THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Engineering yeast for improved recombinant protein production

YANYAN WANG

Department of Life Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023

Engineering yeast for improved recombinant protein production

YANYAN WANG

ISBN 978-91-7905-779-4

© Yanyan Wang, 2023.

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5245
ISSN 0346-718X

Division of Systems and Synthetic Biology
Department of Life Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover design: Four strategies for improving the recombinant protein production capacity of *Saccharomyces cerevisiae*.

Printed by Chalmers Reproservice
Gothenburg, Sweden 2023

Engineering yeast for improved recombinant protein production

Yanyan Wang
Department of Life Sciences
Chalmers University of Technology

Abstract

Recombinant proteins are broadly used from basic research to therapeutic development and include industrial enzymes and pharmaceutical proteins. The increasing demand for improved production and enhanced quality of recombinant proteins requires robust biotech-based strategies to overcome the limitations of protein extraction from natural sources. A variety of cell factories are therefore established for the large-scale production of recombinant proteins of interest. In comparison to other expression systems, the budding yeast *Saccharomyces cerevisiae* is an attractive production platform due to its high tolerance to harsh fermentation conditions, and importantly its capability to perform eukaryotic post-translational modifications and to secrete the biologically active product to the extracellular medium. Thus, many strategies have been applied to engineer this organism for increasing its recombinant protein secretory capacity and productivity.

The major aim of this thesis work was to study and develop efficient yeast platforms for the production of different heterologous proteins for medical or industrial use through diverse engineering strategies. The first part of this work explored in depth a line of previously evolved yeast strains with improved protein secretory capacity. The universal applicability of the evolved strains was evaluated to produce different antibody fragments, but it was concluded that this secretion platform was not suitable for all types of pharmaceutical proteins tested. Furthermore, by re-introducing all 42 protein-sequence-altering mutations identified in the evolved strains into the parental strain using the CRISPR/Cas9 technology, 14 targets were shown to be beneficial for protein production and 11 out of these 14 beneficial targets were newly identified to be related to recombinant protein production. The second part of this work focused on investigating novel targets related to the cellular stress response and the protein secretory process to rationally optimize *S. cerevisiae*. Furthermore, screening for suppressors of amyloid-β cytotoxicity in a yeast Alzheimer's disease model revealed a number of gene targets that reduced oxidative stress and improved production of recombinant proteins. Additionally, a proteome-constrained genome-scale protein secretory model of *S. cerevisiae* (pcSecYeast) was constructed to simulate the secretion of various recombinant proteins and predict system-level engineering targets for increasing protein production. In summary, the work presented in this thesis provides different efficient strategies to develop yeast platforms for the high-level production of valuable industrial or pharmaceutical proteins, and also provides general guidelines for designing other cell platforms for efficient protein production. Integrated application of various engineering approaches will make meaningful advancements in the field of recombinant protein production in the future.

Keywords: *Saccharomyces cerevisiae*, recombinant protein, protein secretion, CRISPR/Cas9, point mutation, omics analysis, genome-scale modeling

List of Publications

This thesis is based on the work contained in the following papers and manuscripts:

Paper I: Expression of antibody fragments in *Saccharomyces cerevisiae* strains evolved for enhanced protein secretion
Yanyan Wang, Xiaowei Li, Xin Chen, Jens Nielsen, Dina Petranovic and Verena Siewers. *Microbial Cell Factories*, 2021, 20(1): 1-17.

Paper II: CRISPR/Cas9-mediated point mutations improve α-amylase secretion in *Saccharomyces cerevisiae*
Yanyan Wang, Xiaowei Li, Xin Chen and Verena Siewers. *FEMS Yeast Research*, 2022, 22(1): foac033.

Paper III: Suppressors of amyloid-β toxicity improve recombinant protein production in yeast by reducing oxidative stress and tuning cellular metabolism
Xin Chen, Xiaowei Li, Boyang Ji, **Yanyan Wang**, Olena P. Ishchuk, Egor Vorontsov, Dina Petranovic, Verena Siewers and Martin K.M. Engqvist. *Metabolic Engineering*, 2022, 72: 311-324.

Paper IIIb: Dataset for suppressors of amyloid-β toxicity and their functions in recombinant protein production in yeast
Xin Chen, Xiaowei Li, Boyang Ji, **Yanyan Wang**, Olena P. Ishchuk, Egor Vorontsov, Dina Petranovic, Verena Siewers and Martin K.M. Engqvist. *Data in Brief*, 2022: 108322.

Paper IV: Improving recombinant protein production by yeast through genome-scale modeling using proteome constraints
Feiran Li, Yu Chen*, Qi Qi*, **Yanyan Wang***, Le Yuan, Mingtao Huang, Ibrahim E. Elsemman, Amir Feizi, Eduard J. Kerkhoven and Jens Nielsen. *Nature Communications*, 2022, 13(1): 1-13.

* Authors contributed equally to this work

Additional papers and manuscripts not included in this thesis:

Paper V: Metabolic network remodelling enhances yeast's fitness on xylose using aerobic glycolysis
Xiaowei Li, **Yanyan Wang**, Gang Li, Quanli Liu, Rui Pereira, Yun Chen and Jens Nielsen. *Nature Catalysis*, 2021, 4(9): 783-796.

Paper VI: Reprogramming methanol utilization pathways to convert *Saccharomyces cerevisiae* to a synthetic methylotroph
Chunjun Zhan, Xiaowei Li, Edward E. K. Baidoo, Yankun Yang, Yang Sun, Shijie Wang, **Yanyan Wang**, Guokun Wang, Jens Nielsen, Jay D. Keasling, Yun Chen and Zhonghu Bai. *Nature Catalysis*, 2023. *Accepted*

Paper VII: Modular deregulation of cellular central carbon metabolism to establish a xylose catalyzing platform in yeast
Xiaowei Li, **Yanyan Wang**, Chunjun Zhan, Xin Chen, Yun Chen and Jens Nielsen. *Manuscript*

Contribution Summary

Paper I

Co-designed the project, carried out the experiments, co-analyzed the data, wrote the manuscript.

Paper II

Co-designed the project, carried out the experiments, co-analyzed the data, wrote the manuscript.

Paper III and **IIIb**

Carried out part of the strain construction and protein quantification experiments, co-analyzed the data and assisted in manuscript preparation.

Paper IV

Carried out part of the overexpression targets verification experiments, co-analyzed the data and assisted in manuscript preparation.

Preface

This dissertation serves as partial fulfillment of the requirements to obtain the degree of Doctor of Philosophy at the Department of Life Sciences at Chalmers University of Technology. The PhD studies were carried out at the division of Systems and Synthetic Biology (SysBio) under the supervision of Verena Siewers and the co-supervision of Xin Chen. The thesis was examined by Ivan Mijakovic. Temporary supervisors were Dina Petranovic and Martin Engqvist and the temporary examiner was Jens Nielsen. The research was funded by the Swedish Foundation for Strategic Research (SB16-017), VINNOVA center CellNova (2017-02105) and the Novo Nordisk Foundation (NNF10CC1016517).

Yanyan Wang

March 2023

Table of Content

Abstract .... **III**

List of Publications .... **IV**

Contribution Summary .... **V**

Preface .... **VII**

Abbreviations .... **XI**

1 Introduction **1**

1.1 Recombinant proteins 1

1.1.1 Proteins 1

1.1.2 Recombinant proteins 1

1.2 Recombinant protein production hosts 2

1.2.1 Host cells 2

1.2.1.1 Mammalian cells 2

1.2.1.2 *E. coli* 3

1.2.1.3 Yeasts 4

1.2.2 *S. cerevisiae* as a model organism 5

1.3 Protein production in *S. cerevisiae* 6

1.3.1 Protein synthesis 6

1.3.1.1 Transcription 7

1.3.1.2 Translation 8

1.3.2 Protein secretion 9

1.3.2.1 Translocation 10

1.3.2.2 Protein glycosylation 11

1.3.2.3 Protein folding and quality control in the ER 12

1.3.2.4 Intracellular vesicle trafficking 14

1.3.2.5 Golgi processing 15

1.3.2.6 Exocytosis 16

1.3.3 Other factors affecting protein production 16

1.4 Genetic engineering 17

1.5 Strain library construction and screening techniques 19

1.5.1 Strain library construction techniques 19

1.5.2 Strain library screening techniques 20

1.6 Systems biology tools 21

1.6.1 Top-down approach 21

1.6.2 Bottom-up approach 22

1.7 Aim and significance 23

2 Model proteins in this thesis **25**

3 Characterization of strains evolved for increased protein secretion **27**

3.1 Evaluating the expression of antibody fragments in the mutant strains 28

3.1.1 Expression of antibody fragments in the high-secretion mutant strain............................29
3.1.2 Assessment of secretion levels of antibody fragments in different mutant strains...........31
3.1.3 Investigating the mechanism for protein-specific performance of the mutant strains 32
3.2 Characterizing the modifications of the mutant strains 34
3.2.1 Protein-altering mutations in the mutant strains 35
3.2.2 Strategies for the introduction of the desired point mutations...36
3.2.3 Beneficial targets for α-amylase production..38
4 Rational optimization of platform strains by investigating novel targets 45
4.1 Novel targets from Aβ42 yeast model 45
4.1.1 Effects of single targets on recombinant protein production 46
4.1.2 Effects of combinatorial targets on recombinant protein production 48
4.1.3 Evaluating the general applicability of effective targets 51
4.2 Novel targets from the pcSecYeast model..51
4.2.1 Brief description of pcSecYeast...52
4.2.2 Verification of the predicted overexpression targets 54
5 Conclusion and perspectives 57
5.1 Conclusion 57
5.2 Perspectives...58
Acknowledgements 61
References..63

Abbreviations

CAGR	Compound annual growth rate
mAb	Monoclonal antibody
PTM	Post-translational modification
CHO	Chinese hamster ovary
HSA	Human serum albumin
AD	Alzheimer's disease
ER	Endoplasmic reticulum
mRNA	Messenger RNA
tRNA	Transfer RNA
GFP	Green fluorescent protein
3'-UTR	3' untranslated region
SRP	Signal recognition particle
PDI	Protein disulfide isomerase
PMTs	Protein O-mannosyl transferases
MNTs	α-1,2-Mannosyltransferases
ROS	Reactive oxygen species
UPR	Unfolded protein response
ERAD	ER-associated degradation
IgG	Immunoglobulin G
COPII	Coat protein complex II
GEF	Guanine nucleotide exchange factor
gpαf	Glycosylated pro-α-factor
SM	Sec1p/Munc18p
COPI	Coat protein complex I
GAP	GTPase activating protein
SNARE	Soluble N-ethylmaleimide-sensitive factor attachment protein receptor
GEM	Genome-scale metabolic model
CRISPR	Clustered regularly interspaced short palindromic repeats
Cas	CRISPR-associated
crRNA	CRISPR RNA
tracrRNA	Trans-activating crRNA
gRNA	Guide RNA
PAM	Protospacer adjacent motif
dCas9	Deactivated Cas9
CRISPRa	CRISPR activation
CRISPRi	CRISPR interference
NHEJ	Non-homologous end joining
HDR	Homology-directed repair
RNAi	RNA interference
UV	Ultraviolet
Aβ	Amyloid-β
SGA	Synthetic genetic array
pcSecYeast	Proteome-constrained genome-scale protein secretory model of yeast
Fab	Antigen-binding fragment
scFv	Single-chain variable fragment
Nan	Nanobody
Pex	Pexelizumab
Ran	Ranibizumab
GLA	Glucan-1,4-α-glucosidase

LA	Low-secretion strain
MA	Medium-secretion strain
HA	High-secretion strain
Ran-H	Heavy chain of Ran
Ran-L	Light chain of Ran
ELISA	Enzyme-linked immunosorbent assay
PC1	First principal component
PC2	Second principal component
GO	Gene Ontology
KOG	euKaryotic orthologous groups
COB	Cytochrome b
MAPK	Mitogen-activated protein kinase
HOG	High osmolarity glycerol
DCW	Dry cell weight
ERMES	ER-mitochondrial encounter structure
GPI	Glycosylphosphatidylinositol
FSEOF	Flux scanning based on enforced objective flux

“The only true wisdom is in knowing you know nothing.”

-Socrates

1 Introduction

1.1 Recombinant proteins

1.1.1 Proteins

Proteins are highly complex molecules that consist of 20 basic building blocks, known as amino acids. They are present in all living organisms as part of natural metabolism, playing critical roles in catalyzing metabolic reactions, acting as construction elements of biological assemblies, charging inter- and intracellular interactions, cell signaling and cell cycle events, etc. (Demain & Vaishnav, 2009, Ferrer-Miralles *et al.*, 2009). However, several factors hinder protein extraction from natural sources (commonly plants and animals). For example, purification of small amounts of proteins requires large amounts of plant and animal tissues and large volumes of biological extraction reagents which leads to long processing times and high production costs (Rosano & Ceccarelli, 2014). Additionally, the extraction process may cause protein denaturation and contamination (Mattanovich *et al.*, 2012). On the other hand, proteins cannot be synthesized chemically due to their complex structure and function (Overton, 2014). To bypass these limitations, researchers have over the past decades been working to produce proteins in the recombinant form in new hosts.

1.1.2 Recombinant proteins

A recombinant protein is produced by cloning a heterologous DNA encoding the protein of interest into a specific expression system. Compared to extraction from natural sources, recombinant protein production has many prominent advantages, such as higher productivity and specificity, better flexibility and stability, superior purity, and easier and cheaper large-scale production (Baeshen *et al.*, 2014). Recombinant proteins have a wide range of industrial applications, such as food and beverages, clothing and cosmetics, detergents and paper, waste management and biofuel industries, as well as medical applications, such as prevention, diagnosis and therapy of human diseases, health care products, medical materials and biological reagents (Holliger & Hudson, 2005, Puetz & Wurm, 2019).

The advent of recombinant DNA technology or genetic engineering in the 1970s opened a new gate to producing a broad spectrum of proteins and biochemicals in naturally non-producing host organisms (Mattanovich *et al.*, 2012). With the development of recombinant DNA technologies, the human hormone somatostatin and human insulin were successfully expressed in *Escherichia coli* by researchers from Genentech in the late 1970s. Subsequently, the first recombinant biopharmaceutical, human insulin produced in *E. coli*, was marketed by Eli Lilly and Company for diabetes treatment in 1982 (Itakura *et al.*, 1977, Wang *et al.*, 2017). Since then, recombinant DNA technology-based products have increased rapidly with their utility and applicability. Of the 71 new biopharmaceutical active ingredients marketed from 2015 to 2018, 62 (87%) are recombinant proteins (Walsh, 2018). In the next few years, recombinant protein-based biopharmaceuticals are expected to continue to dominate the approvals rather than nucleic acid- or cell-based products (Walsh, 2018). According to a recent report by the Mordor Intelligence, the global biopharmaceuticals market has reached USD 407.72 billion in 2023, and the compound annual growth rate (CAGR) is projected to reach 8.03% by

2027 (Available online: https://www.mordorintelligence.com/industry-reports/global-biopharmaceuticals-market-industry (accessed in February 2023)). Especially, the demand for protein subunit-based vaccines and the upcoming monoclonal antibodies (mAbs) and therapeutic drugs against the SARS-CoV-2 virus will promote the significant growth of the biopharmaceutical industry during the ongoing COVID-19 pandemic. In parallel, the industrial enzyme market has reached over USD 60 billion in 2021, and the CAGR is expected to reach more than 6% by 2027 (Available online: https://www.mordorintelligence.com/industry-reports/industrial-enzymes-market (accessed in February 2023)). In this work, we studied both types of recombinant proteins, including biopharmaceutical proteins (**Paper I**) and an industrial enzyme (**Paper II-IV**). The characteristics of these proteins will be discussed in detail in **Chapter 2**. Despite these achievements, current manufacturing technologies are still unable to meet the increasing demand for optimum yield and quality of recombinant proteins, which guides us to explore and further develop efficient cell platforms for the expression of recombinant proteins.

1.2 Recombinant protein production hosts

1.2.1 Host cells

The production of recombinant proteins using different expression systems can be interpreted from a fundamental biological architecture, the bow tie structure (Figure 1A). Various carbon sources, such as glycose and glycerol, together with different nitrogen sources, such as ammonium sulfate and amino acids, are converted into recombinant proteins through a series of cellular processes within the host cells. At present, a wide range of cell factories, including microorganisms (bacteria, yeast and filamentous fungi) and higher organisms (mammalian cells, insect cells and plants) are frequently used as host cells for the production of recombinant proteins (Davy *et al.*, 2017). In addition to protein production and quality, protein stability, downstream processing and production costs are also critical factors for the selection of host cells. Recombinant biopharmaceutical proteins are produced in different expression systems using microbes (*E. coli* (29.8%) and (*Saccharomyces cerevisiae* (18.5%)), mammalian cells (39%), insect cells (0.75%) and other systems (Ferrer-Miralles *et al.*, 2009, Porro *et al.*, 2011) (Figure 1B). On the other hand, industrial enzymes are made by fungi (over 50%), bacteria (30%), animals (8%), plants (4%) and other systems (Demain & Vaishnav, 2009) (Figure 1C). Table 1 lists the species involved, advantages and disadvantages of recombinant protein production in different host cells, of which mammalian cells, *E. coli* and yeasts are the most popular cell factories (Rader, 2008), which will be described in detail below.

