A (1994) What are the basic functions of microfilaments? Insights from studies in budding yeast. *J. Cell Biol.* 126: 821–825.

Brodsky JL, Goeckeler J & Schekman R (1995) BiP and Sec63p are required for both co- and posttranslational protein translocation into the yeast endoplasmic reticulum. *Proc. Natl. Acad. Sci.* 92: 9643–9646.

Bryant NJ & Stevens TH (1998) Vacuole biogenesis in *Saccharomyces cerevisiae*: protein transport pathways to the yeast vacuole. *Microbiol. Mol. Biol. Rev.* 62: 230–247.

Burda P (1998) The *ALG10* locus of *Saccharomyces cerevisiae* encodes the alpha-1,2 glucosyltransferase of the endoplasmic reticulum: the terminal glucose of the lipid-linked oligosaccharide is required for efficient *N*-linked glycosylation. *Glycobiology* 8: 455–462.

Burgoyne RD (1988) Control of exocytosis. *Nature* 328: 112–113.

Burri L & Lithgow T (2004) A complete set of SNAREs in yeast. *Traffic* 5: 45–52.

Bussey H, Steinmetz O & Saville D (1983) Protein secretion in yeast: Two chromosomal mutants that oversecrete killer toxin in *Saccharomyces cerevisiae*. *Curr. Genet.* 7: 449–456.

Butz J a, Niebauer RT & Robinson AS (2003) Co-expression of molecular chaperones does not improve the heterologous expression of mammalian G-protein coupled receptor expression in yeast. *Biotechnol. Bioeng.* 84: 292–304.

Caro LH, Tettelin H, Vossen JH, Ram AF, van den Ende H & Klis FM (1997) In silicio identification of glycosyl-phosphatidylinositol-anchored plasma-membrane and cell wall proteins of *Saccharomyces cerevisiae*. *Yeast* 13: 1477–1489.

Carr CM, Grote E, Munson M, Hughson FM & Novick PJ (1999) Sec1p binds to SNARE complexes and concentrates at sites of secretion. *J. Cell Biol.* 146: 333–344.

Chawla A, Chakrabarti S, Ghosh G & Niwa M (2011) Attenuation of yeast UPR is essential for survival and is mediated by Ire1 kinase. *J. Cell Biol.* 193: 41–50.

Chen Y, Pioli D & Piper PW (1994) Overexpression of the gene for polyubiquitin in yeast confers increased secretion of a human leucocyte protease inhibitor. *Bio/Technology* 12: 819–823.

Chen YA & Scheller RH (2001) SNARE-mediated membrane fusion. *Nat. Rev. Mol. Cell Biol.* 2: 98–106.

Chow TY, Ash JJ, Dignard D & Thomas DY (1992) Screening and identification of a gene, *PSE*-1, that affects protein secretion in *Saccharomyces cerevisiae*. *J. Cell Sci.* 101: 709–719.

Christis C, Lubsen NH & Braakman I (2008) Protein folding includes oligomerization – examples from the endoplasmic reticulum and cytosol. *FEBS J.* 275: 4700–4727.

Ciplys E, Sasnauskas K & Slibinskas R (2011) Overexpression of human calnexin in yeast improves measles surface glycoprotein solubility. *FEMS Yeast Res.* 11: 514–523.

Clerc S, Hirsch C, Oggier DM, Deprez P, Jakob C, Sommer T & Aebi M (2009) Htm1 protein generates the *N*-glycan signal for glycoprotein degradation in the endoplasmic reticulum. *J. Cell Biol.* 184: 159–172.

Copic A, Latham CF, Horlbeck MA, D'Arcangelo JG & Miller EA (2012) ER cargo properties specify a requirement for COPII coat rigidity mediated by Sec13p. *Science* 335: 1359–1362.

Copley KS, Alm SM, Schooley DA & Courchesne WE (1998) Expression, processing and secretion of a proteolytically-sensitive insect diuretic hormone by *Saccharomyces cerevisiae* requires the use of a yeast strain lacking genes encoding the Yap3 and Mkc7 endoproteases found in the secretory pathway. *Biochem. J.* 1340: 1333–1340.

Cuozzo JW & Kaiser CA (1999) Competition between glutathione and protein thiols for disulphide-bond formation. *Nat. Cell Biol.* 1: 130–135.

Cyr DM, Langer T & Douglas MG (1994) DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. *Trends Biochem. Sci.* 19: 176–181.

Damasceno LM, Anderson K, Ritter G, Cregg JM, Old LJ & Batt C a (2007) Cooverexpression of chaperones for enhanced secretion of a single-chain antibody fragment in *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* 74: 381–389.

Darby N & Creighton TE (1995) Characterization of the active site cysteine residues of the thioredoxin-like domains of protein disulfide isomerase. *Biochemistry* 34: 16770–16780.

Davydenko SG, Juselius JK, Munder T, Bogengruber E, Jäntti J & Keränen S (2004) Screening for novel essential genes of *Saccharomyces cerevisiae* involved in protein secretion. *Yeast* 21: 463–471.

Deak PM & Wolf DH (2001) Membrane topology and function of Der3/Hrd1p as a ubiquitin-protein ligase (E3) involved in endoplasmic reticulum degradation. *J. Biol. Chem.* 276: 10663–10669.

Demain AL & Vaishnav P (2009) Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* 27: 297–306.

De Pourcq K, De Schutter K & Callewaert N (2010) Engineering of glycosylation in yeast and other fungi: current state and perspectives. *Appl. Microbiol. Biotechnol.* 87: 1617–1631.

Denic V, Quan EM & Weissman JS (2006) A luminal surveillance complex that selects misfolded glycoproteins for ER-associated degradation. *Cell* 126: 349–359.

Deshaies RJ, Sanders SL, Feldheim DA & Schekman R (1991) Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. *Nature* 349: 806–808.

Deshaies RJ & Scheckman R (1989) *SEC62* encodes a putative membrane protein required for protein translocation into the yeast endoplasmic reticulum. *J. Cell Biol.* 109: 14550–14557.

Dobson CM (2004) Principles of protein folding, misfolding and aggregation. *Semin. Cell Dev. Biol.* 15: 3–16.

Duden R (2003) ER-to-Golgi transport: COP I and COP II function. *Mol. Membr. Biol.* 20: 197–207.

Ecker M, Mrsa V, Hagen I, Deutzmann R, Strahl S & Tanner W (2003) *O*-mannosylation precedes and potentially controls the *N*-glycosylation of a yeast cell wall glycoprotein. *EMBO Rep.* 4: 628–632.

Egel-Mitani M, Andersen A, Diers I, Hach M, Thim L, Hastrup S & Vad K (2000) Yield improvement of heterologous peptides expressed in *yps1*-disrupted *Saccharomyces cerevisiae* strains. *Enzyme Microb. Technol.* 26: 671–677.

Eisenhaber B, Bork P & Eisenhaber F (1998) Sequence properties of GPI-anchored proteins near the ω-site: Constraints for the polypeptide binding site of the putative transamidase. *Protein Eng.* 11: 1155–1161.

Elrod-erickson MJ (1998) Regulation of COPII vesicle formation and protein sorting by the *BST* genes in *Saccharomyces cerevisiae* (Ph.D.Thesis), MIT publishing, 1-184.

Erickson PR & Herzberg MC (1993) Evidence for the covalent linkage of carbohydrate polymers to a glycoprotein from *Streptococcus sanguis*. *J. Biol. Chem.* 268: 23780–23783.

Farquhar MG & Palade GE (1981) The Golgi apparatus (complex) (1954-1981) from artifact to center stage. *J. Cell Biol.* 91: 77-103.

Farquhar R, Honey N, Murant SJ, Bossier P, Schultz L, Montgomery D, Ellis RW, Freedman RB & Tuite MF (1991) Protein disulfide isomerase is essential for viability in *Saccharomyces cerevisiae*. *Gene* 108: 81–89.

Fernández FS, Trombetta SE, Hellman U & Parodi AJ (1994) Purification to homogeneity of UDP-glucose:glycoprotein glucosyltransferase from *Schizosaccharomyces pombe* and apparent absence of the enzyme from *Saccharomyces cerevisiae*. *J. Biol. Chem.* 269: 30701–30706.

Fewell SW, Travers KJ, Weissman JS & Brodsky JL (2001) The action of molecular chaperones in the early secretory pathway. *Annu. Rev. Genet.* 35: 149–191.

Fiedler K & Simons K (1996) Characterization of VIP36, an animal lectin homologous to leguminous lectins. *J. Cell Sci.* 109: 271–276.

Field C, Li R & Oegema K (1999) Cytokinesis in eukaryotes: a mechanistic comparison. *Curr. Opin. Cell Biol.* 11: 68–80.

Finger FP & Novick P (1998) Spatial regulation of exocytosis: Lessons from yeast. *J. Cell Biol.* 142: 609–612.

Finke K, Plath K, Panzner S, Prehn S, Rapoport T, Hartmann E & Sommer T (1996) A second trimeric complex containing homologs of the Sec61p complex functions in protein transport across the ER membrane of *S. cerevisiae*. *EMBO J.* 15: 1482–1494.

Frand AR, Cuozzo JW & Kaiser CA (2000) Pathways for protein disulphide bond formation. *Trends Cell Biol.* 10: 203–210.

Frand AR & Kaiser CA (1999) Ero1p oxidizes protein disulfide isomerase in a pathway for disulfide bond formation in the endoplasmic reticulum. *Mol. Cell* 4: 469–477.

Frieman MB & Cormack BP (2003) The ω-site sequence of glycosylphosphatidylinositol-anchored proteins in *Saccharomyces cerevisiae* can determine distribution between the membrane and the cell wall. *Mol. Microbiol.* 50: 883–896.

Garí E, Piedrafita L, Aldea M & Herrero E (1997) A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in *Saccharomyces cerevisiae*. *Yeast* 13: 837–848.

Gasser B, Maurer M, Gach J, Kunert R & Mattanovich D (2006) Engineering of *Pichia pastoris* for improved production of antibody fragments. *Biotechnol. Bioeng.* 94: 353–361.

Gasser B, Maurer M, Rautio J, Sauer M, Bhattacharyya A, Saloheimo M, Penttilä M & Mattanovich D (2007a) Monitoring of transcriptional regulation in *Pichia pastoris* under protein production conditions. *BMC Genomics* 8: 179-197.

Gasser B, Sauer M, Maurer M, Stadlmayr G & Mattanovich D (2007b) Transcriptomics-based identification of novel factors enhancing heterologous protein secretion in yeasts. *Appl. Environ. Microbiol.* 73: 6499–6507.

Gauss R, Jarosch E, Sommer T & Hirsch C (2006) A complex of Yos9p and the HRD ligase integrates endoplasmic reticulum quality control into the degradation machinery. *Nat. Cell Biol.* 8: 849–854.

Gemmill TR & Trimble RB (1999) Overview of *N*- and *O*-linked oligosaccharide structures found in various yeast species. *Biochim. Biophys. Acta - Gen. Subj.* 1426: 227–237.

Gentzsch M & Tanner W (1997) Protein *O*-glycosylation in yeast: protein-specific mannosyltransferases. *Glycobiology* 7: 481–486.

Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK & Weissman JS (2003) Global analysis of protein expression in yeast. *Nature* 425: 737–741.

Girrbach V, Zeller T, Priesmeier M & Strahl-Bolsinger S (2000) Structure-function analysis of the dolichyl phosphate-mannose: protein *O*-mannosyltransferase ScPmt1p. *J. Biol. Chem.* 275: 19288–19296.

Glick BS & Malhotra V (1998) The curious status of the Golgi apparatus. *Cell* 95: 883–889.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H & Oliver SG (1996) Life with 6000 genes. *Science* 274: 546–567.

Golgi C (1898) On the structure of the nerve cells of the spinal ganglia - 1898. *J. Microsc.* 155: 9–14.

González M, Brito N & González C (2012) High abundance of Serine/Threonine-rich regions predicted to be hyper-*O*-glycosylated in the secretory proteins coded by eight fungal genomes. *BMC Microbiol.* 12: 213-223.

Graham TR & Emr SD (1991) Compartmental organization of Golgi-specific protein modification and vacuolar protein sorting events defined in a yeast *sec18* (NSF) mutant. *J. Cell Biol.* 114: 207–218.

Griffiths G & Simons K (1986) The trans Golgi network: Sorting at the exit site of the Golgi complex. *Science* 234: 438–443.

De Groot PWJ, Ram AF & Klis FM (2005) Features and functions of covalently linked proteins in fungal cell walls. *Fungal Genet. Biol.* 42: 657–675.

Gross E, Sevier CS, Heldman N, Vitu E, Bentzur M, Kaiser CA, Thorpe C & Fass D (2006) Generating disulfides enzymatically: reaction products and electron acceptors of the endoplasmic reticulum thiol oxidase Ero1p. *Proc. Natl. Acad. Sci. U. S. A.* 103: 299–304.

Grote E, Carr CM & Novick PJ (2000) Ordering the final events in yeast exocytosis. *J. Cell Biol.* 151: 439–452.

Guerfal M, Ryckaert S, Jacobs PP, Ameloot P, Van Craenenbroeck K, Derycke R & Callewaert N (2010) The *HAC1* gene from *Pichia pastoris*: characterization and effect of its overexpression on the production of secreted, surface displayed and membrane proteins. *Microb. Cell Fact.* 9: 49-61.

Guo J & Polymenis M (2006) Dcr2 targets Ire1 and downregulates the unfolded protein response in *Saccharomyces cerevisiae*. *EMBO Rep.* 7: 1124–1127.

Hackel BJ, Huang D, Bubolz JC, Wang XX & Shusta EV (2006) Production of soluble and active transferrin receptor-targeting single-chain antibody using *Saccharomyces cerevisiae*. *Pharm. Res.* 23: 790–797.

Halic M & Beckmann R (2005) The signal recognition particle and its interactions during protein targeting. *Curr. Opin. Struct. Biol.* 15: 116–125.

Hamada K, Fukuchi S, Arisawa M, Baba M & Kitada K (1998) Screening for glycosylphosphatidylinositol (GPI)-dependent cell wall proteins in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* 258: 53–59.

Hamada K, Terashima H, Arisawa M & Kitada K (1998) Amino acid sequence requirement for efficient incorporation of glycosylphosphatidylinositol-associated proteins into the cell wall of *Saccharomyces cerevisiae*. *J. Biol. Chem.* 273: 26946–26953.

Hamada K, Terashima H, Arisawa M, Yabuki N & Kitada K (1999) Amino acid residues in the ω-minus region participate in cellular localization of yeast glycosylphosphatidylinositol-attached proteins. *J. Bacteriol.* 181: 3886–3889.

Hammond C, Braakman I & Helenius A (1994) Role of *N*-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proc. Natl. Acad. Sci. U. S. A.* 91: 913–917.

Hammond C & Helenius A (1994) Folding of VSV G protein: Sequential interaction with BiP and calnexin. *Science* 266: 456–458.

Harmsen MM, Bruyne MI, Raué HA & Maat J (1996) Overexpression of binding protein and disruption of the *PMR1* gene synergistically stimulate secretion of bovine prochymosin but not plant Thaumatin in yeast. *Appl. Microbiol. Biotechnol.* 46: 365–370.

Harsay E & Schekman R (2002) A subset of yeast vacuolar protein sorting mutants is blocked in one branch of the exocytic pathway. *J. Cell Biol.* 156: 271–285.

Hatahet F & Ruddock LW (2009) Protein disulfide isomerase: a critical evaluation of its function in disulfide bond formation. *Antioxid. Redox Signal.* 11: 2807–2850.

Hayano T, Hirose M & Kikuchi M (1995) Protein disulfide isomerase mutant lacking its isomerase activity accelerates protein folding in the cell. 377: 505–511.

Hebert DN, Garman SC & Molinari M (2005) The glycan code of the endoplasmic reticulum: asparagine-linked carbohydrates as protein maturation and quality-control tags. *Trends Cell Biol.* 15: 364–370.

Heider MR & Munson M (2012) Exorcising the exocyst complex. *Traffic* 13: 898–907.

Helms JB & Zurzolo C (2004) Lipids as targeting signals: lipid rafts and intracellular trafficking. *Traffic* 5: 247–254.

Herrmann JM, Malkus P & Schekman R (1999) Out of the ER - outfitters, escorts and guides. *Trends Cell Biol.* 9: 5–7.

Hirsch C, Blom D & Ploegh HL (2003) A role for *N*-glycanase in the cytosolic turnover of glycoproteins. *EMBO J.* 22: 1036–1046.

Hirsch C, Gauss R, Horn SC, Neuber O & Sommer T (2009) The ubiquitylation machinery of the endoplasmic reticulum. *Nature* 458: 453–460.

Holst B, Tachibana C & Winther JR (1997) Active site mutations in yeast protein disulfide isomerase cause dithiothreitol sensitivity and a reduced rate of protein folding in the endoplasmic reticulum. *J. Cell Biol.* 138: 1229–1238.

Hong E, Davidson AR & Kaiser C. A (1996) A Pathway for targeting soluble misfolded proteins to the yeast vacuole. *J. Cell Biol.* 135: 623–633.

Hong W & Lev S (2014) Tethering the assembly of SNARE complexes. *Trends Cell Biol.* 24: 35–43.

Hoshida H, Fujita T, Cha-aim K & Akada R (2013) *N*-Glycosylation deficiency enhanced heterologous production of a *Bacillus licheniformis* thermostable α-amylase in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 97: 5473–5482.

Hou J, Tyo KEJ, Liu Z, Petranovic D & Nielsen J (2012) Metabolic engineering of recombinant protein secretion by *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 12: 1–20.

Hu C, Ahmed M, Melia TJ, Söllner TH, Mayer T & Rothman JE (2003) Fusion of cells by flipped SNAREs. *Science* 300: 1745–1749.

Huang D, Gore PR & Shusta EV (2008) Increasing yeast secretion of heterologous proteins by regulating expression rates and post-secretory loss. *Biotechnol. Bioeng.* 101: 1264–1275.

Huo X, Liu Y, Wang X, Ouyang P, Niu Z, Shi Y & Qiu B (2007) Co-expression of human protein disulfide isomerase (hPDI) enhances secretion of bovine follicle-stimulating hormone (bFSH) in *Pichia pastoris*. *Protein Expr. Purif.* 54: 234–239.

Idiris A, Bi K, Tohda H, Kumagai H & Giga-Hama Y (2006a) Construction of a protease-deficient strain set for the fission yeast *Schizosaccharomyces pombe*, useful for effective production of protease-sensitive heterologous proteins. *Yeast* 23: 83–99.

Idiris A, Tohda H, Bi K, Isoai A, Kumagai H & Giga-Hama Y (2006b) Enhanced productivity of protease-sensitive heterologous proteins by disruption of multiple protease genes in the fission yeast *Schizosaccharomyces pombe*. *Appl. Microbiol. Biotechnol.* 73: 404–420.

Idiris A, Tohda H, Kumagai H & Takegawa K (2010a) Engineering of protein secretion in yeast: strategies and impact on protein production. *Appl. Microbiol. Biotechnol.* 86: 403–417.

Idiris A, Tohda H, Sasaki M, Okada K, Kumagai H, Giga-Hama Y & Takegawa K (2010b) Enhanced protein secretion from multiprotease-deficient fission yeast by modification of its vacuolar protein sorting pathway. *Appl. Microbiol. Biotechnol.* 85: 667–677.

Imperiali B & Hendrickson TL (1995) Asparagine-linked glycosylation: Specificity and function of oligosaccharyl transferase. *Bioorganic Med. Chem.* 3: 1565–1578.

Imperiali B & O'Connor SE (1999) Effect of *N*-linked glycosylation on glycopeptide and glycoprotein structure. *Curr. Opin. Chem. Biol.* 3: 643–649.

Imperiali B & Rickert KW (1995) Conformational implications of asparagine-linked glycosylation. *Proc. Natl. Acad. Sci.* 92: 97–101.

Inan M, Aryasomayajula D, Sinha J & Meagher MM (2006) Enhancement of protein secretion in *Pichia pastoris* by overexpression of protein disulfide isomerase. *Biotechnol. Bioeng.* 93: 771–778.

Jahn R, Lang T & Südhof TC (2003) Membrane Fusion. *Cell* 112: 519–533.

Jakob CA, Burda P, te Heesen S, Aebi M & Roth J (1998) Genetic tailoring of *N*-linked oligosaccharides: The role of glucose residues in glycoprotein processing of *Saccharomyces cerevisiae in vivo*. *Glycobiology* 8: 155–164.

Jakob CA, Burda P, Roth J & Aebi M (1998) Degradation of misfolded endoplasmic reticulum glycoproteins in *Saccharomyces cerevisiae* is determined by a specific oligosaccharide structure. *J. Cell Biol.* 142: 1223–1233.

Jungmann J & Munro S (1998) Multi-protein complexes in the *cis* Golgi of *Saccharomyces cerevisiae* with α-1,6-mannosyltransferase activity. *EMBO J.* 17: 423–434.

Kang H, Kim SJ, Choi ES, Rhee SK & Chung BH (1998) Efficient production of intact human parathyroid hormone in a *Saccharomyces cerevisiae* mutant deficient in yeast aspartic protease 3 (Yap3). *Appl. Microbiol. Biotechnol.* 50: 187–192.

Kanjou N, Nagao A, Ohmiya Y & Ohgiya S (2007) Yeast mutant with efficient secretion identified by a novel secretory reporter, Cluc. *Biochem. Biophys. Res. Commun.* 358: 429–434.

Kapteyn JC, Montijn RC, Vink E, De La Cruz J, Llobell A, Douwes JE, Shimoi H, Lipke PN & Klis FM (1996) Retention of *Saccharomyces cerevisiae* cell wall proteins through a phosphodiester-linked β-1,3-/β-1,6-glucan heteropolymer. *Glycobiology* 6: 337–345.

Karpova TS, Reck-peterson SL, Elkind NB, Mooseker MS, Novick PJ & Cooper JA (2000) Role of actin and Myo2p in polarized secretion and growth of *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 11: 1727–1737.

Kerry-Williams SM, Gilbert SC, Evans LR & Ballance DJ (1998) Disruption of the *Saccharomyces cerevisiae YAP3* gene reduces the proteolytic degradation of secreted recombinant human albumin. *Yeast* 14: 161–169.

Kim M, Han K, Kang H & Rhee S (2003) Coexpression of BiP increased antithrombotic hirudin production in recombinant *Saccharomyces cerevisiae*. 101: 81-87.

Kimata Y, Kimata YI, Shimizu Y, Abe H, Farcasanu IC, Takeuchi M, Rose MD & Kohno K (2003) Genetic evidence for a role of BiP/Kar2 that regulates Ire1 in response to accumulation of unfolded proteins. *Mol. Biol. Cell* 14: 2559–2569.

Kimura T, Hosoda Y, Kitamura Y, Nakamura H, Horibe T & Kikuchi M (2004) Functional differences between human and yeast protein disulfide isomerase family proteins. *Biochem. Biophys. Res. Commun.* 320: 359–365.

Klabunde J, Kleebank S, Piontek M, Hollenberg CP, Hellwig S & Degelmann A (2007) Increase of calnexin gene dosage boosts the secretion of heterologous proteins by *Hansenula polymorpha*. *FEMS Yeast Res.* 7: 1168–1180.

Knauer R & Lehle L (1999) The oligosaccharyltransferase complex from yeast. *Biochim. Biophys. Acta - Gen. Subj.* 1426: 259–273.

Knop M, Hauser N & Wolf DH (1996) *N*-Glycosylation affects endoplasmic reticulum degradation of a mutated derivative of carboxypeptidase yscY in yeast. *Yeast* 12: 1229–1238.

Kohno K, Normington K, Sambrook J, Gething MJ & Mori K (1993) The promoter region of the yeast *KAR2* (*BiP*) gene contains a regulatory domain that responds to the presence of unfolded proteins in the endoplasmic reticulum. *Mol. Cell. Biol.* 13: 877–890.

Kroukamp H, den Haan R, van Wyk N & van Zyl WH (2013) Overexpression of native *PSE1* and *SOD1* in *Saccharomyces cerevisiae* improved heterologous cellulase secretion. *Appl. Energy* 102: 150–156.

Larsson S, Cassland P & Jönsson LJ (2001) Development of a *Saccharomyces cerevisiae* strain with enhanced resistance to phenolic fermentation inhibitors in lignocellulose hydrolysates by heterologous expression of laccase. *Appl. Environ. Microbiol.* 67: 1163–1170.

Lederkremer GZ (2009) Glycoprotein folding, quality control and ER-associated degradation. *Curr. Opin. Struct. Biol.* 19: 515–523.

Lederkremer GZ & Glickman MH (2005) A window of opportunity: timing protein degradation by trimming of sugars and ubiquitins. *Trends Biochem. Sci.* 30: 297–303.

Lee MCS & Miller EA (2007) Molecular mechanisms of COPII vesicle formation. *Semin. Cell Dev. Biol.* 18: 424–434.

Leidich SD, Drapp DA & Orlean P (1994) A conditionally lethal yeast mutant blocked at the first step in glycosyl phosphatidylinositol anchor synthesis. *J. Biol. Chem.* 269: 10193–10196.

Leopold PE, Montal M & Onuchic JN (1992) Protein folding funnels: A kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. U. S. A.* 89: 8721–8725.

Liu Z, Liu L, Osterlund T, Hou J, Huang M, Fagerberg L, Petranovic D, Uhlén M & Nielsen J (2014) Improved production of a heterologous amylase in *Saccharomyces cerevisiae* by inverse metabolic engineering. *Appl. Environ. Microbiol.* 80: 5542–5550.

Lodi T, Neglia B & Donnini C (2005) Secretion of human serum albumin by *Kluyveromyces lactis* overexpressing *KlPDI1* and *KlERO1*. 71: 4359-4363.

Lodish HF (1988) Transport of secretory and membrane glycoproteins from the rough endoplasmic reticulum to the Golgi. A rate-limiting step in protein maturation and secretion. *J. Biol. Chem.* 263: 2107–2110.

Loibl M & Strahl S (2013) Protein *O*-mannosylation: what we have learned from baker's yeast. *Biochim. Biophys. Acta* 1833: 2438–2446.

Loubéry S & Coudrier E (2008) Myosins in the secretory pathway: tethers or transporters? *Cell. Mol. Life Sci.* 65: 2790–2800.

Lu CF, Kurjan J & Lipke PN (1994) A pathway for cell wall anchorage of *Saccharomyces cerevisiae* α-agglutinin. *Mol. Cell. Biol.* 14: 4825–4833.

Lu CF, Montijn RC, Brown JL, Klis FM, Kurjan J, Bussey H & Lipke PN (1995) Glycosyl phosphatidylinositol-dependent cross-linking of alpha- agglutinin and beta 1,6-glucan in the *Saccharomyces cerevisiae* cell wall. *J. Cell Biol.* 128: 333–340.

Määttänen P, Gehring K, Bergeron JJM & Thomas DY (2010) Protein quality control in the ER: the recognition of misfolded proteins. *Semin. Cell Dev. Biol.* 21: 500–511.

Malsam J, Kreye S & Söllner TH (2008) Membrane fusion: SNAREs and regulation. *Cell. Mol. Life Sci.* 65: 2814–2832.

Marcinowski M, Höller M, Feige MJ, Baerend D, Lamb DC & Buchner J (2011) Substrate discrimination of the chaperone BiP by autonomous and cochaperone-regulated conformational transitions. *Nat. Struct. Mol. Biol.* 18: 150–158.

Martin JL (1995) Thioredoxin - A fold for all reasons. *Structure* 3: 245–250.

Martin SG & Arkowitz R (2014) Cell polarization in budding and fission yeasts. *FEMS Microbiol. Rev.* 38: 228–253.

Matlack KE, Misselwitz B, Plath K & Rapoport TA (1999) BiP acts as a molecular ratchet during posttranslational transport of prepro-alpha factor across the ER membrane. *Cell* 97: 553–564.

Matsuoka K, Morimitsu Y, Uchida K & Schekman R (1998) Coat assembly directs v-SNARE concentration into synthetic COPII vesicles. *Mol. Cell* 2: 703–708.

Mattanovich D, Gasser B, Hohenblum H & Sauer M (2004) Stress in recombinant protein producing yeasts. *J. Biotechnol.* 113: 121–135.

Mellman I & Warren G (2000) The road taken : Past and future foundations of membrane traffic. *Cell* 100: 99–112.

Meyer H-A, Grau H, Kraft R, Kostka S, Prehn S, Kalies K-U & Hartmann E (2000) Mammalian Sec61 is associated with Sec62 and Sec63. *J. Biol. Chem.* 275: 14550–14557.

Miller EA & Barlowe C (2010) Regulation of coat assembly- sorting things out at the ER. *Curr. Opin. Cell Biol.* 22: 447–453.

Milstein C, Brownlee GG, Harrison TM & Mathews MB (1972) A possible precursor of immunoglobulin light chains. *Nature* 239: 117–120.

Mori K, Sant A, Kohno K, Normington K, Gething MJ & Sambrook JF (1992) A 22 bp *cis*-acting element is necessary and sufficient for the induction of the yeast *KAR2* (*BiP*) gene by unfolded proteins. *EMBO J.* 11: 2583–2593.

Nakatsukasa K & Brodsky JL (2008) The recognition and retrotranslocation of misfolded proteins from the endoplasmic reticulum. *Traffic* 9: 861–870.

Ng DT, Brown JD & Walter P (1996) Signal sequences specify the targeting route to the endoplasmic reticulum membrane. *J. Cell Biol.* 134: 269–278.

Ni Z, Zhou X, Sun X, Wang Y & Zhang Y (2008) Decrease of hirudin degradation by deleting the *KEX1* gene in recombinant *Pichia pastoris*. *Yeast* 25: 1–8.

Nørgaard P, Westphal V, Tachibana C, Alsøe L, Holst B & Winther JR (2001) Functional differences in yeast protein disulfide isomerases. *J. Cell Biol.* 152: 553–562.

Normington K, Kohno K, Kozutsumi Y, Gething MJ & Sambrook J (1989) *S. cerevisiae* encodes an essential protein homologous in sequence and function to mammalian BiP. *Cell* 57: 1223–1236.

Novick P, Ferro S & Schekman R (1981) Order of events in the yeast secretory pathway. *Cell* 25: 461–469.

Novick P, Field C & Schekman R (1980) Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway. *Cell* 21: 205–215.

Ohishi K, Inoue N, Maeda Y, Takeda J, Riezman H & Kinoshita T (2000) Gaa1p and Gpi8p are components of a glycosylphosphatidylinositol (GPI) transamidase that mediates attachment of GPI to proteins. *Mol. Biol. Cell* 11: 1523–1533.

Oka T & Krieger M (2005) Multi-component protein complexes and Golgi membrane trafficking. *J. Biochem.* 137: 109–114.

Orlean P & Menon AK (2007) Thematic review series: lipid posttranslational modifications. GPI anchoring of protein in yeast and mammalian cells, or: how we learned to stop worrying and love glycophospholipids. *J. Lipid Res.* 48: 993–1011.

Osborne AR, Rapoport T & van den Berg B (2005) Protein translocation by the Sec61/SecY channel. *Annu. Rev. Cell Dev. Biol.* 21: 529–550.

Palade G (1975) Intracellular aspects of the process of protein synthesis. *Science* 189: 347–358.

Panzner S, Dreier L, Hartmann E, Kostka S & Rapoport TA (1995) Posttranslational protein transport into the endoplasmic reticulum. *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. 60, pp. 31–40.

Parlati F, Dominquez M, Bergeron JJM & Thomas DY (1995) *Saccharomyces cerevisiae CNE1* encodes an endoplasmic reticulum (ER) membrane protein with sequence similarity to calnexin and calreticulin and functions as a constituent of the ER quality control apparatus. *J. Biol. Chem.* 270: 277–253.

Parts L et al. (2011) Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res.* 21: 1131–1138.

Payne T, Finnis C, Evans LR, Mead DJ, Avery S V, Archer DB & Sleep D (2008) Modulation of chaperone gene expression in mutagenized *Saccharomyces cerevisiae* strains developed for recombinant human albumin production results in increased production of multiple heterologous proteins. *Appl. Environ. Microbiol.* 74: 7759–7766.

Pelham HR (1999) SNAREs and the secretory pathway-lessons from yeast. *Exp. Cell Res.* 247: 1–8.

Petkova MI, Pujol-Carrion N & de la Torre-Ruiz MA (2012) Mtl1 *O*-mannosylation mediated by both Pmt1 and Pmt2 is important for cell survival under oxidative conditions and TOR blockade. *Fungal Genet. Biol.* 49: 903–914.

Petrescu A-J, Milac A-L, Petrescu SM, Dwek RA & Wormald MR (2004) Statistical analysis of the protein environment of *N*-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* 14: 103–114.

Pruyne D & Bretscher A (2000) Polarization of cell growth in yeast I. Establishment and maintenance of polarity states. 375: 365–375.

Pruyne DW, Schott DH & Bretscher A (1998) Tropomyosin-containing actin cables direct the Myo2p-dependent polarized delivery of secretory vesicles in budding yeast. *J. Cell Biol.* 143: 1931–1945.

Qian W, Liu Y, Zhang C, Niu Z, Song H & Qiu B (2009) Expression of bovine follicle-stimulating hormone subunits in a *Hansenula polymorpha* expression system increases the secretion and bioactivity in vivo. *Protein Expr. Purif.* 68: 183–189.

Quan EM, Kamiya Y, Kamiya D, Denic V, Weibezahn J, Kato K & Weissman JS (2008) Defining the glycan destruction signal for endoplasmic reticulum-associated degradation. *Mol. Cell* 32: 870–877.

Raimondi S, Zanni E, Talora C, Rossi M, Palleschi C & Uccelletti D (2008) *SOD1*, a new *Kluyveromyces lactis* helper gene for heterologous protein secretion. *Appl. Environ. Microbiol.* 74: 7130–7137.

Rakestraw JA, Sazinsky SL, Piatesi A, Antipov E & Wittrup KD (2009) Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 103: 1192–1201.

Rapoport T (2007) Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* 450: 663–669.

Rapoport TA, Goder V, Heinrich SU & Matlack KES (2004) Membrane-protein integration and the role of the translocation channel. *Trends Cell Biol.* 14: 568–575.

Robinson AS, Bockhaus J, Voegler AC & Wittrup KD (1996) Reduction of BiP levels decreases heterologous protein secretion in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 271: 10017–10022.

Robinson AS, Hines V & Wittrup KD (1994) Protein disulfide isomerase overexpression increases secretion of foreign proteins in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 12: 381–384.

Robinson AS & Wittrup KD (1995) Constitutive overexpression of secreted heterologous proteins decreases extractable BiP and protein disulfide isomerase levels in *Saccharomyces cerevisiae*. *Biotechnol. Prog.* 11: 171–177.

Robinson NGG, Guo L, Imai J, Toh-E A, Matsui Y & Tamanoi F (1999) Rho3 of *Saccharomyces cerevisiae*, which regulates the actin cytoskeleton and exocytosis, is a GTPase which interacts with Myo2 and Exo70. *Mol. Cell. Biol.* 19: 3580–3587.

Rose MD, Misra LM & Vogel JP (1989) *KAR2*, a karyogamy gene, is the yeast homolog of the mammalian *BiP/GRP78* gene. *Cell* 57: 1211–1221.

Roth J, Zuber C, Park S, Jang I, Lee Y, Kysela KG, Le Fourn V, Santimaria R, Guhl B, Cho JW (2010) Protein *N*-glycosylation, protein folding, and protein quality control. *Mol. Cells* 30: 497–506.

Rothman JE (1981) The Golgi apparatus: Two organelles in tandem. *Science* 213: 1212–1219.

Rudolph HK, Antebi A, Fink GR, Buckley CM, Dorman TE, LeVitre J, Davidow LS, Mao JI & Moir DT (1989) The yeast secretory pathway is perturbed by mutations in *PMR1*, a member of a Ca^{2+} ATPase family. *Cell* 58: 133–145.

Ruohonen L, Toikkanen J, Tieaho V, Outola M, Soderlund H & Keranen S (1997) Enhancement of protein secretion in *Saccharomyces cerevisiae* by overproduction of Sso protein, a late-acting component of the secretory machinery. *Yeast* 13: 337–351.

Sakai, A., Shimizu, Y. & Hishinuma F (1988) Isolation and characterization of mutants which show an oversecretion phenotype in *Saccharomyces cerevisiae*. *Genetics* 119: 499–506.

Sakoh-Nakatogawa M, Nishikawa S-I & Endo T (2009) Roles of protein-disulfide isomerase-mediated disulfide bond formation of yeast Mnl1p in endoplasmic reticulum-associated degradation. *J. Biol. Chem.* 284: 11815–11825.

Sanders SL, Whitfield KM, Vogel JP, Rose MD & Schekman RW (1992) Sec61p and BiP directly facilitate polypeptide translocation into the ER. *Cell* 69: 353–365.

Sato K (2004) COPII coat assembly and selective export from the endoplasmic reticulum. *J. Biochem.* 136: 755–760.

Schekman R (1982) The secretory pathway in yeast. *Trends Biochem. Sci*. 7: 243–246.

Schlenstedt G, Harris S, Risse B, Lill R & Silver PA (1995) A yeast DnaJ homologue, Scj1p, can function in the endoplasmic reticulum with BiP/Kar2p via a conserved domain that specifies interactions with Hsp70s. *J. Cell Biol.* 129: 979–988.

Schott D, Ho J, Pruyne D & Bretscher A (1999) The COOH-terminal domain of Myo2p, a yeast myosin V, has a direct role in secretory vesicle targeting. *J. Cell Biol.* 147: 791–808.

Schroder M & Kaufman R (2006) Divergent roles of IRE1α and PERK in the unfolded protein response. *Curr. Mol. Med.* 6: 5–36.

Schubert U, Antón LC, Gibbs J, Norbury CC, Yewdell JW & Bennink JR (2000) Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* 404: 770–774.

Sekijima Y, Hammarström P, Matsumura M, Shimizu Y, Iwata M, Tokuda T, Ikeda S, & Kelly JW (2003) Energetic characteristics of the new transthyretin variant A25T may explain its atypical central nervous system pathology. *Lab. Investig.* 83: 409–417.

Sekijima Y, Wiseman RL, Matteson J, Hammarström P, Miller SR, Sawkar AR, Balch WE & Kelly JW (2005) The biological and chemical basis for tissue-selective amyloid disease. *Cell* 121: 73–85.

Shen Q, Wu M, Wang H-B, Naranmandura H & Chen S-Q (2012) The effect of gene copy number and co-expression of chaperone on production of albumin fusion proteins in *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* 96: 763–772.

Shen X, Zhang K & Kaufman RJ (2004) The unfolded protein response - a stress signaling pathway of the endoplasmic reticulum. *J. Chem. Neuroanat.* 28: 79–92.

Shusta EV, Holler PD, Kieke MC, Kranz DM & Wittrup KD (2000) Directed evolution of a stable scaffold for T-cell receptor engineering. *Nat. Biotechnol.* 18: 754–759.

Shusta EV, Raines RT, Pluckthun A & Wittrup KD (1998) Increasing the secretory capacity of *Saccharomyces cerevisiae* for production of single-chain antibody fragments. *Nat. Biotechnol.* 16: 773–777.

Shuster JR (1991) Gene expression in yeast: protein secretion. *Curr. Opin. Biotechnol.* 2: 685–690.

Simons JF, Ferro-novick S, Rose MD & Helenius A (1995) BiP/Kar2p serves as a molecular chaperone during carboxypeptidase Y folding in yeast. *J. Cell Biol.* 130: 41–49.

Simons K & Toomre D (2000) Lipid rafts and signal transduction. *Nat. Rev. Mol. Cell Biol.* 1: 31–39.

Smith JD, Tang BC & Robinson AS (2004) Protein disulfide isomerase, but not binding protein, overexpression enhances secretion of a non-disulfide-bonded protein in yeast. *Biotechnol. Bioeng.* 85: 340–350.

Smith RA, Duncan MJ & Moir DT (1985) Heterologous protein secretion from yeast. *Science* 229: 1219–1224.

Sodoyer R (2004) Expression systems for the production of recombinant pharmaceuticals. *BioDrugs* 18: 51–62.

Söllner T, Whiteheart SW, Brunner M, Erdjument-Bromage H, Geromanos S, Tempst P & Rothman JE (1993) SNAP receptors implicated in vesicle targeting and fusion. *Nature* 362: 318–324.

Song Y, Sata J, Saito A, Usui M, Azakami H & Kato A (2001) Effects of calnexin deletion in *Saccharomyces cerevisiae* on the secretion of glycosylated lysozymes. *J. Biochem.* 130: 757–764.

Spang A (2013) Retrograde traffic from the Golgi to the endoplasmic reticulum. *Cold Spring Harb. Perspect. Biol.* 5: pp ii.

Spiro RG (2002) Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* 12:43-56.

Suda Y & Nakano A (2012) The yeast Golgi apparatus. *Traffic* 13: 505–510.

Suzuki T & Lennarz WJ (2002) Glycopeptide export from the endoplasmic reticulum into cytosol is mediated by a mechanism distinct from that for export of misfolded glycoprotein. *Glycobiology* 12: 803–811.

Suzuki T, Park H, Hollingsworth NM, Sternglanz R & Lennarz WJ (2000) *PNG1*, a yeast gene encoding a highly conserved peptide:*N*-glycanase. *J. Cell Biol.* 149: 1039–1052.

Swanson R, Locher M & Hochstrasser M (2001) A conserved ubiquitin ligase of the nuclear envelope/endoplasmic reticulum that functions in both ER-associated and Matalpha2 repressor degradation. *Genes Dev.* 15: 2660–2674.

TerBush DR, Guo W, Dunkelbarger S & Novick P (2001) Regulators and effectors of small GTPases Part E: GTPases involved in visicular traffic. Elsevier pp 100 -110.

Toikkanen JH, Gatti E, Takei K, Saloheimo M, Olkkonen VM, Söderlund H, De Camilli P & Keränen S (1996) Yeast protein translocation complex: isolation of two genes *SEB1* and *SEB2* encoding proteins homologous to the Sec61 beta subunit. *Yeast* 12: 425–438.

Toikkanen JH, Miller KJ, Söderlund H, Jäntti J & Keränen S (2003) The beta subunit of the Sec61p endoplasmic reticulum translocon interacts with the exocyst complex in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 278: 20946–20953.

Toikkanen JH, Sundqvist L & Keränen S (2004) *Kluyveromyces lactis SSO1* and *SEB1* genes are functional in *Saccharomyces cerevisiae* and enhance production of secreted proteins when overexpressed. *Yeast* 21: 1045–1055.

Tokunaga M, Kato S, Kawamura-Watabe A, Tanaka R & Tokunaga H (1998) Characterization of deletion mutations in the carboxy-terminal peptide-binding domain of the Kar2 protein in *Saccharomyces cerevisiae*. *Yeast* 14: 1285–1295.

Tomimoto K, Fujita Y, Iwaki T, Chiba Y, Jigami Y, Nakayama K, Nakajima Y & Abe H (2013) Protease-deficient *Saccharomyces cerevisiae* strains for the synthesis of human-compatible glycoproteins. *Biosci. Biotechnol. Biochem.* 77: 2461–2466.

Travers KJ, Patil CK, Wodicka L, Lockhart DJ, Weissman JS & Walter P (2000) Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. *Cell* 101: 249–258.

Tsai B, Ye Y & Rapoport T (2002) Retro-translocation of proteins from the endoplasmic reticulum into the cytosol. *Nat. Rev. Mol. Cell Biol.* 3: 246–255.

Tsai CW, Duggan PF, Shimp RL, Miller LH & Narum DL (2006) Overproduction of *Pichia pastoris* or *Plasmodium falciparum* protein disulfide isomerase affects expression, folding and *O*-linked glycosylation of a malaria vaccine candidate expressed in *P. pastoris*. *J. Biotechnol.* 121: 458–470.

Tyson JR & Stirling CJ (2000) *LHS1* and *SIL1* provide a lumenal function that is essential for protein translocation into the endoplasmic reticulum. *EMBO J.* 19: 6440–6452.

Uccelletti D, Anticoli S & Palleschi C (2007) The apyrase KlYnd1p of *Kluyveromyces lactis* affects glycosylation, secretion, and cell wall properties. *FEMS Yeast Res.* 7: 731–739.

Uccelletti D, Farina F, Cesca, Rufini S, Magnelli P, Abeijon C & Palleschi C (2006) The *Kluyveromyces lactis* α1,6-mannosyltransferase KlOch1p is required for cell-wall organization and proper functioning of the secretory pathway. *FEMS Microbiol. Rev.* 6: 449–457.

Uccelletti D, Staneva D, Rufini S, Venkov P & Palleschi C (2005) Enhanced secretion of heterologous proteins in *Kluyveromyces lactis* by overexpression of the GDP-mannose pyrophosphorylase, KlPsa1p. *FEMS Yeast Res.* 5: 735–746.

Uccelletti D, Mancini P, Farina F, Morrone S & Palleschi C (1999) Inactivation of the *KIPMR1* gene of *Kluyveromyces lactis* results in defective cell-wall morphogenesis. *Microbiology* 145: 1079–1087.

Vad R, Moe E, Saga K, Kvinnsland M & Oyen TB (1998) High-level production of human parathyroid hormone (hPTH) by induced expression in *Saccharomyces cerevisiae*. *Protein Expr. Purif.* 13: 396–402.

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

Valkonen M, Penttilä M & Saloheimo M (2003) Effects of inactivation and constitutive expression of the unfolded protein response pathway on protein production in the yeast *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol*. 69: 2065-2072.

Van der Heide M, Hollenberg C, van der Klei I & Veenhuis M (2002) Overproduction of BiP negatively affects the secretion of *Aspergillus niger* glucose oxidase by the yeast *Hansenula polymorpha*. *Appl. Microbiol. Biotechnol.* 58: 487–494.

Van Zyl JHD, Den Haan R & Van Zyl WH (2014) Over-expression of native *Saccharomyces cerevisiae* exocytic SNARE genes increased heterologous cellulase secretion. *Appl. Microbiol. Biotechnol.* 98: 5567–5578.

Varki A (1993) Biological roles of oligosaccharides: All of the theories are correct. *Glycobiology* 3: 97–130.

Varki A, Freeze HH & Manzi AE (2009) Overview of glycoconjugate analysis. *Curr. Protoc. Protein Sci.* Chapter 12: 12.1.1–8.

Verghese J, Abrams J, Wang Y & Morano K (2012) Biology of the heat shock response and protein chaperones: budding yeast (*Saccharomyces cerevisiae*) as a model system. *Microbiol. Mol. Biol. Rev.* 76: 115–158.

Vogel JP (1990) Loss of BiP/GRP78 function blocks translocation of secretory proteins in yeast. *J. Cell Biol.* 110: 1885–1895.

Walter P & Johnson E (1994) Signal sequence recognition and protein targeting to the membrane. *Annu. Rev. Cell Biol.* 10: 87-119.

Walter P & Lingappa VR (1986) Mechanism of protein translocation across the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* 2: 499–516.

Wang T-Y, Huang CJ, Chen HL, Ho PC, Ke HM, Cho HY, Ruan SK, Hung KY, Wang IL, Cai YW, Sung HM, Li WH & Shih MC (2013) Systematic screening of glycosylation- and trafficking-associated gene knockouts in *Saccharomyces cerevisiae* identifies mutants with improved heterologous exocellulase activity and host secretion. *BMC Biotechnol.* 13: 71–85.

Weber T, Zemelman B V, McNew JA, Westermann B, Gmachl M, Parlati F, Söllner TH & Rothman JE (1998) SNAREpins: Minimal machinery for membrane fusion. *Cell* 92: 759–772.

Weerapana E & Imperiali B (2006) Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* 16: 91–101.

Wentz AE & Shusta EV (2007a) A novel high-throughput screen reveals yeast genes that increase secretion of heterologous proteins. *Appl. Environ. Microbiol.* 73: 1189–1198.

Wentz AE & Shusta EV (2007b) A novel high-throughput screen reveals yeast genes that increase secretion of heterologous proteins. *Appl. Environ. Microbiol.* 73: 1189–1198.

Wentz AE & Shusta EV (2008) Enhanced secretion of heterologous proteins from yeast by overexpression of ribosomal subunit RRP0. *Biotechnol. Prog.* 24: 748–756.

White RE, Dickinson JR, Semple CM, Powell DJ & Berry C (2011) The retroviral proteinase active site and the *N*-terminus of Ddi1 are required for repression of protein secretion. *FEBS Lett.* 585: 139–142.

Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-MacDonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, Véronneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, Johnston M & Davis RW (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285: 901–906.

Wiseman RL, Powers ET, Buxbaum JN, Kelly JW & Balch WE (2007) An adaptable standard for protein export from the endoplasmic reticulum. *Cell* 131: 809–821.

Wolff AM, Petersen JGL, Nilsson-tillgren T & Din N (1999) The open reading frame YAL048c affects the secretion of proteinase A in *S . cerevisiae*. *Yeast* 15: 427–434.

Wu M, Shen Q, Yang Y, Zhang S, Qu W, Chen J, Sun H & Chen S (2013) Disruption of *YPS1* and *PEP4* genes reduces proteolytic degradation of secreted HSA/PTH in *Pichia pastoris* GS115. *J. Ind. Microbiol. Biotechnol.* 40: 589–599.

Xiao H & Lis JT (1988) Germline transformation used to define key features of heat-shock response elements. *Science* 239: 1139–1142.

Xu L, Shen Y, Hou J, Peng B, Tang H & Bao X (2014) Secretory pathway engineering enhances secretion of cellobiohydrolase I from *Trichoderma reesei* in *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* 117: 45–52.

Xu P, Raden D, Doyle FJ & Robinson AS (2005) Analysis of unfolded protein response during single-chain antibody expression in *Saccaromyces cerevisiae* reveals different roles for BiP and PDI in folding. *Metab. Eng.* 7: 269–279.

Xu X, Azakami H & Kato A (2004a) P-domain and lectin site are involved in the chaperone function of *Saccharomyces cerevisiae* calnexin homologue. *FEBS Lett.* 570: 155–160.

Xu X, Kanbara K, Azakami H & Kato A (2004b) Expression and characterization of *Saccharomyces cerevisiae* Cne1p, a calnexin homologue. *J. Biochem.* 135: 615–618.

Yang B (2000) Sec1 binds a closed conformation of syntaxin1A. *J. Cell Biol.* 148: 247–252.

Yang H-Y, Tatebayashi K, Yamamoto K & Saito H (2009) Glycosylation defects activate filamentous growth Kss1 MAPK and inhibit osmoregulatory Hog1 MAPK. *EMBO J.* 28: 1380–1391.

Yang LL & Haug A (1979) Purification and partial characterization of a procaryotic glycoprotein from the plasma membrane of *Thermoplasma acidophilum*. *BBA - Biomembr.* 556: 265–277.

Young BP, Craven RA, Reid PJ, Willer M & Stirling CJ (2001) Sec63p and Kar2p are required for the translocation of SRP-dependent precursors into the yeast endoplasmic reticulum in vivo. *EMBO J.* 20: 262–271.

Zapun A, Petrescu SM, Rudd PM, Dwek RA, Thomas DY & Bergeron JJM (1997) Conformation-independent binding of monoglucosylated ribonuclease B to calnexin. *Cell* 88: 29–38.

Zhang B, Chang A, Kjeldsen TB & Arvan P (2001) Intracellular retention of newly synthesized insulin in yeast is caused by endoproteolytic processing in the Golgi complex. *J. Cell Biol.* 153: 1187–1198.

Zhang X, Shaw A, Bates PA, Newman R H, Gowen B, Orlova E, Gorman MA, Kondo H, Dokurno P, Lally J, Leonard G, Meyer H, van Heel M, Freemont PS (2000) Structure of the AAA-ATPase. *Mol. Cell* 6: 1473–1484.

Chapter 3

Overexpression of native *PSE1* and *SOD1* in *Saccharomyces cerevisiae* improved heterologous cellulase secretion

Published in Applied Energy (2013) Vol 102: 150-156.

Overexpression of native *PSE1* and *SOD1* in *Saccharomyces cerevisiae* improved heterologous cellulase secretion

Heinrich Kroukamp, Riaan den Haan, Niël van Wyk, Willem H. van Zyl[1]

[1] *Department of Microbiology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa.*

3.1. Abstract:

Engineering cellulolytic ability into the yeast *Saccharomyces cerevisiae* to create an organism for consolidated bioprocessing (CBP) will require the simultaneous production and secretion of a number of heterologous cellulases. In addition, the generally low secretion titers achieved by this yeast will have to be overcome. To this end two native *S. cerevisiae* genes, *PSE1* and *SOD1*, were individually overexpressed by placing each gene under the transcriptional control of the constitutive *PGK1* promoter. The effect of these genes on heterologous protein secretion of three cellulases – an exoglucanase encoded by *cel6A* of *Neocallimastix patriciarum*, a β-glucosidase encoded by *cel3A* of *Saccharomycopsis fibuligera* and an endoglucanase encoded by *cel7B* of *Trichoderma reesei* was investigated by integrating the $PGK1_{P/T}$-*PSE1* and $PGK1_{P/T}$-*SOD1* cassettes into *S. cerevisiae* strains producing the relevant cellulases. Transformants were obtained that showed significantly higher secreted protein yield, with a resulting heterologous protein activity that ranged between 10% to 373% higher compared to the parental strains when grown in complex media. When both *PSE1* and *SOD1* were overexpressed in the yeast that produced Cel3A, a dramatic 447% increase in β-glucosidase activity was observed. This study shows that cellulase secretion in *S. cerevisiae* could be greatly improved with strain engineering. However, it also demonstrated that such strain engineering may have very enzyme specific effects as the induction of Cel3A secretion was far greater than that of the other cellulases investigated. Identifying cellulases amenable to expression in *S. cerevisiae* and engineering strains to maximise heterologous protein secretion may be imperative to creating optimal strains for CBP and may have wider implications for heterologous protein secretion in *S. cerevisiae* in general.

Keywords: *Saccharomyces cerevisiae*; Overexpression; *SOD1*; *PSE1*; Cellulase.

3.2. Introduction

Lignocellulosic biomass is potentially an abundant, renewable feedstock for production of biofuels and chemicals if the main technological barrier that impedes its widespread utilization - the lack of low-cost technologies to overcome its recalcitrance – can be overcome [1,2]. Organisms that hydrolyse the cellulose and hemicelluloses in biomass and produce a commodity product such as ethanol at a high rate and titre would significantly reduce the costs of biomass conversion. This would allow steps that are currently accomplished in different reactors, often by different organisms, to be combined in a consolidated bioprocess (CBP). While there is still no ideal organism to use in one-step biomass conversion, several candidates have been identified that are in various stages of development for establishment of a cellulolytic system and/or improvement of product-forming attributes [2,3]. One of the organisms furthest along in this development process is the yeast *Saccharomyces cerevisiae.* As this yeast is generally robust and well established in commercial ethanol production, engineering cellulolytic ability into it would yield an organism suited to industrial bioethanol production from cellulosic feedstocks.

Full enzymatic hydrolysis of crystalline cellulose requires synergistic action of three major types of enzymatic activity (i) endoglucanases, (ii) exoglucanases, including cellodextrinases and cellobiohydrolases, and (iii) β-glucosidases [4]. There have been several reports showing the expression of one or more cellulase encoding gene(s) in *S. cerevisiae* [5]. Strains of *S. cerevisiae* were created that could ferment cellobiose, the main product of the action of cellobiohydrolases on cellulosic substrates, at approximately the same rate as glucose in anaerobic conditions [6]. Recently, intracellular utilization of cellobiose with concomitant ethanol production, was reported by co-expression of the high affinity cellodextrin transport system of *Neurospora crassa* or the lactose permease encoding gene from *Kluyveromyces lactis* with an intracellular β-glucosidases or cellobiose phosphorilase [7,8]. Conversion of amorphous cellulose to ethanol was also demonstrated using strains co-expressing cellulases tethered to the cell wall [9,10], as free enzymes [11,12] or as part of a mini-cellulosome [13,14].

Recently the expression of relatively high levels of exoglucanases in *S. cerevisiae* was reported for the first time [15,16]. Ilmen et al. [16] reported a large increase in the maximum titer achieved for two critical exocellulases: Cel6A (CBH1) and Cel7A (CBH2). The cellulase expression levels achieved in this study meets the calculated levels for growth on cellulose at rates required for an industrial process [2]. Using these exoglucanases, a yeast

strain was constructed that was able to convert most of the glucan available in paper sludge to ethanol [15]. The strain was also able to displace 60% of the enzymes required to convert the sugars available in pretreated hardwood to ethanol in an SSF configuration. A similar strain expressing three alternative cellulases produced ethanol in one step from pretreated corn stover without the addition of exogenously produced enzymes [17]. These results demonstrate that cellulolytic *S. cerevisiae* strains can be used as a platform for developing an economical advanced biofuel process.

Data for heterologous cellulase production in yeasts in literature in general have been reported for aerobic cultures with high cell densities whereas CBP will involve anaerobic cultures with relatively low cell densities. However, there are some indications that production of CBHs under oxygen-limited conditions occur at roughly equivalent levels on a percentage to total cell protein (TCP) basis to aerobic production [2]. However, for complete cellulase degradation, higher titers of a large set of secreted heterologous enzymes will be required. In general, the secreted protein titers of *S. cerevisiae* is relatively low, especially compared to other yeast expression systems such as *Pichia pastoris, Yarrowia lipolytica*, *Hansenula polymorpha* and *Kluyveromyces lactis* [18]. The potential intrinsic value of secreted proteins has initiated several studies focused on optimizing the titers of proteins secreted by *S. cerevisiae* [19-21]. Although great improvement in secreted protein levels have been achieved in the past with strategies like codon optimization, promoter strengths, variation of secretion signals and strain selection [19], this did not hold true for all protein products and in some cases no significant increases could be obtained [22,23]. Ilmen et al. [16] reported that finding a gene compatible with expression in the host could greatly increase secreted protein levels. However, testing multiple gene candidates for their amenability to expression is not always an option. Another promising strategy to improve secretory production of target proteins is strain engineering, focusing on different aspects of protein synthesis, post-translational modification and protein secretion [21]. Several native proteins of *S. cerevisiae* have been identified that, when either disrupted or overexpressed, resulted in improved secretion [21,24,25,26].

In this study the effects of the overexpression of two native *S. cerevisiae* genes, *PSE1* and *SOD1* on the secretion of three heterologously produced cellulases were evaluated. Chow *et al*. [24] reported enhanced secretion of native proteins when the protein secretion enhancer 1 protein (Pse1) of *S. cerevisiae* was overproduced. The native gene cassette that encoded *PSE1* was placed on an episomal vector and resulted in higher levels of secreted killer toxin, α-factor and acid phosphatase in the resulting transformant. Pse1 is a member of the β-

karyopherin family and has been implicated in the export of mRNA out of the nucleus to the cytosol [27]. The superoxide dismutase encoding gene was recently overexpressed in *Kluyveromyces lactis*, which resulted in improved secreted production of heterologous glucoamylase [28]. It was postulated that the burden experienced by the yeast cell with the increased load of heterologous proteins resulted in the release of more ROS. The effect of an increased level of ROS could be relieved by the presence of more Sod1 protein in the cytosol. *S. cerevisiae* Sod1 is a Cu-Zn superoxide dismutase which is involved in the detoxification of reactive oxygen species (ROS) in the cytosol and to some extent the mitochondrial intermembrane space [29]. Sod1 has been reported to extend cell viability in the stationary phase [30] and is also implicated in copper ion buffering [31]. Zielinski *et al*. [32] proposed that in addition to the superoxide scavenging function; Sod1 has an 80S-ribosome regulatory function under stress conditions, linking it directly to the protein production/secretion pathway. Although the native *SOD1* has been overexpressed previously in *S. cerevisiae* [30], its effect on the phenotype of heterologous protein secretion was not investigated.

The native *S. cerevisiae* genes *PSE1* and *SOD1* were each placed under the transcriptional control of the constitutive *PGK1* promoter and terminator and the gene cassettes were integrated into the recombinant laboratory yeast strains Y294[*cel6A*], Y294[*cel3A*] and Y294[*cel7B*] – which heterologously produced Cel6A of *Neocallimastix patriciarum* (exoglucanase), Cel3A of *Saccharomycopsis fibuligera* (β-glucosidase) and Cel7B of *Trichoderma reesei* (endoglucanase), respectively [6,33,34]. The effect of the overexpression on heterologous protein secretion was evaluated.

3.3. Materials and Methods

3.3.1. Media and culturing conditions

S. cerevisiae strains were routinely cultured in YPD (10 g/L yeast extract, 20 g/L peptone and 20 g/L glucose) medium at 30°C on an orbital shaker at 200 rpm. For enzyme assays, the yeast strains were either cultivated in YPD or double-strength buffered synthetic complete (SC) media. (3.4 g/L yeast nitrogen base (Difco) with amino acids, 20 g/L succinate, 10 g/L ammonium sulphate and 20 g/L glucose pH adjusted to pH 6.0 with 10 N NaOH). For the generation of yeast transformants, cells were selected on YPD plates containing 20 g/L agar and either 200 µg/ml HyQ G418-sulfate (HyClone) or 100 µg/ml hygromycin B (Calbiochem).

3.3.2. Recombinant yeast strain constructions

Standard protocols were used for DNA manipulations [35]. Restriction endonucleases and T4 DNA ligase were purchased from Fermentas or Roche Molecular Biochemicals and used as directed by the manufacturer. Digested DNA was eluted from agarose gels with the Zymoclean Gel DNA Recovery Kit (Zymo Research). For polymerase chain reactions (PCR), Phusion DNA polymerase was purchased from Finnzymes and used as recommended by the manufacturer with a Perkin Elmer GeneAmps PCR System 2400 (The Perkin–Elmer Corporation, Norwalk, CT, USA). Details of the primers used in this study and the source sequences of the relevant genes are given in **Table 3.1**. For the construction of *PSE1* and *SOD1* overexpressing strains, the open reading frames of the *PSE1* and *SOD1* genes of *Saccharomyces cerevisiae* Y294 were amplified using the primer sets PSE1-L/R and SOD1-L/R, respectively. A 3270-bp PCR fragment for *PSE1* and a 489-bp PCR fragment for *SOD1* were digested with *Asc*I and *Pac*I and ligated into the yeast expression vector pBKD1 – to yield pBKD1-PSE1 and pBKD1-SOD1. The primer set hph-L/R was used to amplify the hygromycin resistance gene (*hphNT*) from the plasmid pAG26 (http://web.uni-frankfurt.de/fb15/mikro/euroscarf/data/pAG26.html). The *Mun*I/*Spe*I digested *hphNT* gene cassette was ligated into the *Eco*RI/*Spe*I digested pBKD1-SOD1 and replaced the geneticin resistance marker to generate pBHD-SOD1.

Table 3.1. Primers used in this study. Restriction sites are shown in boldface.

Primer name	**Oligonucleotide sequence (5'-3')**	**Restriction sites**	**Genbank**
PSE1-L	GACT**TTAATTAA**ATGTCTGCTTTACCGGA	*Asc*I	CAA8914
PSE1-R	GCTA**GGCGCGCC**TTATGCAAACCATTTAT	*Pac*I	
SOD1-L	GACT**TTAATTAA**ATGGTTCAAGCAGTCGC	*Asc*I	CAA8963
SOD1-R	GCTA**GGCGCGCC**TTAGTTGGTTAGACCAA	*Pac*I	
PGK1term	CATAGAAATATCGAATGGGAA		CAA4232
hph-L	GATC**CAATTG**CAACCCTTAATATAACTTC	*SpeI*	AAA9225
hph-R	CCTCATTCT**ACTAGT**GGATC	*MunI*	

All integration plasmids were linearized with *Bst*11071 after which transformation of Y294[*cel6A*], Y294[*cel3A*] and Y294[*cel7B*] were conducted according to a LiOAc/DMSO protocol [36]. Transformants were plated out on geneticin-containing plates after an expression step of three hours in liquid YPD medium containing 1 M sorbitol. The

transformants, Y294[*cel3A*]-*PSE1*, Y294[*cel6A*]-*PSE1* and Y294[*cel7B*]-*PSE1*were subsequently transformed with linearized pBHD-SOD1 under hygromycin B selection. The total genomic DNA of each yeast transformant was isolated [37] and successful integration of either the *SOD1* or *PSE1* overexpression cassette was confirmed with PCR analyses using PGK1termR and SOD1-L or the PSE1-L primers for the relevant transformants. Yeast transformants thus possessed the native copy of *PSE1* and *SOD1* plus one or more integrated copies of the gene under constitutive transcriptional regulation. All relevant plasmids and yeast strains used and generated in this study are shown in **Table 3.2**.

3.3.3. Enzyme activity assays

All yeast transformants obtained were initially screened in YPD media for improved heterologous protein production by determining enzyme activity of the yeast supernatant in liquid assays with lichenin, carboxymethylcellulose (CMC) and *p*-nitrophenyl-β-D-glucopyranoside (*p*NPG) as substrates for Y294[*cel6A*], Y294[*cel7B*] and Y294[*cel3A*], based strains respectively, using modified protocols as described previously [6,33,34]. All volumetric values were normalized with the dry cell weight (DCW) of the corresponding yeast cultures in mg/ml [38]. Enzyme activities were expressed as units/mg dry cell weight where a unit was defined as the amount of enzyme required to release one μmol of reducing sugar or equivalent per minute. For lichenin and CMC assays glucose was used to draw a standard curve in the range of 3 mM to 50 mM from which the amount of Units per millilitre (U/mL) of each sample was calculated. For *p*NPG assays a *p*NP standard curve in the range of 1.5 mM to 25 mM was used. Up to 67 transformants of each strain were screened in 5 mL YPD test tube cultures grown for 48 hours on a rotary wheel at 30°C. Transformants with the highest normalized activity compared to the parental strain were subsequently assayed in triplicate, with lichenin as substrate for the Cel6A and Cel7B producing strains and *p*NPG for the Cel3A producing strains. These transformant strains were cultured in 10 mL shake flasks with either YPD or double-strength buffered SC media for 72 hours at 30°C.

Table 3.2. Plasmids and strains constructed and used in the study.

Yeast Strains/Plasmid	Abbreviated name	Relevant genotype	Source/reference
Plasmids:			
pBKD1		*bla δ*-site $PGK1_P$-$PGK1_T$ *kanMX δ*-site	[45]
pBKD1-PSE1		*bla δ*-site $PGK1_P$-*PSE1*-$PGK1_T$ *kanMX δ*-site	This work
pBKD1-SOD1		*bla δ*-site $PGK1_P$-*SOD1*-$PGK1_T$ *kanMX δ*-site	This work
pBHD1-SOD1		*bla δ*-site $PGK1_P$-*SOD1*-$PGK1_T$ *hphNT δ*-site	This work
Parental yeast strains:			
S. cerevisiae Y294:		α *leu2*-3,112 *ura3*-52 *his3 trp1*-289	ATCC 201160
(*fur1*::*LEU2* YEp352)	Y294[REF]	*bla ura3/URA3* $PGK1_P$-*XYNSEC*-$PGK1_T$	[6]
(*fur1*::*LEU2* pNpCel6A)	Y294[*cel6A*]	*bla ura3/URA3* $PGK1_P$-*XYNSEC-N.p.cel6A*-$PGK1_T$	[33]
(*fur1*::*LEU2* pAZ40)	Y294[*cel7B*]	*bla ura3/URA3* $ENO1_P$-*T.r.cel7B*-$ENO1_T$	[34]
(*fur1*::*LEU2* ySFI)	Y294[*cel3A*]	*bla ura3/URA3* $PGK1_P$-*XYNSEC-S.f.cel3A*-$PGK1_T$	[6]
Constructed yeast strains:			
S. cerevisiae Y294 (fur1::LEU2 pNpCel6A):			
Y294_*SOD1* overexpressed	Y294[*cel6A*]-*SOD1*	*bla ura3/URA3* $PGK1_P$-*N.p.cel6A*-$PGK1_T$ *kanMX* $PGK1_P$-*SOD1*-$PGK1_T$	This work
Y294_*PSE1* overexpressed	Y294[*cel6A*]-*PSE1*	*bla ura3/URA3* $PGK1_P$-*N.p.cel6A*-$PGK1_T$ *kanMX* $PGK1_P$-*SOD1*-$PGK1_T$	This work
Y294_*PSE1-SOD1*-overexpressed	Y294[*cel6A*]-*PSE1/SOD1*	*bla ura3/URA3* $PGK1_P$-*N.p.cel6A*-$PGK1_T$ *kanMX* $PGK1_P$-*PSE1*-$PGK1_T$ *hph* $PGK1_P$-*SOD1*-$PGK1_T$	This work
S. cerevisiae Y294 (fur1::LEU2 pAZ40):			
Y294_*SOD1* overexpressed	Y294[*cel7B*]-*SOD1*	*bla ura3/URA3* $ENO1_P$-*T.r.cel7B*-$ENO1_T$ *kanMX* $PGK1_P$-*SOD1*-$PGK1_T$	This work
Y294_*PSE1* overexpressed	Y294[*cel7B*]-*PSE1*	*bla ura3/URA3* $ENO1_P$-*T.r.cel7B*-$ENO1_T$ *kanMX* $PGK1_P$-*PSE1*-$PGK1_T$	This work
S. cerevisiae Y294 (fur1::LEU2 ySFI):			
Y294_*PSE1-SOD1*-overexpressed	Y294[*cel7b*]-*PSE1/SOD1*	*bla ura3/URA3* $PGK1_P$-*T.r.cel7B*-$PGK1_T$ *kanMX* $PGK1_P$-*PSE1*-$PGK1_T$ *hph* $PGK1_P$-*SOD1*-$PGK1_T$	This work
Y294_*SOD1* overexpressed	Y294[*cel3A*]-*SOD1*	*bla ura3/URA3* $PGK1_P$-*XYNSEC-S.f.cel3A*-$PGK1_T$ *kanMX* $PGK1_P$-*SOD1*-$PGK1_T$	This work
Y294_*PSE1* overexpressed	Y294[*cel3A*]-*PSE1*	*bla ura3/URA3* $PGK1_P$-*XYNSEC-S.f.cel3A*-$PGK1_T$ *kanMX* $PGK1_P$-*PSE1-PGK1*	This work
Y294_*PSE1-SOD1* overexpressed	Y294[*cel3A*]-*PSE1/SOD1*	*bla ura3/URA3* $PGK1_P$-*XYNSEC-S.f.cel3A*-$PGK1_T$ *kanMX* $PGK1_P$-*PSE1*-$PGK1_T$ *hph* $PGK1_P$-*SOD1*-$PGK1_T$	This work

3.3.4. SDS-PAGE, Zymograms and *N*-deglycosylation

Electrophoresis in SDS-polyacrylamide gels was carried out as previously described by Laemmli [39]; gels were stained with silver staining [40]. Densitometric analysis, using the ImageJ software (http://rsbweb.nih.gov/ij/) was performed on silver stained gels according to the manufacturer's instructions. Amounts of proteins were estimated from gels using a standard curve set from the values obtained with pure BSA standards (5 ng to 100 ng) from the BCA protein assay kit (Pierce). To establish zymograms of cellulase activity for the strains expressing *cel6A* or *cel7B*, the separating portions of 8% SDS-polyacrylamide gels were polymerized with 0.2% (w/v) CMC (low viscosity - Sigma). Supernatant protein samples employed for zymograms were resuspended in loading buffer containing 125 mM Tris-HCl (pH 6.8) and (w/v) 15% sucrose, 2.5% SDS, and 0.02% bromophenol blue; the resuspended samples were incubated at 80°C for 3 min before loading onto the gels. Renaturation of proteins following electrophoresis was performed by incubating the gels in 0.1 M sodium acetate buffer (pH 5) containing 1% (v/v) deionized Triton X-100 (Calbiochem) for 2 h at room temperature, changing the buffer every 30 minutes. Subsequently, the gels were incubated in fresh 0.1 M sodium acetate/Triton X-100 buffer at 50°C for 2 hours (Cel7B containing gels) or 12 hours (Cel6A containing gels). To identify protein bands with cellulase activity, the gels were stained with 0.1% (w/v) Congo red (Saarchem) for 15 minutes and destained as required with 1 M NaCl after which the gels were photographed. To create a zymogram for the secreted β-glycosidase, the supernatant samples were suspended in loading buffer, incubated at 50°C for 3 minutes and separated on a 6% SDS-PAGE. For renaturation, the gels were incubated in 0.1 M sodium acetate buffer (pH 5) containing 1% (v/v) Triton X-100 for 16 hours at 4°C. An overlay gel was created that contained 2% (w/v) agarose (SeaKem), 0.1 M sodium acetate buffer (pH 5), 0.3% (w/v) esculin (Sigma) and 0.15% (w/v) ferric citrate (Merck). After renaturation the protein containing gel was blotted against the esculin/ferric citrate containing gel and incubated at 50°C for 1 hour after which the gels were photographed.

N-deglycosylation reactions were performed by using the PNGase F kit (New England Biolabs). However, to ensure that active enzymes were obtained the instructions of the manufacturer were adapted as follows: 1 µL 5% (w/v) SDS solution was added to 9 µL of supernatant sample and incubated at 80°C (Cel7B or Cel6A containing samples) or 50°C (Cel3A-containing samples) for 10 minutes. Two µL each of G7 buffer and 10% NP-40 as well as 5 µL of deionized water were added followed by 1µL of PNGase F for the

deglycosylated samples (indicated by a + on the gels and zymograms) or 1 μL deionized water for the non-deglycosylated samples (indicated by a - on the gels and zymograms). The reaction mixes were incubated at 37°C for 16 hours after which 7 μL of loading buffer was added to the samples and 20 uL of this mixture was loaded onto the SDS-PAGE gels described above.

3.4. Results

3.4.1. Screening of *PSE1* and *SOD1* overexpressing strains

The *PSE1* or *SOD1* overexpressing transformants with the highest enzyme activity per gram dry cell weight were selected for the *S. cerevisiae* strains expressing one of the three reporter cellulases. During this screening, a wide range of enzyme activities was observed between transformants with the same constitutively expressed gene, with some values lower than that of the parental strain's enzyme activity (data not shown). The best *PSE1* or/and *SOD1* overexpressing strains were subsequently grown in either YPD or buffered SC media for three days and were assayed in triplicate (**Fig. 3.1**). After 72 hours of growth in YPD media, the Y294[*cel3A*]-*PSE1/SOD1* had the highest β-glucosidase activity compared to the parental strain per gram dry weight of 86.8 U/mg DCW, with the strain overexpressing only the *PSE1* having a β-glucosidase activity of 72.4 U/mg DCW. The Y294[*cel3A*]-*SOD1* strain showed no significant increase in β-glucosidase activity compared to the parental Y294[cel3A] strain's 12.2 and 17.6 U/mg DCW in buffered SC and YPD media, respectively. In the Cel7B producing strains grown in YPD *SOD1* and/or *PSE1* overexpression led to increases of 10% to 15% over the background strain in endoglucanase activity after 72 hours of cultivation. These changes do not appear to be statistically significant. The Cel6A producing strains constitutively expressing *SOD1* and/or *PSE1*, had similar Cel6A activity values compared to the parental strain after 72 hours when grown in SC media, although the co-expression of *SOD1* and/or *PSE1* appeared to diminish Cel6A activity. In YPD media cultivations a 20% to 25% increase in Cel6A activity was observed for strains expressing *PSE1* and the *SOD1*/*PSE1* combination.