1.2.1.1 Mammalian cells

Post-translational modifications (PTMs) such as glycosylation are crucial for ensuring proper protein function and biological activity. Incorrect PTMs of produced pharmaceutical proteins often confer a short half-life, which will reduce the efficacy for use, require a higher dosage and/or even cause immunogenicity (Pirkalkhoran *et al.*, 2023). Mammalian cells, such as Chinese hamster ovary (CHO) cells, have the highest similarity to human cells, and produced proteins are often properly folded and

glycosylated. These will alleviate supply shortages and show good safety for therapeutic use. However, there are many drawbacks. The limited secretory capacity will reduce the efficiency of multiple PTMs and thus lead to low protein production and affect the product quality (Demain & Vaishnav, 2009). The costs are often high due to the expensive media, slow growth, and easy contamination by microorganisms and viruses (Martínez *et al.*, 2012). To avoid these production drawbacks from mammalian cells, the use of microbes has therefore drawn increasing attention.

1.2.1.2 *E. coli*

The bacterium *E. coli* is one of the earliest and most widely used microbial host organisms for recombinant protein production for several reasons. These include annotated information about its genetics, molecular biology, and metabolic pathways; vast availability of genetic tools; ease of culture, survival in various environmental conditions; and high growth rate, high cell densities, and high product yields, achieving 80% of its dry weight (Demain & Vaishnav, 2009). However, some obstacles limit its application. High intracellular densities of produced proteins tend to aggregate and form inclusion bodies in *E. coli*, which are insoluble, inactive and require refolding. If a protein contains many disulfide bonds, it is very difficult to be produced by *E. coli* because it is almost impossible to correctly refold such a protein. In addition, *E. coli* lacks the ability for complex PTMs, such as glycosylation and proteolytic processing (Ferrer-Miralles *et al.*, 2009, Porro *et al.*, 2011).

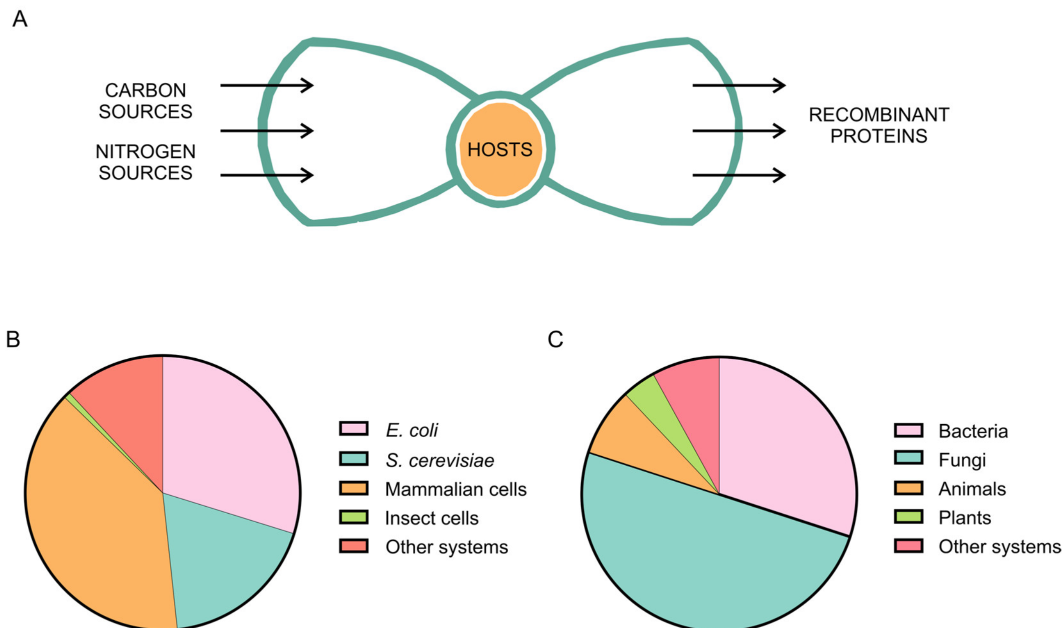

Figure 1. The recombinant protein production hosts. (**A**) Fundamental biological similarities in different expression systems. The bow tie structure shows how the carbon sources and nitrogen sources are converted into a recombinant protein. Inspired by Backman *et al*. (Backman *et al.*, 2018). The percentage of recombinant proteins, including biopharmaceuticals (**B**) and industrial enzymes (**C**), produced by different hosts (Demain & Vaishnav, 2009, Ferrer-Miralles *et al.*, 2009).

Table 1. The species, advantages and disadvantages of different host systems (Martínez *et al.*, 2012, Baeshen *et al.*, 2014, Kesik-Brodacka, 2018, Vieira Gomes *et al.*, 2018).

Hosts	Bacteria	Yeasts	Mammalian cells	Insect cells	Plant cells
Species	*E. coli;* *Corynebacterium glutamicum;* *Streptococci;* *Bacillus subtilis;* *B. licheniformis;* *B. megaterium;* *B. brevis;* *Ralstonia eutropha;* *Pseudomonas fluorescens;* *Staphylococcus carnosus*	Baker's yeast: *S. cerevisiae*; Methylotrophic yeasts: *Pichia pastoris*; *Hansenula polymorpha* Oleaginous yeasts: *Yarrowia lipolytica* Fission yeast: *Schizosaccharomyces pombe*	Nonprimate: Chinese hamster ovary cells; Murine cells Primate: Monkey cells; Human cells	*Spodoptera frugiperda*	Tobacco *Arabidopsis thaliana* Cereals and legumes: Maize Rice Alfalfa Soybean Fruit and vegetables: Tomato Potato Lettuce Banana Turnip
Advantages	Ease of culture; High growth rate; High cell densities; High proteins yields; Cost effective; Rapid growth in chemically defined media; Well-developed molecular tools for genetic manipulation and systems biology.	Ease of culture; High growth rate; High cell densities; High proteins yields; Cost effective; Rapid growth in chemically defined media; Well-developed molecular tools for genetic manipulation and systems biology; Stable production strains; Secretion capacity; Can perform PTMs*.	The highest similarity to human cells; Proper PTMs; Safety.	Cheaper than mammalian cell culture; High cell densities; Safety; Flexibility of protein size; Multiple genes expressed simultaneously; PTMs without complication.	Cost effective; Easy and cheap scalability; Produce proteins in different tissues; Ease of manipulation; Not prone to contamination by human pathogens; The ability to fold and assemble complex proteins; Can perform PTMs.
Disadvantages	Unable to perform complex PTMs; Tend to form inclusion bodies; Proteins produced with endotoxins; Cell toxicity from acetate formation.	High mannose type N-glycosylation; Short glycan structures of O-glycosylation; Secretion pathway differs from human.	Expensive cultivation; limited secretion capacity; Low protein yields.	Time consuming; More expensive than microbes; Tend to form aggregates; Low expression levels; Limitation in PTMs, such as incorrect glycosylation.	Low manufacturing capacity; Lack of defined regulatory pathways; PTMs differs from human; Environmental biosafety risks; Acceptability issues.

*PTM indicates post-translational modification

1.2.1.3 Yeasts

Alternatively, yeasts, the unicellular eukaryotic hosts, are known to perform proper PTMs required for a biologically active protein, which are similar to mammalian cells. They have low nutritional demands and are less easily contaminated than mammalian cells (Vieira Gomes *et al.*, 2018). In addition, yeasts can grow in chemically defined media and own well-established industrial fermentation processes

(Madhavan *et al.*, 2021). Another advantage is that they can secrete recombinant proteins into the extracellular medium facilitating isolation and downstream purification (Wang *et al.*, 2017). One of the limitations with the use of yeast, specifically *S. cerevisiae*, is that it performs high-mannose type N-glycosylation (Martínez *et al.*, 2012). Many engineering studies have been performed on yeast glycosylation to achieve a human-like glycosylation pattern, which will be described in **Chapter 1.3.2.2**.

The production of a recombinant protein usually starts with choosing an appropriate production host from a comparative view. For the commercial production of a protein, quantity and quality are the most dominant factors to be considered. In general, high yields are more critical for producing industrial enzymes, while high quality is essential for producing biopharmaceutical proteins (Porro *et al.*, 2011). Bacteria are generally used to produce intracellular proteins that are not complex and do not require PTMs, while the production of secreted, complex, and PTM-requiring proteins is commonly obtained in yeast and higher eukaryotic systems (Vieira Gomes *et al.*, 2018).

1.2.2 *S. cerevisiae* as a model organism

One of the most popular yeast species, baker's yeast *S. cerevisiae* has been used by humans in daily life since 7000 BCE in China (Sicard & Legras, 2011). With the development of the production of wine, beer and leavened bread, *S. cerevisiae* was continually domesticated and became a valuable microorganism in the food industry. These applications also contribute to its "generally regarded as safe" status. The close relationship with human activities led Louis Pasteur to find the 'hidden' world of yeast activity in alcohol fermentation in 1857 (Cavaillon & Legout, 2022). Due to shortages of raw materials, it was gradually exploited to produce glycerol and ethanol. The parameters that make *S. cerevisiae* a popular organism to be applied in the industry are its robustness, fast growth, resistance to low pH and simple fermentation process (Cakar *et al.*, 2012). In the 1970s, genetic engineering techniques also made it possible to engineer *S. cerevisiae* to produce biopharmaceuticals, such as human insulin. With the rapid expansion as a cell factory, much interest has shifted to basic research, which made this yeast a vital model organism for eukaryotic biology (Nielsen, 2019). Subsequently, the publication of the whole genome sequence in 1996 enabled the development of various genome engineering tools and techniques to make *S. cerevisiae* the preferred eukaryotic platform organism for producing valuable proteins and chemicals (Figure 2).

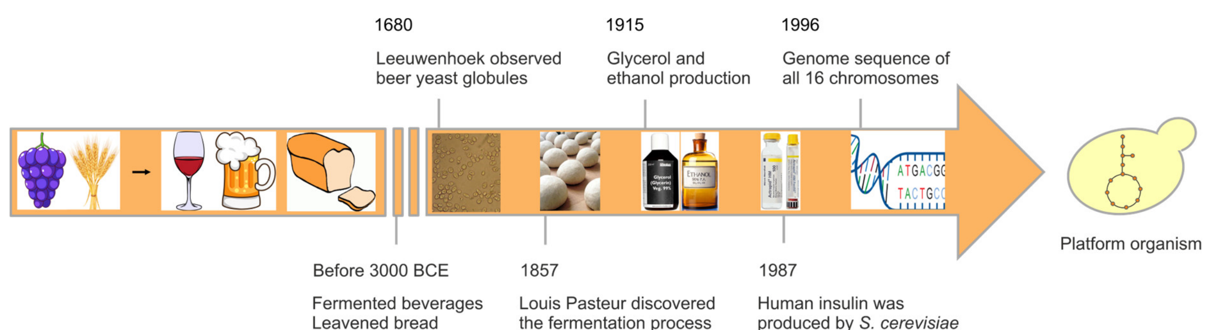

Figure 2. A timeline indicating the use of *S. cerevisiae* for production of different chemicals and proteins.

1.3 Protein production in *S. cerevisiae*

The importance of studying production and secretion of recombinant proteins in *S. cerevisiae* is considered in two main aspects. First, an increasing number of commercialized biopharmaceuticals is produced by *S. cerevisiae* (insulin, glucagon, hirudin, growth hormone, human serum albumin (HSA), hepatitis vaccines, etc.) (Martínez *et al.*, 2012). The growing demand for industrial enzymes also mainly depends on the secretory pathway (Mattanovich *et al.*, 2012). However, as the secretory pathway has not been fully exploited, many recombinant proteins are produced in *S. cerevisiae* at only 1% (or less) of the theoretical capacity. Second, many human diseases, such as Alzheimer's disease (AD), Parkinson's disease and diabetes mellitus, are related to endoplasmic reticulum (ER) stress (Yoshida, 2007). The ER stress can be caused by the formation of misfolded proteins or protein aggregation. Thus, the yeast *S. cerevisiae* can be used as a powerful eukaryotic model to study the implication of protein misfolding in human pathologies (Chen *et al.*, 2020).

Protein production involves numerous intracellular processes, which have been investigated and engineered through many strategies. The protein production process starts with protein synthesis and is followed by PTMs, vesicle trafficking, and secretion. A detailed description of the protein production processes and their engineering strategies are provided in the following section.

1.3.1 Protein synthesis

Protein synthesis starts with transcription and translation processes in cells. This genetic information flow is explained by the central dogma of molecular biology (Crick, 1970). It describes that the coding sequence of the gene of interest (DNA) is transcribed to RNA, and messenger RNA (mRNA) is translated to protein (Figure 3). Unlike prokaryotic cells, where transcription and translation are linked together due to absence of a nuclear compartment, the transcription is separated from the translation process in eukaryotic cells with the compartmentalized and membrane-bound organelles (Gottshall *et al.*, 2014). Mature mRNA must be transported out of the nucleus to the cytoplasm where it can be bound by ribosomes. The aminoacylated transfer RNA (tRNA) is brought into the ribosome-mRNA complex through complementary base-pairing with mRNA. Then, the ribosome-ribosomal RNA (rRNA) complex catalyzes the formation of peptide bonds between the amino acid and adds the new amino acid from tRNA to the growing polypeptide chain (Van Der Kelen *et al.*, 2009). In some cases, the translation occurs at the outer membrane of the rough ER, leading to co-translational translocation of the peptide chain into the ER, which will be described in **Chapter 1.3.2.1**. To effectively reach high levels of recombinant protein expression, several factors affecting the early steps of transcription and translation have been regulated to optimize the expression levels.

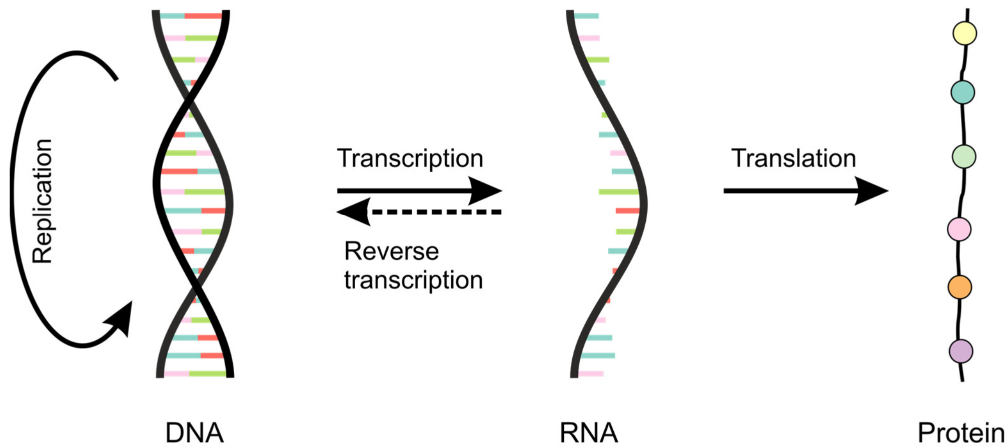

Figure 3. The flow of information from DNA to RNA to protein described by the central dogma of molecular biology.

1.3.1.1 Transcription

The transcriptional efficiency is a central issue for efficient protein production. Using promoters of different strength to control gene expression at the transcriptional level is the most common approach to adjusting protein expression. Various promoters are available and well characterized, including constitutive promoters such as *TEF1p*, *PGK1p*, *ADH1p*, *TPI1p*, *HXT7p*, *TDH3p* and *PYK1p*, and inducible promoters like *GAL1p*, *GAL10p*, *PHO5p* and *MET25p* (Vieira Gomes *et al.,* 2018). The strong constitutive promoters maintain constant and high transcriptional levels and have been regularly applied to direct the expression of heterologous genes in *S. cerevisiae* (Partow *et al.*, 2010). However, sometimes strong constitutive promoters might result in low secretion efficiency due to excessive transcription (and translation) of the coding sequence in a short time and thus aggregation of misfolded proteins (Liu *et al.*, 2012). Synthetic promoter libraries, created through random mutagenesis of a native promoter, were applied to fine-tune gene expression and unravel transcription regulatory mechanisms (Alper *et al.*, 2005, Sharon *et al.*, 2012). Additionally, inducible promoters allow separating the growth stage and production stage. By employing these, it is therefore possible to prevent unintended selection of rapidly growing non-recombinant cells or to enable the production of toxic proteins (Vieira Gomes *et al.*, 2018). For example, under control of *MET25p*, high production of HSA and HSA-fused protein was obtained in the media lacking methionine in a late log phase of the growth (Solow *et al.*, 2005).