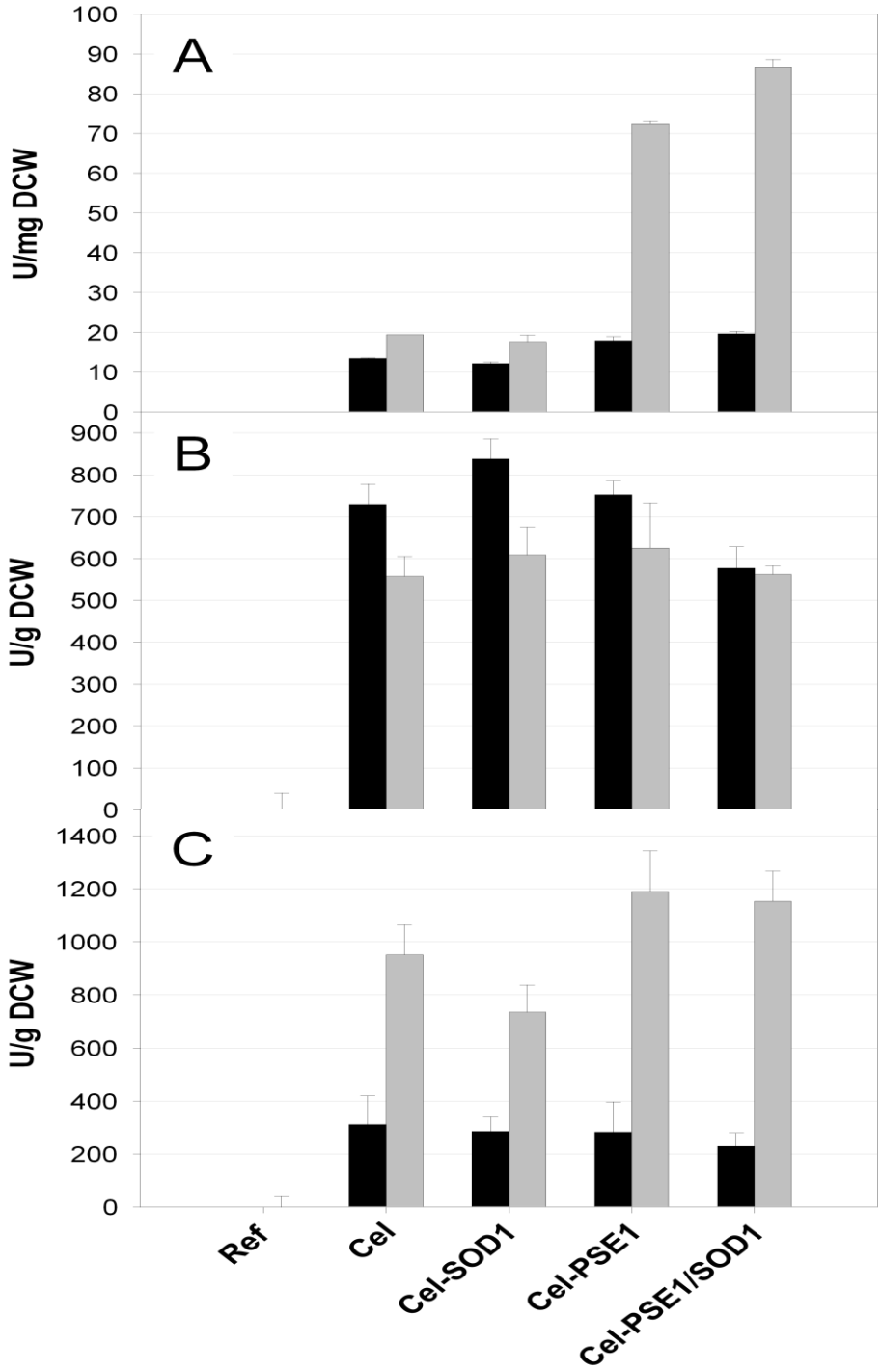


Fig. 3.1. Enzyme activity profiles of *PSE1* and *SOD1* overexpression Y294 strains in comparison to parental Y294 strains. [A] shows β-glucosidase activity of Cel3A producing strains on *p*NPG. [B] and [C] shows the activity of Cel7B and Cel6A producing strains on lichenin. The black and grey bars represent strains grown in SC and YPD media respectively. Values obtained were normalized with the dry cell weight of the yeast. Values were obtained from independently conducted enzyme assays in triplicate. A reference strain expressing no heterologous genes was also cultured in both conditions and assayed on all substrates but no activity could be measured. Error bars indicate standard deviation from the mean value.

3.4.2. Zymogram and densitometry analysis

Densitometric analysis of the silver stained protein gels in **Fig. 3.2** revealed that significant changes in secreted protein levels underlay changes in the secreted protein activities shown in **Fig. 3.2**. The heterologous cellulase proteins were identified on silver stained gels after zymogram analysis of identical gels. To estimate protein levels, single, non-diffuse bands on the silver stained gels were required and therefore supernatant samples were *N*-deglycosylated. Deglycosylated samples from SC medium cultures (72 hours) of the *cel3A* expressing strains were separated on a 6% SDS-PAGE along with purified BSA in known quantities from 5 ng to 100 ng (not shown). Densitometric analysis, using the ImageJ software was performed on this gel and the amount of secreted Cel3A was determined from a standard curve set from the values obtained with the BSA standards. Subsequently the protein levels per μL supernatant were standardized against the amount of DCW produced by the relevant culture. Secreted Cel3A production levels in μg/mg DCW were 1.46 (Y294[*cel3A*]), 1.79 (Y294[*cel3A*]-*PSE1*), 1.46 (Y294[*cel3A*]-*SOD1*) and 3.27 (Y294[*cel3A*]-*PSE1*/*SOD1*), respectively. Therefore the relatively similar β-glucosidase activity levels displayed at 72 hours by the Y294[*cel3A*] and Y294[*cel3A*]-*SOD1* strains (**Fig. 3.2**) are reflected in the similar protein levels secreted by these strains. The ~30% increase in secreted β-glucosidase activity in the Y294[*cel3A*]-*PSE1* strain is mirrored by the ~25% increase of secreted Cel3A protein by this strain. The amount of Cel3A secreted by the Y294[*cel3A*]-*PSE1/SOD1* strain was more than double that of the parental strain leading to a >50% increase in secreted activity. The amounts of Cel7B could not accurately be determined using this method as the deglycosylation reactions did not yield clear, single bands. The disperse bands observed are likely a result of incomplete *N*-deglycosylation or the fact that variable *O*-linked glycosylation may be present. Densitometric analysis of Cel6A producing strains showed that, similarly to the Cel3A producing strains, changes in secreted activity levels could be explained by changes in the amount of secreted protein. Taken together, these data suggest that changes in the levels of secreted cellulases and not in the specific activity of the cellulases resulted in the altered activity levels observed.

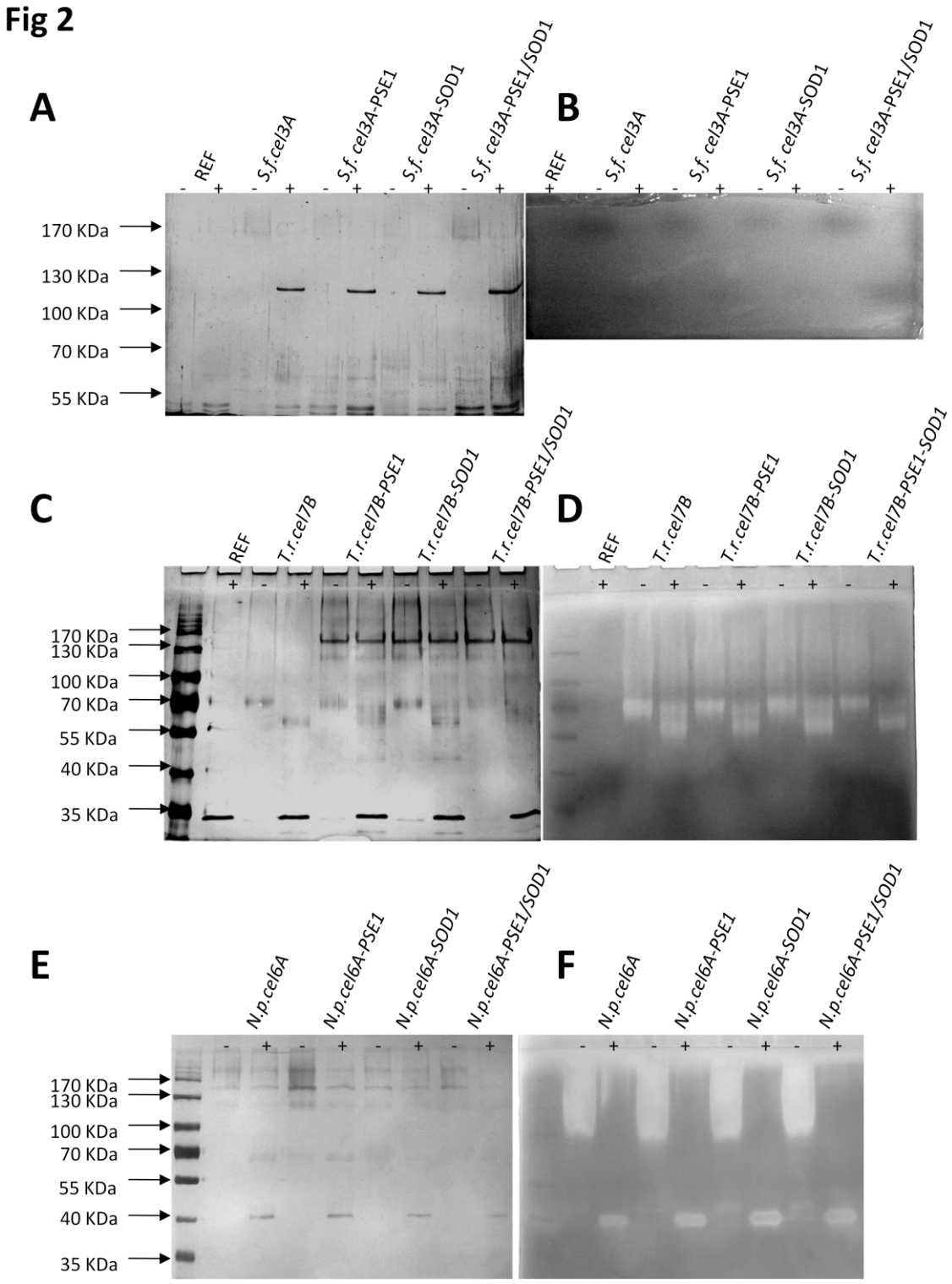


Fig. 3.2. SDS-PAGE and zymogram analysis of the proteins secreted by *S. cerevisiae* Y294 transformants. [A] shows a silver stained 6% SDS-PAGE gel with supernatants of Cel3A producing strains prepared as described in Materials and Methods. [C] and [E] show silver stained 8% SDS-PAGE gels with supernatants of Cel7B and Cel6A producing strains. Zymograms of identical gels containing esculin/ferric citrate [B] or CMC [D, F] are shown. (+) denotes deglycosylated samples and (-) denotes untreated samples.

3.5. Discussion

The initial screening of yeast transformants containing the genes targeted for integration to the delta sequences on the yeast chromosomes showed a wide range of reporter enzyme activity of which some were higher, lower or without change compared to the parental strain (results not shown). This clear example of phenotypic variance between transformants could be ascribed to the putative genotypic variation in the strains, such as copy number and position of integration into different delta sequences present on the host genome. Apart from the noticeable changes in secreted enzyme activity values, no significant differences were observed between the growth rates of the native gene overexpressing strains when compared to their respective parental strains. This implied that in these strains a metabolic burden caused by the overexpression of *SOD1* and/or *PSE1* was relatively small. However, a lower cell density at stationary phase was observed for the Y294[*cel6A*]-*PSE1* and Y294[*cel6A*]-*SOD1* when grown in YPD media.

As shown in **Fig. 3.2**, most of the recombinant strains showed improved secreted enzymatic activity when *PSE1* was overexpressed although *SOD1* overexpression did not seem to have a pronounced effect in most cases. The observed changes in Cel3A and Cel6A enzyme activity correlated with the increase in functional Cel3A and Cel6A protein determined by densitometry studies and it is assumed that this is also true for the observed increases in Cel7B activity when these native genes were overexpressed. When *PSE1* was overexpressed in each of the three cellulase producing strains, the best transformants achieved an increase of 3.7, 1.15 and 1.25 fold in secreted enzyme activity compared to the parental strains of Y294[*cel3A*], Y294[*cel7B*] and Y294[*cel6A*] respectively when grown in YPD, with smaller increases when cultivated in SC media. Chow *et al*. [24] reported an increase of up to fourfold for secreted native *S. cerevisiae* k1-killer toxin, α-factor, and acid phosphatase in a strain overexpressing *PSE1*. It was proposed that the overexpression of *PSE1* led to a general increase in secreted proteins. Although the effect of *PSE1* overexpression differed for each reporter protein, in general our data supports this theory of a general effect on secreted protein quantities with *PSE1* overexpression. One proposed role of Pse1 is that it is involved in the export of mRNA from the nucleus to the cytoplasm [27]. This suggests that the increase in secreted protein quantities is due to changes at the transcription/translation steps in the protein secretion pathway. However, this is in contrast with findings that there was no significant increase in total intracellular proteins when *PSE1* was overexpressed [24]. An alternative theory, is that higher levels of secretion are due to a

more indirect effect of *PSE1* overexpression by increasing the quantitative levels of other proteins involved with secretion, resulting in the observed general enhancement in secretion.

When *SOD1* was overexpressed an increase in secreted enzyme activity was observed in the recombinant strain producing Cel7B grown in minimal media. The overexpression of *SOD1* in the Y294[*cel3A*] or Y294[*cel6A*] strains resulted in no observable increase in secreted activity. However, simultaneous overexpression of *SOD1* with *PSE1* in the Y294[*cel3A*] strain resulted in a 4.5 fold increase in secreted β-glucosidase activity when cultivated in YPD, a remarkable 20% improvement on the Y294[*cel3A*]-*PSE1* strain under the same cultivation conditions. Although no significant increase in secreted Cel3A was observed when only *SOD1* was constitutively expressed in the Y294[*cel3A*] strain, the combined effect could be due to more sufficient oxidative damage reduction with the increased ROS generated by the additional metabolic strain caused by increased Cel3A production in the Y294[*cel3A*]-*PSE1* strain. To our knowledge, this is the first reported case where *SOD1* overexpression in *S. cerevisiae* resulted in higher heterologous protein secretion.

This study substantiates previous reports that a differential effect of the same secretion enhancing native protein on different reporter proteins could be observed [41-43]. This is probably due to the intrinsically different properties of each heterologous protein. One fundamentally different property of different heterologous proteins may be the level of protein glycosylation. Future work in this respect could elucidate the possible effect of glycosylation on activity and secretion levels. Ilmen et al. [16] recently showed that the unfolded protein response (UPR) of *S. cerevisiae* was induced to varying degrees by the expression of heterologous cellobiohydrolases. Even proteins that were closely related in sequence and structure induced dissimilar levels of UPR leading to greatly varying levels of secreted cellulases. This infers that different heterologous proteins may exert diverse amounts of stress on the host cell at different points during gene transcription, translation and maturation. Alleviation of these stresses may therefore require the overproduction of different native or non-native host proteins such as chaperones, foldases or ROS degrading enzymes that may be unique to a specific heterologous protein.

For the cost-efficient production of ethanol from lignocellulosic biomass, it is of great interest to construct *S. cerevisiae* strains capable of hydrolyzing cellulose without a large input of additional cellulases [5]. Although many studies have shown recombinant cellulase production by *S. cerevisiae*, growth of this yeast on a crystalline form of cellulose still requires at least a large increase in secreted cellulase – particularly exoglucanase – activity

[15]. Here, we have shown that the recombinant production of the enzymatic activities required for complete cellulose hydrolysis could be increased with the aid of strain engineering. However, it was also demonstrated that such strain engineering had a very enzyme specific effect as the induction of Cel3A secretion was far greater than that of the other cellulases investigated.

This study shows the tremendous potential for strain engineering of *S. cerevisiae* to attain increased heterologous protein secretion. By overexpressing two different genes individually or simultaneously, it was shown that significant improvements in heterologous protein secretion could be achieved. There are several other gene candidates that have been shown to influence protein secretion [25,26,43]. Optimal combinations of these and other candidates may lead to greatly enhanced secretion of heterologous proteins of interest and could lead to a general enhancement of protein secretion in the yeast *S. cerevisiae.*

3.6. Acknowledgements

The authors would like to thank John McBride, Mascoma Corporation, for providing pBKD1. Furthermore, we would like to thank the National Research Foundation (South Africa) for financial assistance.

3.7. References

[1] Lynd LR, Weimer PJ, Van Zyl WH, Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 2002;66:506-577.

[2] Olson DG, McBride JE, Shaw AJ, Lynd LR. Recent progress in consolidated bioprocessing. Curr.Opin.Biotechnol 2012;23:396-405.

[3] La Grange DC, Den Haan R, Van Zyl WH. Engineering cellulolytic ability into bioprocessing organims. Appl Environ Microbiol 2010;87:1195-1208.

[4] Zhang YH, Lynd LR. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. Biotechnol Bioeng 2004;88:797-824.

[5] van Zyl WH, Lynd LR, den Haan R, McBride JE. Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae*. Biofuels, Vol. 108, pp. 205-235. Springer Berlin / Heidelberg. 2007.

[6] van Rooyen R, Hahn-Hägerdal B, La Grange DC, van Zyl WH. Construction of cellobiose-growing and fermenting *Saccharomyces cerevisiae* strains. J Biotechnol 2005;120: 284-295.

[7] Galazka JM, Tian C, Beeson WT, Martinez B, Glass NL, Cate JHD. Cellodextrin transport in yeast for improved biofuel production. Science 2010;330:84-86.

[8] Sadie CJ, Rose SH, Den Haan R, Van Zyl WH. Co-expression of a cellobiose phosphorylase and lactose permease enables intracellular cellobiose utilisation by *Saccharomyces cerevisiae*. App Microbiol Biotechnol 2011;90:1373-1380.

[9] Fujita Y, Ito J, Ueda M, Fukuda H, Kondo A. Synergistic saccharification, and direct fermentation to ethanol, of amorphous cellulose by use of an engineered yeast strain codisplaying three types of cellulolytic enzyme. Appl Environ Microbiol 2004;70:1207-1212.

[10] Yamada R, Taniguchi N, Tanaka T *et al.* Cocktail δ-integration: A novel method to construct cellulolytic enzyme expression ratio-optimized yeast strains. Microb Cell Fact 2010;9:32.

[11] Den Haan R, McBride JE, Grange DCL, Lynd LR, Van Zyl WH. Functional expression of cellobiohydrolases in *Saccharomyces cerevisiae* towards one-step conversion of cellulose to ethanol. Enzyme Microb Technol 2007;40:1291-1299.

[12] Jeon E, Hyeon JE, Suh DJ *et al.*. Production of cellulosic ethanol in *Saccharomyces cerevisiae* heterologous expressing *Clostridium thermocellum* endoglucanase and *Saccharomycopsis fibuligera* β-glucosidase genes. Mol Cells 2009;28:369-373.

[13] Tsai SL, Oh J, Singh S, Chen R, Chen W. Functional Assembly of Minicellulosomes on the Saccharomyces cerevisiae Cell Surface for Cellulose Hydrolysis and Ethanol Production. Appl Environ Microbiol 2009;75:6087-6093.

[14] Wen F, Sun J, Zhao H. Yeast surface display of trifunctional minicellulosomes for simultaneous saccharification and fermentation of cellulose to ethanol. Appl Environ Microbiol 2010;76:1251-1260.

[15] McBride, J. E., Brevnova, E., Ghandi, C., Mellon, M., Froehlich, A., Delaault, K., Rajgarhia, V., Flatt, J., Van Zyl, W. H., Den Haan, R., La Grange, D. C., Rose, S. H., Penttilä, M., Ilmen, M., Siika-aho, M., Uusitalo, J., Hau, H. H., Rice, C., Villari, J., Stonehouse, E. A., Gilbert, A., Keating, J. D., Xu, H., Willes, D., Shikhare, I., Thorngren, N., Warner, A. K., and Murphy, D. Yeast expressing cellulases for simultaneous saccharification and fermentation using cellulose. PCT/US2009/065571. 5-27-2010.

[16] Ilmen M, Den Haan R, Brevnova E, McBride J, Wiswall E, Froehlich A, Koivula A, Voutilainen SP, Siika-Aho M, la Grange DC, Thorngren N, Ahlgren S, Mellon M, Deleault K, Rajgarhia V, van Zyl WH, Penttilä M. High level secretion of cellobiohydrolases by *Saccharomyces cerevisiae*. Biotechnol Biofuels 2011;4:30.

[17] Khramtsov N, McDade L, Amerik A, Divatia K, Tikhonov A, Minto M, Kabongo-Mubalamate G, Markovic Z, Ruiz-Martinez M, Henck S. Industrial yeast strain engineered to ferment ethanol from lignocellulosic biomass. Bioresour Technol 2011;102:8310-8313.

[18] Müller S, Sandal T, Kamp-Hansen P, Dalbøge H. Comparison of expression systems in the yeasts *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Klyveromyces lactis*,

Schizosaccharomyces pombe and *Yarrowia lipolytica*. Cloning of two novel promoters from *Yarrowia lipolytica*. Yeast 1998;14: 1267-1283.

[19] Fleer R. Engineering yeast for high level expression. Curr Opin Biotech 1992;3: 486-496.

[20] Gasser B, Mattanovich D. Antibody production with yeasts and filamentous fungi: on the road to large scale? Biotechnol Lett 2007;29: 201-212.

[21] Idiris A, Tohda H, Kumagai H, Takegawa K. Engineering of protein secretion in yeast: strategies and impact on protein production. Appl Microbiol Biotechnol 2010;86: 403-417.

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

[23] Porro D, Sauer M, Branduardi P, Mattanovich D. Recombinant protein production in yeasts. Mol Biotechnol 2005;31:245-259.

[24] Chow TY, Ash JJ, Dignard D, Thomas DY. Screening and identification of a gene, *PSE*-1, that affects protein secretion in *Saccharomyces cerevisiae*. J Cell Sci 1992;101:709-719.

[25] Robinson AS, Hines V, Wittrup KD. Protein disulfide isomerase overexpression increases secretion of foreign proteins in *Saccharomyces cerevisiae*. Nat Biotechnol 1994;12:381-384.

[26] Ruohonen L, Toikkanen J, Tieaho V, Outola M, Soderlund H, Keranen S. Enhancement of protein secretion in *Saccharomyces cerevisiae* by overproduction of Sso protein, a late-acting component of the secretory machinery. Yeast 1997;13:337-351.

[27] Seedorf M, Silver PA. Importin/karyopherin protein family members required for mRNA export from the nucleus. Proc Natl Acad Sci U.S.A 1997;94:8590-8595.

[28] Raimondi S, Zanni E, Talora C, Rossi M, Palleschi C, Uccelletti D. *SOD1*, a new *Kluyveromyces lactis* helper gene for heterologous protein secretion. Appl Environ Microbiol 2008;74:7130-7137.

[29] Sturtz LA, Diekert K, Jensen LT, Lill R, Culotta VC. A fraction of yeast Cu,Zn-superoxide dismutase and its metallochaperone, *CCS*, localize to the intermembrane space of mitochondria. J Biol Chem 2001;276:38084-38089.

[30] Harris N, Bachler M, Costa V, Mollapour M, Moradas-Ferreira P, Piper PW. Overexpressed Sod1p acts either to reduce or to increase the lifespans and stress resistance of yeast, depending on whether it is Cu^{2+}-deficient or an active Cu,Zn-superoxide dismutase. Aging Cell 2005;4:41-52.

[31] Culotta VC, Joh H-D, Lin S-J, Hudak Slekar K, Strain J. A physiological role for *Saccharomyces cerevisiae* copper/zinc superoxide dismutase in copper buffering. J Biol Chem 1995;270:29991-29997.

[32] Zielinski R, Pilecki M, Kubinski K, Zien P, Hellman U, Szyszka R. Inhibition of yeast ribosomal stalk phosphorylation by Cu-Zn superoxide dismutase. Biochem Biophys Res Commun 2002;296:1310-1316.

[33] van Wyk N, den Haan R, van Zyl WH. Heterologous production of NpCel6A from *Neocallimastix patriciarum* in *Saccharomyces cerevisiae*. Enzyme Microb Technol 2010;46:378-383.

[34] Den Haan R, Rose SH, Lynd LR, van Zyl WH. Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. Metab Eng 2007;9: 87-94.

[35] Sambrook J, Russel DB. *Molecular cloning: a laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 2001.

[36] Hill J, Donald KAIG, Griffiths DE. DMSO-enhanced whole cell yeast transformation. Nucleic Acids Res 1991;19:5791.

[37] Hoffman CS, Winston F. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformaion of *Escherichia coli*. Gene 1987;57: 267-272.

[38] Meinander N, Zacchi G, Hahn-Hagerdal B. A heterologous reductase affects the redox balance of recombinant *Saccharomyces cerevisiae*. Microbiology 1996;142:165-172.

[39] Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970;227:680-685.

[40] O'Connell KL, Stults JT. Identification of mouse liver proteins on two-dimensional electrophoresis gels by matrix-assisted laser desorption/ionization mass spectrometry of *in situ* enzymatic digests. Electrophoresis 1997;18:349-359.

[41] Robinson AS, Bockhaus JA, Voegler AC, Wittrup KD. Reduction of BiP levels decreases heterologous protein secretion in *Saccharomyces cerevisiae*. J Biol Chem 1996;271:10017-10022.

[42] van der Heide M, Hollenberg C, van der Klei I, Veenhuis M. Overproduction of BiP negatively affects the secretion of *Aspergillus niger* glucose oxidase by the yeast *Hansenula polymorpha*. Appl Microbiol Biotechnol 2002;58:487-494.

[43] Valkonen M, Penttila M, Saloheimo M. Effects of inactivation and constitutive expression of the unfolded- protein response pathway on protein production in the yeast *Saccharomyces cerevisiae*. Appl Environ Microbiol 2003;69:2065-2072.

[44] Rossini D, Porro D, Brambilla L, Venturini M, Ranzi BM, Vanoni M, Alberghina L. In *Saccharomyces cerevisiae*, protein secretion into the growth medium depends on environmental factors. Yeast 1993;9:77-84.

[45] McBride JEE, Deleault K, M., Lynd LR, Pronk JT. Recombinant yeast strains expressing tethered cellulase enzymes. *Patent PCT/US2007/085390*.

Chapter 4

Protein *N*-glycosylation chain length alters the protein secretion capacity and cell wall retention of *Saccharomyces cerevisiae*

Protein *N*-glycosylation chain length alters the protein secretion capacity and cell wall retention of *Saccharomyces cerevisiae*

Heinrich Kroukamp[1], Riaan den Haan[2] and Willem H. van Zyl[1]

[1] *Department of Microbiology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa.*

[2] *Department of Biotechnology, University of the Western Cape, Private Bag X17, Bellville, 7535, South Africa*

4.1. Abstract

In recent years, strain engineering has become a widespread strategy to improve yeast protein production and secretion. Like many other genomic targets that were shown to influence protein secretion in yeast, the manner by which *N*-glycosylation mutants engender higher protein yields is poorly understood. We investigated the potential of *Saccharomyces cerevisiae* Golgi-apparatus *N*-glycosylation mutants to improve cellulase secretion and aimed to elucidate some of the underlying mechanisms responsible for this high secretion phenotype. A 30% increase in secretion of an unglycosylated protein, Cel12A, was obtained with a *MNN2* deletion strain, implying the potential involvement of cell wall alterations in improved protein secretion. In contrast, deletion of the genes encoding the mannan polymerase II complex subunits Mnn10p and Mnn11p resulted in small improvements of heterologous Cel3A production. However, removing mannosyltransferases essential for outer chain elongation resulted in up to 121% higher supernatant Cel3A activities. These results shed some light on the intricate and unique relationships of diverse heterologous proteins with various components of the expression host secretion machinery that inevitably lead to protein specific secretion improvements during strain engineering strategies.

4.2. Introduction

Yeasts have been used for the production of recombinant proteins, such as insulin, from the early 1980's (Johnson 1983). *Pichia pastoris*, *Yarrowia lipolytica*, *Kluyveromyces lactis* and *Hansenula polymorpha* have been shown to be prolific protein producers under laboratory and industrial conditions, however no other yeast has been more intensively studied than *S. cerevisiae* (Domínguez et al. 1998). In addition to many early studies elucidating crucial metabolic pathways, nutrient sensing, cellular stress responses and the fungal secretion pathway; recent genomic, proteomic and metabolomic studies rapidly advanced our understanding of *S. cerevisiae*'s global metabolism and regulation. Additionally, the high

demand for industrial and pharmaceutical proteins means *S. cerevisiae* is a frequent choice for rational design strategies to improve current recombinant protein production and secretion yields (Porro et al. 2011).

Engineering protein glycosylation pathways is a recurrent strategy in order to “humanize” recombinant proteins for pharmaceutical use, enhance specific activity (in the case of enzymes) and to alter protein properties such as stability and solubility (De Pourcq et al. 2010; Beckham et al. 2012; Imperiali & O’Connor 1999). Glycosylation engineering has also been implicated in protein secretion and even the addition of a single *N*-glycosylation site had drastic effects on secreted protein titers (Sagt et al. 2000; Hoshida et al. 2013). By introducing an *N*-glycosylation sequon (Asn-X_{aa}-Ser/Thr) at the N-terminal region of a recombinant cutinase, Sagt et al. (2000) achieved a fivefold increase in the secreted titer thereof.

In *S. cerevisiae,* protein *N*-glycosylation is initiated in the ER by the *en bloc* transfer of a preassembled $Man_8GlcNAc_2$ glycan to selected Asn amino acids of a nascent polypeptide. Shortly after the folded glycoproteins enter the Golgi, Och1p adds an α-1,6-mannosyl residue to the glycan chain. From this point, the *N*-linked glycans can mature into "core-type" structures, consisting of 9 - 13 mannose residues, or may be further processed by the addition of an outer chain, sometimes referred to as "hyperglycosylation-type" structures, of up to 200 mannose residues (Herscovics & Orlean 1993). The stepwise progression of the glycan maturation is depicted in **Fig 4.1**.

Smith et al. (1985) documented the first account of enhanced secretion by a yeast glycosylation mutant. They obtained a 15-fold increase in secreted bovine growth hormone that was attributed to a mutation in the *SSC1* (supersecreting) gene (later revealed to be the Golgi Ca^{2+}/Mn^{2+} ATPase gene, *PMR1*). The deletion of the *PMR1* gene has been shown to reduce Ca^{2+} ion concentrations in the Golgi, an essential co-factor for many of the resident mannosyltransferases. *PMR1* deletion improve the secretion of a wide variety of native and recombinant proteins; however, as with many 'secretion enhancing gene alterations' its effectiveness varies greatly depending on the choice of secreted reporter protein (Harmsen et al. 1996; Rudolph et al. 1989; Agaphonov et al. 2007; Xu et al. 2014).

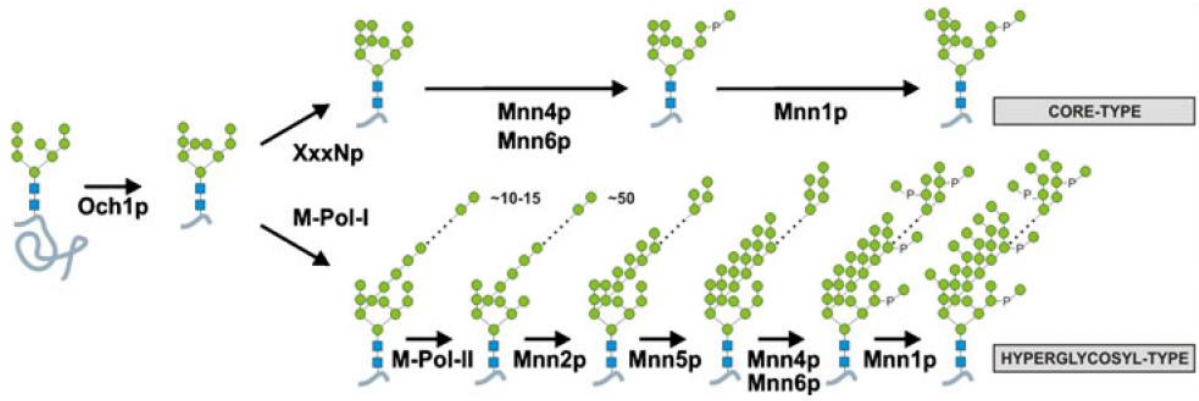


Fig. 4.1. Protein *N*-linked glycan elongation in the yeast Golgi-apparatus. Elongation of the $Man_8GlcNAc_2$ glycan outer chain is initiated by Och1p, adding a single mannose through a α-1,6-glycosidic bond. Mannan polymerase I (consisting of Van1p and Mnn9p) is then able to extend the mannose outer chain with up to 15 mannosyl residues. Mannan polymerase II (Hoc1p, Mnn8p, Mnn9p, Mnn10p and Mnn11p) further extend this outer chain up to 50 mannosyl residues. Mnn2p and Mnn6p are responsible for branching of the outer chain, adding single mannosyl residues through α-1,2-glycosidic and phosphate bonds, respectively. Mnn1p terminates the elongation by adding α-1,3-linked mannosyl residues to chain ends. Adapted from De Pourcq et al. (2010).

Deletion of various other ER and Golgi resident mannosyltransferase genes have been shown to benefit the secretion of heterologous proteins in *S. cerevisiae* and *K. lactis* (Hoshida et al. 2013; Uccelletti et al. 2006; Bartkeviciūte & Sasnauskas 2004; Wang et al. 2013). The knockout mutation of the *K. lactis OCH1* was able to improve the supernatant activity of an *Arxula adeninivorans* glucoamylase 4-fold and the deletion of *MNN10* in *S. cerevisiae* enhanced total secretion per culture optical density of a *Phanerochaete chrysoporium* cellobiohyrdolase 6-fold. This emphasized the possible benefits of glycosylation alterations for biofuel production, but failed to elucidate the mechanisms responsible for the observed phenotype (Uccelletti et al. 2006; Wang et al. 2013).

In this study the effect of *N*-glycosylation deficiencies on heterologous cellulase secretion and cell wall retention of enzymes were evaluated. Since protein *N*-glycosylation could affect both the cell wall integrity and the reporter protein properties, we chose the *Saccharomycopsis fibuligera* Cel3A, containing 14 potential *N*-glycosylation sequons and the *Neosartorya fischeri* Cel12A, containing no *N*-glycosylation sequons (and thus only influenced by cell wall changes) as reporter proteins. Like many other exoglucanases and *β*-glucosidases of the glycosylhydrolase family 3 (GH3), the *S. fibuligera* Cel3A also possesses a putative fibronectin-like carbohydrate binding domain involved in cell wall retention of these enzymes (Marín-Navarro et al. 2011). The Cel3A and Cel12A reporter enzymes were

individually expressed in yeast mutants, with gene knockouts perturbing different steps in Golgi *N*-glycan chain elongation. Previous studies suggest different mechanisms by which *N*-glycosylation may influence secretion; either via increased cell wall permeability for macromolecules or by causing an overall higher secretion rate due to aberrant polarization. We compared the relative degree of cell wall disruption and reporter protein *N*-linked glycan alterations in order to link these characteristics with the enhanced secretion phenotypes. We also showed the cell wall integrity dependent nature of the Cel3A cell wall associated characteristic.

4.3. Materials and methods

4.3.1. Media and culturing conditions

Saccharomyces cerevisiae strains were routinely cultured in YPD (10 g/L yeast extract, 20 g/L peptone and 20 g/L glucose) medium at 30°C, unless otherwise specified. Putative *S. cerevisiae* transformants were selected on YPD agar plates, supplemented with 300 µg/mL hygromycin B (InvivoGen, San Diego, USA). For enzyme assays, the yeast strains were cultivated in 24-well plates containing 1 mL YPD medium per well, supplemented with 200 µg/mL hygromycin B for selection. The 24-well plates were sealed with AeraSeal™ breathable film (Sigma-Aldrich, Germany) to prevent excessive evaporation. For partial purification and SDS-PAGE analysis, strains were cultivated in double-strength buffered synthetic complete (2xSC) media (3.4 g/L yeast nitrogen base with amino acids (Difco), 10 g/L ammonium sulfate, 20 g/L succinate and 20 g/L glucose; pH adjusted to pH 6.0 with 10N NaOH). *Escherichia coli* was used for plasmid propagation and was routinely cultured in Luria-Bertani broth (Sambrook & Russel 2001) at 37°C, supplemented with 100 µg/mL ampicillin for plasmid selection.

4.3.2. Construction of *N*-glycosylation deficient yeast strains.

Standard protocols were used for DNA manipulations (Sambrook & Russel 2001). Restriction endonucleases, T4 DNA ligase, Phusion DNA polymerase were purchased from Thermo Scientific and used as directed by the manufacturer. Zymoclean Gel DNA Recovery kit (Zymo Research) was used to elute digested DNA from agarose gels. For the construction of the episomal expression plasmids, the hygromycin B resistance (*hph*) cassette was obtained from the pBHD1_SOD1 plasmid (Kroukamp et al. 2013) by digesting with *Bam*HI

and *Spe*I and ligated into the pMU1531 (designated pHK); replacing the zeocin resistance cassette. *Neosartorya fischeri* Cel12A and *Saccharomycopsis fibuligera* Cel3A genes were obtained by digesting pRDH166 and pRDH135 with *Pac*I and *Asc*I, respectively. The isolated reporter genes were subsequently ligated into *Pac*I and *Asc*I digested pHK (**Fig. 4.2**), yielding pHK_CEL3A and pHK_CEL12A. These plasmids were introduced separately into the BY4742 deletion strain series (Brachmann et al. 1998), as indicated in **Table 4.1**, according to a LiOAc/DMSO protocol (Hill et al. 1991). Transformants were plated out on plates containing 300 µg/mL hygromycin B after a recovery step of 4 h in liquid YPD medium. The presence of the expression plasmid in putative transformants was confirmed with polymerase chain reaction (PCR) analysis using the following primers; 5'-GGATCCACTAGTCTTCTAGGCGGGTTATC-3' and 5'-GACTGGCGCGCCTTA CAAACATTGAGAGTAGTATGGG-3'.

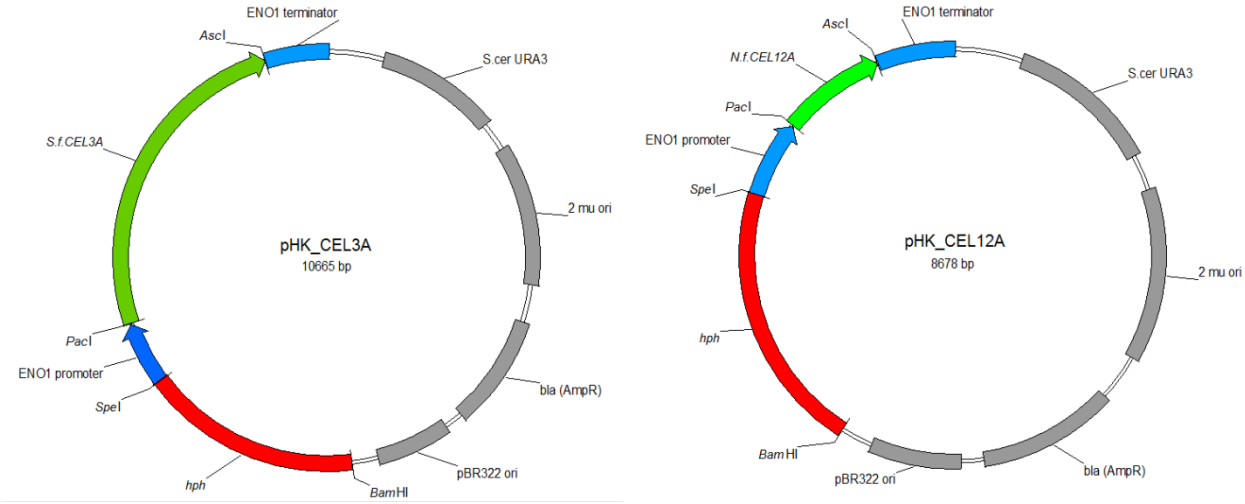


Fig. 4.2. A graphical representation of the episomal expression plasmids used in this study. The CEL12A open reading frame was ligated into *Pac*I and *Asc*I digested pHK, placing it under the transcriptional regulation of the *S. cerevisiae ENO1* promoter, yielding pHK_CEL12A. pHK_CEL3A was constructed in a similar fashion.

Table 4.1. List of strains and vectors used in this study

Constructed yeast strains/plasmids	Abbreviated name	Relevant genotype	Reference/Source
Plasmids			
pBHD1_SOD1		*bla δ PGK1p-S.c.SOD1-PGK1t hph δ*	(Kroukamp et al. 2013)
pHK		*bla URA3 hph ENO1p – XYNSEC - ENO1t*	This work
pHK_CEL3A		*bla URA3 hph ENO1p – XYNSEC-S.f.CEL3A - ENO1t*	This work

pHK_CEL12A		*bla URA3 hph ENO1p – XYNSEC-N.f.CEL12A - ENO1t*	This work
S. cerevisiae strains			
BY4742		*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	(Brachmann et al. 1998)
BY4742 + pHK	BYRef	*bla ura3/URA3 hph ENO1p – XYNSEC - ENO1t*	This work
BY4742 + pHK_CEL3A	BY4742B	*bla ura3/URA3 hph ENO1p – XYNSEC - S.f.CEL3A - ENO1t*	This work
BY4742 *och1Δ*::kanMX4 + pHK_CEL3A	och1B	*bla och1Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-S.f.CEL3A - ENO1t*	This work
BY4742 *mnn1Δ*::kanMX4 + pHK_CEL3A	mnn1B	*bla mnn1Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-S.f.CEL3A - ENO1t*	This work
BY4742 *mnn2Δ*::kanMX4 + pHK_CEL3A	mnn2B	*bla mnn2Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-S.f.CEL3A - ENO1t*	This work
BY4742 mnn6*Δ*::kanMX4 + pHK_CEL3A	mnn6B	*bla mnn6Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-S.f.CEL3A - ENO1t*	This work
BY4742 *mnn9Δ*::kanMX4 + pHK_CEL3A	mnn9B	*bla mnn9Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-S.f.CEL3A - ENO1t*	This work
BY4742 *mnn10Δ*::kanMX4 + pHK_CEL3A	mnn10B	*bla mnn10Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-S.f.CEL3A - ENO1t*	This work
BY4742 *mnn11*::kanMX4 + pHK_CEL3A	mnn11B	*bla mnn11Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-S.f.CEL3A - ENO1t*	This work
BY4742 *pmr1Δ*::kanMX4 + pHK_CEL3A	pmr1B	*bla pmr1Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-S.f.CEL3A - ENO1t*	This work
BY4742 *hac1Δ*::kanMX4 + pHK_CEL3A	hac1B	*bla hac1Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-S.f.CEL3A - ENO1t*	This work
BY4742 + pHK_CEL12A	BY4742E	*bla ura3/URA3 hph ENO1p – XYNSEC-N.f.CEL12A - ENO1t*	This work
BY4742 *och1Δ*::kanMX4 + pHK_CEL12A	och1E	*bla och1Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-N.f.CEL12A - ENO1t*	This work
BY4742 *mnn1Δ*::kanMX4 + pHK_CEL12A	mnn1E	*bla mnn1Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-N.f.CEL12A - ENO1t*	This work
BY4742 *mnn2Δ*::kanMX4 + pHK_CEL12A	mnn2E	*bla mnn2Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-N.f.CEL12A - ENO1t*	This work
BY4742 *mnn6Δ*::kanMX4 + pHK_CEL12A	mnn6E	*bla mnn6Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-N.f.CEL12A - ENO1t*	This work
BY4742 *mnn9Δ*::kanMX4 + pHK_CEL12A	mnn9E	*bla mnn9Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-N.f.CEL12A - ENO1t*	This work
BY4742 *mnn10Δ*::kanMX4 + pHK_CEL12A	mnn10E	*bla mnn10Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-N.f.CEL12A - ENO1t*	This work
BY4742 *mnn11Δ*::kanMX4 + pHK_CEL12A	mnn11E	*bla mnn11Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-N.f.CEL12A - ENO1t*	This work
BY4742 *pmr1Δ*::kanMX4 + pHK_CEL12A	pmr1E	*bla pmr1Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-N.f.CEL12A - ENO1t*	This work
BY4742 *hac1Δ*::kanMX4 + pHK_CEL12A	hac1E	*bla hac1Δ:: kanMX ura3/URA3 hph ENO1p – XYNSEC-N.f.CEL12A - ENO1t*	This work

4.3.3. Enzymatic activity determination

Yeast strains were assayed in quadruplicate for total secreted enzyme activity (cells and supernatant) and supernatant activity (cells removed by centrifugation) after two days of cultivation in 24-well plates. *p*-Nitrophenyl-*β*-D-glucopyranoside (*p*NPG; Sigma-Aldrich, Germany) and carboxymethyl-cellulose (CMC) were used as substrates for the Cel3A and Cel12A producing strains, respectively, using modified protocols as described previously (van Rooyen et al. 2005; Den Haan et al. 2007). Volumetric values were normalized with dry cell weight (DCW) where indicated. DCW was determined by weighing desiccated pellets of 1 mL yeast culture. The background enzyme activity value of the control strains, containing the pHK vector, was subtracted from the enzyme activity values. Enzyme activities were expressed as units/mg dry cell weight where a unit was defined as the amount of enzyme required to release 1 µmol of reducing sugar or equivalent per minute. For *p*NPG assays a *p*-nitrophenol (*p*NP) standard curve in the range of 6.13 - 100 µM was used. For CMC assays, units per milliliter (U/mL) were calculated using a glucose standard curve in the range of 0.78 - 12.5 mM.

4.3.4. SDS-PAGE analysis and silver-staining

Concentrated supernatant proteins from 2x SC media grown cultures were separated with SDS-polyacrylamide gel electrophoresis as previously described by Laemmli (1970) and visualized with silver-staining (Kroukamp et al. 2013). Yeast supernatants were concentrated up to 60 times with Vivaspin6 columns (10 000Da MWCO; Sartorius, Germany) using a Heraeus Biofuge stratos centrifuge. *N*-deglycosylation of the supernatant proteins was performed according to the manufacturer's instructions using the PNGase F kit (New England Biolabs). Twenty microliter supernatant or PNGase F-treated supernatant was boiled with seven microliters of loading buffer for three minutes, before loading onto the SDS-PAGE gels (Kroukamp et al. 2013).

4.3.5. Sensitivity to cell wall integrity dependent growth inhibitors

A two-fold serial dilution of yeast cultures (without any episomal plasmids), grown in YPD until saturated, were spotted onto YPD agar plates containing selected inhibitors. Resistance to Tunicamycin (0.1 ug/mL; 0.5 ug/mL; 1.0 ug/mL), hygromycin B (20 µg/mL; 50 µg/mL; 80 ug/mL), sodium orthovanadate (6 mM; 8 mM; 10 mM; 12 mM), dithiothreitol (1 mM; 5 mM; 10 mM; 15 mM) and Congo Red (25 µg/mL; 50 µg/mL; 75 µg/mL) was evaluated.

4.4. Results

4.4.1. *N*-glycosylation deficient yeast strains

The BY4742 deletion strains were transformed with either pHK_CEL3A or pHK_CEL12A, and confirmed with PCR analysis. Transformants were subsequently named for the deletion it harbored; those expressing the *CEL3A* (also known as *BGL1*) were indicated with a 'B' and strains expressing the *CEL12A* (also known as *EGIII*) were indicated with an 'E'.

4.4.2. Extracellular glycoprotein activity

N-glycosylation mutants (*och1*, *mnn1*, *mnn2*, *mnn6*, *mnn9*, *mnn10*, *mnn11* and *pmr1*) derived from the isogenic *S. cerevisiae* BY4742, expressing either a recombinant β-glucosidase (Cel3A) or endoglucanase (Cel12A), were assayed for the respective activities. Similar DCW values were observed for the BY4742 and the glycosylation mutants, with the exception of the *pmr1* deletion strain, which produced ~52% less biomass compared to the isogenic strain. There were significant differences between the volumetric supernatant β-glucosidase activities of the mnn10B, mnn11B and pmr1B mutant strains and the isogenic strain, exhibiting 201%, 221% and 54% the activity of the BY4742B strain, respectively (**Addendum B, Fig. B1**). Similar, but more subdued trends were observed for the secreted volumetric β-glucosidase activity, with lowered values observed for the och1B and mnn9B strains compared to BY4742B. The total secreted and supernatant cell specific activities of the BY4742B strain, without any *N*-glycosylation gene deletions, were 5.3 U/g DCW and 1.7 U/g DCW, respectively (**Fig. 4.3A**). No significant difference was observed between the cell specific activities of the BY4742B and the unfolded protein response (UPR) defective hac1B strain. Enhanced total secreted β-glucosidase activity per DCW was obtained for the mnn10B, mnn11B and pmr1B, however of these only mnn10B and mnn11B showed increased supernatant activity of up to 228% of the isogenic strain.

Approximately a third of the total β-glucosidase activity was found in the cell free supernatant of strain BY4742B (**Fig. 4.3B**). Similar activity values as the isogenic strain were obtained for the mnn1B, mnn2B, mnn6B, pmr1B and hac1B strains. These strains also had optical density values which directly correlated to their respective dried cell weight (**Addendum B, Fig. B2**). This was not the case for och1B, mnn9B, mnn10B and mnn11B, with a 24 - 36% reduction on optical density compared to the isogenic strain with a similar dry cell weighs.

activity of 285 U/g DCW. The pmr1E strain had 13% less volumetric activity than the isogenic strain and was produced by approximately 53% of the cell matter based on cell dry weights (data not shown). The relationship between the cell dry weight and the optical density observed for all the evaluated Cel12A producing strains was similar to their corresponding Cel3A producing strains.

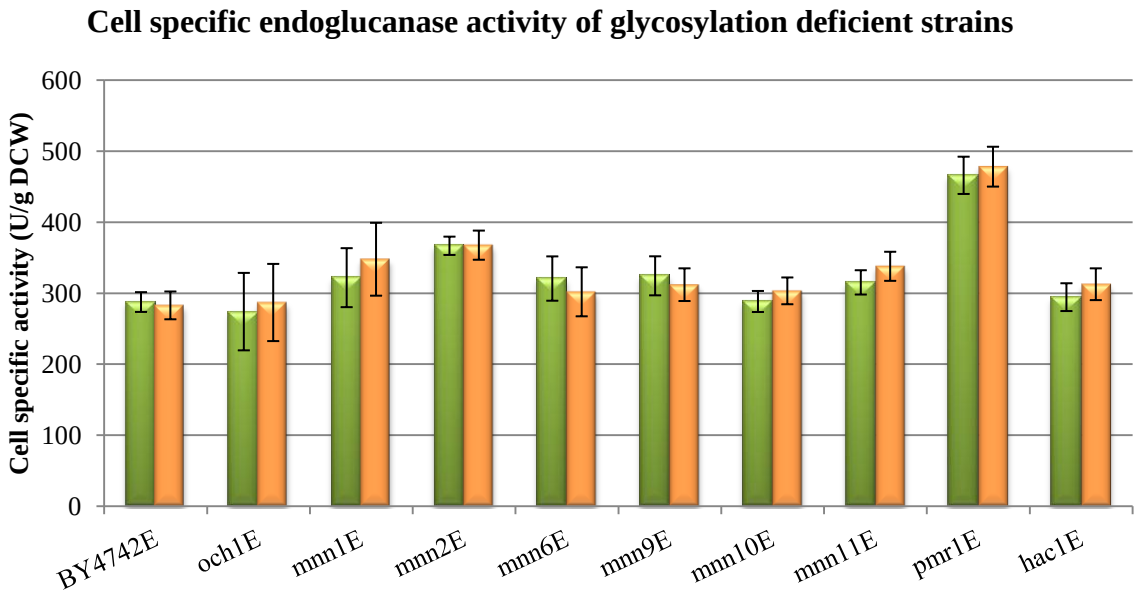


Fig. 4.4. The cell specific supernatant (green) and total secreted (orange) endoglucanase activity of the evaluated deletion strains, displayed as enzymatic units per gram dry cell weight. Values shown are the mean values of four biological repeats and the error bars indicate standard deviation.

4.4.3. Reporter protein migration size

The silver-stained SDS-PAGE gels containing PNGase F treated supernatant samples of the Cel3A producing strains produced distinct bands of the expected size, 120 kDa of deglycosylated Cel3A (**Fig. 4.5**), as previously described by Kroukamp et al. (2013). A very diffuse band above 170 kDa was detected for the glycosylated Cel3A produced by the BY4742B strain. Similar diffuse bands were detected for the mnn1B, mnn2B, mnn6B strains, although a slight shift could be detected for the mnn2B Cel3A. The lanes containing mnn10B and mnn11B supernatants had diffuse bands between 150 - 180 kDa. Distinct bands of 130 kDa were observed for the Cel3A in the lanes containing the och1B and mnn9B supernatants. The och1B and mnn9B produced Cel3A was expected to contain a glycan structure similar to the $Man_{9\text{-}13}GlcNAc_2$ core structure. SDS-PAGE analysis of the Mnn10 and Mnn11 deficient strains revealed a diffuse Cel3A band, between 150 – 170 kDa. As expected, Cel12A

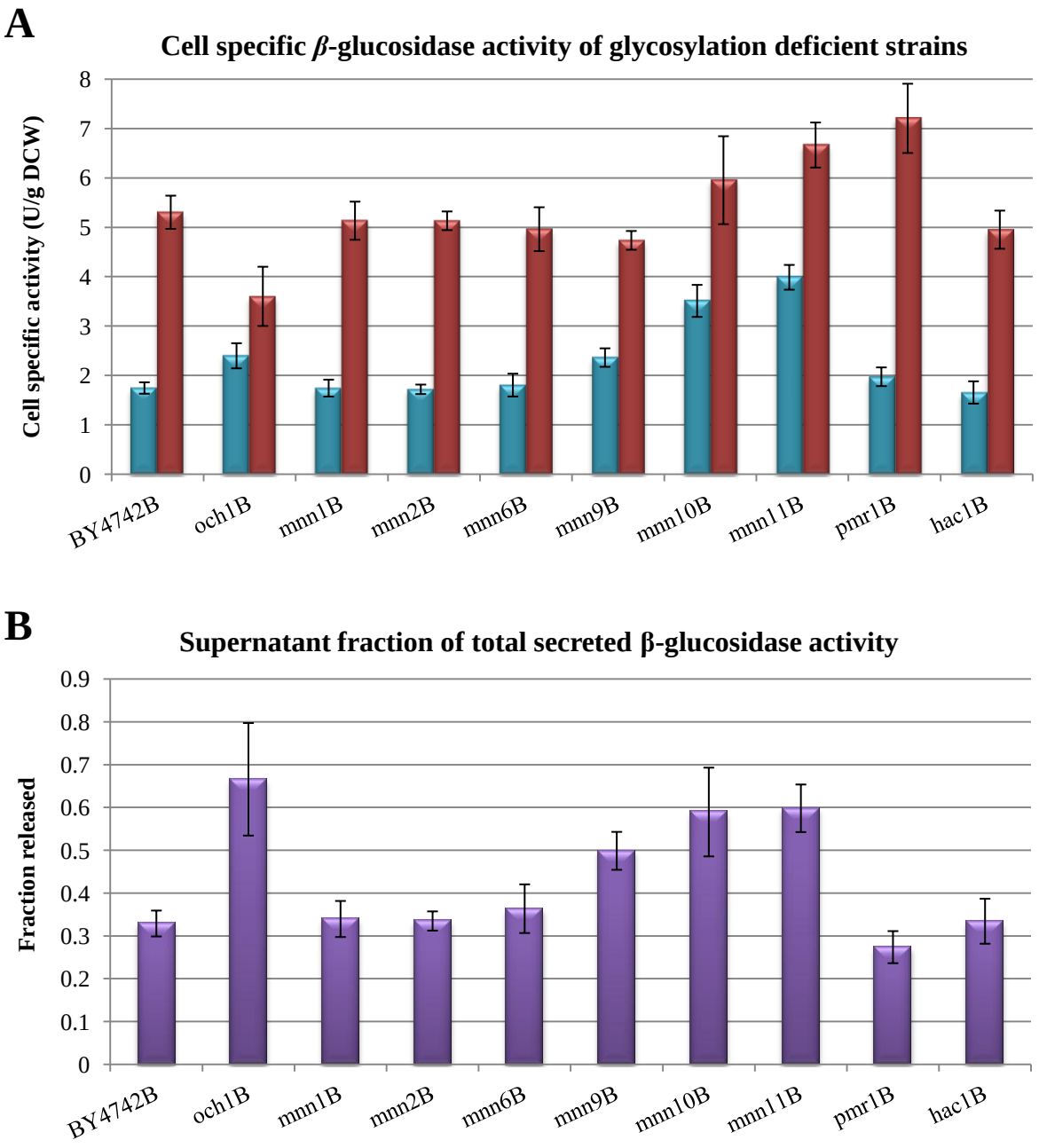


Fig. 4.3. (**A**) The cell specific supernatant (blue) and total secreted (red) β-glucosidase activity of the evaluated deletion strains, displayed as enzymatic units per gram dry cell weight. (**B**) The fraction of the total β-glucosidase activity released from the cell. Significantly higher percentages of β-glucosidase activity were found in the supernatant of the outer chain extending deficient strains och1B, mnn9B, mnn10B and mnn11B. Values shown are the mean values of four biological repeats and the error bars indicate standard deviation.

No significant difference was observed between the supernatant and total secreted cell specific activities of the Cel12A producing BY4742E strains (**Fig. 4.4**). The similar supernatant and total secreted activities indicate that most, if not all, of the secreted Cel12A gets released into the supernatant. The mnn2E and pmr1E strains had secreted endoglucananse activities of 30% and 69% higher than the isogenic strain which had and

producing strains revealed no apparent change in the mobility of the Cel12A produced by the glycosylation mutants and that of the isogenic BY4742E strain, producing a ~25kDa band in all lanes except the reference strain (**Addendum B, Fig. B3**).

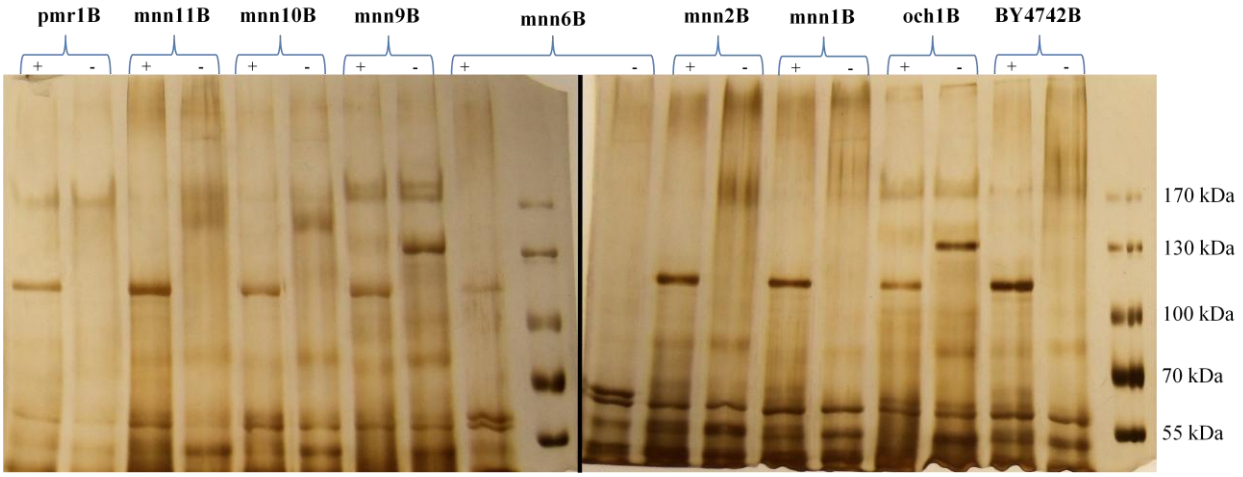


Fig. 4.5. SDS-PAGE analysis of the glycosylation-deficient strain supernatants producing Cel3A. (+) Denotes deglycosylated samples and (-) denotes untreated samples.

4.4.4. Cell morphology and cell wall integrity analysis of *N*-glycosylation mutant strains

Microscopic analysis of the *och1*, *mnn9, mnn10 and mnn11* deletions strains revealed a high frequency of enlarged cells compared to the BY4742 strains (**Addendum B, Fig. B4**). Large clumps of cells, which could not be separated by 0.5 mM EDTA or 2% mannose, were prominent in these four strains (data not shown). Deletion of the *OCH1*, *MNN2*, *MNN9, MNN10* and *MNN11* genes resulted in sensitivity to 75 µg/mL of hygromycin B, although *mnn2Δ* and *mnn11Δ* still showed some growth at the lower dilution ranges (**Fig. 4.6**). Except for *pmr1Δ*, similar results were obtained with the cell wall binding Congo Red. Strains deficient in outer chain elongation were vanadate resistant. This was also partially the case for the *PMR1* deletion strain. Complete inhibition of growth was observed for the *hac1Δ* strain in the presence of 6 mM tunicamycin. The *OCH1*, *MNN9* and *PMR1* deletion strains had lowered viability under tunicamycin induced stress.

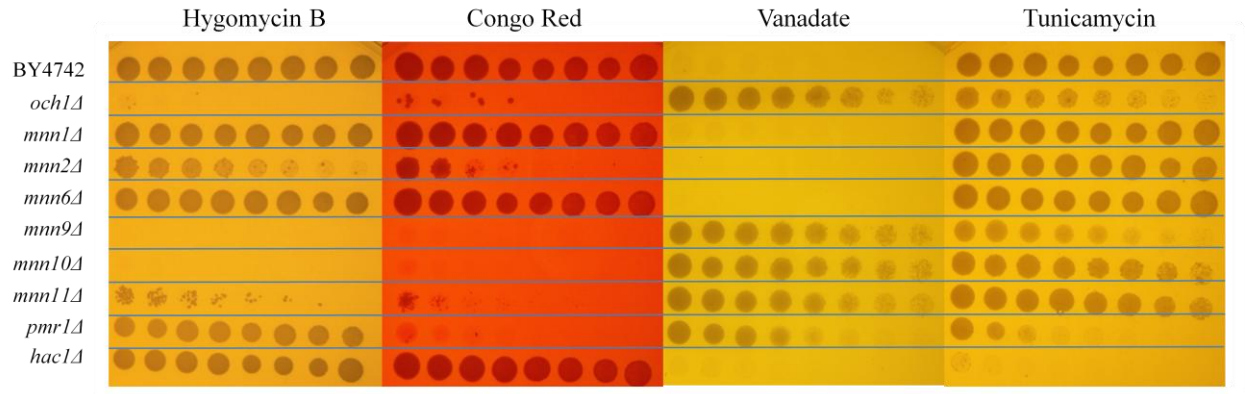


Fig. 4.6. Analysis of 2-fold serial dilutions of deletion strains in a plate viability assay after 24 hours. Resistance to 80 µg/mL Hygromycin B, 75 µg/mL Congo Red, 0.5 µg/mL Sodium vanadate and 6 mM tunicamycin was evaluated. Growth for all deletion strains was identical to the BY4742 strain, on YPD agar plates without inhibitor supplementation (**Addendum B, Fig. B5**).

4.5. Discussion

Strain engineering has become a valuable tool to enhance the production of pharmaceutical and industrially relevant proteins; inevitably, it also serves as a means to elucidate the mechanisms by which certain genomic alterations influence protein secretion (Idiris et al. 2010; den Haan et al. 2013). In order to evaluate the secretion enhancing effect of Golgi *N*-glycosylation mutants, we used enzymes relevant to second generation bio-fuel production as reporter proteins. We were able to obtain a 22% increase in total secreted cell specific β-glucosidase activity with the deletion of *MNN11* (**Fig. 4.3A**). Up to 221% increase in supernatant volumetric β-glucosidase activity was obtained for the mnn10B and mnn11B strains (**Addendum B, Fig. B1**). These increased supernatant titers are in agreement with that of previous studies on *K. lactis MNN10* deletion strains (Bartkeviciūte & Sasnauskas 2004). This increased enzymatic activity of the Cel3A, represents a direct increase in secreted protein, since no significant change was found between the specific activities of the native enzyme and its deglycosylated form (Marín-Navarro et al. 2011). Although volumetric activity increase is less ideal for biofuel production purposes, since ethanol production is cell biomass depenent, it holds relevance in the pharmaceutical industry by lowering protein purification cost. The proteins produced by these strains also lack extensive mannosylation, a potential allergenic or immunogenic agent, making their use as therapeutics more attractive.

The pmr1B and pmr1E strains had the highest total secretion cell specific activities of Cel3A and Cel12A, respectively; however these strains also had the lowest biomass production, making *PMR1* deletion a less attractive option as genetic modification in industrial

production hosts (**Fig. 4.3**; **Fig. 4.4**; **Addendum B, Fig. B2**). Due to the similarities in dry cell weights for all glycan elongating mutants (excluding the *PMR1* mutant) and the isogenic BY4742 strain, the mnn10B and mnn11B strains also had higher cell specific activities than the BY4742B strain. Most of the Cel12A expressing strains, with the exception of pmr1E and mnn2E, had similar supernatant and secreted cell specific activities compared to the BY4742E strain. Since no *N*-glycosylation sites were present on the Cel12A protein, the 30% increase in secreted activity of the mnn2E strain is possibly due to a perturbed cell wall structure. From the Congo Red and vanadate plate growth assays, it is clear that the *OCH1*, *MNN9*, *MNN10* and *MNN11* deletion strains had impaired cell wall integrity (**Fig. 4.6**). In contrast to previous studies on *MNN10* and *MNN11,* which speculated that the cell wall changes might be the primary cause for the enhanced secretion phenotype (Bartkeviciūte & Sasnauskas 2004; Uccelletti et al. 2006), our results would suggest that the supernatant activity increases of the mnn10B and mnn11B strains were a direct result of altered reporter protein glycosylation, since no change in Cel12A levels were obtained in similar deletion strains. However, more reporter proteins should be evaluated to confirm this, due to the protein specific nature of protein secretion enhancing genes.

Mutant strains with deletions crucial for outer-chain elongation, displayed abnormal cell morphologies (**Addendum B, Fig. B4**). As previously reported for glycosylation mutants, enlarged, non-flocculant clumps of cells were detected in the *och1Δ*, *mnn9Δ*, *mnn10Δ*, *mnn11Δ* strains, as a result of incomplete separation of mother and daughter cells. This clumpy phenotype is strikingly similar to that of mutants with perturbed polarization and actin cytoskeleton assembly (Karpova et al. 2000). Although protein glycosylation and the secretion machinery interact directly through mannosyltransferases (Alg7p and Mnn10p) and cell polarization components, this might not be the sole contributor to this phenotype, since cell wall integrity perturbing chemicals such as Calcofluor White are also able to induce this phenotype (Kukuruzinska & Lennon 1998; Mondésert & Reed 1996; Roncero & Duran 1985). The clumpy phenotype might not notably alter protein secretion, but its influence on the optical density of the yeast cultures was evident (**Addendum B, Fig. B2**). These clear discrepancies between the relative culture OD and DCW were not due to bound enzymes, as might be envisioned for the Cel3A producing strains, since similar DCW/OD ratios were obtained for the Cel12A producing strain which had almost no cell wall bound activity. These low OD values might have been perceived as poor growth, as many earlier glycosylation

studies would suggest and might result in inflated cell specific activity values (Nagasu et al. 1992; Wang et al. 2013).

Interestingly, the deletion of the two mannosyltransferases, *MNN10* and *MNN11*, responsible for the most extensive additions to the glycan outer chain resulted in enhanced secretion of the highly glycosylated Cel3A; and the deletion of the genes crucial for the outer chain elongations, namely *OCH1*, *MNN9*, *MNN10* and *MNN11*, impaired the yeast cell wall and resulted in a significant increase in the Cel3A present in the supernatant. Up to 2-fold increase in cell wall released Cel3A was obtained for these four strains; a useful feature for downstream purification. In a study by Marín-Navarro et al. (2011), the authors postulated that the putative carbohydrate recognition domain of the Cel3A might be influenced by cell wall composition. This was supported by a follow-up study showing different cell wall retention levels of the Cel3A between different *S. cerevisiae* strains (Gurgu et al. 2011). Our demonstration of deceased cell wall association in strains with an impaired cell wall structure supports their hypothesis.

The unfolded protein response is induced in yeast cells as a cytoprotective response to the accumulation of misfolded proteins and polypeptide aggregates (Shen et al. 2004). Protein *N*-glycosylation directly assists protein folding in the ER and mutations in ER *N*-glycosylation genes have been shown to induce the unfolded protein response (UPR), leading to increased protein degradation, impaired cell growth and even cell death (Khan & Schröder 2008; Bull & Thiede 2012; Trombetta 2003; Herscovics & Orlean 1993). Induction of the UPR has also been shown to have beneficial or detrimental effects on secreted protein yields (Xu & Robinson 2009; Valkonen et al. 2003). No indication of UPR influence was evident from the secreted enzyme activities or the biomass yields due to overexpressing either of the two cellulase reporter proteins, when comparing the *HAC1* deletion strain and the BY4742 strain (**Fig. 4.3A**; **Fig. 4.4**). No reduction in fitness was observed for the BY4742 strain in the presence of 6 mM tunicamycin, but it completely inhibited the *hac1Δ* strain which is unable to activate the UPR. The majority of the Golgi *N*-glycosylation mutants were unaffected by tunicamycin induced ER stress, emphasizing the implausability of UPR influenced secretion alterations in these strains.

4.6. Conclusion

The growing demand for recombinant proteins drives the rapid development of more efficient expression vectors, optimized culturing conditions to allow high density fermentations and expression hosts with high product yields. Protein *N*-glycosylation plays a major role in both product maturation and cell physiology, making it a promising target for expression host engineering strategies. We were able to obtain enhanced secretion of reporter proteins by altering the degree of protein *N*-glycosylation. To our knowledge, this is the first report showing that *MNN2* deletion could enhance secretion in yeast. We also demonstrated increased supernatant levels of cell wall associated enzymes form the cell wall impaired strains, with up to 121% more enzyme being available for purification from the supernatant. This study emphasize the impact of the cell wall on enzyme secretion and cell wall retention, and how the stepwise alterations in *N*-linked glycan elongation influenced product yields, allowing for more directed strain and protein engineering in future studies.

4.7. Acknowledgements

The authors would like to thank the National Research Fund (South Africa) for financial support.

4.8. References

Agaphonov M, Plotnikova T, Fokina V, Romanova N, Packeiser N, Kang H & Ter-Avanesyan MD (2007) Inactivation of the *Hansenula polymorpha PMR1* gene affects cell viability and functioning of the secretory pathway. *FEMS Yeast Res.* 7: 1145–1152.

Bartkeviciūte D & Sasnauskas K (2004) Disruption of the *MNN10* gene enhances protein secretion in *Kluyveromyces lactis* and *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 4: 833–840.

Beckham GT, Dai Z, Matthews JF, Momany M, Payne CM, Adney WS, Baker SE & Himmel ME (2012) Harnessing glycosylation to improve cellulase activity. *Curr. Opin. Biotechnol.* 23: 338–345.

Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P & Boeke JD (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C : a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* 132: 115–132.

Bull VH & Thiede B (2012) Proteome analysis of tunicamycin-induced ER stress. *Electrophoresis* 33: 1814–1823.

Den Haan R, Kroukamp H, Mert M, Bloom M, Görgens JF & van Zyl WH (2013) Engineering *Saccharomyces cerevisiae* for next generation ethanol production. *J. Chem. Technol. Biotechnol.* 88: 983–991.

Den Haan R, Rose SH, Lynd LR & van Zyl WH (2007) Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. *Metab. Eng.* 9: 87–94.

Domínguez A, Fermiñán E, Sánchez M, González FJ, Pérez-Campo FM, García S, Herrero AB, San Vicente A, Cabello J, Prado M, Iglesias FJ, Choupina A, Burguillo FJ, Fernández-Lago L, López MC (1998) Non-conventional yeasts as hosts for heterologous protein production. *Internatl Microbiol* 1: 131–142.

De Pourcq K, De Schutter K & Callewaert N (2010) Engineering of glycosylation in yeast and other fungi: current state and perspectives. *Appl. Microbiol. Biotechnol.* 87: 1617–1631.

Gurgu L, Lafraya Á, Polaina J & Marín-Navarro J (2011) Fermentation of cellobiose to ethanol by industrial *Saccharomyces* strains carrying the β-glucosidase gene (BGL1) from *Saccharomycopsis fibuligera*. *Bioresour. Technol.* 102: 5229–5236.

Harmsen MM, Bruyne MI, Raué HA & Maat J (1996) Overexpression of binding protein and disruption of the *PMR1* gene synergistically stimulate secretion of bovine prochymosin but not plant Thaumatin in yeast. *Appl. Microbiol. Biotechnol.* 46: 365–370.

Herscovics A & Orlean P (1993) Glycoprotein biosynthesis in yeast. *FASEB J.* 7: 540–550.

Hill J, Donald KAG & Griffiths DE (1991) DMSO-enhanced whole cell yeast transformation. *Nucleic Acids Res.* 19: 6688.

Hoshida H, Fujita T, Cha-aim K & Akada R (2013) *N*-Glycosylation deficiency enhanced heterologous production of a *Bacillus licheniformis* thermostable α-amylase in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 97: 5473–5482.

Idiris A, Tohda H, Kumagai H & Takegawa K (2010) Engineering of protein secretion in yeast: strategies and impact on protein production. *Appl. Microbiol. Biotechnol.* 86: 403–417.

Imperiali B & O'Connor SE (1999) Effect of *N*-linked glycosylation on glycopeptide and glycoprotein structure. *Curr. Opin. Chem. Biol.* 3: 643–649.

Johnson IS (1983) Human insulin from recombinant DNA technology. *Science* 219: 632–637.

Karpova TS, Reck-peterson SL, Elkind NB, Mooseker MS, Novick PJ & Cooper JA (2000) Role of actin and Myo2p in polarized secretion and growth of *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 11: 1727–1737.

Khan SU & Schröder M (2008) Engineering of chaperone systems and of the unfolded protein response. *Cytotechnology* 57: 207–231.

Kroukamp H, den Haan R, van Wyk N & van Zyl WH (2013) Overexpression of native *PSE1* and *SOD1* in *Saccharomyces cerevisiae* improved heterologous cellulase secretion. *Appl. Energy* 102: 150–156.

Kukuruzinska M & Lennon K (1998) Protein *N*-glycosylation: molecular genetics and functional significance. *Crit. Rev. Oral Biol. Med.* 9: 415–448.

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680–685.

Marín-Navarro J, Gurgu L, Alamar S & Polaina J (2011) Structural and functional analysis of hybrid enzymes generated by domain shuffling between *Saccharomyces cerevisiae* (var. *diastaticus*) Sta1 glucoamylase and *Saccharomycopsis fibuligera* Bgl1 β-glucosidase. *Appl. Microbiol. Biotechnol.* 89: 121–130.

Mondésert G & Reed SI (1996) *BED1*, a gene encoding a galactosyltransferase homologue, is required for polarized growth and efficient bud emergence in *Saccharomyces cerevisiae*. *J. Cell Biol.* 132: 137–151.

Nagasu T, Shimma Y, Nakanishi Y, Kuromitsu J, Iwama K, Nakayama K, Suzuki K & Jigami Y (1992) Isolation of new temperature-sensitive mutants of *Saccharomyces cerevisiae* deficient in mannose outer chain elongation. *Yeast* 8: 535–547.

Porro D, Gasser B & Fossati T (2011) Production of recombinant proteins and metabolites in yeasts. *Metab. Eng.* 939–948.

Roncero C & Duran A (1985) Effect of Calcofluor white and Congo red on fungal cell wall morphogenesis: in vivo activation of chitin polymerization. *J. Bacteriol.* 163: 1180–1185.

Rudolph HK, Antebi a, Fink GR, Buckley CM, Dorman TE, LeVitre J, Davidow LS, Mao JI & Moir DT (1989) The yeast secretory pathway is perturbed by mutations in *PMR1*, a member of a Ca2+ ATPase family. *Cell* 58: 133–145.

Sagt CMJ, Kleizen B, Verwaal R, de Jong MD, Müller WH, Smits A, Visser C, Boonstra J, Verkleij AJ & Verrips CT. (2000) Introduction of an *N*-glycosylation site increases secretion of heterologous proteins in yeasts. *Appl. Environ. Microbiol.* 66: 4940–4944.

Sambrook J & Russel DB (2001) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Shen X, Zhang K & Kaufman RJ (2004) The unfolded protein response: a stress signaling pathway of the endoplasmic reticulum. *J. Chem. Neuroanat.* 28: 79–92.