On the other hand, increasing the gene copy number is a straightforward approach to increase the transcriptional level of recombinant proteins. In *S. cerevisiae*, there are mainly two types of plasmids, including centromeric plasmids and episomal plasmids (Mumberg *et al.*, 1995). A centromeric plasmid, incorporating a centromeric sequence and an autonomously replicating sequence, maintains 1-2 copy per cell (Vieira Gomes *et al.*, 2018). Even though this plasmid is mitotically stable without chromosomal integration, low gene expression levels limit its use. An episomal plasmid is usually employed to achieve high copy numbers, such as the native 2 μ-derived plasmid maintaining 10 - 40

copies per cell (Mumberg *et al.*, 1995). However, such plasmid-based systems can suffer from genetic instability, including segregational instability and structural instability, which results in cell-to-cell variations in protein production (Tyo *et al.*, 2009). The stability and copy number of the expression plasmid can be influenced by the choice of the selection markers. According to the literature, various markers have been widely employed in *S. cerevisiae*, such as auxotrophic markers (e.g., *URA3*, *HIS3*, *ADE2*, *LEU2* and *TRP1*), C/N source-related markers (e.g., *amdS* and *GAP1*), autoselection systems (e.g., glycolytic pathway genes such as *FBA1*, *TPI1* and *PGI1*, essential cell division cycle genes such as *CDC4*, *CDC9* and *CDC28*), and resistance markers (e.g., heterologous gene *cat* conferring resistance to chloramphenicol, *kan* conferring resistance to G418, and *nat1* conferring resistance to nourseothricin) (Siewers, 2014). Auxotrophic markers are the most commonly used markers. In particular, defective auxotrophic markers, such as truncated promoters, are helpful in modulating the plasmid copy number. Combining destabilization of the marker protein with decreasing promoter strength of the marker gene increased the 2 μ-based plasmid copy number by 3-fold and thus enhanced the protein expression (Chen *et al.*, 2012). However, auxotrophic marker systems have to be maintained in synthetic medium, which is not practical for large-scale and long-term industrial fermentation. The autoselection systems are a preferable alternative for studying protein secretion, such as the secretion of glucagon (Egel-Mitani *et al.*, 2000), β-glucosidase (Tang *et al.*, 2013), insulin precursor and α-amylase (Liu *et al.*, 2012). As the cell lacks the marker gene, essential for cell viability, a plasmid carrying this essential gene will maintain high stability even in a complex medium. For example, the widely used CPOTud plasmid, using a triose phosphate isomerase gene (*POT1*) from *Schizosaccharomyces pombe* to complement a *TPI1* deletion in *S. cerevisiae*, has a high plasmid copy number due to the lower expression level of *POT1* in *S. cerevisiae*. Compared to the auxotrophic *URA3* marker, *POT1* as a selection marker led to production of 26.6-fold more insulin precursor and 4.79-fold more α-amylase (Liu *et al.*, 2012). In this study, we use these four aforementioned types of markers in the genetic manipulation, including autoselection markers (**Paper I**-**IV**), C/N source-related markers (**Paper I**), resistance markers (**Paper II**-**IIIb**) and auxotrophic markers (**Paper III**-**IV**).

1.3.1.2 Translation

Translation consists of three stages: initiation, elongation and termination. In eukaryotes, the Kozak sequence, which occupies approximately the first six nucleotides upstream of the START codon, as a translation initiation site mediates translation initiation efficiency (Kozak, 1986). The strength of the Kozak sequence is related to the favorability of initiation, affecting the amount of protein being synthesized from a given mRNA. The consensus Kozak sequence differs among organisms in length and nucleotide composition (Cavener, 1987, Hamilton *et al.*, 1987). The *S. cerevisiae* Kozak consensus sequence was identified as (A/T)A(A/C)A(A/C)A**ATG**TC(T/C) among the highly expressed genes (Hamilton *et al.*, 1987). A previous study showed that point mutations within the Kozak sequence regulated the production of proinsulin over a 20-fold range (Kozak, 1986). In addition, through screening Kozak libraries in *S. cerevisiae*, the expression level of yeast-enhanced green fluorescent protein (yEGFP) showed an almost 10-fold change between the weakest and strongest Kozak sequences (Petersen *et al.*, 2018).

Codon usage varies among different organisms, called codon usage bias. Heterologous genes may be hard to express in another host due to their rare codons, which need to be read by complementary rare tRNAs that have been assigned amino acids. This will reduce the rate of translation elongation and also incite translation errors (Ikemura, 1981). Codon optimization of heterologous genes to match the host codon bias is an initial strategy for improving the efficiency of translation and accuracy of protein synthesis and therefore achieving effectively high levels of recombinant protein production (Al-Hawash *et al.*, 2017). In recent years, a number of cases showed enhanced levels of protein expression through codon optimization based on the individual codon usage bias approach and codon-pair context bias approach (Gustafsson *et al.*, 2004, Lanza *et al.*, 2014, Ahn *et al.*, 2016). For example, an improved translational efficiency of the catechol 1,2-dioxygenase gene was obtained by codon-pair context bias optimization, which yielded higher protein activity in *S. cerevisiae* (Lanza *et al.*, 2014). However, considering that the rapid translation rate may cause protein misfolding, an alternative strategy such as codon coordination will help identify and maintain slow translation regions (Villalobos *et al.*, 2006).

Terminator regions, located downstream of coding sequences, contain the 3' untranslated region (3'-UTR), which will affect the processing, localization, stability, half-life and translational efficiency of mRNA (Kuersten & Goodwin, 2003). A genome-wide assessment of terminator regions in *S. cerevisiae* showed the activity of terminators varied over a 70-fold range (Yamanishi *et al.*, 2013). Compared to commonly used *PGK1* and *CYC1* terminators, 1.2-fold and 1.5-fold higher production of GFP was obtained using the *TPS1* terminator (Yamanishi *et al.*, 2011), which indicated that the *TPS1* terminator could be generally used in yeast high-production expression systems. When coupled with a low-strength promoter, the expression-enhancing terminator led to an 11-fold expression difference of yellow fluorescent protein over the *CYC1* terminator (Curran *et al.*, 2013). Therefore, the terminator activity is critical in determining the heterologous protein expression level, although not the major determinant.

1.3.2 Protein secretion

Even if the transcription and translation levels are optimized sufficiently for protein overproduction in a suitable host, many proteins are still secreted at relatively low levels. To meet the increasing production demands, the most direct way is to increase protein secretion capacity. However, protein secretion in *S. cerevisiae* is a quite complex process as it comprises more than 550 proteins that carry a signal peptide, which takes up about 9.2% of the entire proteome (Delic *et al.*, 2013). Only a very small amount of these proteins is secreted outside the cell as different organelles such as ER, Golgi, endosome, vacuole and cytoplasmic membrane are employed before the mature proteins reach their destination. Proteins located in these organelles and some other native proteins are also processed and transported via the secretory pathway. As shown in Figure 4, the protein secretion pathway covers many steps from mRNAs/peptides to mature proteins, including co- or post-translational translocation of proteins from the cytoplasm into the ER, post-translational glycosylation in the ER and Golgi apparatus, protein folding and quality control in the ER, intracellular vesicle trafficking between ER

and Golgi, Golgi processing, and exocytosis (Idiris *et al.*, 2010). The minimum requirement for this process is a secretion signal peptide at the N-terminus of the nascent polypeptide.

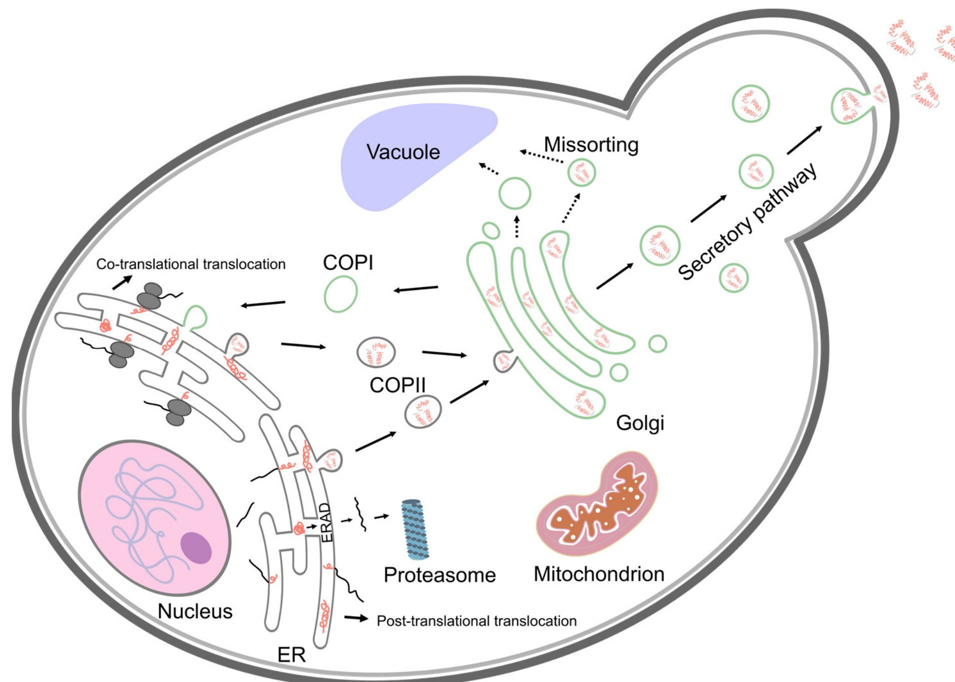

Figure 4. Schematic illustration of the protein secretory pathway in yeast. COP I: coat protein complex I; COP II: coat protein complex II.

1.3.2.1 Translocation

The secretion pathway starts by translocation of the (nascent) polypeptide from the cytoplasm into the ER lumen. Depending on the hydrophobicity of the pre-signal peptide, the translocation process can occur either post-translationally or co-translationally (Zimmermann *et al.*, 2011). Co-translational translocation (ribosome-coupled) is initiated by highly hydrophobic pre-signals, such as the signal sequence of dipeptidyl aminopeptidase B, repressible alkaline phosphatase and invertase (Zheng & Gierasch, 1996). This process directly links translation and translocation processes. Upon the emergence of the pre-signal sequence from the ribosome, Srp54p, a subunit of the signal recognition particle (SRP), recognizes and binds to the signal peptide. This combination pauses translation, and the ribosome-protein-SRP complex is transferred to the ER membrane (Delic *et al.*, 2013). GTP-containing Srp54p binds to the SRP receptor, forming a GTP-stabilized complex, which directs the ribosome-nascent chain complex to the translocation site – the Sec61p or Ssh1p translocon pore. Following GTP hydrolysis, SRP is released from the signal peptide and later from the SRP receptor while the ribosome-nascent chain complex is still bound to the translocon pore (Osborne *et al.*, 2005). Once the signal peptide is successfully recognized by the Sec61p or Ssh1p complex, the nascent chain will be inserted into the complex, and then the ribosome will also be tightly bound to the complex (Kalies & Hartmann, 1998). The polypeptide is synthesized as it passes through the translocon pore into the ER, and GTP hydrolysis provides driving energy. Post-translational translocation (ribosome-uncoupled) is initiated by weakly hydrophobic pre-signals, such as the signal sequence of secreted alpha mating

factor, protein disulfide isomerase (PDI) and carboxypeptidase Y (Zheng & Gierasch, 1996). The translation is completed in the cytoplasm. The polypeptide chain needs to be maintained in an unfolded state, and the unfolded polypeptide is stabilized by cytosolic chaperones, which will spontaneously leave before transport into the ER (Plath & Rapoport, 2000). The SEC complex, including the Sec61p complex and Sec62p/Sec63p complex, directly interacts with the pre-signal sequence, which happens in an SRP-independent manner (Kalies & Hartmann, 1998). Sec62p recognizes the signal peptide and simultaneously, the signal peptide is bound to Sec61p and Sec62p (Plath *et al.*, 2004). The unfolded polypeptide chain is pulled into the ER by binding to the ER luminal chaperone Kar2p (BiP). The ATPase activity of Kar2p provides a driving force for post-translational translocation (Matlack *et al.*, 1999).

The commonly used signal peptide is composed of a pre-region and a pro-region, which determines the translocation mode, protein solubility and trafficking efficiency (Rakestraw *et al*., 2009). The signal peptide can be a native, a heterologous, or a synthetic (designed) one. It is so far not possible to forecast which signal peptide is the most effective in secreting a specific protein. For example, comparing two different signal peptides, the *S. cerevisiae* native α-factor leader was more effective for the secretion of α-amylase, while the insulin precursor had a higher final yield using the synthetic leader YAP3-TA57 (Liu *et al*., 2012). Both signal peptides will be described in **Paper I**. Combination of four mutations in the native α-factor leader (A9D, A20T, L42S, D83E) resulted in significantly improved secretion of several fungal oxidoreductases and hydrolases in *S. cerevisiae*. Based on the above optimized signal leader, the additional introduction of the randomization of positions 86 and 87 was able to further increase protein secretion yields (Aza *et al.*, 2021). In another approach, the α-factor pre-signal sequence was replaced with the *S. cerevisiae* Ost1p signal sequence (the alpha-subunit of the oligosaccharyltransferase complex), which conversed post-translational translocation to co-translational translocation (Besada-Lombana & Da Silva, 2019). This approach improved the secretion of monomeric superfolder GFP (Fitzgerald & Glick, 2014). In addition, overexpressing the translocation components, such as Srp54p and Srp14p, and introducing *S. cerevisiae* cytosolic chaperones to *P. pastoris*, such as Ssa1p and Ydj1p, had a positive effect on protein secretion (Zhang *et al*., 2006, Tang *et al*., 2015). Engineering Hsp70 chaperone cycles can provide both pushing force on the cytosolic side and pulling force on the ER side to increase the flux of recombinant protein through the translocation pore (Zahrl *et al.*, 2022).

1.3.2.2 Protein glycosylation

Most of the secreted proteins are glycosylated, and the early glycosylation occurs during translocation. In yeast, there are two types of glycosylation, N-linked and O-linked. N-linked glycosylation can be divided into three processes: assembly of the lipid-linked oligosaccharide precursor Glc3Man9GlcNAc2 (Glu represents glucose, Man represents mannose, and NAc represents N-acetylglucosamine) at the ER membrane mediated by multiple mannosyltransferases (encoded by *ALG* genes), recognition of the asparagine residue within the sequence asparagine-X-serine/threonine (X can be any amino acid except proline), and transfer of the completely assembled Glc3Man9GlcNAc2

to the carboxamido nitrogen on the asparagine residue of the nascent polypeptide mediated by the oligosaccharyltransferase complex (Burda & Aebi, 1999). Further N-glycosylation, which occurs in the Golgi apparatus, varies with the proteins and species, that is, the N-linked glycan undergoes further maturation by adding an outer chain containing up to 200 mannose residues (Jungmann & Munro, 1998). For O-linked glycosylation, a single mannose from dolichyl phosphate-bound mannose is transferred to the hydroxyl groups of serine and threonine of the nascent polypeptide, mediated by protein O-mannosyl transferases (PMTs) in the ER. Subsequently, this mannosylated protein is further processed to form a linear chain of up to 5 mannose residues Manα1-3Manα1-3Manα1-2Manα1-2Man-O in the Golgi apparatus mediated by α-1,2-mannosyltransferases (MNTs) and α-1,3-MNTs (Goto, 2007).

Protein glycosylation has many advantages, such as increasing protein stability, solubility and folding; protecting proteins against proteases, and playing a general role in the quality control of protein folding (Ellgaard & Helenius, 2003, Parthasarathy *et al.*, 2006, Demain & Vaishnav, 2009, Roth *et al.*, 2010). It should be noted that the production of human therapeutic glycoproteins with human-like N-glycosylation currently accounts for 70% of all approved therapeutic recombinant proteins (Idiris *et al.*, 2010). However, yeast performs non-human N-glycosylation of the high-mannose type (see above). Given the therapeutic value of glycoproteins, humanizing yeast glycosylation systems has turned out to be one of the main glycosylation engineering strategies since humanized N-glycan production in yeast was first attempted in 1992 (Nagasu *et al.*, 1992). Mammals and yeasts share the same glycosylation pathway in the ER, while they have a distinctly different series of glycosyltransferase reactions in the Golgi (Chiba & Akeboshi, 2009). Many optimizations focused on engineering these reactions to introduce human-like glycosyl structures (Chiba *et al.*, 1998, Hamilton *et al.*, 2006, Hamilton & Gerngross, 2007). For example, through the deletion of *OCH1* and thus eliminating the endogenous glycosylation pathway and the introduction of five active eukaryotic glycosylation-related proteins, including MnsI, MnsII, GnTI, GnTII and uridine 5'-diphosphate-GlcNAc transporter, Hamilton *et al*. successfully secreted a human glycoprotein with complex N-glycosylation in *P. pastoris* (Hamilton *et al.*, 2003). Besides the N-glycosylation system, mammalian-type O-glycosylation has also been established in *S. cerevisiae* (Oka & Jigami, 2006, Chigira *et al.*, 2008).

1.3.2.3 Protein folding and quality control in the ER

After translocation, the protein needs to experience a range of modifications in the ER, including glycosylation, signal peptide cleavage, protein folding and disulfide bond formation. These processes are strictly regulated by a quality control system that determines if a protein is correctly folded before it is allowed to leave the ER into the Golgi. Upon extrusion into the ER, the pre-signal sequence is cleaved by signal peptidase immediately. The unfolded protein exposes the hydrophobic amino-acid residues that should be buried inside the folded protein. Therefore, the ER chaperone Kar2p recognizes transiently and covers hydrophobic patches to prevent protein aggregation (Simons *et al.*, 1995). Next, the unfolded chaperoned protein will perform cycles of binding and releasing from Kar2p until it no longer presents any Kar2p-binding motifs. At each release, the protein will have the

opportunity to advance along the folded path (Gething, 1999). Disulfide bond formation between thiol groups in two cysteine residues of the polypeptide chain is an important part of assisting protein folding and stabilizing the protein in its correct conformation. This process requires transferring electrons to PDI (Pdi1p in *S. cerevisiae*), a protein catalyzing disulfide formation, disulfide reduction, and rearranging incorrect disulfides. Subsequently, the electrons pass further to the FAD-dependent ER oxidase Ero1p, and then to molecular oxygen. The oxygen as the final electron acceptor can result in oxidative stress through the production of reactive oxygen species (ROS) (Sevier & Kaiser, 2008). Until now, the enzymes detoxifying ROS in the ER of yeasts have not been identified. PDI is re-oxidized by Ero1p. PDI and its four homologues (Eug1p, Mpd1p, Mpd2p, and Eps1p) are responsible for disulfide rearrangement and reduction (Tu & Weissman, 2004).

If the peptide is synthesized too fast and protein folding exceeds the normal capacity of the processing machinery of the ER, this will form misfolded or unfolded proteins. These proteins will expose hydrophobic amino acid residues and then form stable complexes with Kar2p and other molecular chaperones. Decreased concentrations of free Kar2p will induce an unfolded protein response (UPR) (Kohno *et al.*, 1993). UPR is a transcriptional response that adjusts secretory sources to assist correction of damage caused by unfolded proteins, clears misfolded proteins, and increases oxidative stress on a large scale. This signal transduction process from the ER to the nucleus involves three major proteins, including ER transmembrane protein Ire1p, which senses abnormally high levels of unfolded ER proteins and stimulates its endoribonuclease activity; transcriptional activator Hac1p, which directly activates the transcription of UPR target genes by binding to the UPR elements; and tRNA ligase Rlg1p, which rejoins the *HAC1* exons allowing translation to continue after a nonconventional intron has been removed by Ire1p (Patil & Walter, 2001). Hac1p activates approximately 380 genes, including those encoding multiple ER chaperones and PDIs, enzymes for protein transport compartments, as well as proteins involved in phospholipid biosynthesis and ER-associated degradation (ERAD) (Travers *et al.*, 2000). If proteins reside in the ER for too long, they will be recognized as aberrant via the detection of some special structures such as exposed hydrophobic regions, unpaired cysteine residues and immature glycans. Then, these misfolded proteins will be targeted to the ERAD system to perform their clearance. If misfolding occurs in the ER lumen, ER transmembrane protein Der1p initiates the export of misfolded proteins from the ER lumen by recruiting and inserting them into the ER membrane and then routing them to ubiquitin ligase Hrd1p for polyubiquitylation. These ubiquitinated proteins are pulled out of the ER membrane to the cytoplasm through a driving force provided by ubiquitin-binding factors, such as the Cdc48p-Npl4p-Ufd1p complex, and then targeted for degradation by the proteasome (Meusser *et al.*, 2005, Mehnert *et al.*, 2014, Wu & Rapoport, 2018). As one another response pathway to ER stress, protein synthesis will be transiently attenuated to relieve the burden on the ER (Yoshida, 2007).

Protein folding and quality control tend to be the most rate-limiting step in protein secretion as they determine whether the protein should enter the following secretory process, for properly folded proteins, or should be targeted to the degradation pathway, for misfolded proteins. Thus, engineering protein folding and quality control systems has turned out to be one of the most helpful ways to

enhance heterologous protein secretion. A straightforward and effective strategy is to overexpress multiple molecular chaperones, co-chaperones, PDIs and other folding helpers in *S. cerevisiae*, such as Kar2p (Hou *et al.*, 2012, Huang *et al.*, 2014), Sil1p, Lhs1p, Scj1p and Jem1p (Payne *et al.*, 2008), Pdi1p and Ero1p (Hou *et al.*, 2012, Huang *et al.*, 2014, Huang *et al.*, 2018, Qi *et al.*, 2020), Hac1p (Valkonen *et al.*, 2003, Lee *et al.*, 2012), and Hsf1p (heat shock response transcription factor), which regulates the transcription of hundreds of genes coding for molecular chaperones (Hou *et al.*, 2013). Since heterologous proteins can be misrecognized as aberrant, disruption of the ERAD machinery by knocking out Der1p could minimize the retrotranslocation of proteins into the cytoplasm and thereby increase protein secretion capacity (Besada-Lombana & Da Silva, 2019). In addition, strains deleted in other ERAD components such as Δ*yos9* Δ*ire1* and Δ*htm1* Δ*ire1* showed an increased immunoglobulin G (IgG) ER residence time and therefore reduced the IgG clearance from the cells by ERAD (de Ruijter & Frey, 2015). Many studies showed that increasing ER size could effectively reduce ER stress caused by protein overexpression (van Anken *et al.*, 2003, Schuck *et al.*, 2009). Indeed, the secretion capacity of IgG, when expressed in an ER-enlarged *S. cerevisiae* strain by deleting the negative lipid regulatory gene *OPI1*, was 4.8-fold higher than in the wild-type (de Ruijter *et al.*, 2016). However, these general engineering strategies are often protein- or host-specific and not beneficial for the production of all proteins (Davis *et al.*, 2000, van der Heide *et al.*, 2002, Korpys-Woźniak *et al.*, 2021).

1.3.2.4 Intracellular vesicle trafficking

Correctly folded proteins are transported from the ER to the Golgi apparatus to be further processed. This transportation process is mediated by coat protein complex II (COPII) vesicles. GTPase Sar1p triggers the formation of the COPII vesicle at the ER exit sites. The ER transmembrane protein Sec12p acting as a guanine nucleotide exchange factor (GEF), activates Sar1p-GDP by exchanging GDP with GTP. The activated GTP-bound Sar1p is inserted into the ER membrane and then recruits the Sec23p/Sec24p complex to form the pre-budding complex, which is the inner layer of the coat of which, Sec23p interacts with Sar1p, while Sec24p recognizes and captures the cargo proteins into the nascent vesicle (Wendeler *et al.*, 2007). Sec16p stabilizes the pre-budding complex to facilitate the vesicle coat assembly (Supek *et al.*, 2002). Then, the pre-budding complex (Sar1p/Sec23p/Sec24p) recruits the Sec13p/Sec31p complex that polymerizes to form the outer layer of the coat. After the intact COPII vesicle forms, it will leave the ER membrane and move toward the Golgi membrane. This will pass through tethering, docking and fusion processes.

The precise selection system and efficient trafficking process of COPII vesicles affect the efficiency of ER export of the cargo proteins. Much work has demonstrated the potential to improve protein secretion by engineering protein transport. Moderate overexpression of Sec16p facilitated the translocation of different model proteins (Bao *et al.*, 2017). Overexpression of the protein components of COPII vesicles, such as Sec13p, improved the secretion of heterologous proteins (Tang *et al.*, 2017). The well-characterized transmembrane receptor Erv29p selects and packages glycosylated pro-α-factor (gpαf) fusion cargo proteins into COPII vesicles, and its overexpression increased the secretion of several recombinant proteins (Huang *et al.*, 2018, Besada-Lombana & Da Silva, 2019). It was also

shown that overexpression of components of some trafficking complexes, such as Sec1p/Munc18p (SM) proteins and conserved oligomeric Golgi complex protein Cog5p, improved protein production (Hou *et al.*, 2012, Huang *et al.*, 2018). Engineering vesicular trafficking pathway between ER and Golgi will also be discussed in **Paper II** (**Chapter 3.2**).

1.3.2.5 Golgi processing

In addition to hyper-mannosylation of glycoproteins (mentioned in **Chapter 1.3.2.2**), protein maturation requires cleaving the pro-leader-fusion protein by Golgi resident proteases before it is trafficked to the cell membrane. This determines protein quality and the sorting process. Corresponding to the recognition sites, there are three types of proteases, including Kex1p, Kex2p and Ste13p (Hou *et al.*, 2012). The most well-studied protease Kex2p cleaves dibasic Lys/Arg-Arg sites (Le Marquer *et al.*, 2019). Many strategies have been applied to increase the downstream process efficiency and decrease the secretion of unprocessed pro-proteins. Overexpression of Kex2p protease in *P. pastoris* enhanced the secretion of recombinant two-chain insulin glargine (Sreenivas *et al.*, 2015). In some cases, a spacer is added at the N-terminus to the secreted protein, which could provide a hydrophilic environment and therefore facilitate Kex2p cleavage (Liu *et al.*, 2012, Wang *et al.*, 2021).

After Golgi processing, the return of membrane components, cargo adaptor proteins, trafficking relevant proteins and escaped ER resident proteins from Golgi to the ER is crucial to ensure the supply of these resources in the ER. This retrograde transport process is mediated by coat protein complex I (COPI) vesicles (Szul & Sztul, 2011). Similar to COPII vesicles, GTPase Arf1p triggers the formation of the COPI vesicle at the *cis*-Golgi. Gea1p/Gea2p acting as a GEF activates Arf1p-GDP by exchanging GDP with GTP (Peyroche *et al.*, 1996). The conformational change of activated GTP-bound Arf1p renders the hydrophobic N-terminus more exposed, which then interacts with the Golgi membrane. The activated GTP-bound Arf1p recruits the coatomer, including seven different protein subunits α, β, β', γ, δ, ε and ζ (Cop1p, Sec26p, Sec27p, Sec21p, Ret2p, Sec28p and Ret3p) (Gaynor *et al.*, 1998). After the intact COPI vesicle forms, it will leave the Golgi membrane and move toward the ER membrane. Once the COPI vesicle arrives at the ER membranes, it will be disassembled and release the cargo proteins. The process is mediated by the inactivation of Arf1p through GTP hydrolysis to GDP by GTPase activating proteins (GAPs) Gcs1p, Glo3p, Age1p and Age2p (Poon *et al.*, 1999, Benjamin *et al.*, 2011). The selection process of cargo proteins is mediated by different retrieval systems (Gaynor *et al.*, 1998, Arakel & Schwappach, 2018). 1) For direct retrieval, dilysine (KKXX)-containing proteins are directly packaged into vesicles by interacting with the COPI subunits. 2) Soluble ER resident HDEL-containing proteins and Sec12p interact with specific receptors Erd2p and Rer1p in the *cis*-Golgi, respectively, and are then packaged into vesicles. 3) Other cargo proteins such as adaptor or receptor proteins and trafficking proteins need to continuously cycle between the ER and *cis*-Golgi. It has been shown that overexpression of Gcs1p and Glo3p could expand the ER membrane and increase the secretion of heterologous proteins (Bao *et al.*, 2018).

1.3.2.6 Exocytosis

After maturation, the proteins will exit the Golgi and be sorted to the many different final destinations, including reverse travel to ER by COPI vesicles, entry into the vacuole either directly (like alkaline phosphatase pathway) or passing through endosomes (vacuolar protein sorting), and entry into secretory vesicles to arrive at the plasma membrane or extracellular environment (exocytosis/secretory pathway) (Feyder *et al.*, 2015). After vesicle trafficking, tethering and docking processes, secretory vesicles from the Golgi will be fused with the plasma membrane. This fusion process is mediated by the exocyst complex composed of Sec3p, Sec5p, Sec6p, Sec8p, Sec10p, Sec15p, Exo70p and Exo84p, Rab family GTPase Ypt31p/Ypt32p and Sec4p, SM protein Sec1p, and soluble N-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs) including Sec9p, Snc1p/Snc2p, Spo20p and Sso1p/Sso2p (Delic *et al.*, 2013). After being secreted from the plasma membrane, proteins diffuse through the cell wall to the external medium.

The exocytosis process represents another possible bottleneck for protein secretion. One study demonstrated that overexpression of Sec1p, involved in docking and fusion of exocytic vesicles, increased the secretion of heterologous and endogenous proteins (Hou *et al.*, 2012). Tang *et al.* upregulated several proteins concerned with different stages of exocytosis, including Sso1p, Snc2p, Sec1p, Exo70p, Ypt32p and Sec4p, which had positive effects on the secretion of β-glucosidase from *Sacchromycopsis fibuligera*. In addition, this engineering strategy also enhanced the surface display efficiency of heterologous proteins (Tang *et al.*, 2017). Many other studies improved the secretion capacity by overexpression of different exocytosis targets (Ruohonen *et al.*, 1997, Toikkanen *et al.*, 2003, Liu *et al.*, 2005, Gasser *et al.*, 2007). On the other hand, manipulating cell wall-related proteins can be a new strategy for improving protein production. Disruption of Cwp2p, a mannoprotein stabilizing the cell wall, increased extracellular cellobiohydrolase activity by 85.9%. Furthermore, the authors observed that disruption of Cwp2p could upregulate some translation and secretory pathway-relevant genes (Li *et al.*, 2020). Deletion of β-1,3-glucanosyltransferase gene *GAS1*, required for cell wall assembly, resulted in the improved secretion of several recombinant proteins (Vai *et al.*, 2000, Marx *et al.*, 2006, Passolunghi *et al.*, 2010). If recombinant proteins accumulate in the periplasmic space bounded by the cell wall, endocytosis might significantly reduce protein productivity. Therefore, knock-down of endocytosis genes *RVS161* and *END3* resulted in a 2-fold and 3-fold increase in secreted α-amylase compared to controls, respectively (Rodríguez-Limas *et al.*, 2015).

1.3.3 Other factors affecting protein production

In addition to optimization of protein synthesis and secretion, successful improvements of protein titers to a considerably high level require the adjustment of environmental conditions and fermentation processes. They are closely related to cell growth, secretion mechanisms and stability in the extracellular medium. The temperature has an important effect on cellular metabolism and regulates folding-related proteins. It has been reported that protein titers increased when the cultivation temperature was lowered from 30 ℃ to 20 - 25 ℃ (Shusta *et al*., 1998, Li *et al*., 2001, Hackel *et al*., 2006). An optimized amino acid supplement medium led to 8-fold higher *S. pombe* phosphatase

levels compared to the unsupplemented synthetic dextrose medium (Wittrup & Benig, 1994). A study indicated that adding BSA to the medium could effectively prevent human insulin degradation (Tyo *et al.*, 2014). Controlling the dilution rate of a chemostat culture is also a good strategy to improve protein production. For example, α-amylase has a high yield in a low specific growth rate, while the production of human insulin precursors occurs in a growth-associated manner (Liu *et al.*, 2012, Liu *et al.*, 2013). Based on previous studies, oxygen-limiting or hypoxic conditions can be favorable for protein secretion by activating hypoxia-induced genes (Liu *et al.*, 2013, Liu *et al.*, 2015, Martínez *et al.*, 2015, Huang *et al.*, 2017). As mentioned in **Chapter 1.3.2.3**, during the formation of disulfide bonds, electrons are eventually transferred to oxygen, leading to the induction of ROS and thus ER oxidative stress in oxygen-rich conditions. A previous work demonstrated that under anaerobic conditions, fumarate replaces oxygen as the final electron acceptor in *S. cerevisiae*, which reduced cellular stress (Liu *et al.*, 2013). Thus, analyzing the external and internal factors is of great importance for obtaining high-level protein production.