Smith RA, Duncan MJ & Moir DT (1985) Heterologous protein secretion from yeast. *Science* 229: 1219–1224.

Trombetta ES (2003) The contribution of *N*-glycans and their processing in the endoplasmic reticulum to glycoprotein biosynthesis. *Glycobiology* 13: 77–91.

Uccelletti D, Farina F, Cesca, Rufini S, Magnelli P, Abeijon C & Palleschi C (2006) The *Kluyveromyces lactis* α1,6-mannosyltransferase KlOch1p is required for cell-wall organization and proper functioning of the secretory pathway. *FEMS Microbiol. Rev.* 6: 449–457.

Valkonen M, Penttilä M & Saloheimo M (2003) Effects of inactivation and constitutive expression of the unfolded protein response pathway on protein production in the yeast *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 69: 2065–2072.

Van Rooyen R, Hahn-Hägerdal B, La Grange DC & van Zyl WH (2005) Construction of cellobiose-growing and fermenting *Saccharomyces cerevisiae* strains. *J. Biotechnol.* 120: 284–295.

Wang T-Y, Huang CJ, Chen HL, Ho PC, Ke HM, Cho HY, Ruan SK, Hung KY, Wang IL, Cai YW, Sung HM, Li WH, Shih MC (2013) Systematic screening of glycosylation- and trafficking-associated gene knockouts in *Saccharomyces cerevisiae* identifies mutants with improved heterologous exocellulase activity and host secretion. *BMC Biotechnol.* 13: 71–85.

Xu L, Shen Y, Hou J, Peng B, Tang H & Bao X (2014) Secretory pathway engineering enhances secretion of cellobiohydrolase I from *Trichoderma reesei* in *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* 117: 45–52.

Xu P & Robinson AS (2009) Decreased secretion and unfolded protein response up-regulation are correlated with intracellular retention for single-chain antibody variants produced in yeast. *Biotechnol. Bioeng.* 104: 20–29.

Chapter 5

Identifying genes conferring high cellobiohydrolase secretion in *Saccharomyces cerevisiae*

Identifying genes conferring high cellobiohydrolase secretion in *Saccharomyces cerevisiae*

Heinrich Kroukamp[1], Riaan den Haan[2] and Willem H. van Zyl[1]

[1]*Department of Microbiology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa.*
[2]*Department of Biotechnology, University of Western Cape, Private Bag X17, Bellville 7530, South Africa.*

5.1. Abstract

The yeast *Saccharomyces cerevisiae* has a long association with alcoholic fermentation industries and has received renewed interest as a biocatalyst for second generation bioethanol production. Enabling *S. cerevisiae* to hydrolyze lignocellulose, requires the heterologous expression of three groups of hydrolytic enzymes, namely β-glucosidases, endoglucanases and cellobiohydrolases. Although significant improvements were seen in recent years with the expression of these activities in yeast, cellobiohydrolase titers still remain relatively low. With the availability of high-throughput next generation sequencing technologies, it is now possible to identify host specific factors that are beneficial or detrimental to protein secretion; and provide the tools for rational strain design strategies. Through classical breeding and selection, we obtained *S. cerevisiae* strains with up to 3.5-fold increased cellobiohydrolase secretion titers and demonstrated the polygenic nature of the high protein secretion phenotype. Due to chromosomal instabilities, we were unable to identify the causative single nucleotide polymorphisms conferring high cellobiohydrolase secretion. However, we were able to show that methyl methanesulfonate (used for selection) induced these chromosomal abnormalities, linked cell flocculation to secretion, and established a functional pipeline to identify quantitative trait loci (QTL) and strain specific genes involved in the high secretion phenotype for future studies.

5.2. Introduction

Bioethanol produced from lignocellulosic waste is a promising renewable energy source (van Zyl et al. 2007). One of the main production expenses of this second generation technology, is the cost of the cellulolytic enzymes needed for the hydrolysis of the polymeric sugar chains of the cellulosic substrate, to produce free sugars for yeast to ferment to ethanol (Lynd et al. 2005). The

concept of Consolidated Bioprocessing (CBP) could reduce this expense by consolidating the substrate hydrolysis and ethanol fermentation in one step (Lynd et al. 2005). *Saccharomyces cerevisiae* strains have been developed for CBP by heterologously producing the three main classes of recombinant cellulases, namely exoglucanases (i.e. cellobiohydrolases), endoglucanases and β-glucosidases (van Rooyen et al. 2005; Den Haan et al. 2007; Ilmén et al. 2011). Significant progress has been made with the construction of recombinant cellulolytic yeast strains. Examples of these include the first report of a recombinant *S. cerevisiae* strain, expressing a *Saccharomycopsis fibuligera* β-glucosidase I (CEL3A), being able to grow on cellobiose as sole carbohydrate source at similar rates as on glucose (van Rooyen et al. 2005). By expressing an endoglucanase I from *Trichoderma reesei*, in combination with the Cel3A, Den Haan et al. (2007) enabled the yeast to hydrolyze even more complex cellulosic substrates, demonstrating the hydrolysis and fermentation of phosphoric acid swollen cellulose (PASC) to ethanol for the first time. Recently, Ilmén et al. (2011) reported high levels of cellobiohydrolase I (Cel7) and cellobiohydrolase II (Cel6) production in *S. cerevisiae*. They were also able to achieve the fermentation of microcrystalline cellulose (Avicel^{TM}) to ethanol by a CBH-producing strain and β-glucosidase addition. Even though the study of Ilmén et al. (2011) reported the highest levels of CBH production in *S. cerevisiae* to date, the yeast could only hydrolyze 23% of the Avicel to fermentable sugars and it is believed that more enzyme could increase the hydrolysis rate and efficiency.

With the accumulating information on the cellular mechanisms involved in secretion, host strain rational design strategies have been a frequent choice to improve secretion of heterologous proteins (Idiris et al. 2010). These strategies are often limited by the protein specific nature of the secretion enhancing gene alteration. Protein secretion is also a polygenic phenotype, further complicating the identification of these enhancing elements (Parts et al. 2011). With the availability of high-throughput next generation sequencing technologies and increasing computing capacity for large data sets, it is now possible to pinpoint multiple genes that are collectively responsible for a specific phenotype. One of these techniques, namely pooled-segregant whole-genome sequencing analysis (PSWGSA), has been demonstrated to identify genes responsible for polygenic traits such as ethanol tolerance (Swinnen et al. 2012) and heat tolerance (Parts et al. 2011). This bioinformatic technique relies on the identification of

quantitative trait loci (QTLs), contributing to a specific phenotype, by analyzing the segregation of shuffled genomes in selected progeny.

In this study, we aimed to identify alleles or genomic alterations conferring high cellobiohydrolase I secretion in yeast, using a pooled-segregant whole-genome sequence analysis strategy. Although we identified multiple chromosomal regions where these putative enhancing elements reside, we were unable to elucidate the specific causative alteration due to unexpected changes in the yeast ploidy state. We were able to elucidate the reason for these chromosomal abnormalities, link cell flocculation to secretion, and established a functional pipeline to identify quantitative trait loci (QTL) and strain specific genes for future studies.

5.3. Materials and methods

5.3.1. Media and growth conditions

The *S. cerevisiae* strains were routinely cultured in YPD (10 g/L yeast extract, 20 g/L peptone and 20 g/L glucose) medium at 30°C on an orbital shaker at 200 rpm. Putative parental strains, containing the integrated *CEL7A* expression cassette, were selected on yeast nitrogen base (YNB) media (1.7 g/L yeast nitrogen base (Difco) with amino acids (excluding isoleucine and valine), 5 g/L ammonium sulphate and 20 g/L glucose) supplemented with 40 ug/mL sulfometuron methyl (SMM, Sigma-Aldrich). The *S. cerevisiae* F1 progeny were screened in 24-well tissue culture plates (Greiner bio-one, Germany), covered with an AeraSeal™ breathable film (Sigma-Aldrich, Germany). *Escherichia coli* DH5α (*deoR endA1 gyrA96 hsdR17* 6[*lac*]*U169 recA1 relA1 supE44 thi*-1 [*q80 lacZ6M15*]) was grown in either Terrific Broth (12 g/L tryptone, 24 g/L yeast extract and 4 ml/L glycerol) or on Luria-Bertani agar plate supplemented with 100 µg/mL Ampicillin for selecting plasmid containing cells.

5.3.2. Plasmid and recombinant strain construction

Standard protocols were followed for DNA manipulations (Sambrook & Russel 2001). Restriction endonucleases, DNA polymerases and T4 DNA ligase were purchased from Thermo Scientific (Waltham, MA, USA) and used as directed by the manufacturer. Digested DNA was eluted from agarose gels with the Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, CA, USA).

5.3.2.1. Construction of plasmids, yeast transformant and parental strains

Plasmid descriptions are given in **Table 5.1**. For the construction of the pRDH219 integrative plasmid, coding for the engineered *Talaromyces emersonii* Cel7A (Ilmén et al. 2011), pRDH144 was digested with *Bgl*II and *Bam*HI, and the resulting *ENO1*$_P$-*CEL7A*-*ENO1*$_T$ containing fragment cloned into the *Bam*HI site of the pCEL2 vector. The pRDH219 was linearized by digesting with *Not*I with and subsequently transformed into the *S. cerevisiae* strains NI-C-D4, W303, M0341, Y294, CEN.PK2-1 and BY4742 strains, targeted to the *LEU2* locus. These strains are briefly described in **Addendum C, Table C1**. Putative transformants were selected on agar plates containing 40 µg/mL sulfometuron methyl (SMM). The integration and the location of the *CEL7A*-expression cassette were confirmed with PCR, using the oligonucleotides interSMR_L and A7708 (**Addendum C, Table C2**). A single integrated copy of the CEL7A-expressing cassette was detected with realtime-PCR (qPCR) (Applied Biosystems StepOne Real-time PCR System) using the comparative C_t method (**Addendum C, Fig. C2**). Primers for the qPCR were designed to amplify ~100 bp-homologous regions shared between the native *TFC1* and *TAF10* genes, and the *CEL7A* of both parental strains (**Addendum C, Table C2**). The Y294 strain was converted to the *MAT**a*** mating type, by overexpressing the *S. cerevisiae HO* site-specific endonuclease gene by means of the pFL39GAL1 HO Kan MX episomal plasmid and subsequent induction with 2% galactose. After the *MAT**a*** mating type was confirmed by PCR (primer sequences in **Addendum C, Table C2**) and the pFL39GAL1 HO Kan MX plasmid cured from the Y294[CEL7A] strain, it was mated with the M0341[CEL7A] strain to produce the diploid *S. cerevisiae* H3 (Treco and Winston, 2008). A first generation (F1) haploid progeny were obtained through random spore isolation (increased β-mercaptoethanol to 40 ul) (Treco and Winston, 2008), after the H3 strain was induced to sporulate on a 10 g/l potassium acetate agar plate at 25°C. Homozygous diploids of the M0341[CEL7A] and Y294[CEL7A] were also produced by swapping the mating type as described above, and subsequently mated to produce M0341[CEL7A](2n) and Y294[CEL7A](2n).

5.3.2.2. FLO8 deletion strains

The clonat resistance marker was amplified with primers designed with 30 bp overhangs homologous to the promoter and terminator regions of the *FLO8* gene (**Addendum C, Table C2**). The PCR product was used to transform the flocculating M0341 strain using the

LiOAc/DMSO method (Hill et al. 1991), and selected on YPD agar plates containing 200 ng/mL clonat. Putative transformants were confirmed visually for lack of the flocculation phenotype and with PCR analysis.

Table 5.1. Plasmids and strains constructed and used in the study.

Yeast Strains/Plasmid	Relevant genotype	Source/reference
Plasmids:		
pRDH219	*bla leu2:: ENO1$_P$-T.e.Cel7A-T.r.CBH_cbm-ENO1$_T$ SMR*	This study
pRDH224	*bla Sh ble URA3 ENO1$_P$-T.e.Cel7A-T.r.CBH_cbm-ENO1$_T$*	(Ilmén et al. 2011)
ySFI	*bla URA3 PGK1$_P$-S.f.Cel3A-PGK1$_T$*	(van Rooyen et al. 2005)
pMU1531_EGIII	*bla Sh ble URA3 ENO1$_P$-N.f.Cel12A-ENO1$_T$*	This study
***S. cerevisiae* strains:**		
Y294:	*MATα leu2-3 leu2-112 ura3-52 his3Δ trp1-289 GAL+ [cir+]*	ATCC 201160
M0341:	*MATα gre3*::*kanMX*	(Ilmén et al. 2011)
W303-1a:	*MATα leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15* [*phi*$^+$]	(Voth et al. 2005)
NI-C-D4:	*MATα trp1 ura3 pep4*	(Wang et al. 2001)
BY4742:	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	(Brachmann et al. 1998)
CEN.PK2-1:	*MATα ura3-52 trp1-289 leu2-3_112 his3Δ1 MAL2-8C SUC2*	(van Dijken et al. 2000)
M0341Δflo8:	*MATα gre3*::*kanMX flo8*::*natMX*	This study
H3	*MATα/MAT***a** *LEU2*::CEL7A/*leu2-112*::CEL7A *URA3*/*ura3*-52 *HIS3* /*his3 TRP1/trp1*-289 *gre3*::*kanMX/gre3*	This study
H3K27	*MAT***a**	This study
H3O23	*MATα ura3-52 his3Δ trp1-289*	This study
H3M1	*MAT***a** *ura3-52 his3Δ trp1-289*	This study
H3M28	*MAT***a** *his3Δ gre3*::*kanMX/gre3*	This study
H3H29	*MATα trp1-289*	This study

5.3.2.3. Expression of cellulase genes in hybrid strains

The integrated Cel7A expressing cassette was removed by restoring the *LEU2* allele. PCR products containing the functional *LEU2* promoter and ORF was used to remove the *CEL7A* cassette from the M0341[CEL7A], Y294[CEL7A], H3K27, H3O23, H3M28, H3M1 and H3H29 strains by direct replacement. Putative transformants were selected on SC^{-LEU} agar plates and the absence of cellobiohydrolase I activity was confirmed. The functional *URA3* allele of the H3K27, H3M28, H3H29 and M0341 strains was subsequently removed by transforming with PCR amplified *ura3* ORFs from the Y294. The Y294 strain contains a point mutation, rendering the *URA3* gene non-functional. Transformants were selected on SC media plates containing 12 mg/L uracil and 1 mg/L 5-Fluoroorotic acid (FOA). The selected strains containing the non-functional Y294 *ura3-52* allele was subsequently transformed with ySFI, pRDH224 and

pMU1531_EGIII, in order to secrete the *S. fibuligera* Cel3A, *T. emersonii* Cel7A and *N. fischeri* Cel12A, respectively. These plasmids contained a functional *URA3* gene as selectable marker, as was selected on SC^{-URA} agar plates. The resulting strains were renamed in the following fashion: hybrid progeny name + cellulase name, e.g. H3K27+CEL7A.

5.3.3. Enzymatic assays

The potential parental strains were grown for five days in 10 mL YPD and assayed daily in triplicate for CBH1 activity. 50 µL of culture supernatant were mixed with 50 µl of 4 mM 4-methylumbelliferyl-β-D-lactoside (MULac; in DMSO; Sigma-Aldrich) and incubated at 37°C for 10 minutes. To terminate the reactions, 100 µL of 1M Na_2CO_3 were added and the fluorescence was measured with a Thermo Scientific Varioskan™ fluorometer (excitation at 355 nm and emission at 460 nm). A standard curve ranging between 0.63 – 20.0 µM 4-methylumbelliferone (Sigma-Aldrich) was used to calculate enzymatic activity. For the screening of the first generation (F1) progeny, up to 530 different progeny colonies were cultivated in 1 mL YPD in 24-well tissue culture plates, covered with AeraSeal™ film and assayed after 48 h of growth at 200 rpm and 30°C. The MULac-assay previously described for the parental strains was downscaled in equal ratios, to 100 µL, for the screening of the F1 progeny. After selecting 28 high secreting progeny from the first screen, the assays were repeated in duplicate to confirm that the secreted Cel7A activities per OD_{600} was higher than the secreted activity of the M0341 parental strain. The Cel7A segregants of the selected and control pools were re-evaluated after several rounds of propagation in YPD media to access the stability of the secretion phenotype.

To evaluate the secreted β-glucosidase, cellobiohydrolase and endoglucanase activities of episomal plasmid containing strains, cells were cultivated under similar conditions as for the F1 segregants in triplicate, but 1 mL SC^{-URA} media was used. *p*-Nitrophenyl based assays were carried out using *p*-nitrophenyl-α-D-glucopyranoside (*p*NPG; Sigma-Aldrich) as substrate for Cel3A producing strain supernatants (1 h at 50°C) or with *p*-nitrophenyl-*β*-D-cellobioside (*p*NPC; Sigma-Aldrich) as substrate for Cel7A producing strain supernatants (3h at 50°C) as previously described (Kroukamp et al. 2013; Ilmén et al. 2011). Reducing sugar assays with carboxymethyl cellulose (CMC; Sigma-Aldrich) as substrate were used to determine supernatant activity of the Cel12A producing strains (30 min at 50°C) as previously described (Kroukamp et al. 2013). In all cases, one enzyme unit (U) was defined as the amount of enzyme required to

produce 1 µmol glucose or equivalent (i.e. *p*NP) in one minute under assay conditions. Student's t-test method was used to determine significant differences between triplicate sample values, using a 95% confidence level.

5.3.4. Evaluation of prominent phenotypes

To determine whether flocculation was linked to high Cel7A secretion, the *FLO8* gene, encoding a transcription factor for several other flocculin genes, was deleted in the flocculation M0341[CEL7A] strain as described in **section 5.3.2.2**.. The flocculation of the M0341 strain was completely relieved with the addition of mannose or glucose to a concentration of 20 g/L, corresponding to the NewFlo-flocculation phenotype (Soares 2011). The addition of mannose did not reverse flocculation in all the F1 segregants (referred to as the Flo1-phenotype), although 2 mM EDTA was shown to be effective.

The auxotrophic requirements, inherited from the Y294 strain, of the selected high-secretion and control segregants were determined by growth assays on SC agar plates lacking either histidine or tryptophan. The *GRE3* gene, coding for an aldose reductase, of the M0341 strain had previously been deleted with a *kanMX* marker cassette. The presence of the *GRE3* deletion in the segregants was determined by growth on YPD agar plates supplemented with 200 µg/mL G418 (Melford, Ipswich). The mating type of ~500 segregants was determined with PCR analysis using the Mat**a**, Matα and Mat primers (**Addendum C, Table C2**).

5.3.5. Pooled DNA preparation and whole-genome sequencing

The workflow for pooled-segregant whole-genome sequencing used in this study was similar to what was previously described (Swinnen et al. 2012). Twenty eight first generation meiotic segregants with the highest levels of recombinant CBH1 secretion were selected and individually grown until stationary phase in YPD media. EDTA was added to a final concentration of 2 mM to alleviate cell aggregation before pooling the strains based on equal OD_{600} values. Genomic DNA was extracted (Johnston 1994), of which at least 3 µg was provided to Beijing Genomics Institute (BGI, Hong Kong) for sequencing analysis using the Illumina HiSeq2000 next generation sequencing technology. The same was done for 28 randomly selected meiotic segregants (to serve as reference for random segregation), the Y294[CEL7A] and

M0341[CEL7A] parental strains. Up to 5.5 million Paired-end short reads of ± 90 bp were generated for each of the four samples with an average coverage of ~35 times.

5.3.6. Genome analysis, SNP detection and bioinformatics

Java Eclipse software with the NGSEP plugin (Duitama et al. 2012) was used to determine single nucleotide polymorphisms (SNPs) in the parental strains and the single nucleotide variance (SNV) in the pooled samples. The sequence reads were aligned to the *S. cerevisiae* S288C (sacCer3, April 2011) reference genome, allowing subsequent genotyping of SNPs and insertion/deletion mutations of more than one base pair (indels). This NGSEP plugin provides a user friendly interface for Bowtie2, a memory-efficient aligning tool for aligning sequence reads to long reference sequences using the Burrows-Wheeler transform algorithm (Langmead & Salzberg 2012). Using custom Linux Bash scripts, the individual SNPs (per chromosome), in the pooled samples were matched to the parental SNPs, in order to determine from which parental strain they were inherited, and the frequency at which this occurred. For simplicity, indel frequencies were not calculated. Values for a trend line, representing the average SNP frequencies per 5kb genomic region (bin), were determined using the custom scripts Simple_SNP_call_file.sh, Frequency_calculations.sh and Trendline.sh (**Addendum C, Fig. C5**). The standard deviations were calculated for the SNP frequencies for each 5kb bin using the STDevChr script (**Addendum C, Fig. C5**).

In order to determine the prominent genomic differences between the parental strains Y294[CEL7A] and M0341[CEL7A], the sequencing reads of both strains were aligned to the S288C reference genome using SOAP2 v2.21 (implementing two-way Burrows-Wheeler transform algorithm) (Li et al. 2009) and *de novo* assembled using Velvet v1.2.10 (implementing de Bruijn graphs manipulation)(Zerbino 2010). The soap2sam Perl script (obtained from http://soap.genomics.org.cn/soapaligner.html#down2) was used to convert the soap alignment format to the standard Sequence Alignment/Map (SAM) format. These SAM files were manipulated with standard samtools and bcftools (samtools.sourceforge.net/samtools.shtml) for downstream processing. Integrative Genomic Viewer (IGV) v2.3.20 was used to visualize the SOAP2 aligned genomes (Thorvaldsdóttir et al. 2013). Augustus v2.5.5, implementing hidden Markov models, was used to predict putative *S. cerevisiae* open reading frames (ORFs) and subsequently the proteins they encode, for the SOAP2 alignments and the Velvet assembled

contigs (Stanke et al. 2004). The predicted ORFs were compared and aligned to protein sequences from 18 different *S .cerevisiae* strains, using BLAST+ v2.2.29 (Altschul et al. 1990). Only sequences with a homology of at least 95% were retained as putative full length proteins. A list of the *S. cerevisiae* strains, from diverse industrial and natural sources, used as protein database is given in **Table 5.2**.

Table 5.2. List of *S. cerevisiae* strains used as protein database for BLAST search as described on the *Saccharomyces* Genome Database (http://www.yeastgenome.org/).

S. cerevisiae strain	Strain description
AWRI1631	Haploid derivative of a South African commercial wine strain N96
AWRI796	South African red wine strain
CBS7960	Brazilian bioethanol factory isolate
CEN.PK113-7D	Laboratory strain
CLIB215	New Zealand bakery isolate
EC9-8	Haploid derivative of an Israeli canyon isolate
Foster's B	Commercial ale strain
JAY291	Haploid derivative of the Brazilian industrial bioethanol strain PE-2
Kyokai7	Japanese sake yeast
LalvinQA23	Portuguese Vinho Verde white wine strain
S288C	Laboratory strain, genomic reference strain for *S .cerevisiae*
Sigma1278b	Laboratory strain
T73	Spanish red wine strain
Vin13	South African white wine strain
W303	Laboratory strain
Y10	Philippine coconut isolate
YJM789	Haploid derivative of an opportunistic human pathogen
YPS163	Pennsylvania woodland isolate

The average read coverage of each chromosome of each parental strain and pooled samples was determined by calculating the number of reads which start within 1kb bins spanning the entire length of each chromosome. This was achieved with the Coverage_chr script (Example script S5). The average read coverage was determined similarly for the SOAP aligned sequences of

both pooled genomic DNA samples. The bioinformatics workflow is illustrated in **Addendum C, Fig. C1**.

5.3.7. Flow cytometry

The parental strains and the 28 high secretion progeny were cultivated to mid-exponential phase in 5 mL YPD tubes at 30°C. Washed cells were fixed and stained with propidium iodide (PI, Sigma-Aldrich) as described by Popolo et al. (1982). At least 10 000 events were detected per sample with a BD biosciences FACSAria™ flow cytometer. A linear relationship between DNA content and PI fluorescence was demonstrated for haploid, diploid and tetraploid transition states of a S288C control cells.

5.4. Results

5.4.1. Screening yeast strains for high Cel7A activity

In order to select high and reference cellobiohydrolase I secreting parental strains, the secretion capacity of different laboratory and industrial strains, containing a single integrated *CEL7A* expression cassette, were detemined. The *CEL7A* producing transformants were evaluated based on the extracellular cellobiohydrolase activity per dry cell weight (DCW), over five days of growth in YPD media (**Fig. 5.1**). The M0341[CEL7A] strain had the highest activity per DCW for the time monitored, reaching 0.58 U/gDCW. With less than half the activity per DCW, the NI-C-D4[CEL7A] strain produced the second highest activity per DCW. The BY4742[CEL7A] strain had almost no cellobiohydrolase activity, similar to that of the reference strain without the *CEL7A* expression cassette. There was a clear distinction between the M0341[CEL7A] strain's activity and the other transformants and it was therefore chosen as the "high secretion" parent strain, with the Y294[CEL7A] strain selected as the "reference secretion" parent strain.

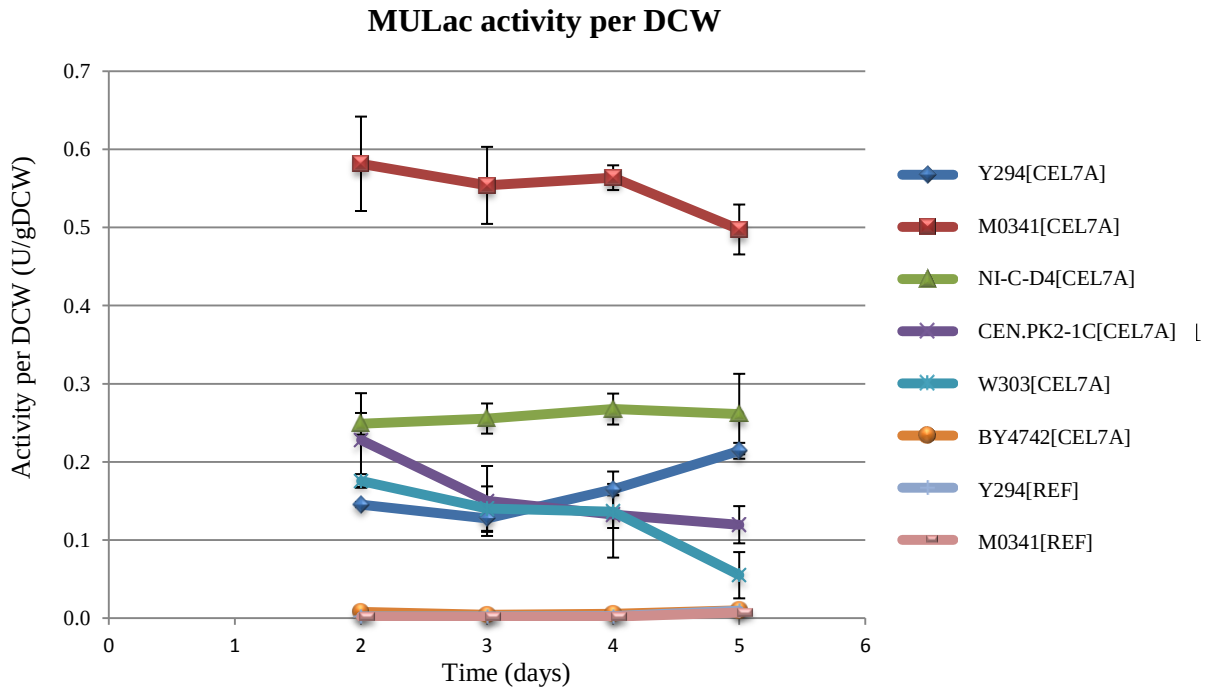


Fig. 5.1. The extracellular cellobiohydrolase I MULac activities of the putative parental strains over five days. The values were normalized with each strain's respective dry cell weight. Error bars indicate the standard deviation from the mean value of biological triplicates.

After mating type switching of the Y294[CEL7A] strain and the subsequent mating to M0341[CEL7A], extracellular Cel7A activity of the resulting diploid strain (designated H3) was measured. Since the resulting diploid strain contained two copies of the *CEL7A*-expression cassette, the activity per OD_{600} of the homozygous diploid versions of the Y294[CEL7A] and M0341[CEL7A] stains were compared to the H3 hybrid strain (**Fig. 5.2**). No significant difference was observed between either the haploid and diploid versions of the respective parental strains. The H3 strain had more than twice the extracellular activity of the diploid Y294CEL7A](2n), but approximately three times less activity than the diploid high secretion parent, M0341[CEL7A](2n).

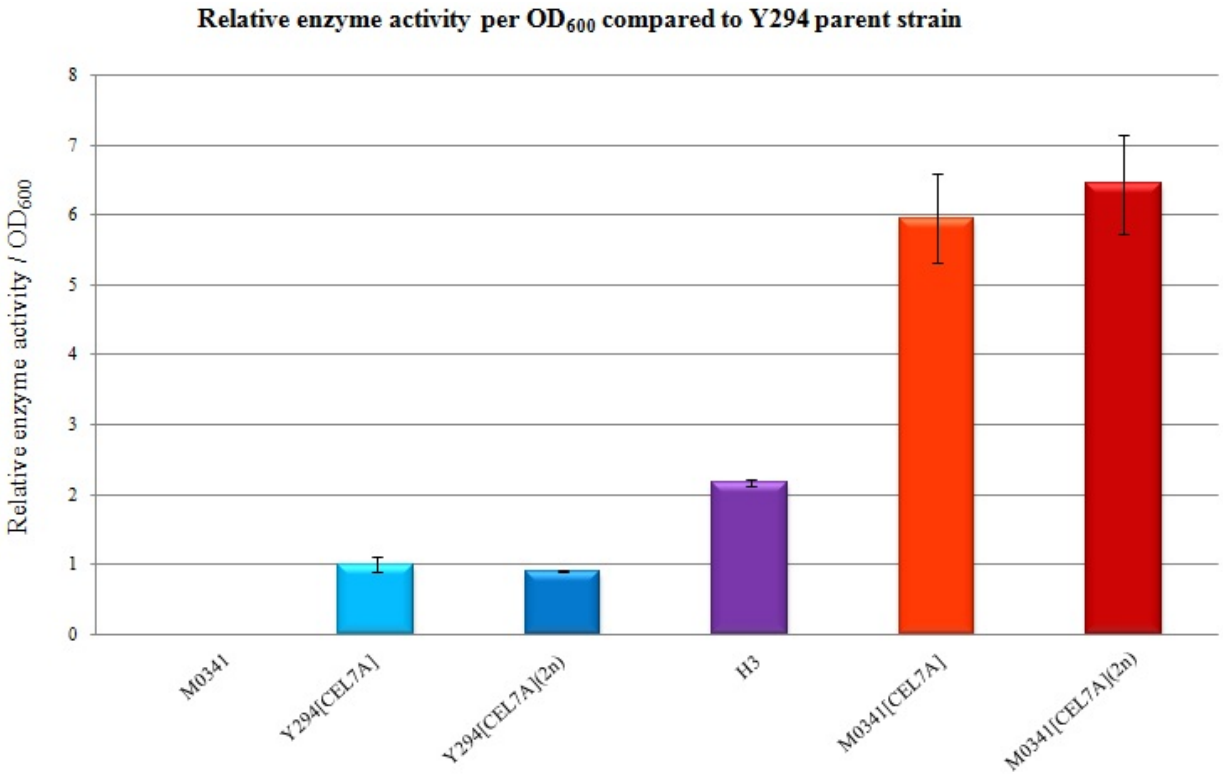
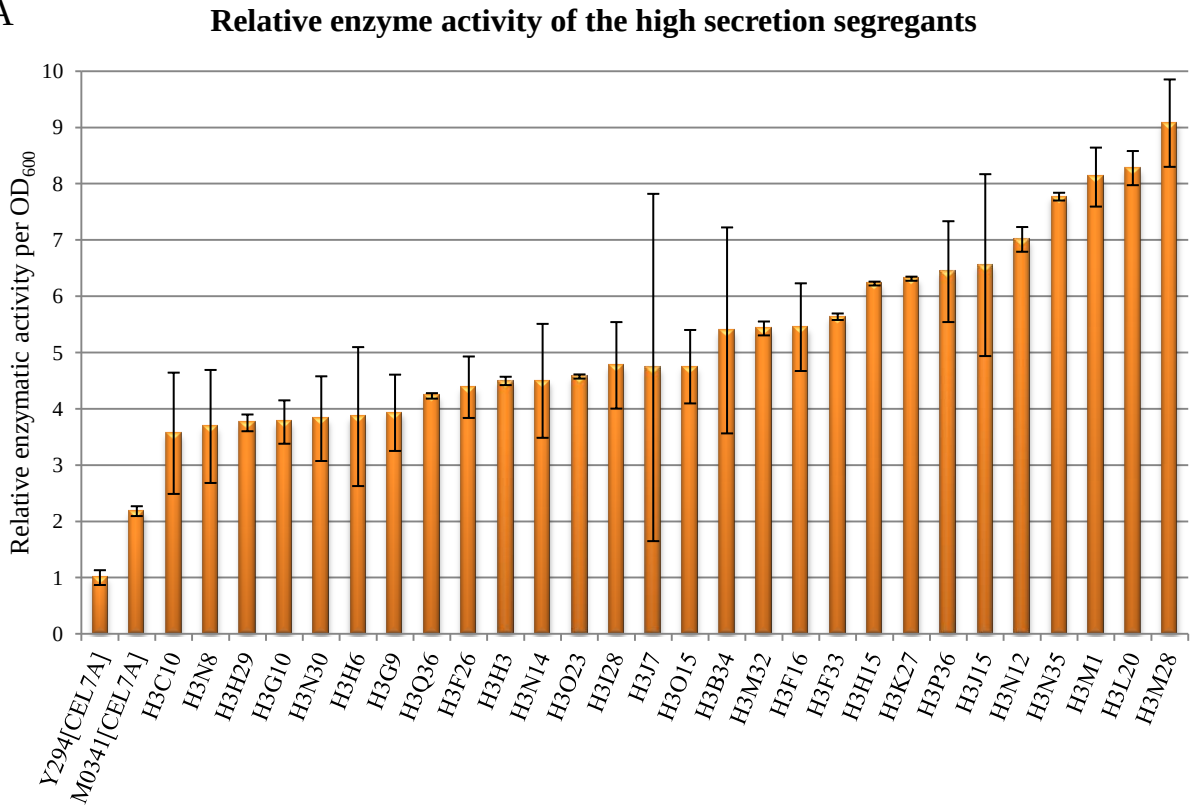


Fig. 5.2. The relative extracellular Cel7A activity on MU-Lac per optical density unit. All activities are represented relative to the haploid reference secretion strain Y294[CEL7A]. Error bars indicate the standard deviation from the mean value of biological quadruplicates.

Subsequently the H3 strain was induced to sporulate on a potassium acetate agar plate followed by random spore isolation. Five hundred and thirty F1-generation meiotic segregants of the H3 strain were screened for high extracellular cellobiohydrolase I activity (**Addendum C, Fig. C3**). The 28 segregants with the highest activity from this initial screen, were re-cultivated and assayed in duplicate (**Fig. 5.3A**). All 28 selected segregants had higher enzyme activities than the high secreting parental strain. However, these segregants had reduced supernatant activities after several rounds of sub-cultivation (**Fig. 5.3B**). At this point, a third of the strains had supernatant CEL7A activity per OD_{600} values below that of the Y294[CEL7A] strain. The H3F16, H3N14, H3I28, H3C10 and H3O23 strains had higher supernatant activities of up to twice that of the M0341[CEL7A] strain. This result indicated some possible genetic instability in the F1 progeny segregants.

A



B

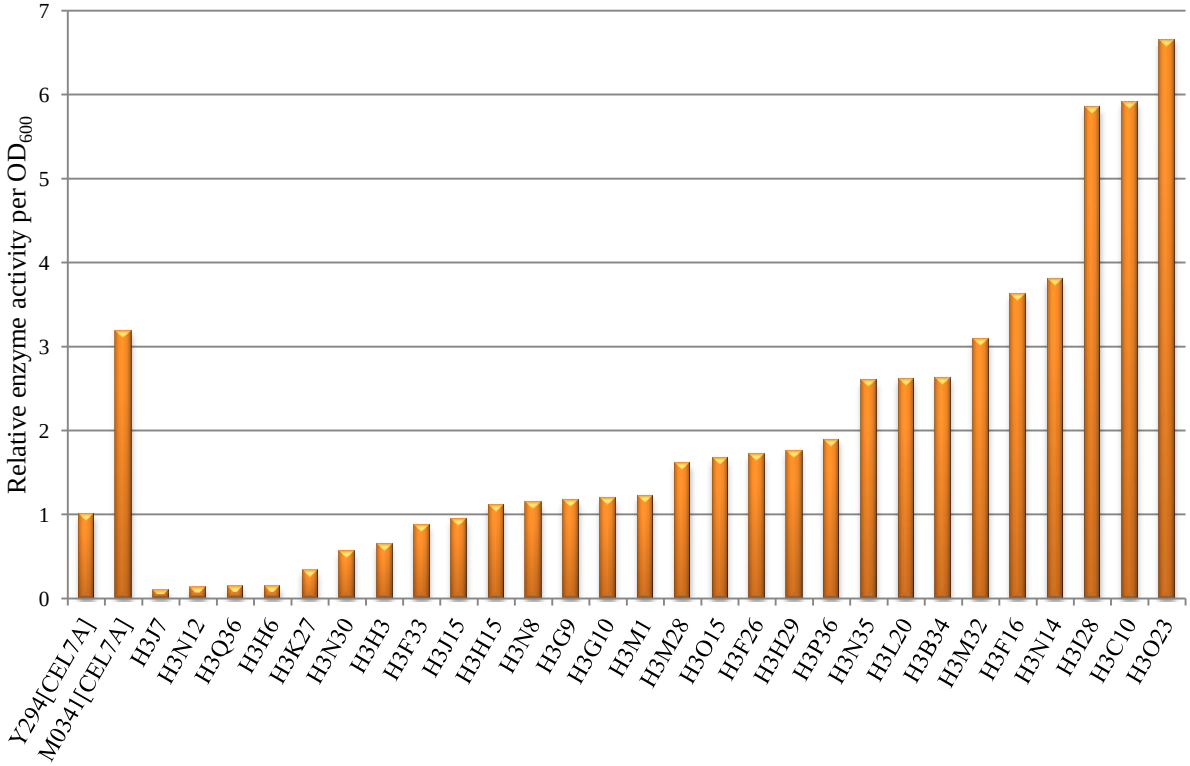


Fig. 5.3. (Previous page) (**A**) The relative extracellular enzyme activity per optical density unit of the 28 selected high secreting segregants that were initially selected from the F1 progeny after the sporulation of the H32 diploid strain. All activities are represented relative to the haploid reference secretion strain Y294[CEL7A]. Error bars indicate the standard deviation from the mean value of duplicates. (**B**) Relative enzyme activities per OD600 of the high secretion pool after several rounds of sub-cultivation.