1.4 Genetic engineering

Heterologous protein synthesis and secretion compete with endogenous proteins that need to be processed through the protein secretion pathway. This competition might cause an unbalanced allocation of cellular resources and therefore limit the protein secretory capacity of host cells. As described in **Chapter 1.3**, many efforts aim at customizing the subsystems of protein synthesis and secretion to overcome possible obstacles and achieve the desired product quality. High-level production of these proteins can require gene modifications including gene overexpressions, gene deletions, or point mutations (Mans *et al.*, 2015). Gene overexpression is a genetic manipulation that increases gene expression levels beyond the regular pattern. The target gene can be transcribed and translated more efficiently using plasmids-based and chromosomal integration-based approaches (Prelich, 2012). The relevant engineering strategies have been described in **Chapter 1.3.1**. Gene deletion is a genetic manipulation that involves the complete inactivation of a gene of interest due to the loss of genetic material. Gene overexpression and deletion have become routine strategies for designing cell factories to improve recombinant protein production (Thak *et al.*, 2020). Moreover, systems biology approaches significantly contribute to the understanding of the underlying cellular mechanisms and thus facilitate the prediction of cell behavior and the selection of deletion and overexpression targets (Choi *et al.*, 2010, Nocon *et al.*, 2014). For example, 84 gene targets were predicted for deletion or overexpression by the genome-scale metabolic model (GEM) Yeast8, and the combinatorial engineering of the 13 validated positive targets led to a 70-fold increase in intracellular heme production in *S. cerevisiae* (Ishchuk *et al.*, 2022). Despite many successful cases, gene overexpression and deletion can sometimes be detrimental to cellular growth. For instance, gene overexpression may trigger stoichiometric imbalances, resource overload, and disruption of interactions and pathway modulation (Kintaka *et al.*, 2016, Eguchi *et al.*, 2018). Gene deletion may result in lethal or substantial genetic damage. In addition, if the gene to be deleted overlaps with the regulatory or genetic information of neighboring genes, the expression of neighboring genes will be affected, therefore generating an erroneous conclusion (Stern, 2014). Point mutation is a genetic

manipulation where one or few nucleotide bases are substituted, inserted, and deleted from the DNA or RNA sequence (Levi *et al.*, 2020). An effective point mutation within the coding or even non-coding region of a specific gene can be used to fine-tune the expression level, alter protein activity or protein regulation (e.g. changing phosphorylation sites), which has been proved to be an efficient tool in inverse metabolic engineering (Hong *et al.*, 2011, Bricker *et al.*, 2012, Liu *et al.*, 2014). For example, α-amylase production was improved 35% by introducing a S196I mutation in the *VTA1* gene involved in endosomal protein sorting in *S. cerevisiae* (Liu *et al.*, 2014). Similarly, the mutation of *RAS2*(D112Y), involved in the Ras/PKA signaling pathway, showed a faster galactose utilization and a 10% higher specific growth rate than the control (Hong *et al.*, 2011). Gene editing on the chromosome is often performed by homologous recombination. High-frequency recombination is usually obtained using double-strand DNA fragments with longer upstream and downstream homologous arms (Solis-Escalante *et al.*, 2013).

Since five direct clustered repeats in *E. coli* were observed in 1987 (Ishino *et al.*, 1987), the CRISPR (Clustered regularly interspaced short palindromic repeats)/Cas (CRISPR-associated) system was first discovered and subsequently triggered a revolution in the genetic engineering field (Figure 5). The widely used type II CRISPR/Cas9 system derived from *Streptococcus pyogenes* is one of the best-characterized systems, containing only one single protein - the nuclease Cas9, and two noncoding RNAs - crRNA (CRISPR RNA) and tracrRNA (trans-activating crRNA). These two noncoding RNAs have been fused into a chimeric single guide RNA (gRNA), which can bind to Cas9 protein and guide Cas9 to its target site flanked by the protospacer adjacent motif (PAM) and therefore cleave the DNA sequence and introduce double-stranded breaks (DSBs) (Jinek *et al.*, 2012). These DSBs can be repaired by two cellular DNA repair pathways: the error-prone non-homologous end joining (NHEJ) pathway, which frequently leads to insertion or deletion mutations in the absence of repair templates, or the high-fidelity homology-directed repair (HDR) pathway, which mediates precise DNA modifications in the presence of exogenous repair templates (Sánchez-Rivera & Jacks, 2015). CRISPR/Cas9-mediated gene deletion and point mutation strategies will be described in detail in **Paper II** (**Chapter 3.2**). Due to its advantages of convenient operation, high mutation efficiency and low cost, the CRISPR/Cas9 system has become the most rapidly developed targeted gene editing technology in various organisms (Hsu *et al.*, 2014). Similar to the original Cas9, the deactivated version (dCas9) generated by point mutations D10A and H840A is still able to bind to its gRNA and the targeted DNA strand. If a transcriptional activator (CRISPRa), such as *VP64-p65-Rta*, is attached to dCas9 or gRNAs, it will facilitate the recruitment of transcription factors and RNA polymerase for transcription of the target gene and therefore increase gene expression (Gilbert *et al.*, 2014). On the other hand, if dCas9 binds to a transcriptional initiation site or gene coding sequence, the transcriptional initiation or elongation will be blocked, thus repressing gene expression. This is called CRISPR interference (CRISPRi) (Larson *et al.*, 2013). Both are fine-tuning orthogonal transcriptional regulation systems compared to the traditional methods of controlling gene expression by changing promoter strength.

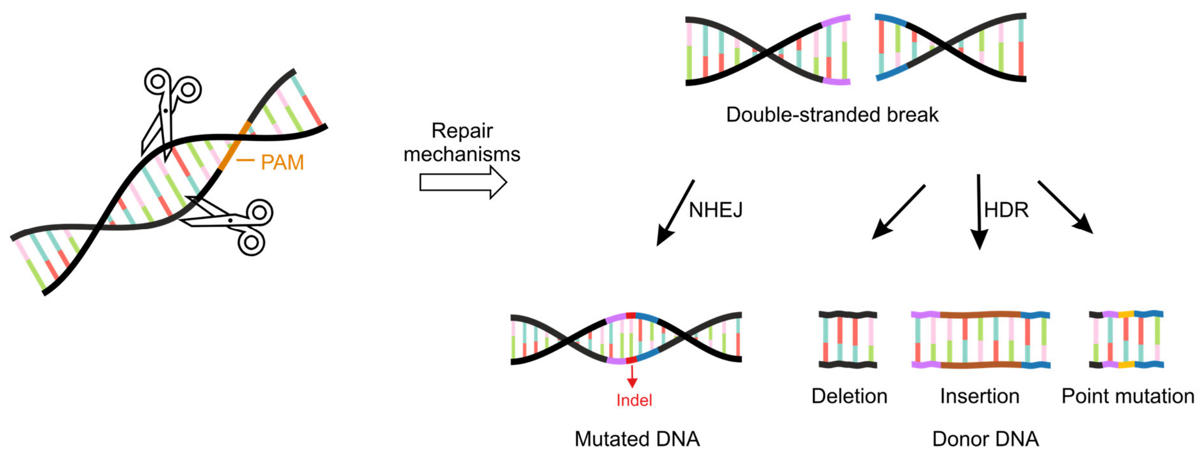

Figure 5. Genome editing using CRISPR/Cas9 technology. NHEJ: Non-homologous end joining; HDR: Homology-directed repair.

1.5 Strain library construction and screening techniques

Since there are multiple processes involved in protein secretion, there may be many potential variables that need to be optimized. However, reasonably predicting or engineering these gene targets is very challenging due to lack of the clear understanding of the protein secretion process. Library construction combined with efficient screening methods can identify effective gene targets and further help understanding the protein secretion mechanism.

1.5.1 Strain library construction techniques

In general, libraries can be divided into three categories, including element libraries, multi-element combination libraries, and genome-scale libraries. (I) An element library characterizes a single genetic element. For example, promoter libraries and terminator libraries with a wide dynamic range were developed to express genes at different levels (Alper *et al.*, 2005, Rajkumar & Maerkl, 2012, Yamanishi *et al.*, 2013, Liu *et al.*, 2021). A promoter library containing 30,000 mutants was generated via random mutagenesis of the constitutive *GAP1* promoter in *P. pastoris*, among these mutants, 33 mutants were collected, which spanned a yEGFP fluorescence range between approximate 0.6% and 19.6-fold compared to the wild-type with *GAP1p* (Qin *et al.*, 2011). As mentioned in **Chapter 1.3.1.2**, by evaluating 5302 terminators, compared to the control strain with *PGK1t*, 5 terminators (*DIT1t*, *RPL41Bt*, *RPL15At*, *RPL3t*, and *IDP1t*) increased GFP expression more than twice (Yamanishi *et al.*, 2013). Among them, the strain with *DIT1t* resulted in a 2.2-fold higher endoglucanase II secretion than the control strain with *PGK1t* (Ito *et al.*, 2013). In addition to promoters and terminators, the use of different signal peptides can have a significant impact on the efficiency of recombinant protein secretion. For instance, directed evolution of α-factor leader peptides has led to the increased secretion of the antibody fragment scFv and the full-size antibody IgG in *S. cerevisiae* (Rakestraw *et al.*, 2009). Gene libraries have been created through many approaches (Packer & Liu, 2015), such as random mutagenesis (Wilson & Keefe, 2001), focused mutagenesis (Wells *et al.*, 1985, Li *et al.*, 2020), and genetic recombination technique (Coco *et al.*, 2001, Wang *et al.*, 2017). (II) A multi-element combination library considers multiple gene expression controlling elements in one system (Choi *et*

al., 2003, Wingler & Cornish, 2011). A previous study constructed a library with 17 control elements, including 4 promoters, 10 signal peptides, 1 terminator, and 2 origins of replication, and 2 fluorescent proteins, to evaluate their effect on protein secretion (Obst *et al.*, 2017). (III) A genome-scale library investigates variants from the entire genome. For instance, a cDNA library is a collection of cloned DNA open reading frames that are complementary to mRNA extracted from an organism, and such a cDNA library can be cloned into expression vectors and thus used as genome-scale library (Soares *et al.*, 1994, Si *et al.*, 2017). In our lab, genome-wide random mutagenesis libraries were generated by exposing yeast strains to different doses of UV light (Liu *et al.*, 2014, Huang *et al.*, 2015). Another study combined a yeast gene deletion library and an Aβ42 expression strain to select gene targets that mitigate cytotoxicity (Chen *et al.*, 2020). Moreover, *Wang et al.* combined a genome-scale RNA interference (RNAi) library to explore the down-regulated gene targets for increased protein production (Wang *et al.*, 2019). As described in **Chapter 1.4**, the CRISPR system has been modified and developed into different libraries, such as CRISPRa libraries, CRISPRi libraries, and CRISPR disruption (CRISPRd) libraries (Shariati *et al.*, 2019). Other methods to generate genome-scale libraries are CRISPR–Cas9- and homology directed-repair (HDR)-assisted genome-scale engineering (CHAnGE) (Bao *et al.*, 2018), multiplex automated genome engineering (MAGE) (Wang *et al.*, 2009), trackable multiplex recombineering (TRMR) (Warner *et al.*, 2010). In our work, a genomic random mutagenesis library and a yeast single gene deletion library were used, with the latter two libraries being mentioned in **Paper I, III** and **IIIb**.

1.5.2 Strain library screening techniques

Notably, libraries, especially genome-scale libraries, often generate large numbers of clones, which require a high-throughput screening method to evaluate these clones. Here, I divided the library screening methods for protein production into two categories, including growth-coupled screening methods, and fluorescence-coupled screening methods. (I) Growth-coupled screening methods use growth as a simple selection criterion to select optimized strain (Alter & Ebert, 2019). For example, the accumulation of toxic intermediates or final products can decrease cell growth. Synthetic genetic arrays (SGA) can be combined with visual examination of the size of the resulting colonies on plates by comparison with the parental strain, thereby screening for mutants that exhibit the desired phenotype (Tong & Boone, 2006). Liu *et al.* performed a screening method that combined UV random mutagenesis and growth on starch plates to select mutant colonies with high α-amylase secretion, which showed an up to 5-fold higher production of α-amylase than the control strain (Liu *et al.*, 2014). (II) Fluorescence-coupled screening methods use the strength of a fluorescence signal as indicator for the phenotype of interest. If the protein of interest is an enzyme, the fluorescent signal can be derived from the direct enzymatic reaction that occurs between the target protein and its substrate, or chemicals or tagged groups that are added into the enzymatic reaction (Fu *et al.*, 2021). Fluorescence-activated cell sorting (FACS) is a method that sorts heterogeneous mixture of cells into two or more groups for downstream experimentation, and was used to isolate genotypes that resulted in higher levels of proteins (Bleichrodt & Read, 2019, Grzeschik *et al.*, 2019). Droplet microfluidics has been developed as a high-throughput method for screening single cells with increased protein secretion

capacity, which can be used for measuring secreted proteins and cell proliferation (Sjostrom *et al.*, 2014, Huang *et al.*, 2015, Wang *et al.*, 2019). The schematic workflow of droplet microfluidics screening and the SGA screening methods, using in this work, will be presented in detail in **Chapter 3** and **Chapter 4**, respectively.

1.6 Systems biology tools

An ideal expression platform maximizes the conversion of substrates to final products. By studying the literature and performing preliminary experiments, we are able to obtain some useful information about how to optimize the properties of host cells, such as the rate-limiting steps, the genes to be manipulated, and the robustness of the host. However, the same engineering strategy may not be valid for the promising production of all proteins. There can be specific challenges that exist for specific proteins due to the complexity of protein synthesis and secretion pathways. For example, simple and small proteins often fold fast and spontaneously, while complex and multi-domain proteins may need more assistance, such as molecular chaperones, to reach their native states (Liu *et al.*, 2012). In addition, a complex structural and regulatory network in yeast limits the rational metabolic engineering for generating optimal phenotypes. Therefore, it is desirable to understand the individual processes in detail and analyze the interactions between these processes.

Systems biology can be a holistic and predictive approach to biological research, and is a biology-based interdisciplinary field focusing on interactions within complex biological systems. It comprises computational and mathematical analysis and modeling of the target systems (Nielsen & Jewett, 2008, Tavassoly *et al.*, 2018). That is, it allows for analyzing the intracellular behavior by integrating omics data and mathematical modeling. In the past few decades, systems biology approaches have been developed as increasingly valuable tools for metabolic engineering of cell factories aiming to improve the production of a metabolite or a protein in *S. cerevisiae* (Stephanopoulos *et al.*, 2004, Tavassoly *et al.*, 2018). There are two main approaches when dealing with a systems biology problem: the top-down and the bottom-up approach. The top-down approach depends mainly on experimental results based on omics data and high-throughput analysis. The bottom-up approach is used to generate and construct mathematical models. The bottom-up approach often incorporates experimental data to achieve global information and reconstruct the dynamic systems for optimization of cell factories (Nielsen & Jewett, 2008).

1.6.1 Top-down approach

To model the interactions of biological processes in a system, the experimental techniques should be system-wide and as complete as possible. Therefore, high-throughput and omics techniques (i.e., genomics, transcriptomics, proteomics, metabolomics and fluxomics) are usually applied to collect quantitative data for constructing models (Romualdi & Lanfranchi, 2009). With the development, omics analysis has been widely applied for engineering host strains to indicate cellular responses to the production of recombinant proteins. The genome storing the genetic information can provide the basis for understanding the complexity of biological systems, and genome sequencing is often applied

after random mutagenesis and adaptive evolution experiments. For example, our laboratory previously identified 146 amino acid sequence altering mutations through the combination of high-throughput droplet microfluidic screening and whole-genome sequencing (Huang *et al.*, 2015). The transcriptome is a dynamic pattern of gene expression, which is affected by the environment (Velculescu *et al.*, 1997). Many studies showed a comprehensive transcriptome map by RNA-seq in yeast (Nagalakshmi *et al.*, 2008). In our lab, Huang *et al*. used RNA-seq to study the transcriptional response in ultraviolet (UV) mutant strains with high protein production (Huang *et al.*, 2017). Chen *et al*. indicated that FMN supplementation could reduce the misfolded protein load and increase the resistance to oxidative stress and therefore reduce the amyloid-β (Aβ) toxicity through a combination of SGA-based screening with RNA-seq (Chen *et al.*, 2020). Proteomics can investigate protein translation efficiency, PTMs and protein-protein interactions, and can in some cases be used as a validation of transcriptomics results (Di Bartolomeo *et al.*, 2020). Metabolomics measures metabolite levels to provide a more detailed landscape of cellular states under different environments. Generally, these omics data need to be combined in an integrated analysis, which can help to understand the protein production machinery and cell metabolism (Björkeroth *et al.*, 2020, Campbell *et al.*, 2020, Qi *et al.*, 2020, Yu *et al.*, 2020, Malina *et al.*, 2021).

1.6.2 Bottom-up approach

Mathematical models can investigate metabolism and predict the cellular behavior of yeast under different conditions. The first yeast GEM was reconstructed in 2003 (Förster *et al.*, 2003), and it was used as a scaffold for further metabolic model refinements that introduced more biological processes and constraints (Herrgård *et al.*, 2006, Nookaew *et al.*, 2008). Since the first reconstruction, multiple updates have been made for the *S. cerevisiae* GEM (Domenzain *et al.*, 2021), such as Yeast1 to Yeast8 (Chen *et al.*, 2022), addition of transcriptional regulatory constraints (Herrgård *et al.*, 2006), iron metabolism (Chen *et al.*, 2021), or proteome constraints (Sánchez *et al.*, 2017). These concepts can be applied to improve recombinant protein secretion. Several studies have shown models describing the protein secretion pathway in various organisms (Krambeck & Betenbaugh, 2005, Feizi *et al.*, 2013, Irani *et al.*, 2016, Gutierrez *et al.*, 2020). However, they either only simulate a part of the secretory pathway, such as N-linked glycosylation, or are not able to perform simulations. Thus, to develop optimal yeast platforms with high-level recombinant protein production, it is desirable to reconstruct a model that contains a description of the complete protein secretory pathway.

1.7 Aim and significance

The overall goal of this thesis was to improve recombinant protein production by optimizing *S. cerevisiae* platform strains. This thesis mainly focused on two parts: exploring in depth previously isolated mutant strains and investigating novel targets to rationally optimize *S. cerevisiae*.

Previously, our laboratory isolated a line of yeast mutant strains with different secretory capacities after UV random mutagenesis. And the best-producing strain B184 showed an increased production of different heterologous proteins. Therefore, in **Chapter 3**, we characterized these mutant strains evolved for increased protein secretion. Firstly, as described above, therapeutic protein production plays an extremely important role in the biopharmaceutical industry. We thus explored whether the secretion capacities of mutant strains were generally suitable for producing different pharmaceutical proteins, such as antibody fragments (**Paper I**). On the other hand, given the fact that UV mutagenesis led to the improved protein production of these mutant strains, to systematically investigate beneficial targets that are responsible for the increased production, we re-introduced a total of 42 protein-sequence-altering point mutations into the parental strain (**Paper II**).

As novel targets can serve as guidelines for designing cell platforms, to further improve protein production, **Chapter 4** is dedicated to investigating novel targets to rationally optimize *S. cerevisiae* strains. High protein production often induces cellular stress, which is supposedly similar to a humanized AD model with the accumulation of misfolded Aβ peptides. Previously, genes that suppress Aβ42 aggregation-related cytotoxicity have been identified in *S. cerevisiae*. We therefore examined whether these suppressors could alleviate cellular stress and increase recombinant protein production in *S. cerevisiae* (**Paper III** and **IIIb**). Since the complexity of the protein secretory pathway often limits its rational engineering, we reconstructed a more systematic model, namely a proteome-constrained genome-scale protein secretory model of *S. cerevisiae* (pcSecYeast), to predict overexpression targets for improving recombinant protein production (**Paper IV**).

This thesis provides various efficient strategies to develop yeast platforms for the high-level production of valuable industrial or pharmaceutical proteins, and also provides general guidelines for designing other cell platforms for efficient protein production.

2 Model proteins in this thesis

As already mentioned above, the application of recombinant proteins is rapidly expanding. This thesis mainly studied two types of heterologous proteins with different properties: antibody fragments as biopharmaceutical proteins (**Paper I**) and α-amylase as an industrial enzyme (**Paper II-IV**). The mainly used proteins are shown in Figure 6.

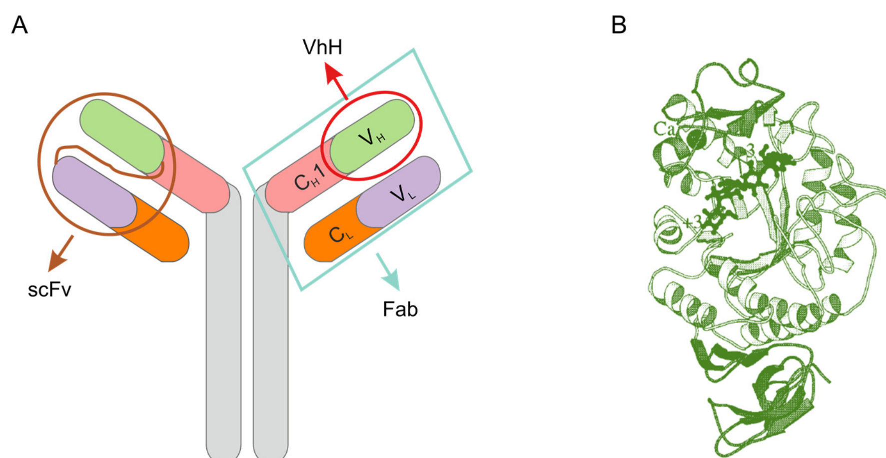

Figure 6. The structure of the mainly used proteins. (**A**) Three antibody fragments. V_H: heavy chain variable domain; V_L: light chain variable domain; C_H1: heavy chain constant domain; C_L: light chain constant domain; Fab: antigen-binding fragment containing 4 domains V_H, V_L, C_H1 and C_L; scFv: single-chain variable fragment, containing 2 domains V_H and V_L; VhH: single V-type domain, containing 1 domain V_H. (**B**) α-Amylase. Adapted from Brzozowski *et al*. (Brzozowski & Davies, 1997).

A typical intact antibody is a large Y-shaped multi-domain protein containing a basic structure of two heavy chains (H) and two light chains (L) (Figure 6). Each heavy chain comprises one variable (V_H) and three constant domains (C_H1, C_H2, C_H3), and each light chain consists of one variable (V_L) and one constant domain (C_L). The antigen-binding sites are located on the variable regions at the tips of the Y-arms. Although antibody fragments are structurally different from conventional mAbs, they still retain the antigen-binding regions. Various specific antibody fragments could be obtained by proteolysis, such as antigen-binding fragment (Fab, containing V_H, V_L, C_H1 and C_L), single-chain variable fragment (scFv, contains V_H and V_L connected by a flexible polypeptide linker), and single V-type domain (a single monomeric V_H). Three classes of mAb-derived fragments represent successive waves of antibody fragment technologies (Holliger & Hudson, 2005, Nelson, 2010). In **Paper I**, we focused on three different antibody fragments, including a nanobody consisting of a single V-type domain (Nan) from the Camelidae family, the scFv peptide Pexelizumab (Pex) from human, and Fab fragment Ranibizumab (Ran) from human. Their characteristics are listed in Table 2.

Table 2. The characteristics of three antibody fragments used in this study.

Generic name	Type	Size	Disulfide bonds	Target	Indication
Nanobody (Nan, Camelid)	VhH	17 kDa	2	Lysozyme	Rheumatoid arthritis and Crohn´s disease (Muyldermans *et al.*, 2009)
Pexelizumab (Pex, Humanized)	scFv	31 kDa	Uncertain	Complement C5	Coronary artery bypass and Angioplasty (Testa *et al.*, 2008)
Ranibizumab (Ran, Humanized)	Fab	51 kDa	5	VEGF-A	Macular degeneration (Martin *et al.*, 2011)

α-Amylase is one of the most popular industrial enzymes. In our studies (**Paper II-IV**), the selected α-amylase is from *Aspergillus oryzae* containing 3 domains with 478 amino acids, 4 disulfide bonds and 1 glycosylation site (Randez-Gil & Sanz, 1993). The biggest advantage of α-amylase is the ability to degrade starch, and therefore some well-developed assays can easily measure α-amylase activity. This provides a convenient detection method and facilitates the analysis of multiple strains at once.

3 Characterization of strains evolved for increased protein secretion

Our laboratory previously isolated 9 yeast mutant strains through several rounds of UV random mutagenesis and two different screening approaches, including manual screening (through directly measuring protein concentrations of all selected colonies from starch plates after UV treatment) and high-throughput droplet microfluidic screening (through first enriching strains with fluorescence intensity beyond a defined threshold by droplet microfluidics and then measuring protein concentration of the enriched strains) (Figure 7). In detail, the starting strain (AAC) expressing α-amylase was spread on starch plates and then immediately subjected to UV random mutagenesis (Liu *et al.*, 2014). Assessing a total of 591 colonies, the first mutant strain (M715) was obtained by manual screening. After that, M715 was subjected to UV random mutagenesis and the second (MH23) and third (MH34) mutant strains were selected by a high-throughput droplet microfluidic system (Huang *et al.*, 2015). Subsequently, MH23 and MH34 were subjected to another round of UV random mutagenesis and droplet microfluidic screening. Six strains, named F83, C120, A64, D5, B130, and B184, were obtained. Through whole-genome sequencing of the original strain AAC together with the nine mutants, 146 protein-sequence-altering mutations were identified, and a duplication of entire chromosome III was also found in MH34 and its descendants (Huang *et al.*, 2015). Among these mutant strains, B184 showed the best performance in α-amylase secretion (5.03-fold compared to AAC) and its ancestral strains are MH34 (2.20-fold compared to AAC), M715 (1.33-fold compared to AAC) and the parental AAC, respectively (Figure 7). In **Paper I** and **II**, we focused on studying this evolutionary path leading from AAC to B184.

B184 showed advantages for the production of different recombinant proteins, such as α-amylase, the *Rhizopus oryzae* glucan-1,4-α-glucosidase (GLA), HSA, and *Trichoderma reesei* endo-1,4-beta-xylanase II in previous studies (Huang *et al.*, 2015, Huang *et al.*, 2017). We thus proposed the question: could these engineered strains be generally suitable for pharmaceutical protein production? The study is presented in **Paper I**. Although strain B184 significantly increased the production of most recombinant proteins tested, the responsible genes have not been investigated systematically. As the increased production is to the result of UV random mutagenesis, we were curious: which mutations are beneficial for α-amylase production? A set of mutated genes selected from all mutant strains in this evolutionary line was subsequently tested. The study is presented in **Paper II**.

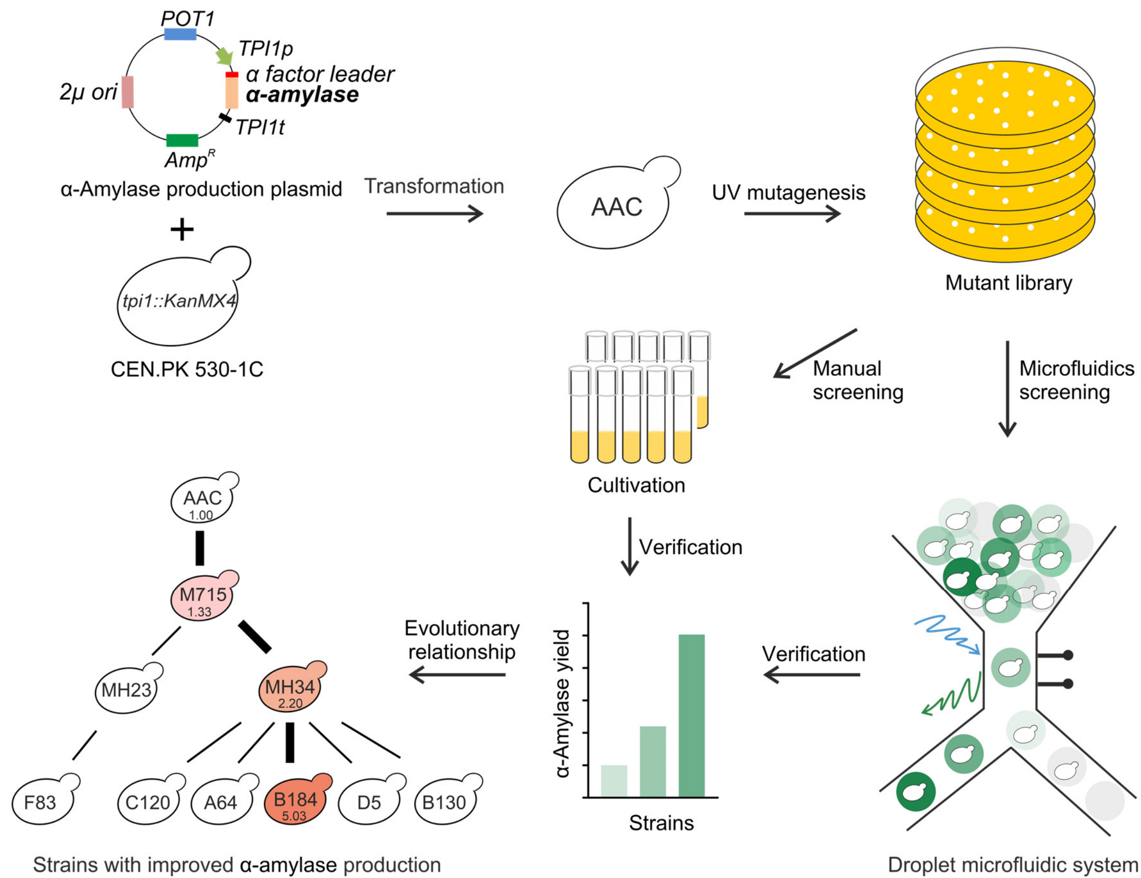

Figure 7. Schematic workflow for screening of *S. cerevisiae* mutant libraries for improved α-amylase production. In the first step, the starting strain AAC was constructed by introducing a COPTud-based α-amylase expression plasmid into a CEN.PK strain. Then, mutant libraries were generated by UV irradiation with different doses. Subsequently, single cells from these mutant libraries were screened according to two approaches. M715 was obtained by manual screening, which was carried out by evaluating the α-amylase concentration of 591 colonies from starch plates. On the other hand, to achieve high-throughput screening from large libraries, single cells were encapsulated to form droplets. These cells were able to grow in the droplets. An α-amylase substrate containing a starch backbone with multiple quenched fluorophores was added, which resulted in fluorescent products after conversion. Then, the droplets were sorted according to their fluorescent signals. Validation of α-amylase production capacity was followed. Finally, 9 mutant strains with improved protein production were selected and whole-genome sequencing was performed.

3.1 Evaluating the expression of antibody fragments in the mutant strains

This chapter is based on three antibody fragments described in **Chapter 2**, including Nan, Pex and Ran, and three strains, including parental strain AAC as well as mutant strains MH34 and B184. AAC is a CEN.PK 530-1C-based strain that is deleted in the *TPI1* gene, and expresses the α-amylase gene from a CPOTud plasmid that is often used for expression of recombinant proteins (Liu *et al.*, 2012) (Figure 7). As mentioned in **Chapter 1.3.1.2**, the *POT1* gene from *S. pombe* with its own promoter and terminator on the CPOTud plasmid complements the absence of the *TPI1* gene. Loss of the α-amylase expression plasmid from corresponding strains (AAC, MH34 and B184) will generate three host strains, including low-secretion strain (LA), medium-secretion strain (MA) and high-secretion strain (HA), respectively. These three plasmid-depleted host strains do not grow on glucose as the sole carbon

source and can obtain stable expression in rich medium and at high copy numbers when (re-)introducing the CPOTud plasmids containing the gene of interest.

3.1.1 Expression of antibody fragments in the high-secretion mutant strain

Since B184 was the best producer among all isolated mutant strains, I first tested whether the selected antibody fragments can be secreted from the high-secretion strain HA. As the antibody fragments Nan and Pex only contained one peptide, the synthesized sequences (α-factor leader-Nan/Pex-his tag) were directly integrated into the expression cassette of the CPOTud plasmid under control of the *TPI1* promoter (Figure 8A and 8D). However, Ran consists of two separate chains: heavy chain (Ran-H) and light chain (Ran-L). The balanced expression of these two chains is important for forming complete constructs. Since 2A-peptide-linked polypeptides are "self-cleaved" during translation, resulting in an equal ratio of the co-expressed proteins (Van der Weken *et al.*, 2019), I introduced a 2A peptide coding sequence between Ran-H and Ran-L (Figure 8G). To increase 2A "cleavage" efficiency, a GSG linker was added in front of the 2A peptide (Chng *et al.*, 2015). To avoid homologous recombination, two different signal peptides, α-factor leader and synthetic leader Yap3-TA57 were added at the N-termini of Ran-H and Ran-L, respectively (Kjeldsen *et al.*, 1999). The synthesized sequence for Ran expression was also integrated as an expression cassette into the CPOTud vector under control of the *TPI1* promoter (Figure 8G). The resulting three plasmids and the empty CPOTud were introduced into HA, generating different recombinant strains HA.Nan, HA.Pex, HA.Ran and HA.CPOTud, respectively.

Western blot was performed to confirm the expression and secretion of the different antibody fragments. Protein signals for the detection of Nan and Pex agreed with the expected sizes of 17 kDa and 31 kDa (Figure 8B and 8E). An additional band at ca. 14 kDa in the HA.Pex strain was probably due to protein hydrolysis (Figure 8E). Ran was detected under reducing and non-reducing conditions. Under reducing conditions, protein bands corresponding to the size of Ran-H (26 kDa) and Ran-L (25 kDa) were detected (Figure 8H). Under non-reducing conditions, there was a band corresponding to the full size of Ran (51 kDa). An additional band at the respective gel corresponding to the size of Ran-H and Ran-L was probably due to the incomplete assembly of the heavy and light chain fragments (Figure 8H). Western blot results indicated that all three antibody fragments could be expressed and secreted in the mutant strain.