5.4.2. Evaluating potential linkage of major phenotypes and cellobiohydrolase activity

A small, but significant difference (p = 0.0056) was found in extracellular cellobiohydrolase I activity between the non-flocculating M0341[CEL7A]flo8Δ strain and the flocculating isogenic M0341[CEL7A] strain (**Fig. 5.4**). The flocculation phenotype segregates in a ~1:1 ratio, however the characteristics of the flocculation observed in F1 progeny differed from that of the flocculating M0341 parent strain. Approximately 27% of the strains displayed a flocculation phenotype of dense cell aggregates which could attach to the surface of the 24-well plates (data not shown). No pseudohyphae were observed in these segregants. No clear linkage of flocculation and high CEL7A secretion were observed in the F1 progeny.

In randomly selected progeny the segregation ratio of the individual Y294 derived auxotrophies was expected to be approximately 1:1. This was indeed the case for uracil (*ura3*-52 locus) and tryptophan (*trp1*-289 locus) dependencies displayed by the F1 progeny; however, in the high secretion pool the *URA3* and *trp1*-289 alleles were favored. The expected 1:1 segregation of the *his3* locus was not observed in the randomly selected segregants, with ~74% of the strains requiring histidine. In the high secretion segregants this frequency was lower, with only 54% of the progeny requiring histidine supplementation.

The mating type ratio of the F1 progeny also deviated from the expected 1:1 ratio. Only 26% of the F1 progeny were confirmed to be *MATα*. A high percentage of diploid segregants were obtained after random spore isolation, even after increasing the *β*-mercaptoethanol concentration to enhance the lysis of vegetative cells.

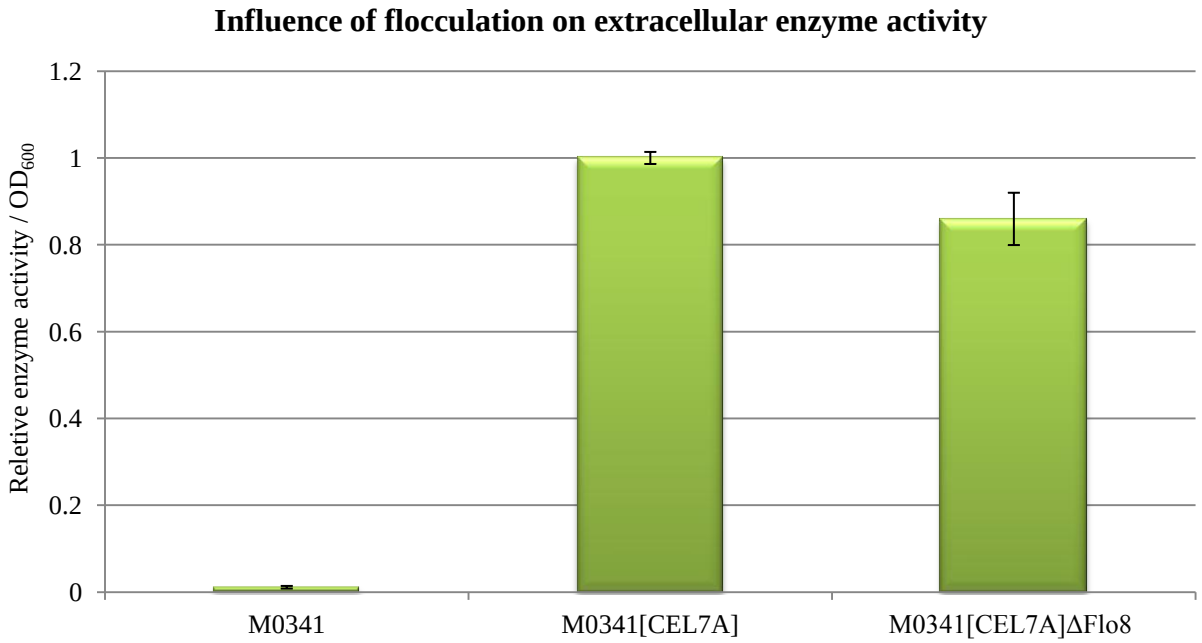


Fig. 5.4. The influence of cell flocculation on supernatant cellobiohydrolase I activity of the M0341[CEL7A] strain. Error bars indicate the standard deviation from the mean value biological triplicates.

5.4.3. SNP analysis and QTL identification

Aligning the M0341[CEL7A] and Y294[CEL7A] parental genomes to the S288C reference genome, revealed ~69 000 and ~17 000 small insertions / deletions and base pair changes, respectively (summarized in **Table 5.3**). The Y294[CEL7A] genome was similar to the reference genome, with only 0.15% base pair differences, with some chromosomes only differing by 0.01%.

Table 5.3. Comparison of genomic differences between the M0341[CEL7A] and Y294[CEL7A] strain.

	S2884 genome	Number of SNPs and indels		Percentage difference		Number of reads		Average base coverage*			
	Chr length	M0341	Y294	M0341	Y294	M0341	Y294	M0341	Y294	Selected pool	Control pool
Chr I	230218	2295	373	1.00	0.16	72472	78352	28.48	30.53	31.69	35.72
Chr II	813184	4217	1554	0.52	0.19	329557	318625	36.48	35.23	31.85	32.60
Chr III	316620	2048	875	0.65	0.28	117443	110394	33.34	31.34	33.72	32.95
Chr IV	1531933	8203	1327	0.54	0.09	595284	634053	34.97	37.25	32.65	33.96
Chr V	576874	2940	349	0.51	0.06	228189	230433	35.59	35.94	33.76	34.34
Chr VI	270161	1742	39	0.64	0.01	108553	96153	36.05	31.93	30.00	30.81
Chr VII	1090940	5767	1686	0.53	0.15	425752	441122	35.12	36.39	32.32	33.06
Chr VIII	562643	3600	219	0.64	0.04	218928	220714	34.99	35.28	35.49	33.80
Chr IX	439888	3257	1228	0.74	0.28	164205	162602	33.59	33.26	37.99	42.20
Chr X	745751	4400	1180	0.59	0.16	285808	304446	34.48	36.73	33.07	32.47
Chr XI	666816	3765	1104	0.56	0.17	255510	262236	34.48	35.44	40.32	37.96
Chr XII	1078177	5761	2420	0.53	0.22	747311	805628	37.11	35.64	34.96	35.76
Chr XIII	924431	4937	1117	0.53	0.12	365649	377379	35.58	36.76	35.45	33.73
Chr XIV	784333	4115	146	0.52	0.02	319753	315660	36.71	36.24	36.35	33.69
Chr XV	1091291	6103	1845	0.56	0.01	427979	446591	35.27	36.81	32.91	33.49
Chr XVI	948066	6176	1403	0.65	0.15	381816	390988	36.25	37.12	35.31	34.81
	11123260	69326	16865	0.62	0.15	5044209	5195376	34.91	35.12	34.24	34.46

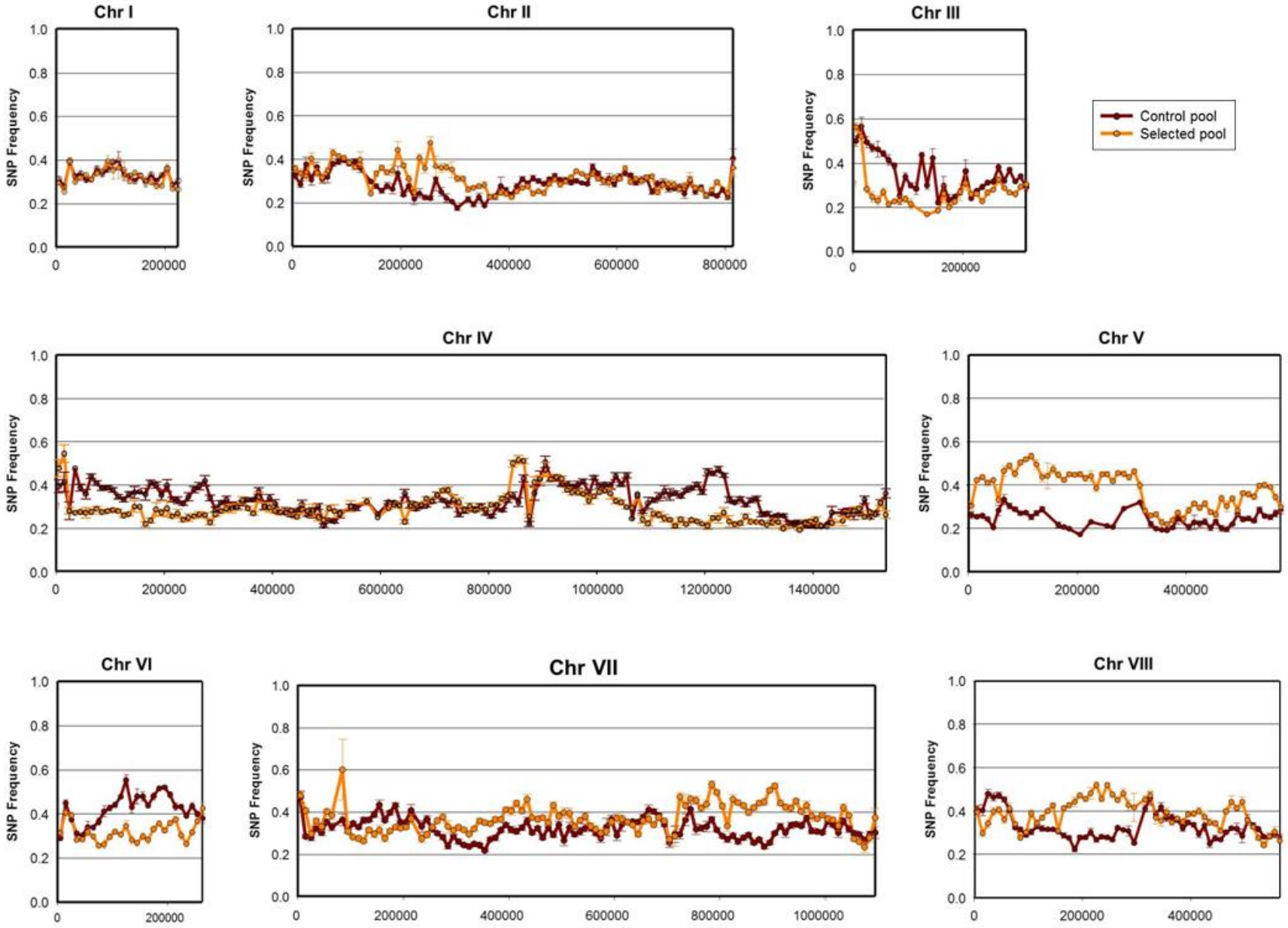
*Chromosome XII rDNA coding coverage values were removed due to high copy numbers

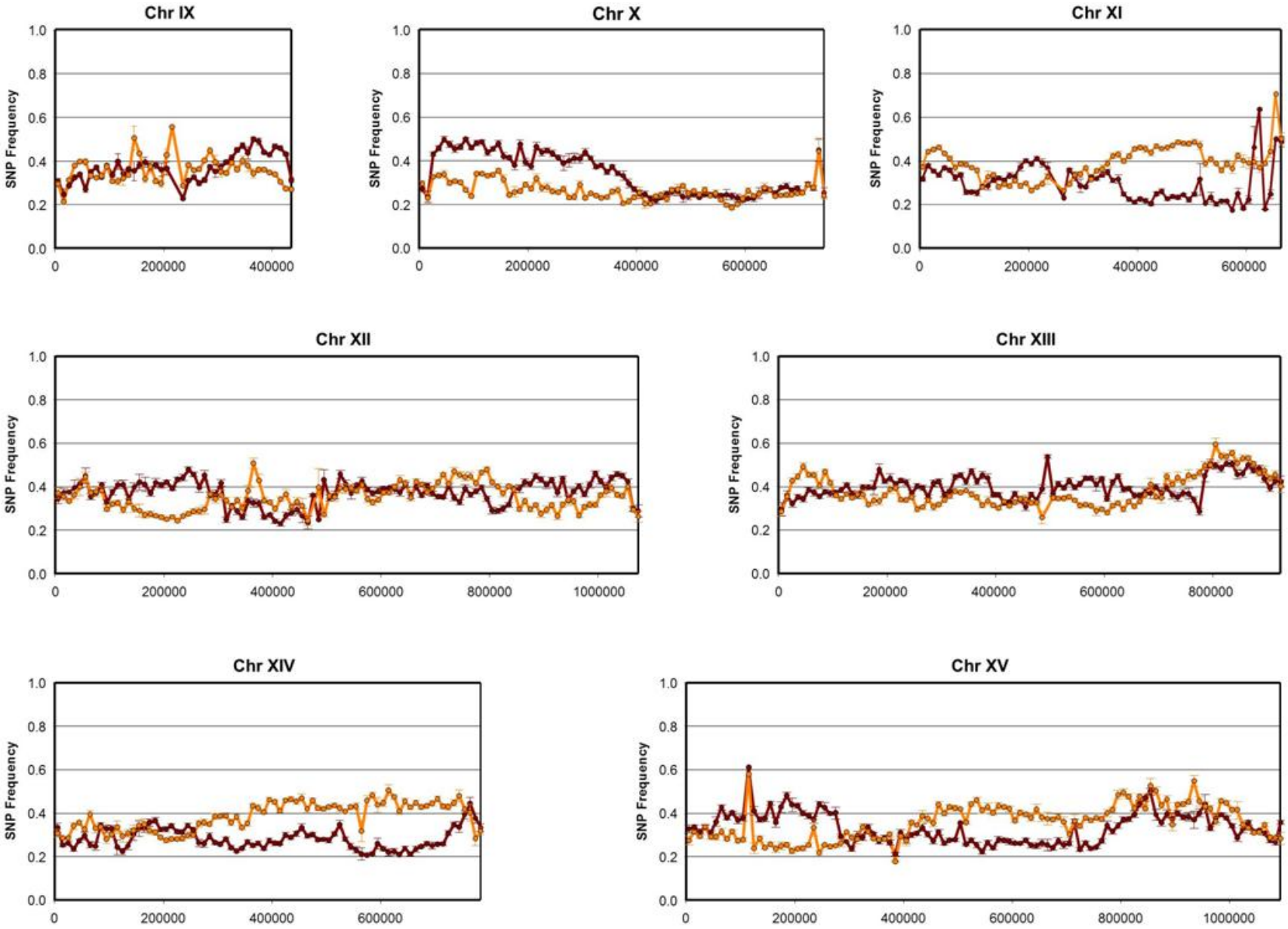
It was also apparent that the genomic differences observed were mostly conserved to certain defined genomic regions, suggesting that the Y294 strain might have S288C or a closely related strain, in its parentage (data not shown). Conversely, genomic changes were distributed throughout the genome of the M0341[CEL7A] strain, with chromosome I being the most divergent compared to the reference chromosome with 1% variation in base pairs. In order to identify genomic regions in which the CEL7A secretion enhancing genetic element lays, the frequency of each SNP was determined and the average frequency per 5 kb region used to draw a trend line of each chromosome, for both the selected and control pooled samples. The frequency graphs are displayed in **Fig. 5.5**.

Multiple genomic regions of the selected pool differed significantly from the control pool SNP frequencies. Six genomic regions differed with more than 20% and were considered major QTLs. The chromosomal regions Chr III:20-50kb, Chr V:90-270kb, Chr VII:880-910kb, Chr XI:390-510kb and Chr XIV:550-710kb had high frequencies of the M0341[CEL7A] SNPs, while there was no selection for the M0341[CEL7A] Chr IV:1200-1230kb region, although this region was more prevalent in the control group. These QTLs were designated QTL_III, QTL_V, QTL_VII, QTL_XI and QTL_XIV, as reference to their chromosomal location. The five QTL regions with a preference for the M0341[CEL7A] SNPs contains ~306 ORFs.

5.4.4. Genome analysis and bioinformatics to identify alleles conferring high secretion

In addition to the prominent phenotypic differences between the two parental strains, we determined which unique, potentially secretion enhancing, ORFs are present in each strain. A total of 5553 and 5529 putative ORFs were identified in the M0341[CEL7A] and Y294[CEL7A] strains, respectively. On average ~60 more ORFs were identified using the Soap aligned reads, compared to the Velvet assembled contigs. 61 and 37 ORFs were identified to be unique to the M0341 and Y294 strains, respectively; most of which are present in the genetically diverse sub-telomeric and centromeric regions of the chromosomes. Only one gene of the strain specific ORFs, namely *HOR7*, was previously implicated in secretion of heterologous proteins (Wang et al. 2013).





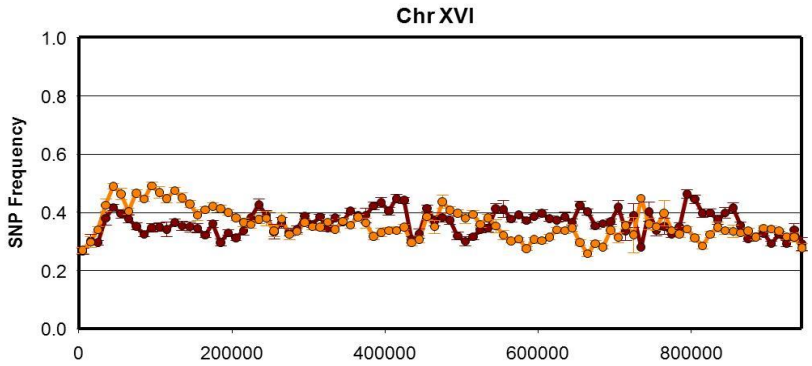


Fig. 5.5. Single nucleotide frequencies of M0341[CEL7A] SNPs across the pooled genomes. A value of 1 represents the M0341[CEL7A] genome, while a value of 0 represents the Y294[CEL7A] genome. Error bars represent the standard deviation of the SNP frequencies present within the same 5 kb genomic regions used to determine the average SNP frequencies.

From the large chromosomal regions of the QTLs, five ORFs that were previously reported to enhance protein secretion were identified. Of these, only the *ALG12* (QTL_XIV) and *GPA2* (QTL_V) alleles of the M0341 strain, coded for an altered protein product as compared to the S288C reference genome. No ORFs previously implied in enhancing protein secretion were found within QTL_III, QTL_VII and QTL_IV.

An average chromosome sequencing coverage of 34.7 was obtained, however some individual chromosomes deviated from this value by more than 10%. For both the pooled samples, Chr VI had lower sequence coverage, while Chr IX and Chr XI had a higher coverage. For the Y294[CEL7A] genome, chromosomes Chr I, Chr III and Chr VI had lower coverages. As with the Y294[CEL7A], M0341[CEL7A] also had a lower sequence coverage for Chr I; indicating possible anueploidy for these chromosomes.

Flow cytometry analysis of the DNA content revealed that Y294[CEL7A] had a genome resembling that of a diploid strain, while the M0341[CEL7A] strain's DNA content mostly represented a haploid strain's, although cells with more than haploid DNA content were detected (**Fig. 5.6**). The individual cells of the H3 strain had low DNA content variation and had a PI fluorescence signature of a triploid strain.

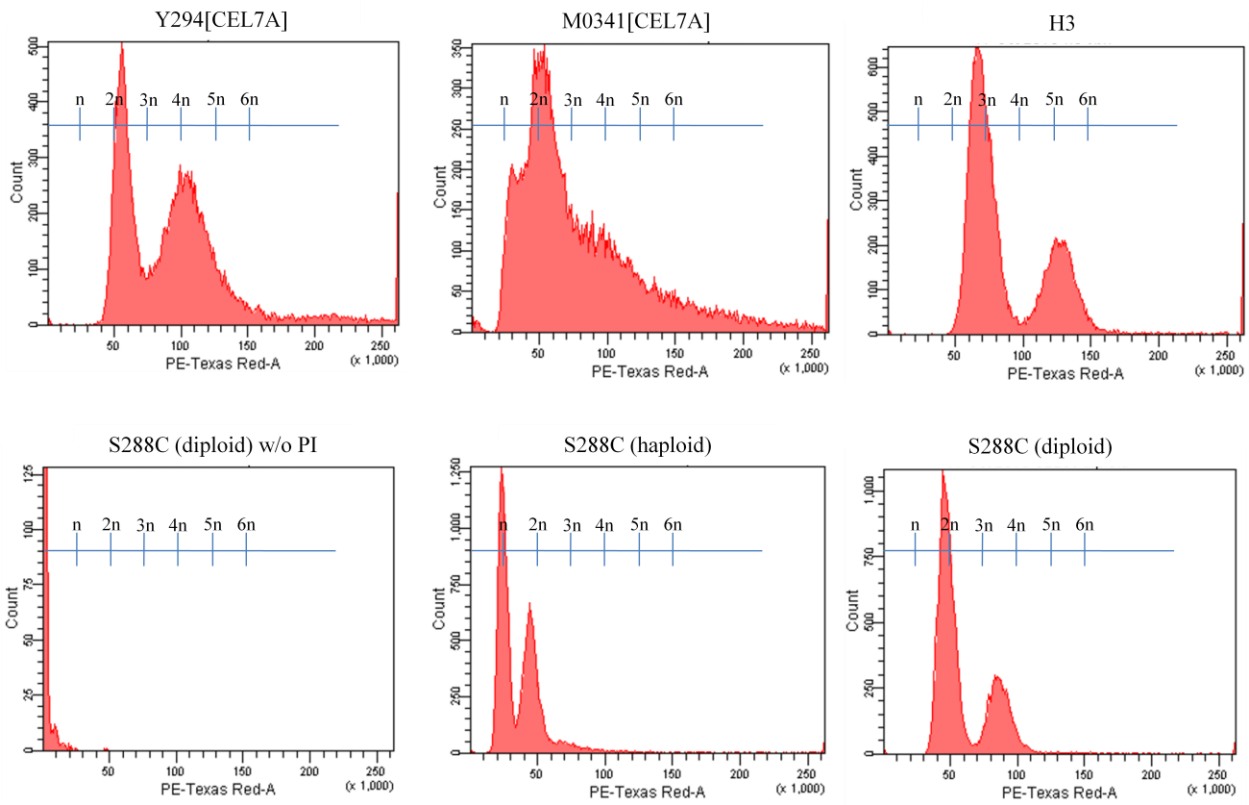


Fig. 5.6. Flow cytometry histograms of the parental stains and the H3 hybrid strain, depicting histograms of cell counts against relative propidium iodide fluorescence. Two peaks were obtained for the reference haploid (n) and diploid (2n) S288C strain; the first peak representing non-dividing cells and the second peak cells in the process of dividing that underwent karyokinesis, but not cytokinesis.

5.4.5. Evaluating high copy cellulase expressing strains

To study the high secretion phenotype of the F1 progeny, five strains were selected from the high CEL7A secretion pool after several rounds of subcultivation. The integrated *CEL7A*-cassette was removed and the non-functional Y294 *ura3* allele introduced into these strains to allow the expression of episomal expression plasmids using *URA3* as the auxotrophy complementation marker. Except for the H3K27+CEL3A which had ~30% lower activity relative to the Y294+CEL3A strain, no significant differences were observed in CEL3A activity per OD_{600} (**Fig. 5.7**). Similarly, no significant difference in CEL12A supernatant activity was observed, except for the H3M1+CEL12A strain. Three of the F1 progeny, namely H3H29, H3M1 and H3M28 displayed enhanced supernatant cellobiohydrolase activity when *CEL7A* was expressed in high copy number from the pRDH224 plasmid.

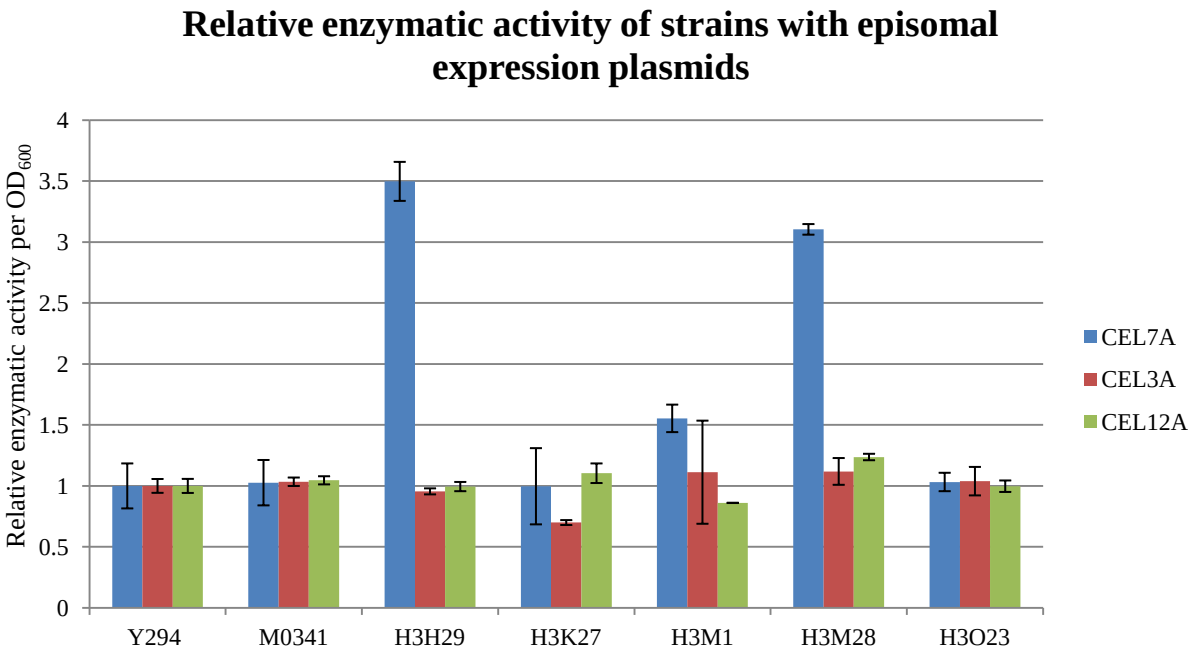


Fig. 5.7. The relative extracellular enzyme activity per optical density unit of strains expression *CEL3A*, *CEL7A* or *CEL12A* from episomal plasmids. All activities were expressed relative to the activities of the Y294 strain. Error bars indicate the standard deviation from the mean value of quadruplicate samples.

5.5. Discussion

Over the last three decades, it became apparent through strain engineering studies that the high protein secretion in yeast is a polygenic phenotype, with an ever expanding list of alleles being implicated in this phenotype (Idiris et al. 2010). Pooled-segregant whole-genome sequencing analysis, relies on genome shuffling during meiosis and is able to narrow down the genomic locations of multiple alleles contributing to the same phenotype (Parts et al. 2011). To this end we evaluated putative parental strains based on their ability to secrete *Talaromyces emersonii* CEL7A, expressed in *S. cerevisiae* from a single, integrated copy. A clear distinction was observed between the M0341[CEL7] strain, derived from an industrial bio-fuel strain, and the evaluated laboratory strains (**Fig. 5.1**). The Y294[CEL7] strain had more than three times lower supernatant cellobiohydrolase activity than the M0341[CEL7] and was subsequently selected for the control parent strain. The H3 hybrid strain, the product of mating the Y294[CEL7A] and M0341[CEL7A], had an intermediate supernatant CEL7A activity compared to the diploid versions of the parental strains (**Fig. 5.2**). This intermediate level of secretion suggested that (1) there were at least two M0341 alleles conferring high secretion and (2) at least one of these

alleles has a dominant phenotype, since the H3 strain had a higher supernatant activity than the Y294 strain. Similar to the study by Egel-Mitani et al. (2000), no activity enhancements were obtained by altering the ploidy of the parental strains, even though these strains contained double the *CEL7A* copies per cell.

Another indication of multiple alleles being responsible for the high secretion of the M0341[CEL7A] strain was revealed by the enzyme activities of the F1 progeny. The gradual increase from those that were poor producers to those with activities more than twice of the M0341[CEL7A] secreting parent strain, is indicative of multiple small enhancements that contribute to the phenotype (Demeke et al. 2013b). The selected 28 high secretion segregants all had higher supernatant activities compared to the M0341[CEL7A] strain. Whether this phenomenon was a result of aneuploidy, acquisition of beneficial genes from the Y294[CEL7A] or the removal of a genomic region, which had a negative impacted on secretion of M0341[CEL7A], warrants further investigation.

The genome sequencing and subsequent SNP detection revealed that ~70-80% of the pooled segregants' genomes originated from the Y294[Cel7A] strain, in stark contrast to the expected 1:1 segregation of chromosomes during meiosis in a diploid H3 strain. Interestingly, an unusual 1:3-ratio was observed for the mating types of the F1 segregants. We postulated that this ratio might indicate an increased ploidy state of the Y294[CEL7A] strain, as this would explain both deviations in the mating type and SNP ratios, since more genomic DNA from the Y294 strain would be distributed to the progeny, as demonstrated in the diagram below (**Fig. 5.8**). This also correlates with the high number of nonviable spores and diploid cells detected after sporulation, which was originally contributed to inefficient lysis of vegetative H3 cells (St Charles et al. 2010).

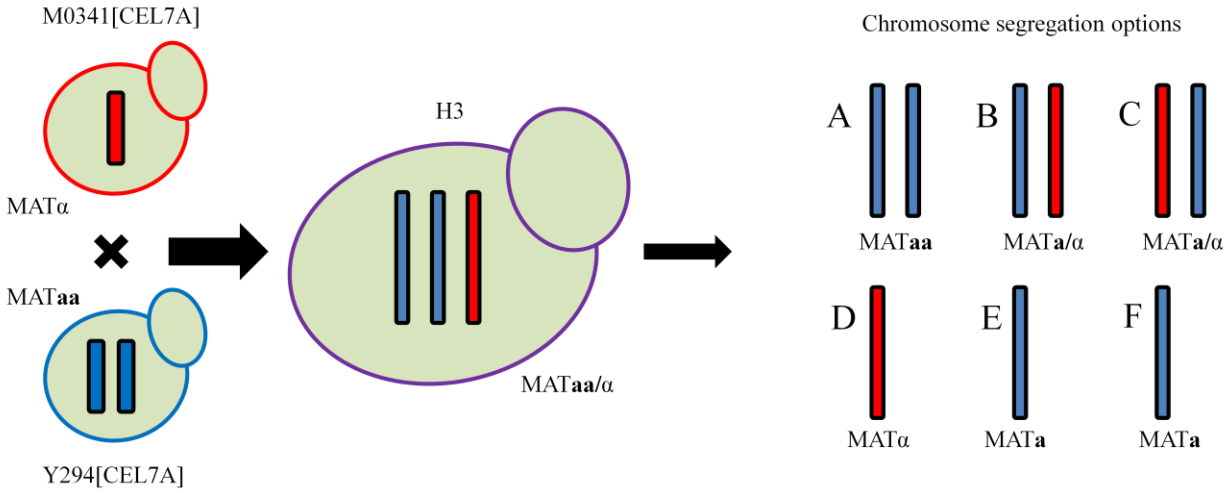


Fig. 5.8. Diagram illustrating the segregation of chromosomes of triploid strains. Chr III was used as an example to illustrate the mating type ratios and for simplicity, crossover events were excluded. Progeny with the chromosome combinations A, D, E and F would all test haploid with a PCR screen, with only D being of the MATα type. If the genomic DNA quantity is taken into account, there would be on average four Y294 chromosomes per M0341 chromosome in the unselected pool, since the MAT**a/a** contribute two chromosomes per cell.

This theory is consistent with the results obtained with flow cytometry, showing an approximate two-fold increase of the Y294[CEL7A] DNA content, a high number of the selected high secretion segregants presenting as aneuploids and the H3 histogram resembling that of a near triploid strain (**Fig. 5.6**). The flow cytometry histogram of the M0341[CEL7A] stain revealed a sloping shoulder which might indicate the presence of an unstable population of aneuploid cells or cell aggregation. In addition to the variation in the ploidy of the F1 progeny, indications of partial aneuploidy were seen for Chr VI:13kb-30kb and Chr XI:512kb-667kb in both pooled gDNA samples. This aberrant ploidy state of the parental strains, especially that of the Y294[CEL7A] was probably due to prolonged exposure to SMM during selection, since the Y294 strain without the CEL7A-SMR cassette had a similar PI-fluorescence profile as that of the S288C haploid control (**Addendum C, Fig. C4**). An aberration of ploidy state in *S. cerevisiae* after EMS-induced random mutagenesis was previously observed (Demeke et al. 2013a). Supporting our findings, Jia et al. (2000) evaluated the effect of different concentrations of SMM on yeast global gene expression levels, detecting significant up-regulation of genes involved in DNA repair. The chromosomal instability of these aneuploid strains helps to explain the dramatic reduction in the supernatant activity of the high CEL7A secreting segregants after several sub-cultivations (St Charles et al. 2010).

This genomic variability of the segregants hampered the detection of minor QTLs, but five major QTLs could still be identified. These QTLs covers large chromosomal regions, requiring fine mapping with SNP specific primers to more accurately pinpoint the causative alleles, however this was not a viable option due to the aneuploid nature of the segregants. Mass mating of the segregants, to reduce the QTL lengths, also proved problematic due to the low frequency of MATα progeny and the enumeration of diploid cells. Although, the alleles conferring enhanced secretion could not be identified, an easily automated genome analysis pipeline was developed that can be used for future whole genome sequencing work. A flow diagram of the pipeline is illustrated in **Addendum C, Fig. C1**.

Using this pipeline, unique ORFs were identified in both parental strains and the ORFs present in the identified QTLs. The unique Y294[CEL7A] ORFs included several protein targeting (*TIM18*), folding (*EMC6*) and vacuolar transport genes (*VTH1* and *VTH2*), however only *HOR7*, an osmotic stress induced allele coding for a protein of unknown function, was previously reported to enhance secretion (Wang et al. 2013). No previously reported secretion genes were detected in the unique M0341[CEL7A] ORFs. It is however clear that the functional M0341[CEL7A] *FLO8* allele (functional) was able to contribute to an enhanced secretion phenotype. Six genes within the QTLs were previously reported to enhance secretion, although only the M0341 *ALG12* (QTL_XIV) and *GPA2* (QTL_V) alleles coded for altered protein products compared to the Y294 alleles.

Although there was a dramatic reduction in the supernatant activity of the high CEL7A secreting segregants after several sub-cultivations, 66% of the pool still had higher secretion values than that of the Y294[CEL7A] strain (**Fig. 5.3B**). In order to further elucidate the selected high secretion phenotype, we introduced high copy number episomal plasmids expressing the CEL7A, CEL3A or CEL12A. Interestingly, we found that the selection for alleles allowing high CEL7A, did not result in a general enhancement of the strains' ability to secrete heterologous proteins. It did however, improve the secretion of the episomally expressed CEL7A, even though the selection was done for single copy integrated CEL7A, suggesting the involvement of protein (and not expression level) specific mechanisms for this phenotype.

5.6. Conclusion

Strain engineering is a promising strategy to enhance protein secretion in yeast. This strategy is however, limited by the reporter protein specific characteristics of previously identified gene targets. Five genomic regions of a high CEL7A secretion strain were identified to contain putative secretion enhancing alleles, although we were unable to pinpoint these enhancing elements due to unstable yeast ploidy. By evaluating the major phenotypic differences of the parental strains, we identified flocculation as a potential secretion enhancer, albeit a small contributing factor to the M0341 high secretion phenotype. These enhancing elements also proved to be CEL7A specific and could enhance its secretion if expressed on high copy plasmids. In this study we also established an easily automated bioinformatics pipeline for future genomic analysis of important polygenic phenotypes and touched on the potential of these genomic approaches in rational yeast design.

5.7. Acknowledgements

The authors would like to thank Prof. Johan M. Thevelein and Dr. María R. Foulquié-Mornero (KU Leuven, Belgium) for insightful discussions about the genomic analysis and the National Research Fund (South Africa) for financial support.

5.8. References

Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic Local Alignment Search Tool. *J. Mol. Biol.* 215: 403–410.

Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P & Boeke JD (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C : a Useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast 132: 115–132.

Demeke MM, Dietz H, Li Y, Foulquié-Moreno MR, Mutturi S, Deprez S, Den Abt T, Bonini BM, Liden G, Dumortier F, Verplaetse A, Boles E & Thevelein JM (2013) Development of a D-xylose fermenting and inhibitor tolerant industrial *Saccharomyces cerevisiae* strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering. *Biotechnol. Biofuels* 6: 89-113.

Demeke MM, Dumortier F, Li Y, Broeckx T, Foulquié-Moreno MR & Thevelein JM (2013) Combining inhibitor tolerance and D-xylose fermentation in industrial *Saccharomyces cerevisiae* for efficient lignocellulose-based bioethanol production. *Biotechnol. Biofuels* 6: 120-137.

Den Haan R, Rose SH, Lynd LR & van Zyl WH (2007) Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. *Metab. Eng.* 9: 87–94.

Duitama J, Srivastava PK & Măndoiu II (2012) Towards accurate detection and genotyping of expressed variants from whole transcriptome sequencing data. *BMC Genomics* 13 Suppl 2: S6.

Egel-Mitani M, Andersen A, Diers I, Hach M, Thim L, Hastrup S & Vad K (2000) Yield improvement of heterologous peptides expressed in *yps1*-disrupted *Saccharomyces cerevisiae* strains. *Enzyme Microb. Technol.* 26: 671–677.

Hill J, Donald KAG & Griffiths DE (1991) DMSO-enhanced whole cell yeast transformation. *Nucleic Acids Res.* 19: 6688.

Idiris A, Tohda H, Kumagai H & Takegawa K (2010) Engineering of protein secretion in yeast: strategies and impact on protein production. *Appl. Microbiol. Biotechnol.* 86: 403–417.

Ilmén M, den Haan R, Brevnova E, McBride J, Wiswall E, Froehlich A, Koivula A, Voutilainen SP, Siika-Aho M, la Grange DC, Thorngren N, Ahlgren S, Mellon M, Deleault K, Rajgarhia V, van Zyl WH, Penttilä M (2011) High level secretion of cellobiohydrolases by *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 4: 30. Jia MH, Larossa RA, Lee J, Rafalski A, Derose E, Gonye G & Xue Z (2000) Global expression profiling of yeast treated with an inhibitor of amino acid biosynthesis , sulfometuron methyl. *Physiol. Genomics* 3: 83–92.

Kroukamp H, den Haan R, van Wyk N & van Zyl WH (2013) Overexpression of native *PSE1* and *SOD1* in *Saccharomyces cerevisiae* improved heterologous cellulase secretion. *Appl. Energy* 102: 150–156.

Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9: 357–359.

Li R, Yu C, Li Y, Lam T-W, Yiu S-M, Kristiansen K & Wang J (2009) SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25: 1966–1967.

Lynd LR, van Zyl WH, McBride JE & Laser M (2005) Consolidated bioprocessing of cellulosic biomass: an update. *Curr. Opin. Biotechnol.* 16: 577–583.

Parts L, Cubillos F, Warringer J, Jain K, Salinas F, Bumpstead SJ, Molin M, Zia A, Simpson JT, Quail M, Moses A, Louis EJ, Durbin R, Liti G (2011) Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res.* 21: 1131–1138.

Popolo L, Vanoni M & Alberghina L (1982) Control of the yeast cell cycle by protein synthesis. *Exp. Cell Res.* 142: 69–78.

Sambrook J & Russel DB (2001) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Soares EV (2011) Flocculation in *Saccharomyces cerevisiae*: a review. *J. Appl. Microbiol.* 110: 1–18.

St Charles J, Hamilton ML & Petes TD (2010) Meiotic chromosome segregation in triploid strains of *Saccharomyces cerevisiae*. *Genetics* 186: 537–550.

Stanke M, Steinkamp R, Waack S & Morgenstern B (2004) AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* 32: 309–312.

Swinnen S, Schaerlaekens K, Pais T, Claesen J, Hubmann G, Yang Y, Demeke M, Foulquié-Moreno MR, Goovaerts A, Souvereyns K, Clement L, Dumortier F, Thevelein JM (2012) Identification of novel causative genes determining the complex trait of high ethanol tolerance in yeast using pooled-segregant whole-genome sequence analysis. *Genome Res.* 22: 975-984.

Thorvaldsdóttir H, Robinson JT & Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14: 178–192.

Treco DA and Winston F (2008) Growth and manipulation of yeast. *Curr. Protoc. Mol. Biol*. 82:13.2.1-13.2.12.

Van Dijken J. Bauer J, Brambilla L, Duboc P, Francois JM, Gancedo C. Giuseppin MLF, Heijnen JJ, Hoare M, Lange HC, Madden E, Niederberger P, Nielsen J, Parrou JL, Petit T, Porro D, Reuss M, van Riel N, Rizzi M, Steensma HY, Verrips CT, Vindeløv J & Pronk JT (2000) An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb. Technol.* 26: 706–714.

Van Rooyen R, Hahn-Hägerdal B, La Grange DC & van Zyl WH (2005) Construction of cellobiose-growing and fermenting *Saccharomyces cerevisiae* strains. *J. Biotechnol.* 120: 284–295.

Van Zyl WH, Lynd LR, den Haan R & McBride JE (2007) Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae*. *Adv. Biochem. Eng. Biotechnol.* 108: 205–235.

Voth WP, Olsen AE, Sbia M, Freedman KH & Stillman DJ (2005) *ACE2*, *CBK1*, and *BUD4* in budding and cell separation. *Eukaryot. Cell* 4: 1018–1028.

Wang B-D, Chen D-C & Kuo T-T (2001) Characterization of a *Saccharomyces cerevisiae* mutant with oversecretion phenotype. *Appl. Microbiol. Biotechnol.* 55: 712–720.

Wang T-Y, Huang CJ, Chen HL, Ho PC, Ke HM, Cho HY, Ruan SK, Hung KY, Wang IL, Cai YW, Sung HM, Li WH, Shih MC (2013) Systematic screening of glycosylation- and trafficking-associated gene knockouts in *Saccharomyces cerevisiae* identifies mutants with improved heterologous exocellulase activity and host secretion. *BMC Biotechnol.* 13: 71–85.

Zerbino DR (2010) Using the Velvet *de novo* assembler for short-read sequencing technologies. *Curr Protoc Bioinformatics*, pp. 1–13.

Chapter 6

General discussion and conclusions

6.1. General discussion and conclusions

The secretion pathway of *S. cerevisiae* can be described as the path through which a newly synthesized peptide is channeled to reach the cell membrane (Novick et al. 1981). This pathway starts after peptide synthesis is initiated with the translocation of a polypeptide into the lumen of the endoplasmic reticulum (ER), where protein folding is directed through the action of multiple chaperones and folding factors. Disulfide bridge formation and protein glycosylation is also crucial and functions within the reducing environment of the ER. Correctly folded proteins are transported via the COPII (coat protein) mediated transport vesicles requiring specialized vesicle fusion complexes in order to fuse and deliver their cargo to the Golgi apparatus. After maturation in the Golgi apparatus, vesicles of the late-secretion pathway deliver the mature proteins to the plasma membrane, where another set of vesicle targeting and fusion complexes facilitate their release into the environment. Although this general sequence of events has been elucidated, much is still unknown about this complex and interconnected pathway.

Microorganisms, especially yeasts, have been found to be prolific production hosts for the ever increasing demand for industrial and pharmaceutical proteins (Corchero et al. 2013; Çelik & Çalık 2012). To improve cell-specific productivity of a host organism, initial approaches focused on increasing the cellular abundance of a certain protein product, thus ensuring that more of this product is available for secretion (Nevoigt 2008). This increased abundance of synthesized protein is commonly achieved by ensuring efficient transcription along with subsequent efficient translation. Significant increases in heterologous protein titers have been obtained by codon optimization (Wiedemann & Boles 2008), increasing the gene copy number (Lopes et al. 1989) and using modified or strong constitutive promoters (Alper et al. 2005; Mumberg et al. 1995). The abundance of the intracellular product, however, may lead to secretion bottlenecks and cellular stress that impairs the secretion rate or induce protein degradation pathways (Mattanovich et al. 2004). It is postulated that protein secretion is often stalled during the quality control steps of protein folding and membrane crossing events, contributing to the observation that yeast secretory expression is 100- to 1000-fold lower than the theoretical protein yield potential (Schröder 2006).

Through systematic screening and semi-rational design of recombinant yeast strains, many genetic targets have been identified that were able to enhance the secretion of native and

heterologous proteins by relieving secretion bottlenecks. The secretion enhancing ability of many genetic alterations vary depending on the reporter protein characteristics, although the properties of the protein products that influence their transit through the secretion pathway and the pathways indirectly interacting with secretion machinery are still poorly understood (Robinson et al. 1994; Wentz & Shusta 2008; Kroukamp et al. 2013).

The main aim of this study was to evaluate the potential use of strain engineering approaches to improve the secreted yields of industrial relevant proteins, identify possible new genetic targets to apply as strain engineering tools and to elucidate some of the underlying mechanisms leading to protein secretion enhancement.

6.1.1. Recombinant protein production in yeast.

One of the distinctive characteristics of fungi is their ability to digest insoluble organic compounds by secreting hydrolases into the surrounding habitat and absorbing the solubilised nutrients. This characteristic of fungi to secrete proteins has led to the biotechnological exploitation thereof, to produce industrially and pharmaceutically relevant proteins with reduced expenses involving the purification of these protein products (Porro et al. 2005). Fungi, especially yeasts, have been demonstrated to be capable of producing an ever expanding array of commercially valuable proteins, natively and heterologously (Schmidt 2004; Macauley-Patrick et al. 2005; Porro et al. 2005; van Ooyen et al. 2006).

In answer to the growing demand for recombinant proteins, efficient vector systems and transformation methods were developed to produce foreign proteins in robust fast growing organisms (Porro et al. 2011). Even though the onset of molecular biology granted a rapid means to exploit host cells for the production of industrial and pharmaceutical proteins, expression of heterologous proteins in fungi (and other expression host) does not always result in reliable product yields. There are a number of factors that need to be considered in the design and development of efficient protein production systems, that includes the growth requirements of the production host, the efficient translation and transcription of the desired recombinant genes, the characteristics of the desired product, proper folding and posttranslational modifications of the protein of interest, prevention of product degradation and purification methods. In some cases, optimizing each of these factors individually has been shown to be effective; however, the

dynamic interaction between these factors poses recurring counteracting problems. By optimizing one aspect of protein production, another factor might be impaired. It is thus imperative to consider all these factors in combination to obtain maximal product yields.

Although the focus of this study was to evaluate the potential benefits of strain engineering for cellulase secretion and identifying beneficial alleles for secretion, the contribution of the growth media was evident with the secretion of the *Neocallimastix patriciarum* Cel6A in *S .cerevisiae* (**Chapter 3**). Even though the secretion of proteins is dependent on polarized growth (Martin & Arkowitz 2014; Liu et al. 2013), and thus higher yields were expected from cultures grown in complex media compared to minimal media, higher cell specific yields of Cel6A were obtained from cells grown on SC (synthetic complete) media. This phenomenon is probably linked to the reduced protein synthesis rate, allowing proteins to fold properly, similar to protein secretion enhancements for growth at low temperature (Hackel et al. 2006).

6.1.2. Strain engineering as a means to improve recombinant cellulase secretion in yeast.

Improving the rate and yield of secreted proteins has been the focus of many studies, and several strategies have been successfully applied to enhance secretion as mentioned in **section 6.1**. The selection of appropriate strong promoters and efficient secretion signals or the use of complex media for cultivation, drastically improved the production and secretion of certain proteins, and usually served as the first steps to improve protein yields. Taking into account the time consuming nature of strain engineering approaches and the lack of a generalized strategy for the improved production of most proteins of interest, the efficacy of such approaches could be questioned. Limited knowledge on the mechanisms at work and the protein specific nature displayed in many strain engineering strategies hampers implementation of strain engineering as a general production host engineering tool.

Despite the limitations of strain engineering as a means to improve secretion, it is becoming a frequently applied procedure, in combination with conventional approaches to improve product synthesis or when the initial methods mentioned previously fail to produce sufficient product yields. The current study showed the potential of strain engineering in the secretion of five different recombinant cellulases in *S. cerevisiae*, with noteworthy increases of more than 300% obtained for the *Saccharomycopsis fibuligera* β-glucosidase Cel3A (**Chapter 3**). It was also

shown that when some of these genetic alterations were combined, they were able to further enhance secreted yields compared to the individual overexpressions; even though none of the genetic alterations were directly linked to the secretion pathway. Strain engineering is of great importance for industrial strains, since general strategies such as high copy episomal plasmids and complex media is not always feasible methods to obtain high yields on an industrial scale.

Regarding strain engineering strategies to improve cellulase secretion for second generation biofuel production - some caution should be taken not to impair other desired characteristics of the production host when implementing any genetic alterations to enhance a desired phenotype, such as high secretion of cellulases. Alterations affecting biomass yield can for instance directly impact on ethanol yields, while membrane and cell wall changes could lower the robustness of the organism in industrial applications. These limitations are not limited to biofuel production and should be taken into consideration on a case by case basis.

6.1.3. Identifying genetic targets to apply as a strain engineering tools.

The rational design of high protein production hosts usually relies on some prior knowledge of possible bottlenecks within the secretion pathway or previous reports of beneficial effect instilled by the presence (or absence) of specific native gene products. With the secretion pathway as a guide, limitations in protein folding have been a prominent area of engineering, although transport vesicle fusion events, protein glycosylation and general cellular stress response genes are also frequently targeted. Some of these strategies are supported by physiological and transcription data of host organisms producing heterologous proteins of interest (Gasser et al. 2007), while others rely on large scale screening of gene libraries in order to identify gene alterations which benefit their protein of interest (Wentz & Shusta 2007). In recent years, the advances in the fields of genomics, transcriptomics and proteomics allow researchers to more accurately identify beneficial and deleterious effects of native gene overexpression or deletion.

Although classical breeding is not generally considered as strain engineering, the principle of adding beneficial factors to the desired host strain is still relevant and has the benefit of transferring multiple (potentially interdependent) genetic elements at once. It also adds an additional level of variation, by introducing gene allele variations into engineering approaches. These beneficial allelic differences between different strains of a potential protein production

organism, is a valuable, rarely explored source of genetic tools for future engineering strategies. Using segregation frequencies and whole genome sequencing, polygenic beneficial alleles can be simultaneously identified, circumventing the risks of conventional semi-rational design strategies when no prior data exists for a given protein product. The use of segregation frequencies is dependent on the availability of large strain collections with observable variation between strains in the level of secretion of the desired protein product and an accurate high throughput screening method for the protein of interest.

Although these strain engineering strategies have been mainly focusing on native genetic targets, or those of closely related organisms, it is crucial to consider the potential of introducing genes from unrelated organisms which could significantly expand the genetic tools available. Even in the intensively studied yeast *S. cerevisiae*, many gene products in its genome are yet to be assigned physiological functions and their role in secretion remain unexplored.

6.1.4. Elucidation of the underlying mechanisms of secretion enhancing genetic elements.

Most studies on the enhancement of protein production and secretion focused on the benefits of the application of the protein of interest and not on the elucidation of the molecular mechanisms responsible for the enhancement. It is imperative for the advancement of protein secretion studies, and its application for other protein products, that these mechanisms are identified. In order to accurately model cellular metabolism and more specifically the secretion pathway, the direct and indirect interactions between pathways must be apparent. It is relatively easy to envision the beneficial effect of reducing intracellular oxidative stress on the secretion pathway, but some mechanisms like that of the non-secretion related protein import/mRNA export by Pse1 are more elusive. Understanding these mechanisms and their impact on the secretion pathway could help explain the elusive protein specific nature that limits wide scale application of strain engineering strategies.

Although this study focused on evaluating and discovering strain engineering genetic tools, relevant to cellulase secretion, we were also able to implicate the involvement of the yeast cell wall in the secretion and cell wall adhesion of proteins in *N*-glycosylation mutants. This study was also able to identify flocculation as a novel secretion enhancing factor; however, the mechanism by which flocculation impacts secretion is a subject for future studies.

From the data presented in this study, the following can be concluded:

- Overexpression of the native *S .cerevisiae PSE1* gene led to an increase in secreted yields of recombinant *S. fibuligera* Cel3A and *Trichoderma reesei* Cel7B in *S. cerevisiae*.
- Overexpression of the native *S. cerevisiae SOD1* gene led to an increase in secreted yields of recombinant *S. fibuligera* Cel3A and *Neocallimastix patriciarum* Cel6A in *S. cerevisiae*.
- Overexpression of the native *S. cerevisiae PSE1* and *SOD1* genes in combination led to a synergistic increase in secreted yields of recombinant *S. fibuligera* Cel3A.
- Deletion of the native *S. cerevisiae MNN2* and *PMR1* genes led to an increase in secreted recombinant *Neosartorya fischeri* Cel12A in *S. cerevisiae.*
- Heterologous protein secretion can be increased by altering the cell wall integrity of the yeast *S .cerevisiae.*
- Deletion of the native *S. cerevisiae MNN10*, *MNN11* and *PMR1* led to an increase in secreted recombinant *S. fibuligera* Cel3A in *S. cerevisiae.*
- Deletion of the native *S. cerevisiae OCH1*, *MNN9*, *MNN10* and *MNN11* led to a decrease in cell wall associated recombinant *S. fibuligera* Cel3A and decreased cell wall integrity in *S. cerevisiae*.
- Deletion of the native *S. cerevisiae PMR1* gene conferred sensitivity to cell wall disrupting agents, but not cell wall integrity dependent inhibitors.
- The unfolded protein response was not required for the production of *S. fibuligera* Cel3A or *N. fischeri* Cel12A in the *S. cerevisiae* BY4742 strain.
- Deletion of the genes encoding Golgi resident mannosyltransferases, responsible for outer *N*-glycan chain elongation, resulted in altered cell morphology and optical density characteristics of the *S. cerevisiae* strains.
- The *S. cerevisiae* M0341 and Y294 strains differed significantly on genomic level.
- High protein secretion in *S. cerevisiae* is a polygenic phenotype.
- The presence of flocculation in *S. cerevisiae* enhanced the secretion of recombinant *Talaromyces emersonii* Cel7A in *S. cerevisiae.*
- Increasing the ploidy of recombinant strains did not significantly increase the secretion of heterologous *T. emersonii* Cel7A in *S. cerevisiae.*

- The *S. cerevisiae* M0341 strain had five selected chromosomal regions that were beneficial for the secretion of heterologous *T. emersonii* Cel7A.
- Secretion of low copy number and overexpressed heterologous *T. emersonii* Cel7A in *S. cerevisiae* was enhanced by classical breeding strategies.
- Improvement of cellulase secretion in *S .cerevisiae* is reporter protein specific.

In this study we have shown the potential of strain engineering as a molecular tool to enhance heterologous cellulase secretion in the yeast *S. cerevisiae.* We identified a previously unreported secretion enhancer and demonstrated the influence of the cell wall on protein secretion. We created *S. cerevisiae* strains with significantly improved secretion of each of the major cellulase groups required for crystalline cellulose utilization, which represents a step toward realizing cost-effective second generation biofuel production.

6.1.5. References

Alper H, Fischer C, Nevoigt E & Stephanopoulos G (2005) Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci. U. S. A.* 102: 12678–12683.

Çelik E & Çalık P (2012) Production of recombinant proteins by yeast cells. *Biotechnol. Adv.* 30: 1108–1118.

Corchero JL, Gasser B, Resina D, Smith W, Parrilli E, Vázquez F, Abasolo I, Giuliani M, Jäntti J, Ferrer P, Saloheimo M, Mattanovich D, Schwartz S Jr, Tutino ML & Villaverde A (2013) Unconventional microbial systems for the cost-efficient production of high-quality protein therapeutics. *Biotechnol. Adv.* 31: 140–153.

Gasser B, Maurer M, Rautio J, Sauer M, Bhattacharyya A, Saloheimo M, Penttilä M & Mattanovich D (2007) Monitoring of transcriptional regulation in *Pichia pastoris* under protein production conditions. *BMC Genomics* 8: 179-197.

Hackel BJ, Huang D, Bubolz JC, Wang XX & Shusta EV (2006) Production of soluble and active transferrin receptor-targeting single-chain antibody using *Saccharomyces cerevisiae*. *Pharm. Res.* 23: 790–797.

Kroukamp H, den Haan R, van Wyk N & van Zyl WH (2013) Overexpression of native *PSE1* and *SOD1* in *Saccharomyces cerevisiae* improved heterologous cellulase secretion. *Appl. Energy* 102: 150–156.

Liu Z, Hou J, Martínez JL, Petranovic D & Nielsen J (2013) Correlation of cell growth and heterologous protein production by *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 97: 8955–8962.

Lopes TS, Klootwijk J, Veenstra AE, Van der Aar PC, Van Heerikhuizen H, Raue HA & Planta RJ (1989) High-copy-number integration into the ribosomal DNA of *Saccharomyces cerevisiae*: A new vector for high-level expression. *Gene* 79: 199–206.

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

Martin SG & Arkowitz R (2014) Cell polarization in budding and fission yeasts. *FEMS Microbiol. Rev.* 38: 228–253.

Mattanovich D, Gasser B, Hohenblum H & Sauer M (2004) Stress in recombinant protein producing yeasts. *J. Biotechnol.* 113: 121–135.

Mumberg D, Müller R & Funk M (1995) Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* 156: 119–122.

Nevoigt E (2008) Progress in metabolic engineering of *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* 72: 379–412.

Novick P, Ferro S & Schekman R (1981) Order of events in the yeast secretory pathway. *Cell* 25: 461–469.

Porro D, Gasser B & Fossati T (2011) Production of recombinant proteins and metabolites in yeasts. *Metab. Eng.* 939–948.

Porro D, Sauer M, Branduardi P & Mattanovich D (2005) Recombinant protein production in yeasts recombinant protein production in yeasts. *Mol. Biotechnol.* 31: 245–259.

Robinson AS, Hines V & Wittrup KD (1994) Protein disulfide isomerase overexpression increases secretion of foreign proteins in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 12: 381–384.

Schmidt FR (2004) Recombinant expression systems in the pharmaceutical industry. *Appl. Microbiol. Biotechnol.* 65: 363–372.

Schröder M (2006) The unfolded protein response. *Mol. Biotechnol.* 34: 279–290.

Van Ooyen AJJ, Dekker P, Huang M, Olsthoorn MM, Jacobs DI, Colussi P & Taron CH (2006) Heterologous protein production in the yeast *Kluyveromyces lactis*. *FEMS Yeast Res.* 6:381-392.

Wentz AE & Shusta EV (2007) A novel high-throughput screen reveals yeast genes that increase secretion of heterologous proteins. *Appl. Environ. Microbiol.* 73: 1189–1198.

Wentz AE & Shusta EV (2008) Enhanced secretion of heterologous proteins from yeast by overexpression of ribosomal subunit RRP0. *Biotechnol. Prog.* 24: 748–756.

Wiedemann B & Boles E (2008) Codon-optimized bacterial genes improve L-arabinose fermentation in recombinant *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 74: 2043–2050.

Addendum

Addendum A: Results not included in previous chapters

Aim:

To evaluate the effect of strain engineering on the secretion of recombinant *Talaromyces emersonii Cel7A* and native invertase secretion in *Saccharomyces cerevisiae*.

Materials and methods:

Saccharomyces cerevisiae BY4742 strains with deletions (Δ) for *SOD1*, *HAC1*, *DDI1*, *MNN10* and *MNN11* were transformed to express the *T. emersonii Cel7A* from the episomal plasmid pRDH224 (Ilmén et al. 2011) and was selected on synthetic complete media, without uracil. The isogenic BY4742 strain was transformed with the pRDH224 and pRDH122 (plasmid backbone with no insert; Ilmén et al., 2011), serving as the reference secretion (BY[CBH]) and assay control (BY[REF]) strains, respectively. Native *PSA1*, *SOD1*, *PSE1*, *SEC1*, mature *HAC1*, *SSO1* and *SSO2* were individually overexpressed in the BY[CBH] strain using the pBHD1 vector, as described in Kroukamp et al. (2013). The *FUR1* gene of all the episomal plasmid containing strains was disrupted to allow autoselection of the plasmid in the absence of selective pressure. Descriptions of the native genes that were overexpressed or disrupted can be found on *Saccharomyces* Genome Database (http://www.yeastgenome.org/).

Strains were cultivated for three days at 30°C in shake flasks containing 10 mL YP (1% yeast extract, 2% peptone powder, Merck) with either 2% glucose (YPD) or 2% galactose (YPG). Supernatants of YPD grown cultures were assayed using *p*NPC as substrate to determine the secreted Cel7A activities, as previously described by Ilmén et al. (2011). Reducing sugar assys were preformed to determine supernatant invertase (native Suc2) activities as previously described (Kroukamp et al. 2013), using sucrose as substrate instead. Enzyme activities were expressed as enzymatic units per culture OD_{600}, relative to the BY[CBH] strain.

Results and discussion:

Small but significant improvements in Cel7A secretion of up to 1.28-fold, was observed for the SOD1[CBH] (p=0.033), SSO1[CBH] (p=0.004) and DDI1[CBH] (p=0.017), after 72 hours of cultivation (**Fig. A1**). Although the mnn10Δ[CBH] and mnn11Δ[CBH] strains had higher volumetric activity values (data not shown) than the BY[CBH] strain, it should be taken into account that these strains had lower OD600 to cell dry weight ratios (refer to Chapter 4) and can't be directly compared to the OD_{600} normalized BY[CBH] strain. The *MNN*-deleted strains also enhanced the supernatant activities of the secreted native invertase (**Fig. A2**). All the recombinant strains had similar OD_{600} values, with the exception of mnnΔ10[CBH], mnnΔ11[CBH] and sod1Δ[CBH]. The *SOD1* deletion strain had a growth defect, resulting in artificially high cell specific activities due to low culture optical densities. No significant change in secreted Cel7A were observed by either inducing (HAC1i[CBH1]; p=0.06) or inhibiting (hac1Δ[CBH]; p=0.47) the unfolded protein response (UPR). This results is in line with previous studies, where only weak induction of the UPR (based on *HAC1* mRNA splicing) was observed in strains producing the *T. emersonii* Cel7A, compared to other enzymes of the same family (Ilmén et al. 2011).

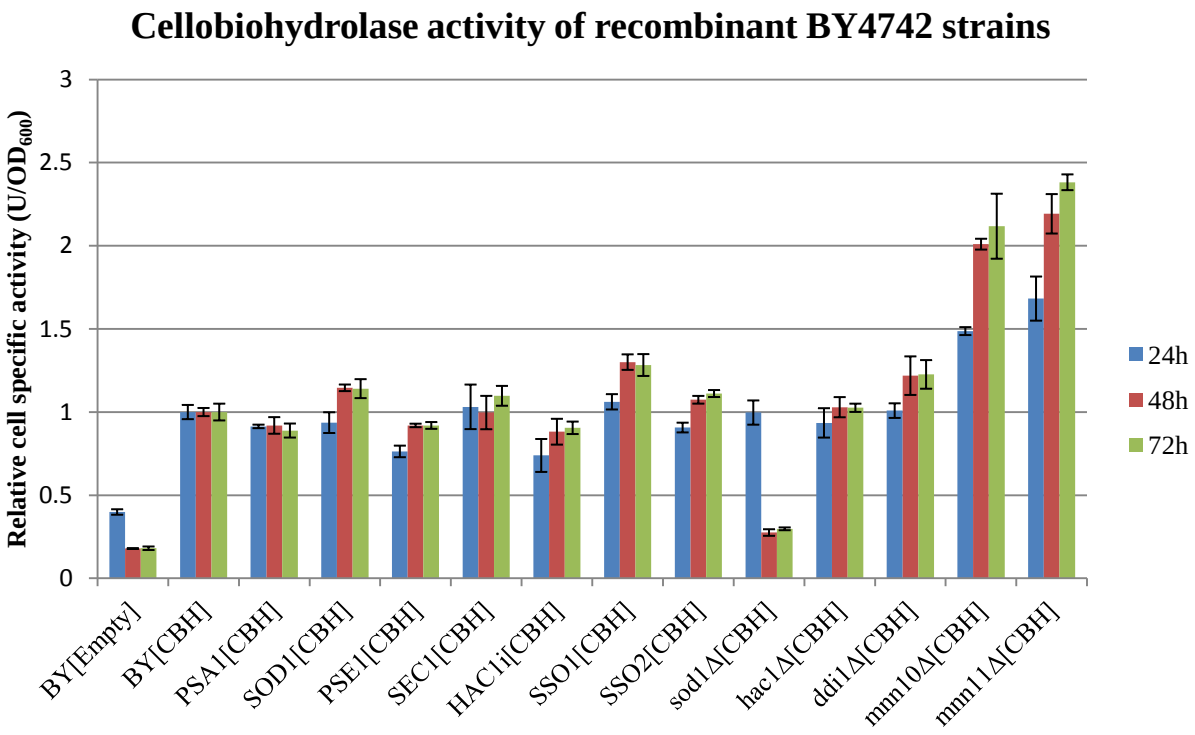


Fig. A1. The relative cell specific Cel7A activities of the recombinant yeast strains. All values are relative to the BY[CBH] strain. Values shown are the mean values of triplicate biological repeats and the error bars indicate standard deviation.

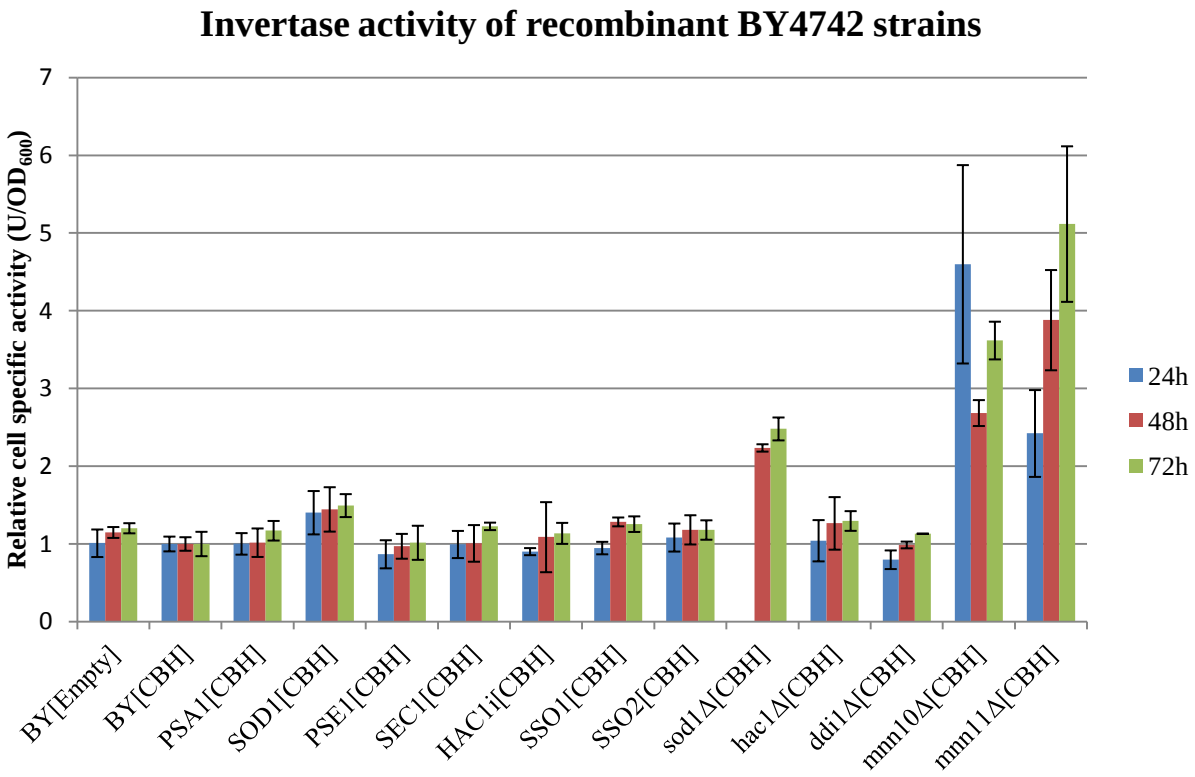


Fig. A2. The relative cell specific invertase activity of the recombinant yeast strains. All values are relative to the BY[CBH] strain. Values shown are the mean values of triplicate biological repeats and the error bars indicate standard deviation.

References

Ilmén M, den Haan R, Brevnova E, McBride J, Wiswall E, Froehlich A, Koivula A, Voutilainen SP, Siika-Aho M, la Grange DC, Thorngren N, Ahlgren S, Mellon M, Deleault K, Rajgarhia V, van Zyl WH & Penttilä M (2011) High level secretion of cellobiohydrolases by *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 4: 30.

Kroukamp H, den Haan R, van Wyk N & van Zyl WH (2013) Overexpression of native *PSE1* and *SOD1* in *Saccharomyces cerevisiae* improved heterologous cellulase secretion. *Appl. Energy* 102: 150–156.

Addendum B: Supplimentary data of Chapter 4

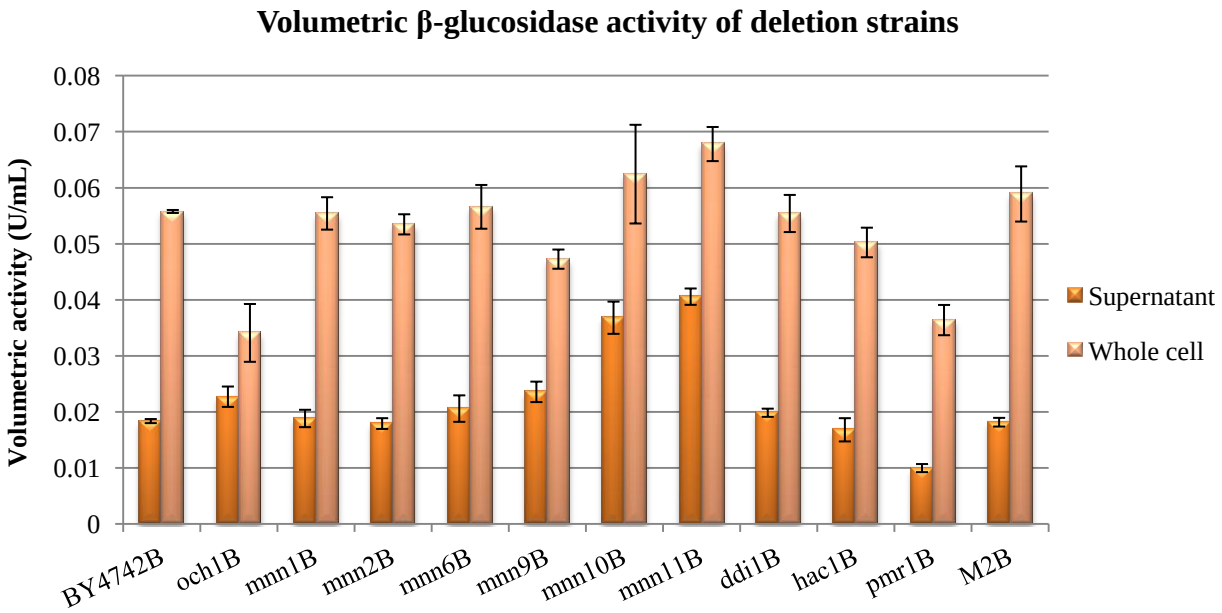


Fig. B1. The volumetric supernatant and total cell β-glucosidase activity of the evaluated deletion strains. ddi1B, containing a *DDI1* deletion that was previously described as a conditional secretion enhancer, was added as a control strain for cellular growth purposes. Although no positional variation in values was observed in prior quality control screens of the 24-well plates, M2B (is identical to the mnn2B strain) was added as variation control between plates. Values shown are the mean values of four biological repeats and the error bars indicate standard deviation.

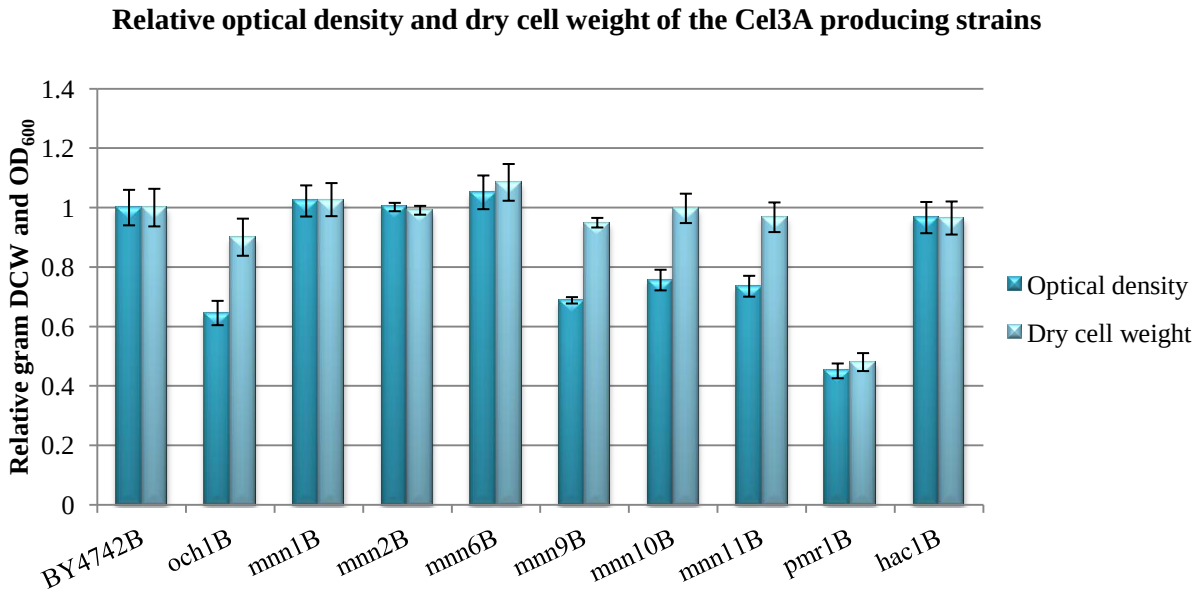


Fig. B2. The relative optical density (at OD of 600nm) and dry cell weight (in g/mL) of the mutant strains, compared to the isogenic BY4742B. Values shown are the mean values of four biological repeats and the error bars indicate standard deviation.

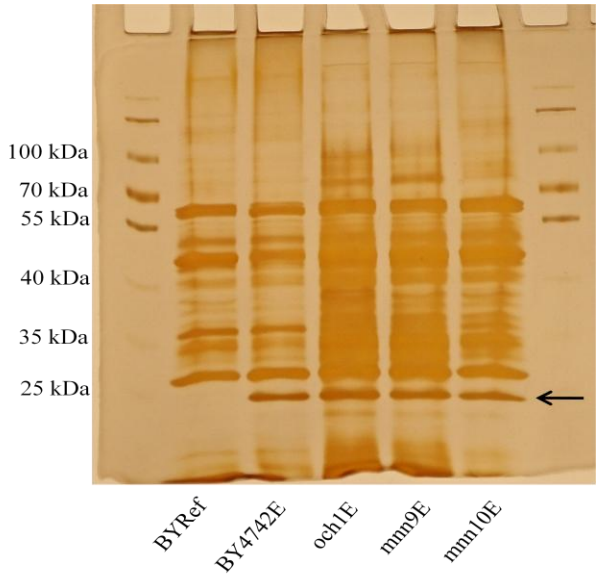


Fig. B3. SDS-PAGE analysis of selected glycosylation-deficient strain supernatants producing Cel12A. No change in Cel12A mobility was observed. (+) Denotes deglycosylated samples and (-) denotes untreated samples.