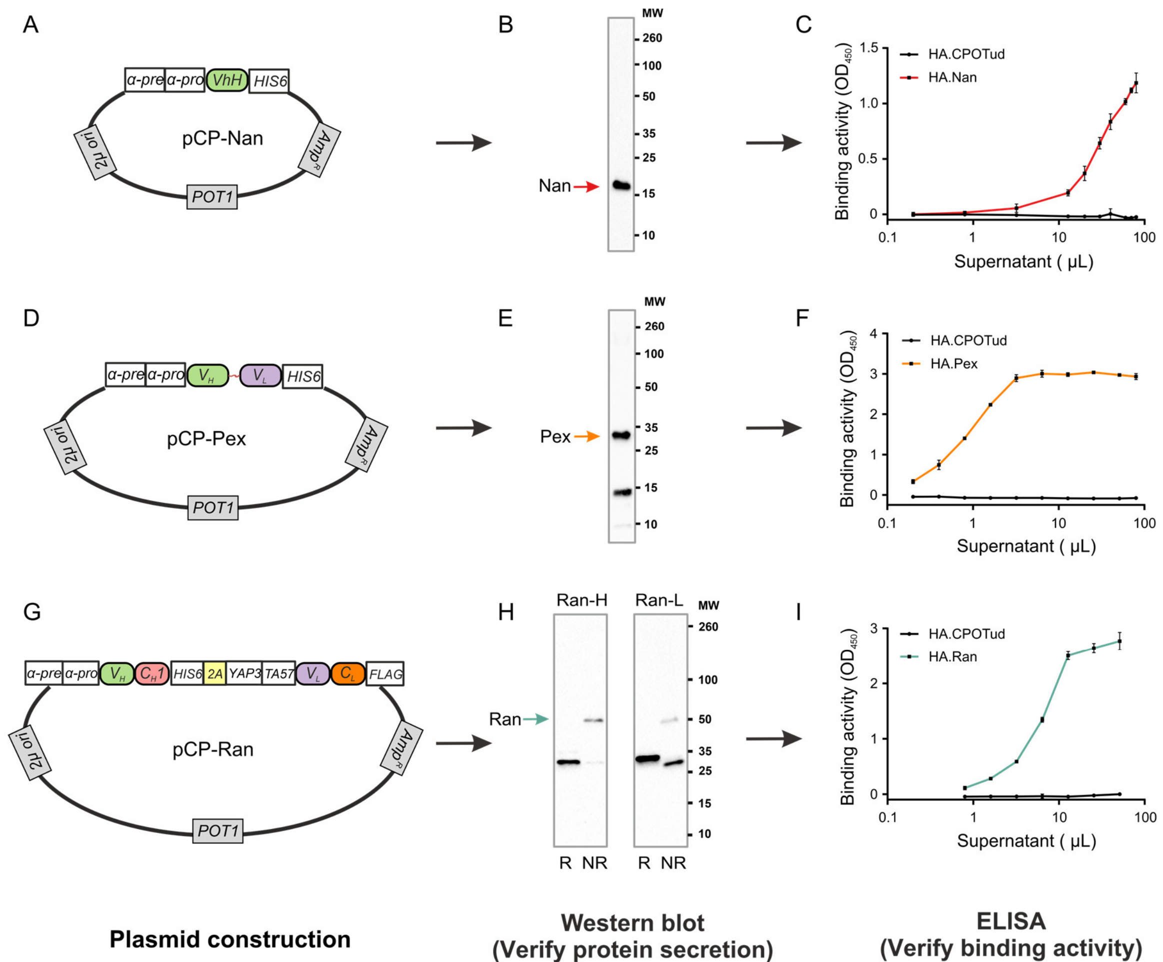

Figure 8. Expression and biological activity detection of antibody fragments. (**A**, **D**, **G**) Schematic illustration of the construction of CPOTud-based plasmids for the expression of proteins. *α-pre* and *α-pro*, components of α-factor leader; *HIS6*, 6xHis-tag coding sequence; *2A*, 2A peptide coding sequence; *YAP3* and *TA57*, components of the synthetic secretion leader; FLAG, FLAG-tag coding sequence. (**B**, **E**, **H**) Western blots of supernatant from recombinant strains, after cultivation in SD-2xSCAA (an optimized culture medium supplemented with 14 amino acids) without BSA medium, to detect the expression of proteins. Nan, Pex and Ran-H were detected with an anti-6x-His-tag monoclonal antibody, while Ran-L was detected with an anti-FLAG-tag monoclonal antibody. Western blot for Nan and Pex was carried out under reducing conditions, while for Ran, it was performed under both reducing (R) and nonreducing (NR) conditions. (**C**, **F**, **I**) A sandwich ELISA format to measure the binding activity of antibody fragments to their respective antigens after cultivation in SD-2xSCAA medium. ELISA signals (absorbance values are displayed as OD_{450}) were generated using an anti-6x-His-tag monoclonal antibody. n = 3 (Nan) and n = 2 (Pex and Nan), error bar = ± SD. Adapted from **Paper I**.

To examine whether the secreted antibody fragments have biological activities, an enzyme-linked immunosorbent assay (ELISA) was adjusted to measure the interaction between antibody fragments and respective antigens (Figure 8C, 8F and 8I). I first confirmed that the impurities in the medium did not cause non-specific reactions in the ELISA assay. Then, the harvested culture supernatant was serially diluted and tested by ELISA after different processing steps (dilution for the supernatant of Nan, concentration for the supernatant of Pex, neither dilution nor concentration for the supernatant of Ran). Unlike the control strain, for each antibody fragment, there was a significant increase in absorbance values (OD_{450}) with increased protein concentration (Figure 8C, 8F and 8I). A pull-down

assay also supported our ELISA results (Figure S5 in **Paper I**). These results indicated that all three secreted antibody fragments have biological activities in binding to their respective antigens.

3.1.2 Assessment of secretion levels of antibody fragments in different mutant strains

To investigate whether the improved secretion capacity of mutant strains is generally suitable for the production of pharmaceutical proteins, three antibody fragment expression plasmids were introduced into different mutants (LA, MA and HA), generating nine strains, including LA.Nan, MA.Nan, HA.Nan, LA.Pex, MA.Pex, HA.Pex, LA.Ran, MA.Ran and HA.Ran. Sometimes, improved protein production is accompanied by reduced growth (Li & Rinas, 2020). Therefore, I first measured the cellular growth rate (Figure 9). Compared to the respective control strains, the maximal specific growth rate decreased in protein expression strains. For each antibody fragment, the higher the secretion capacity, the faster the growth, which indicated that increased secretion capacity in the mutant strains did not adversely affect cell growth. This conclusion was also supported by the previous work (Huang *et al.*, 2017).

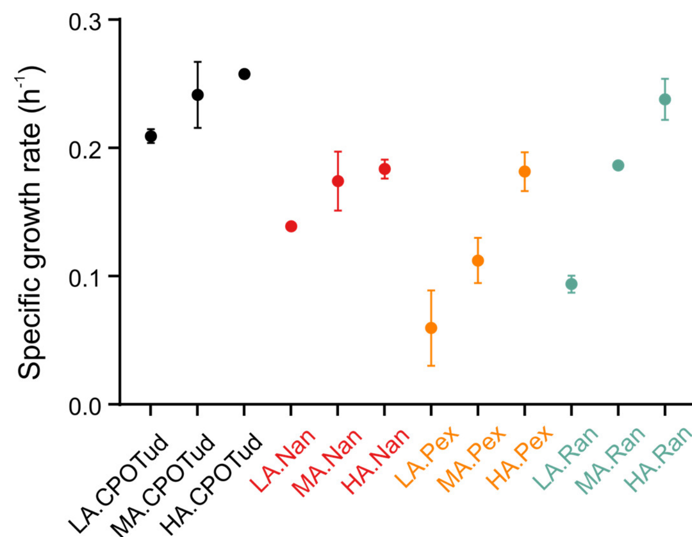

Figure 9. Maximum specific growth rates of different mutant strains expressing different antibody fragments. Recombinant strains were cultivated in a 96-well microtiter plate. A growth profiler was used to record the cell density. n = 3, error bar = ± SD. Adapted from **Paper I**.

Subsequently, I tested the secretion level of antibody fragments in the different mutant strains using a western blot. Compared with LA.Nan, MA.Nan showed a 0.9-fold increase and HA.Nan had a 2.8-fold increase in the amount of Nan (Figure 10A). Similarly, compared with LA.Ran, MA.Ran showed a 0.8-fold increase and HA.Ran had a 9.1-fold increase in the amount of Ran (Figure 10C). However, the pattern of Pex production was opposite to that of Nan and Ran. LA.Pex had a 2.5-fold higher and MA.Pex showed a 1.3-fold higher production of Pex than HA.Pex (Figure 10B). Besides, I also confirmed that the intensity levels of accompanying additional bands were consistent with those of Pex and Ran. Taken together, with the increase of the secretory capacity, the amount of Nan and Ran increased, while the production of Pex decreased.

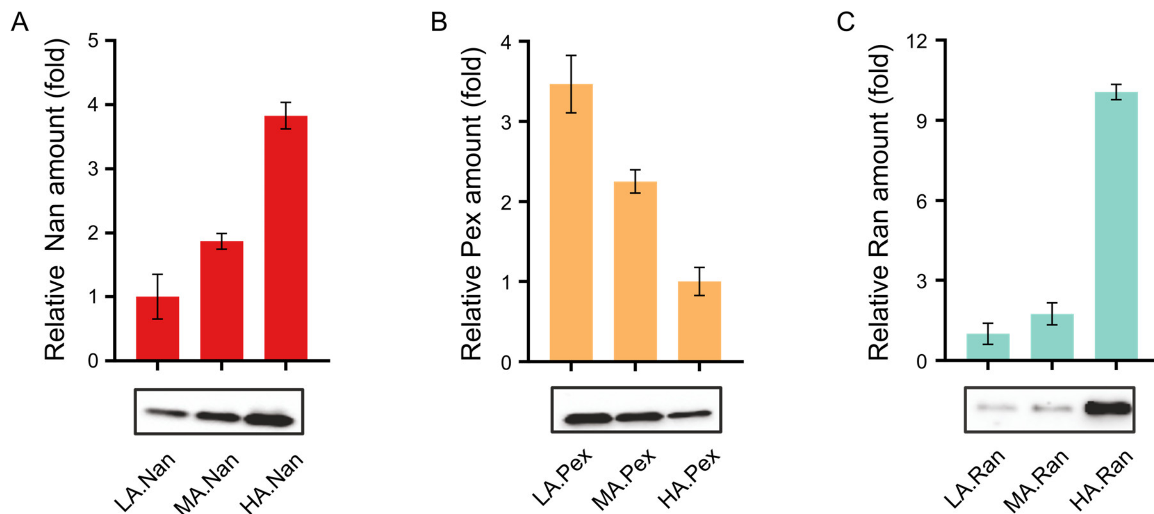

Figure 10. Comparison of secretion capacity of mutant strains expressing each antibody fragment. A western blot using anti-6x-His-tag antibody was carried out under reducing (**A** and **B**) or nonreducing (**C**) conditions. n = 3, error bar = ± SD. Adapted from **Paper I**.

3.1.3 Investigating the mechanism for protein-specific performance of the mutant strains

To investigate the possible reasons for differential secretion levels of antibody fragments, I performed RNA-seq analysis for four strains based on the opposite secretory trend, that is, Nan-expressing strains (LA.Nan and HA.Nan) and Pex-expressing strains (LA.Pex and HA.Pex). To verify how our samples for RNA-seq correlated with each other, principal component analysis was applied to characterize the global expression pattern (Figure 11). The first principal component (PC1) strongly separated strains HA and LA, which was not related to the produced proteins, while PC2 clearly isolated Nan and Pex in the LA strain but not in the HA strain. This demonstrates a high degree of reproducibility.

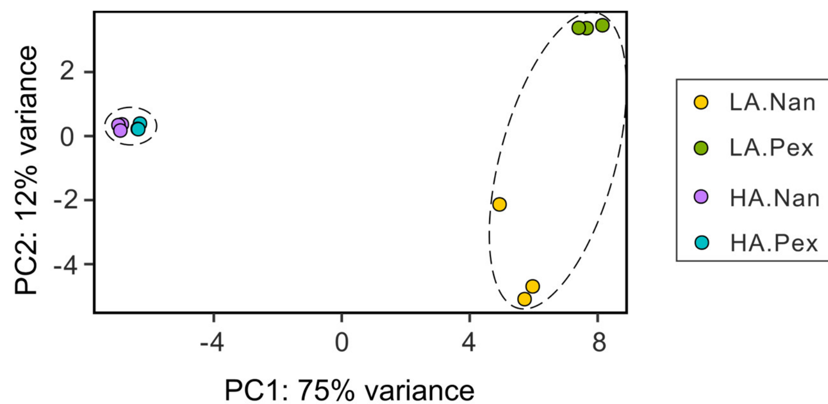

Figure 11. Principal component analysis of transcriptome data from the expression of Nan and Pex in LA and HA strains. n = 3. Adapted from **Paper I**.

To examine whether the protein secretion pathway was affected by the expression of different antibody fragments, we test the expression levels of genes in this pathway (Figure 12). We did not observe significant changes in genes expression when comparing HA.Nan vs LA.Nan and HA.Pex vs

LA.Pex, which proved that different protein secretion levels imposed similar effects on gene expression in the secretory pathway of the evolved strains.

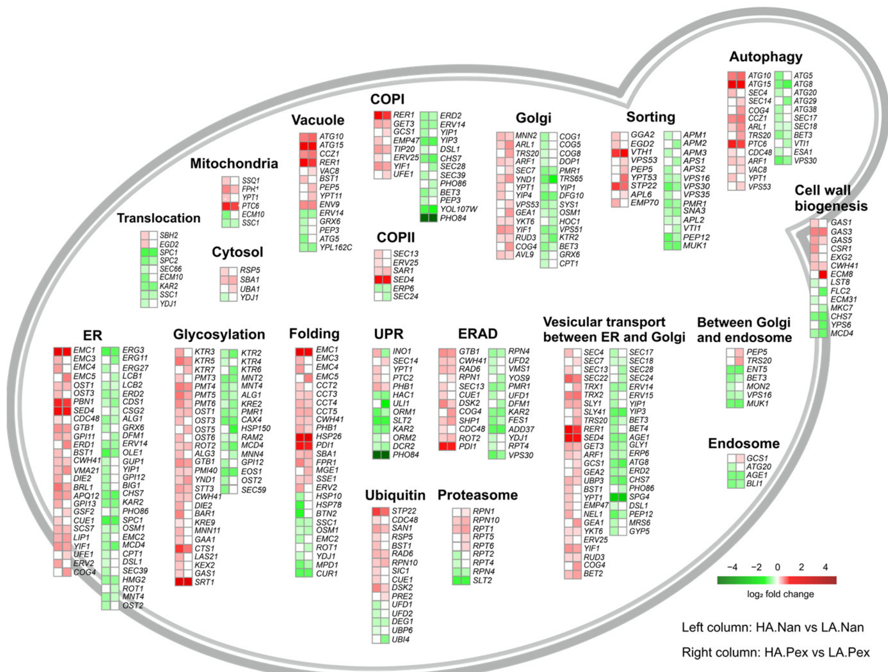

Figure 12. Transcriptional changes of genes in the protein secretory pathway. Red and green boxes represent the corresponding value of the log_2 fold change (p-adj < 0.05), and white boxes represent p-adj ≥ 0.05. Adapted from **Paper I**.

To identify the significant differences in cellular bioprocesses, we carried out the Gene Ontology (GO) term analysis in HA.Nan vs LA.Nan and HA.Pex vs LA.Pex (Figure 13). We focused on those processes in sector I and sector II that showed significant change (p-adj < 0.01) during the expression of only one protein. Regarding up-regulation, four processes, including glycine catabolic process, one-carbon metabolic process, glycine metabolic process and glycine decarboxylation via glycine cleavage system, were significantly up-regulated (p-adj < 0.01) only in HA.Nan vs LA.Nan, but not in HA.Pex vs LA.Pex (Figure 13A). However, experiments with addition of various concentrations of glycine to the medium did not show a clear effect on the production of the two antibody fragments (Figure S10 in **Paper I**). Regarding down-regulation, stress response-related processes, including the cellular response to drugs, cellular iron ion homeostasis, peroxisome organization and proteolysis, were significantly downregulated (p-adj < 0.01) only in HA.Nan vs LA.Nan, while there was no significant down-regulation during the expression of Pex (Figure 13B). This indicated that compared to Nan, the expression of Pex might lead to comparatively high cell stress in strain HA, which was further supported by the result of the increased Pex production in LA.Pex by overexpression of a protein kinase gene *KNS1*, involved in regulating ribosome and tRNA synthesis in response to cellular stress (Figure S9 in **Paper I**) (Lee *et al.*, 2012).