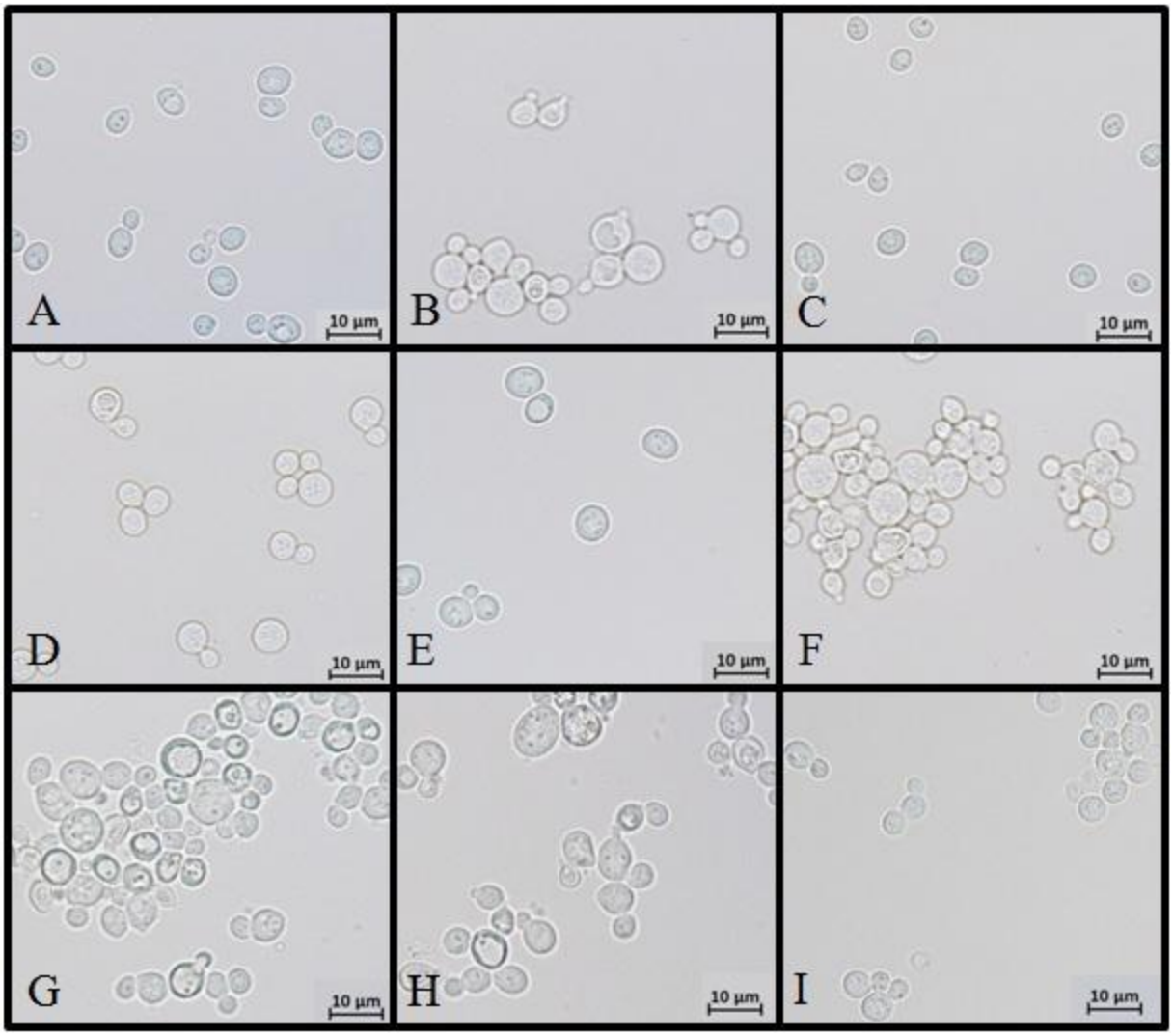


Fig. B4. Light microscope photos of the glycosylation deficient strains. (A) BY4742 (B) *och1Δ* (C) *mnn1Δ* (D) *mnn2Δ* (E) *mnn6Δ* (F) *mnn9Δ* (G) *mnn10Δ* (H) *mnn11Δ* (I) *pmr1Δ*. Enlarged cells and clumping is clearly visible in B, F, G, H. Clumping was also observed in the *pmr1Δ*, although no enlarged cells were observed.

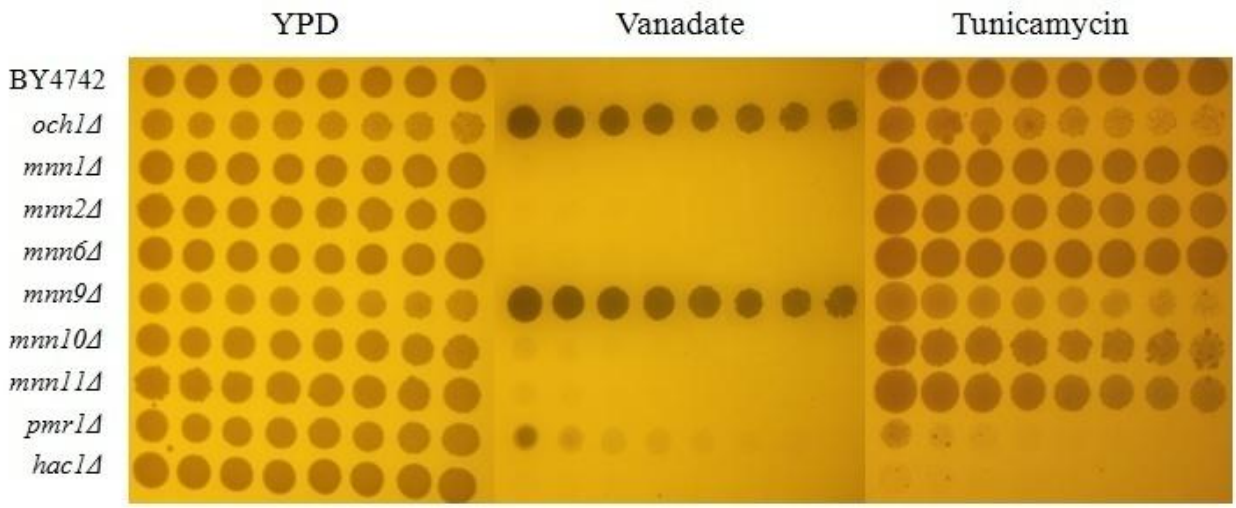


Fig. B5. Analysis of 2-fold serial dilutions of deletion strains in a plate viability assay. Normal cell viability is shown on YPD agar plates agter 24 hours, without inhibitor supplementation, for all strains evaluated. After 48 hours, only *och1Δ* and *mnn9Δ* was able to grow in the presence of 1.0 µg/mL Sodium vanadate. Similar strain viability was observed on 0.5 µg/mL tunicamycin containing plates (24 hrs), shown in **Fig. 4.7** and 1.0 µg/mL tunicamycin plates after 48 hrs.

Addendum C: Supplimentary data of Chapter 5

Table C1. Short description of the haploid yeast strains, to be used as potential parental strains.

S. cerevisiae strain	Background information of the yeast strain	Reference
BY4742	Based on the *S. cerevisiae* strain of which the genome was sequenced strain (S288c); used in the EUROSCARF yeast gene deletion library	(Brachmann et al. 1998)
W303	Widely used lab strain for the production of recombinant proteins and yeast physiology studies	(Voth et al. 2005)
Y294	Lab strain generated by random mutagenesis and selected for its high levels of secreted proteins	ATCC 201160
CEN.PK2	Selected as the overall best lab strain in a study comparing different lab strains on growth rate, transformation efficiency etc. (ref)	(van Dijken et al., 2000)
NI-C-D4	High secretion mutant resulting from mating the HBsAg-mutated NI-C strain (*ΔPEP4*) with wild type strain. The phenotype is linked with a mutation in the *OSA1* gene	(Wang et al., 2001)
MO341	Haploid version of a high secretion diploid industrial strain used for lignocellulosic ethanol production	(Ilmén *et al.*, 2011)

Table C2. List of primers used in this study.

Primer name	Function	Primer sequence (5'→3')
Conventional PCR		
interSMR_L	confirm *CEL7A*-cassette genomic position	TGACACGGTCGTCGAATCTA
A7708	confirm *CEL7A*-cassette genomic position	TTTTGTAATTTCGTGTCGTTTCTATTATGAATTTCAT
MAT**a**	Mating type determination	ACTCCACTTCAA GTAAGAGTTTG
MATα	Mating type determination	GCACGGAATATGGGACTACTTCG
MATlocus	Mating type determination	AGTCACATCAAGATCGTTTATGG
A8577	*FLO8* deletion	TATTTTTAATTCTTGTCACCAGTAAACAGAACATCCAA AAATGGGTACCACTCTTGACGACACGG
A8578	*FLO8* deletion	AGAAAAGATACACGATACGTAAAAAGAACGCGAATTT TATTAGGGGCAGGGCATGCTCCATG
Quantitative PCR		
TFC1 set1	Quantifying *TFC1* copy number	CCAACTTGGCCAACGAAGAT
		GAATTGGCACCTTCATCCCC
TAF1 set1	Quantifying *TAF1* copy number	ACTCCAGGCGGTATTGACAG
		TCTGCAATGTTTGGCTCACC
CBH1 set1	Quantifying *CEL7A* copy number	CCGACATCTCTGGTGTCACT
		ACCGTGTTGAGAGAAGTCGT

Bioinformatics pipeline

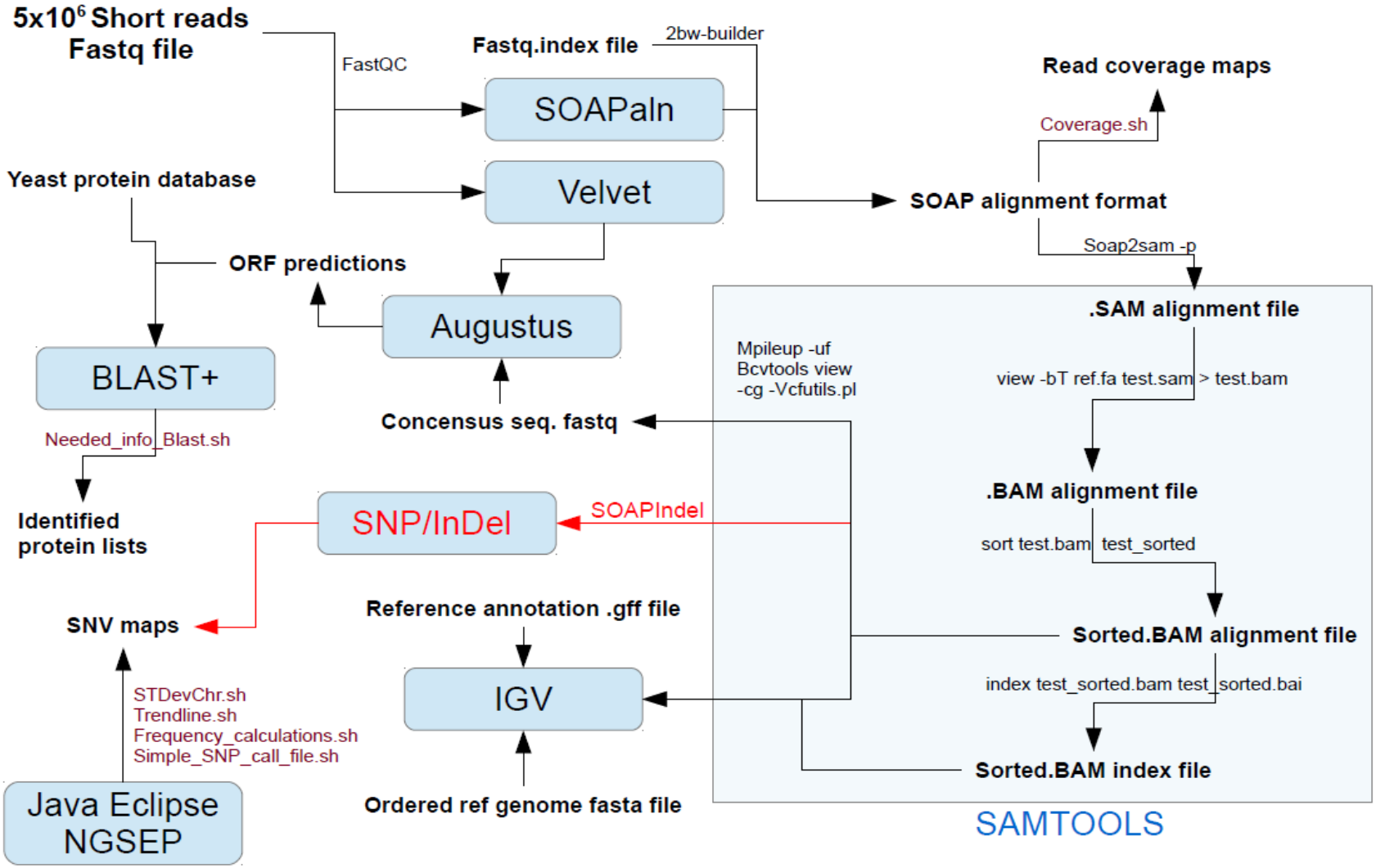


Fig. C1. (Previous page) Flow diagram of the bioinformatic pipeline used in this study. By incorporation of the SNPIndel tool (route in red), full automation could be achieved.

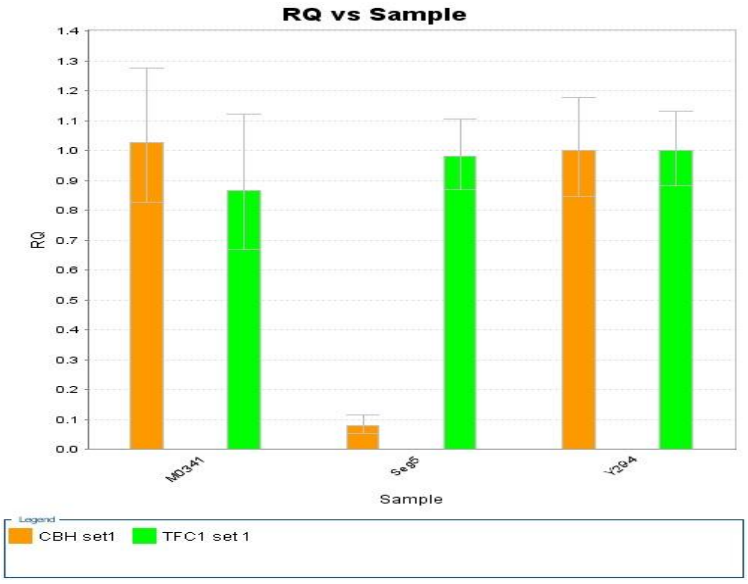


Fig. C2. Gene copy number per sample of either the CEL7A (CBH set 1) or TFC1 (TFC1 set 1), relative to TAF12 copy number. Only one copy of either TAF12 or TFC1 is expected per genome. The haploid Seg5 strain was included as negative control for the CEL7A

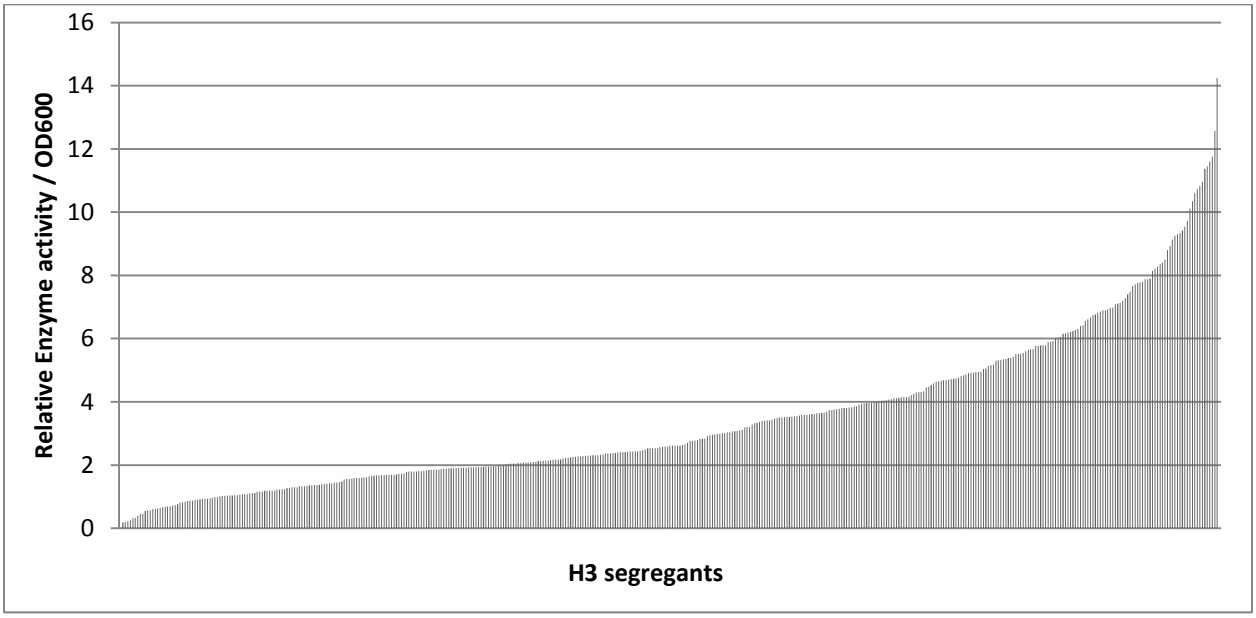


Fig. C3. Enzyme activities per OD_{600} (relative to the Y294[CEL7A] strain) of the 530 H3 segregants.

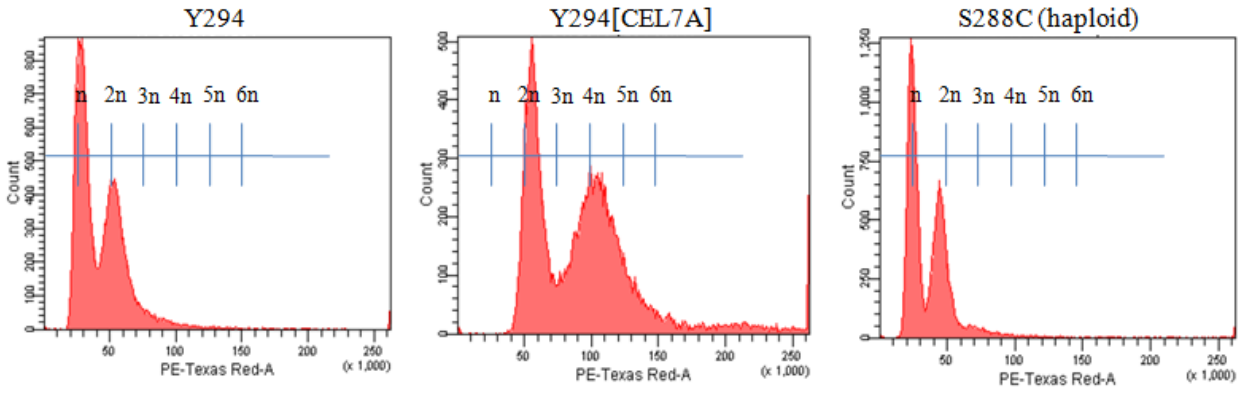


Fig. C4. Cell count against relative fluorescence histograms of the isogenic Y294 strain, the SMM selected Y294[CEL7A] strain and a S288C control haploid strain.

Fig. C5. .Examples of custom the scripts developed in this study.

***Name*:** Simple_SNP_Call_file.sh:

***Function*:** For extracting each base alteration, position and SNP coverage from the .vcf files, generated by the Java Eclipse NGSEP software.

***Possible improvement*:** Place in a 'for-loop' to reduce code length.

***Sample script for Chr I*:**

```
# Chromosome I

grep -w "chrI" Samplen71_.vcf | awk -F:  '{print $1 "\t" $10 "\t" $11}' | awk -F, '{print $1 "\t" $2 "\t" $3 "\t" $4}' | awk '{print $1 "\t" $2 "\t" $4 }' > Simplified71chrI &&

grep -w "chrI" Samplen72_.vcf | awk -F:  '{print $1 "\t" $10 "\t" $11}' | awk -F, '{print $1 "\t" $2 "\t" $3 "\t" $4}' | awk '{print $1 "\t" $2 "\t" $4 }' > Simplified72chrI &&

grep -w "chrI" Samplen73_.vcf | awk -F:  '{print $1 "\t" $10 "\t" $11}' | awk -F, '{print $1 "\t" $2 "\t" $3 "\t" $4}' | awk '{print $1 "\t" $2 "\t" $4 "\t" $5 "\t" $10 "\t" $11 "\t" $12 "\t" $13 "\t" $14}' > Simplified73chrI &&

grep -w "chrI" Samplen74_.vcf | awk -F:  '{print $1 "\t" $10 "\t" $11}' | awk -F, '{print $1 "\t" $2 "\t" $3 "\t" $4}' | awk '{print $1 "\t" $2 "\t" $4 "\t" $5 "\t" $10 "\t" $11 "\t" $12 "\t" $13 "\t" $14}' > Simplified74chrI &&

awk '{print $1 "\t" $2  "\t" $3}' Simplified71chrI | grep -F -f - Simplified73chrI > SupParSNP73chrI && awk '{print $1 "\t" $2  "\t" $3}' Simplified72chrI | grep -F -f - Simplified73chrI > InfParSNP73chrI &&

awk '{print $1 "\t" $2  "\t" $3}' Simplified71chrI | grep -F -f - Simplified74chrI > SupParSNP74chrI && awk '{print $1 "\t" $2  "\t" $3}' Simplified72chrI | grep -F -f - Simplified74chrI > InfParSNP74chrI &&

echo "Chromosome I done!"
```

***Name*:** Frequency_calculations.sh

***Function*:** Determining the parental origin of each SNP, counting SNP occurrence and calculating the SNP frequency using the base coverage data.

***Possible improvement*:** Place in a 'for-loop' to reduce code length.

***Sample script for Chr I*:**

```
# create headings file
echo -e 'Chr\tPosition\tRef\tAlt\tCoverage\tA\tC\tG\tT\tRef Count\tAlt Count\tTotal\tFrequency' >headings_temp &&
#chrI
# Print ref count
#echo "Ref Count" >RefCounttemp
awk '{if ($3 == "A") { print $6} else if ($3 == "C") {print $7} else if ($3 == "G") {print $8} else if ($3 == "T") {print $9} else {print "lol"}}' InfParSNP73chrI >> RefCountInfPar73chrI_temp &&
awk '{if ($3 == "A") { print $6} else if ($3 == "C") {print $7} else if ($3 == "G") {print $8} else if ($3 == "T") {print $9} else {print "lol"}}' InfParSNP74chrI >> RefCountInfPar74chrI_temp &&
awk '{if ($3 == "A") { print $6} else if ($3 == "C") {print $7} else if ($3 == "G") {print $8} else if ($3 == "T") {print $9} else {print "lol"}}' SupParSNP73chrI >> RefCountSupPar73chrI_temp &&
awk '{if ($3 == "A") { print $6} else if ($3 == "C") {print $7} else if ($3 == "G") {print $8} else if ($3 == "T") {print $9} else {print "lol"}}' SupParSNP74chrI >> RefCountSupPar74chrI_temp &&

# Print alt count
#echo "Alt Count" >AltCounttemp
awk '{if ($4 == "A") { print $6} else if ($4 == "C") {print $7} else if ($4 == "G") {print $8} else if ($4 == "T") {print $9} else {print "lol"}}' InfParSNP73chrI >> AltCountInfPar73chrI_temp &&
awk '{if ($4 == "A") { print $6} else if ($4 == "C") {print $7} else if ($4 == "G") {print $8} else if ($4 == "T") {print $9} else {print "lol"}}' InfParSNP74chrI >> AltCountInfPar74chrI_temp &&
awk '{if ($4 == "A") { print $6} else if ($4 == "C") {print $7} else if ($4 == "G") {print $8} else if ($4 == "T") {print $9} else {print "lol"}}' SupParSNP73chrI >> AltCountSupPar73chrI_temp &&
awk '{if ($4 == "A") { print $6} else if ($4 == "C") {print $7} else if ($4 == "G") {print $8} else if ($4 == "T") {print $9} else {print "lol"}}' SupParSNP74chrI >> AltCountSupPar74chrI_temp &&
```

```
paste -d'\t' RefCountInfPar73chrI_temp AltCountInfPar73chrI_temp > InfParSNP73chrICalctemp && # add Ref and Alt SNP count to same file
paste -d'\t' RefCountInfPar74chrI_temp AltCountInfPar74chrI_temp > InfParSNP74chrICalctemp && # add Ref and Alt SNP count to same file
paste -d'\t' RefCountSupPar73chrI_temp AltCountSupPar73chrI_temp > SupParSNP73chrICalctemp && # add Ref and Alt SNP count to same file
paste -d'\t' RefCountSupPar74chrI_temp AltCountSupPar74chrI_temp > SupParSNP74chrICalctemp && # add Ref and Alt SNP count to same file

awk '{print $1+$2}' InfParSNP73chrICalctemp > calc1Inf73chrItemp && # calculate total coverage (sum ref and alt)
awk '{print $1+$2}' InfParSNP74chrICalctemp > calc1Inf74chrItemp && # calculate total coverage (sum ref and alt)
awk '{print $1+$2}' SupParSNP73chrICalctemp > calc1Sup73chrItemp && # calculate total coverage (sum ref and alt)
awk '{print $1+$2}' SupParSNP74chrICalctemp > calc1Sup74chrItemp && # calculate total coverage (sum ref and alt)

paste -d'\t' InfParSNP73chrICalctemp calc1Inf73chrItemp > calc2Inf73chrItemp && # add total coverage to file with Ref and alt snp values
paste -d'\t' InfParSNP74chrICalctemp calc1Inf74chrItemp > calc2Inf74chrItemp && # add total coverage to file with Ref and alt snp values
paste -d'\t' SupParSNP73chrICalctemp calc1Sup73chrItemp > calc2Sup73chrItemp && # add total coverage to file with Ref and alt snp values
paste -d'\t' SupParSNP74chrICalctemp calc1Sup74chrItemp > calc2Sup74chrItemp && # add total coverage to file with Ref and alt snp values

awk '{print $2/$3}' calc2Inf73chrItemp > calc3Inf73chrItemp && # calculate frequency (alt/total snp)
awk '{print $2/$3}' calc2Inf74chrItemp > calc3Inf74chrItemp && # calculate frequency (alt/total snp)
awk '{print $2/$3}' calc2Sup73chrItemp > calc3Sup73chrItemp && # calculate frequency (alt/total snp)
awk '{print $2/$3}' calc2Sup74chrItemp > calc3Sup74chrItemp && # calculate frequency (alt/total snp)

paste -d'\t' calc2Inf73chrItemp calc3Inf73chrItemp > finalInf73chrItemp &&
paste -d'\t' calc2Inf74chrItemp calc3Inf74chrItemp > finalInf74chrItemp &&
paste -d'\t' calc2Sup73chrItemp calc3Sup73chrItemp > finalSup73chrItemp &&
paste -d'\t' calc2Sup74chrItemp calc3Sup74chrItemp > finalSup74chrItemp &&
```

```
paste -d'\t' InfParSNP73chrI finalInf73chrItemp > calc4Inf73chrItemp &&
paste -d'\t' InfParSNP74chrI finalInf74chrItemp > calc4Inf74chrItemp &&
paste -d'\t' SupParSNP73chrI finalSup73chrItemp > calc4Sup73chrItemp &&
paste -d'\t' SupParSNP74chrI finalSup74chrItemp > calc4Sup74chrItemp &&

cat headings_temp calc4Inf73chrItemp > InfPar73chrI_calc && # add new columns to headings file
cat headings_temp calc4Inf74chrItemp > InfPar74chrI_calc && # add new columns to headings file
cat headings_temp calc4Sup73chrItemp > SupPar73chrI_calc && # add new columns to headings file
cat headings_temp calc4Sup74chrItemp > SupPar74chrI_calc && # add new columns to headings file
echo "Chromosome I done!"
```

***Name*:** Trendline.sh

***Function*:** Remove SNPs identical between Y294[CEL7A] and M0341[CEL7A] and calculate the average SNP frequency over 5 kb chromosomal regions for the SNP coverage graph.

***Possible improvement*:** Place in a 'for-loop' to reduce code length. Use R-programming to visualize all SNPs.

***Sample script for Chr I*:**

```
declare -i j
# chrI
j=0
>InfParSNP73ChrI_trend
>InfParSNP74ChrI_trend
>SupParSNP73ChrI_trend
>SupParSNP74ChrI_trend

for (( i=0; i<=150; i++))
do
awk '$13<=0.90 && $13>0.05 {print $0}' InfPar73chrI_calc > InfPar73chrI_calcmodTemp && #new - non-unique snp remover - cheat (need to remove similar snips with grep after compareing 73 with 74
awk '$2<= ('$j' +10000) && $2>'$j' { print $0 }' InfPar73chrI_calcmodTemp > BlockInfPar73chrITemp &&
```

```
awk '$2<= ('$j' +10000) && $2>'$j' { print $0 }' BlockInfPar73chrITemp | wc -l > BlockCountInfPar73chrITemp &&

awk '{ s+=$13 } END { print s }' BlockInfPar73chrITemp > BlockTotalFreqInfPar73chrITemp &&

paste -d' ' BlockTotalFreqInfPar73chrITemp BlockCountInfPar73chrITemp > CombinedBlockInfPar73chrITemp &&

awk '{ print ('$j' +5000) "\t" $1/$2 }' CombinedBlockInfPar73chrITemp >> InfParSNP73ChrI_trend &&


awk '$13<=0.90 && $13>0.05 {print $0}' InfPar74chrI_calc > InfPar74chrI_calcmodTemp && #new - non-unique snp remover - cheat (need to remove similar snips with grep after compareing 73 with 74

awk '$2<= ('$j' +10000) && $2>'$j' { print $0 }' InfPar74chrI_calcmodTemp > BlockInfPar74chrITemp &&

awk '$2<= ('$j' +10000) && $2>'$j' { print $0 }' BlockInfPar74chrITemp | wc -l > BlockCountInfPar74chrITemp &&

awk '{ s+=$13 } END { print s }' BlockInfPar74chrITemp > BlockTotalFreqInfPar74chrITemp &&

paste -d' ' BlockTotalFreqInfPar74chrITemp BlockCountInfPar74chrITemp > CombinedBlockInfPar74chrITemp &&

awk '{ print ('$j' +5000) "\t" $1/$2 }' CombinedBlockInfPar74chrITemp >> InfParSNP74ChrI_trend &&

awk '$13<=0.90 && $13>0.05 {print $0}' SupPar73chrI_calc > SupPar73chrI_calcmodTemp && #new - non-unique snp remover - cheat (need to remove similar snips with grep after compareing 73 with 74

awk '$2<= ('$j' +10000) && $2>'$j' { print $0 }' SupPar73chrI_calcmodTemp > BlockSupPar73chrITemp &&

awk '$2<= ('$j' +10000) && $2>'$j' { print $0 }' BlockSupPar73chrITemp | wc -l > BlockCountSupPar73chrITemp &&

awk '{ s+=$13 } END { print s }' BlockSupPar73chrITemp > BlockTotalFreqSupPar73chrITemp &&

paste -d' ' BlockTotalFreqSupPar73chrITemp BlockCountSupPar73chrITemp > CombinedBlockSupPar73chrITemp &&

awk '{ print ('$j' +5000) "\t" $1/$2 }' CombinedBlockSupPar73chrITemp >> SupParSNP73ChrI_trend &&

awk '$13<=0.90 && $13>0.05 {print $0}' SupPar74chrI_calc > SupPar74chrI_calcmodTemp && #new - non-unique snp remover - cheat (need to remove similar snips with grep after compareing 73 with 74

awk '$2<= ('$j' +10000) && $2>'$j' { print $0 }' SupPar74chrI_calcmodTemp > BlockSupPar74chrITemp &&

awk '$2<= ('$j' +10000) && $2>'$j' { print $0 }' BlockSupPar74chrITemp | wc -l > BlockCountSupPar74chrITemp &&
```

```
awk '{ s+=$13 } END { print s }' BlockSupPar74chrITemp > BlockTotalFreqSupPar74chrITemp &&
paste -d' ' BlockTotalFreqSupPar74chrITemp BlockCountSupPar74chrITemp > CombinedBlockSupPar74chrITemp &&
awk '{ print ($j' +5000) "\t" $1/$2 }' CombinedBlockSupPar74chrITemp >> SupParSNP74ChrI_trend &&

j=$j+10000
done
echo "chrI is done!"
```

***Name*:** STDev.sh

***Function*:** Calculates the standard deviation of the SNP frequencies over 5 kb chromosomal regions for the SNP coverage graph.

***Possible improvement*:** Place in a 'for-loop' to reduce code length. Being an altered form of the Trendline.sh script, it contains some code redundancy which can be removed for increased calculation speed.

***Sample script for Chr I*:**

```
>BlockTemp
>SupParSNP73ChrI_trend
>SupParSNP74ChrI_trend
declare -i j
# chrI
j=0
for (( i=0; i<=170; i++))
do
awk '$13<=0.90 && $13>0.05 {print $0}' SupPar73chrI_calc > SupPar73chrI_calcmodTemp &&
awk '$2<= ($j' +10000) && $2>'$j' { print $0 }' SupPar73chrI_calcmodTemp > BlockSupPar73chrITemp &&
awk '$2<= ($j' +10000) && $2>'$j' { print $0 }' BlockSupPar73chrITemp | wc -l > BlockCountSupPar73chrITemp &&
```

```
awk '{ s+=$13 } END { print s }' BlockSupPar73chrITemp > BlockTotalFreqSupPar73chrITemp &&
paste -d' ' BlockTotalFreqSupPar73chrITemp BlockCountSupPar73chrITemp > CombinedBlockSupPar73chrITemp
&&
awk '{ print ('$j' +5000) "\t" $1/$2 }' CombinedBlockSupPar73chrITemp >> SupParSNP73ChrI_trend &&
awk '{ print $1/$2 }' CombinedBlockSupPar73chrITemp > 73test1 &&
cat BlockSupPar73chrITemp | awk '{print $13}' > 73test2 &&
paste -d' ' 73test2 73test1 > 73test3
awk 'BEGIN{getline;x=$2}{$2= (($1-x)^2)}1' 73test3 > 73test4
awk '{ s+=$2 } END { print s }' 73test4 > 73test5 &&
awk '{print sqrt($1)}' 73test5 > 73test6 &&
paste -d' ' 73test6 BlockCountSupPar73chrITemp > 73test7
awk '{ print ('$j' +5000) "\t" $1/$2 }' 73test7 >> finalSTDev_SupPar73chrI &&
awk '$13<=0.90 && $13>0.05 {print $0}' SupPar74chrI_calc > SupPar74chrI_calcmodTemp &&
awk '$2<= ('$j' +10000) && $2>'$j' { print $0 }' SupPar74chrI_calcmodTemp > BlockSupPar74chrITemp &&
awk '$2<= ('$j' +10000) && $2>'$j' { print $0 }' BlockSupPar74chrITemp | wc -l > BlockCountSupPar74chrITemp
&&
awk '{ s+=$13 } END { print s }' BlockSupPar74chrITemp > BlockTotalFreqSupPar74chrITemp &&
paste -d' ' BlockTotalFreqSupPar74chrITemp BlockCountSupPar74chrITemp > CombinedBlockSupPar74chrITemp
&&
awk '{ print ('$j' +5000) "\t" $1/$2 }' CombinedBlockSupPar74chrITemp >> SupParSNP74ChrI_trend &&
######
awk '{ print $1/$2 }' CombinedBlockSupPar74chrITemp > 74test1 &&
cat BlockSupPar74chrITemp | awk '{print $13}' > 74test2 &&
paste -d' ' 74test2 74test1 > 74test3
awk 'BEGIN{getline;x=$2}{$2= (($1-x)^2)}1' 74test3 > 74test4
awk '{ s+=$2 } END { print s }' 74test4 > 74test5 &&
awk '{print sqrt($1)}' 74test5 > 74test6 &&
awk '{ print ('$j' +5000) "\t" $1/$2 }' 74test7 >> finalSTDev_SupPar74chrI &&
```

```
#echo $j
j=$j+10000
done
echo "chrI is done!
```

***Name*:** Coverage.sh

***Function*:** Calculates the average read coverage over 1 kb chromosomal regions for the read coverage graphs.

***Possible improvement*:** Place in a 'for-loop' to reduce code length. Add calculation for reads starting within one bin, but have majority of its bases in a neighboring bin.

***Sample script for Chr I*:**

```
>coverage_50
declare -i j
# chrV
j=0
for (( i=0; i<=3000; i++))
do
awk '$2<= ('$j' +1000) && $2>'$j' { print $0 }' chrV.readPositions > temp1 &&
awk '$2<= ('$j' +1000) && $2>'$j' { print $0 }' temp1 | wc -l > temp2 &&
awk '{ print ('$j' +1000) "\t" (($1*90)/1000) }' temp2 >> coverage_50 &&
cat temp2
j=$j+1000
done
echo "chrV is done!"
```

***Name*:** Needed_for_blast.sh

***Function*:** Extracting the names of blast identified genes form the alignment file.

***Sample script*:**

```
#change *.velvet in 1ln and the QC stepne with *.soap when needed

grep -E  ">|Identities" *.velvet | grep -A1 ">" > protein_names

grep -w ">" protein_names | awk -F" " '{print $3 "\t" $2 }' | sed 's/,//g' > Chr_less

 grep -w "Identities" protein_names | awk -F" " '{print $3 "\t" $4 "\t" $11}' | sed 's/(//g' | sed 's/)//g' | sed 's/,//g' | sed 's/%//g' > Iden_less

paste -d'\t' Chr_less Iden_less > Saam

grep -w "ORF" Saam | awk -F'_' '{print $1}' | awk -F' ' '{print $2}'  > orf_names

grep -w "ORF" Saam > orf_lank

paste -d'\t' orf_names orf_lank | awk -F' ' '{print $1 "\t" $3 "\t" $4 "\t" $5 "\t" $6 }' > orf_locations

grep -v "ORF" Saam > Saam_minORF

cat Saam_minORF orf_locations | sort > final

##### QC #############

grep -w ">" *.velvet | wc -l && grep -w ">" protein_names | wc -l &&  cat Saam | wc -l && cat final | wc -l

echo "script finished"
```