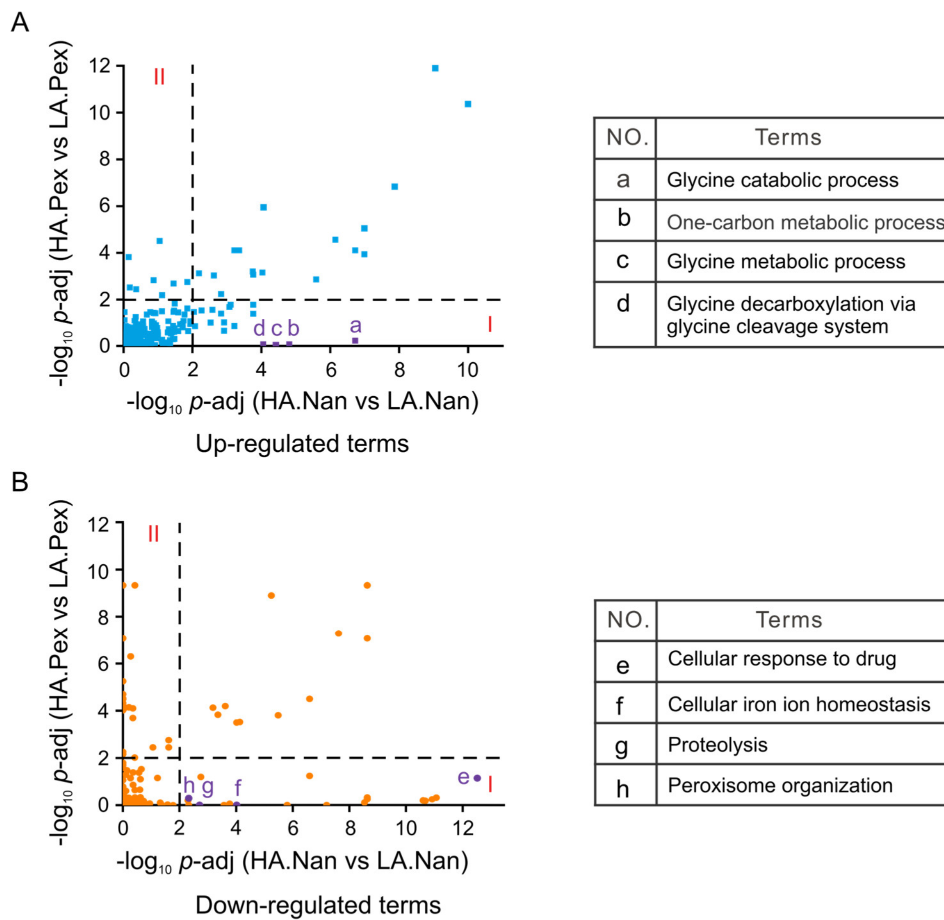

Figure 13. GO term analysis of RNA-seq data of LA and HA strains expressing Nan or Pex. The up-regulated (**A**) or down-regulated (**B**) terms between HA.Nan vs LA.Nan and HA.Pex vs LA.Pex were compared. I and II represent GO terms with significant differences in the expression of only Nan and Pex, respectively. 0-2 indicates *p*-adj > 0.01, i.e. no significant differences and > 2 indicates *p*-adj < 0.01, i.e. a significant difference. Adapted from **Paper I**.

In this chapter, I presented that three antibody fragments with different configurations were successfully expressed and secreted in the high-level secretory platform strain HA. I then evaluated their biological specificity with the respective antigens. I found the secretion of Nan and Ran was positively correlated with the secretory capacity of mutant strains, while the secretion of Pex showed the opposite trend. We revealed that the cellular stress response caused by Pex expression might be the limiting factor for the low protein production efficiency.

3.2 Characterizing the modifications of the mutant strains

The previous chapter reflected that B184 is a suitable candidate for producing pharmaceuticals, albeit to different extents. In this chapter, I will explore the responsible genes resulting in increased protein production along the evolutionary line towards B184. This work will expand the research based on one main model protein, α-amylase, and four strains: parental strain AAC as well as mutant strains M715, MH34 and B184 (Figure 7 and 14A).

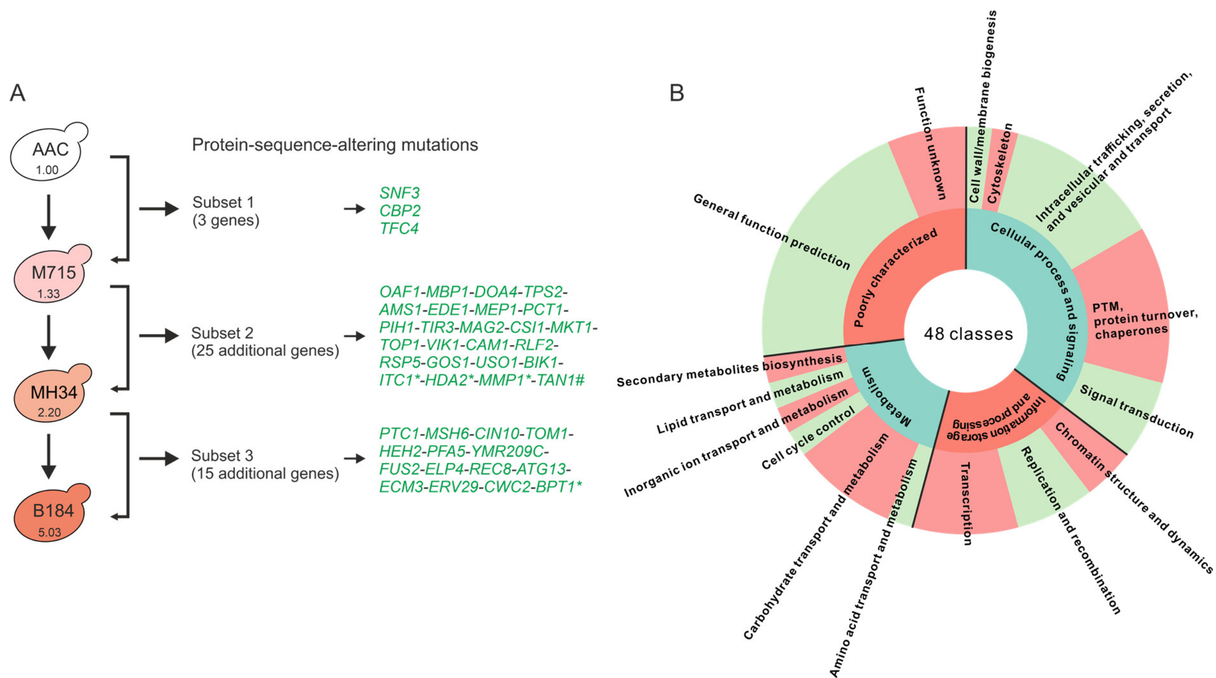

Figure 14. The distribution of protein-altering mutations in the evolutionary line towards B184. (**A**) Protein-altering mutations after each round of mutagenesis and selection. The mutated genes are listed on the right. * indicates nonsense mutations; # indicates a frameshift mutation. (**B**) Functional enrichment analysis of mutations according to the KOG database. The 48 classes were obtained because each of *TIR3*, *EDE1*, *BIK1*, *TOM1* and *ELP4* correspond to 2 classes. The middle arc indicates the KOG group. The outer arc indicates the KOG class. Adapted from **Paper II**.

3.2.1 Protein-altering mutations in the mutant strains

Through whole-genome sequencing, 43 amino acid sequence altering mutations were identified in this strain line. These mutations were divided into 3 types: 38 missense mutations (amino acid substitution), 4 nonsense mutations (stop codon generation) and 1 frameshift mutation (insertion or deletion of a number of nucleotides that is not divisible by three); and 3 subsets: subset 1 (from AAC strain to M715 strain) consisted of 3 missense mutations, subset 2 (from M715 strain to MH34 strain) consisted of 21 missense mutations, 3 nonsense mutations and 1 frameshift mutation, and subset 3 (from MH34 strain to B184 strain) consisted of 14 missense mutations and 1 nonsense mutation (Figure 14A). To examine which biological processes these genes are involved in, I first performed a functional enrichment analysis according to the euKaryotic Orthologous Groups (KOG) database (Tatusov *et al.*, 2003). These genes were divided into 16 subclasses from 4 groups (Figure 14B). It was previously shown that most of the subclasses showed significant transcriptional up-regulation in B184 compared to AAC (Huang *et al.*, 2017). Four subclasses, including 1) intracellular trafficking, secretion, and vesicular transport; 2) PTM, protein turnover, chaperones (mainly ubiquitination); 3) transcription; and 4) carbohydrate transport and metabolism, were most enriched.

As there may be multiple effects after amino acid substitutions, deletion or overexpression of single genes cannot always reveal the linkage between the function of mutated genes and protein production (Thak *et al.*, 2020). Point mutations, an alternative approach, can fine-tune the protein

activity or gene expression level and in turn help to understand the regulatory system involved in enhanced protein production capacities (Wang *et al.*, 2019). As the high secretion level of the mutant strains was caused by gene mutagenesis, it is critical to introduce the exact point mutations into the parental strain to identify potentially beneficial gene targets.

3.2.2 Strategies for the introduction of the desired point mutations

The single or multiple-nucleotide mutations were introduced into the parental strain by the CRISPR/Cas9 system. Multiple mutation sites were introduced simultaneously when genes contained multiple nucleotide mutations. Supplementarily, I performed gene deletions for the corresponding target genes as a comparison to further explore the effect of the mutations. According to the location of the point mutation in association with the nearest PAM site, mutations were introduced into the chromosome, adopting three allele replacement strategies (Figure 15 and 16).

In the first strategy, the mutated site was located within the gRNA recognition sequence (20 bp) and the gRNA expressing plasmid worked well (Figure 15). The point mutation or gene deletion was performed directly through co-transforming the corresponding gRNA expressing plasmid and the repair fragment containing the substituted base pair into the parental strain K01. It is worth pointing out that the only difference between K01 and AAC is that K01 does not contain the genomically integrated *KanMX* marker.

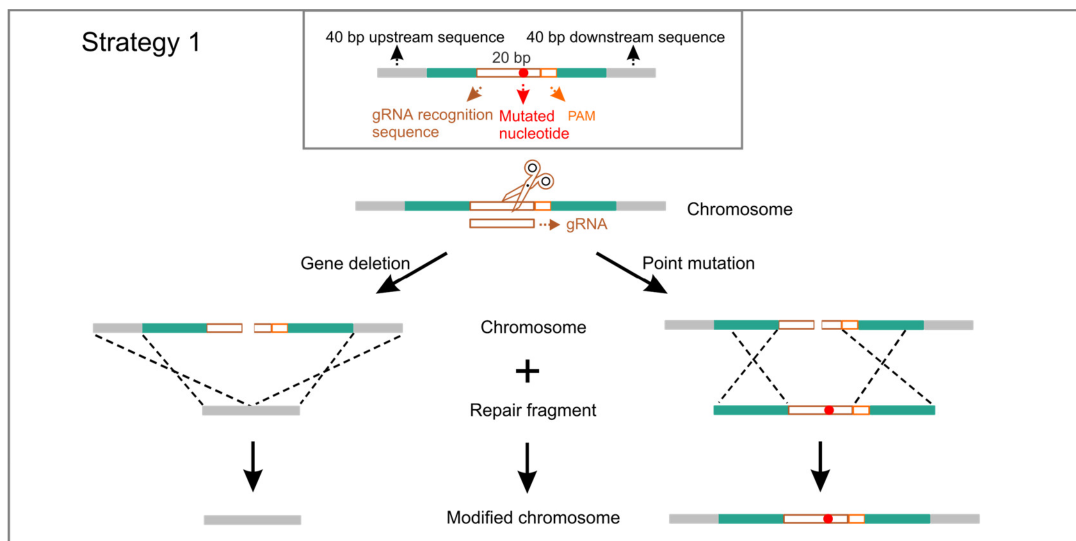

Figure 15. Schematic diagram of gene deletion and point mutation corresponding to strategy 1. For the gene deletion, the repair fragment consists of a 40-bp upstream sequence and a 40-bp downstream sequence of the target gene. For the point mutation, the repair fragment consists of a 40-bp sequence upstream of the gRNA recognition sequence, the 20-bp mutated gRNA recognition sequence and a 40-bp sequence downstream of the gRNA recognition sequence. Adapted from **Paper II**.

The second strategy was applied when strategy 1 was impossible and the gene was essential (Figure 16A). A gRNA recognition sequence near the mutation site was selected and the repair fragment containing both the substituted base pair and one or two mutations in the PAM sequence that corresponded to the same amino acid sequence. In this work, 4 essential genes, including *TFC4*, *RSP5*,

USO1 and *CWC2* were not deleted as their deleted genotypes are not viable. To increase the cutting efficiency, the Cas9 expression cassette (*TEF1-cas9-CYC1t*) was integrated into the chromosome in K01, obtaining KC01 (Figure 16B). The third strategy was applied when strategy 1 was impossible and the gene was non-essential (Figure 16C). A two-step approach was performed. As first step, the original gene in KC01 was replaced by a heterologous gRNA recognition sequence and its PAM sequence. Then, this introduced heterologous sequence was replaced by the mutated gene fragment containing the substituted base pair with the help of the corresponding heterologous gRNA expression plasmid.

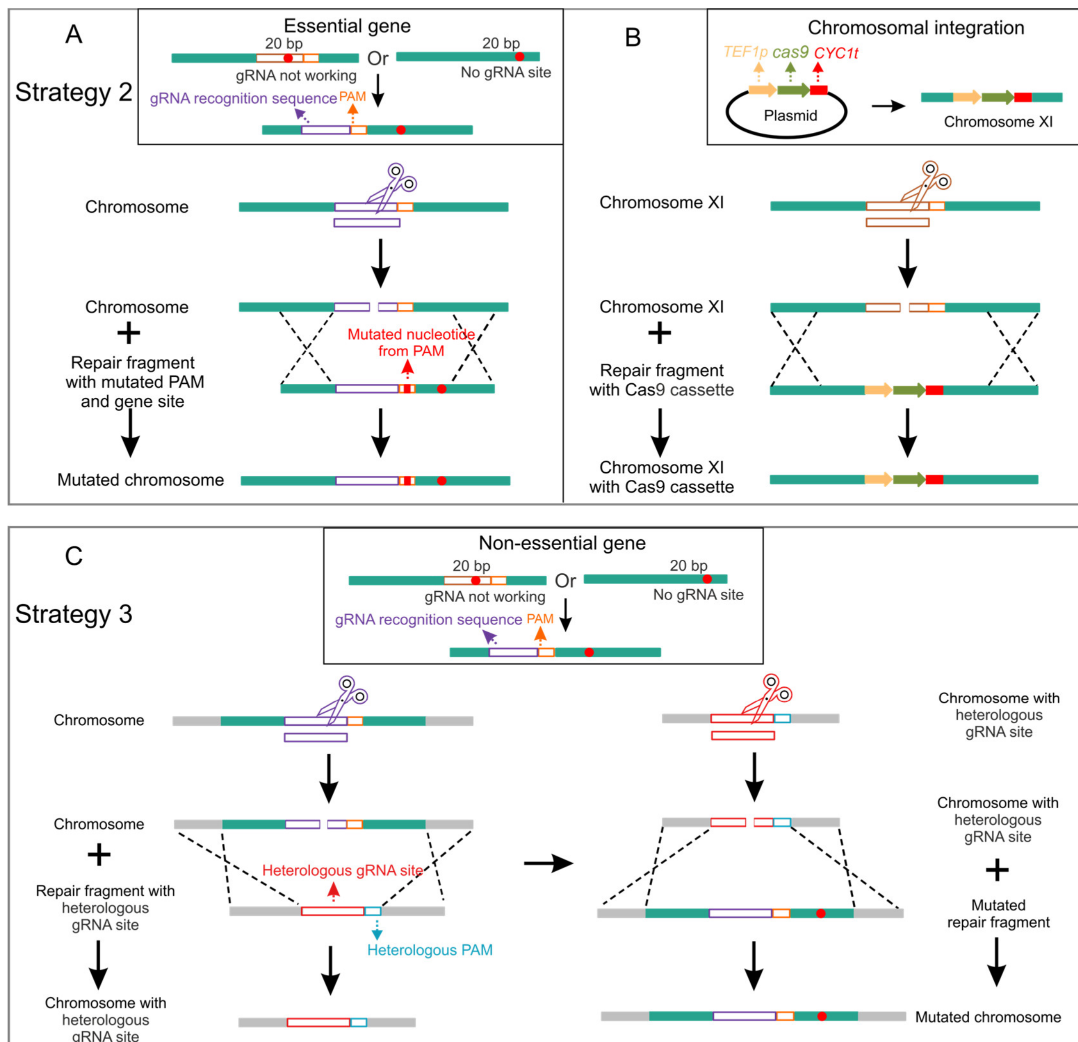

Figure 16. Schematic diagram of the introduction of point mutations corresponding to strategies 2 and 3. (**A**) Strategy 2: The gRNA selected for strategy 1 did not work, or the mutated site was not within the gRNA recognition sequence, and the gene is essential. A gRNA recognition sequence near the mutation site was selected. In addition to the desired mutated site, the repair fragment also contained one or two changed bases of the PAM sequence, but this substitution did not change the amino acid sequence. (**B**) KC01 strain construction. The *cas9* gene under control of the *TEF1* promoter and *CYC1* terminator was integrated into chromosome XI-3 locus of K01. (**C**) Strategy 3: The gRNA selected for strategy 1 did not work, or the mutated site was not within the gRNA recognition sequence, and the gene is non-essential. Firstly, a heterologous gRNA recognition sequence and its PAM sequence replaced the original gene in KC01. Secondly, the mutated gene fragment replaced the introduced heterologous sequence aided by the corresponding gRNA expression plasmid. Adapted from **Paper II**.

3.2.3 Beneficial targets for α-amylase production

To evaluate the individual contribution of the amino acid sequence altering mutations on α-amylase production, I carried out point mutations and single gene deletions. I named the point mutation strain and deletion strain as *GENEm* and *Δgene* in the following description, respectively. In this thesis, I only expand on one gene for each subset in detail to explore the possible mechanisms involved in enhanced protein production after the respective modification.

In subset 1, three point mutation strains (*SNF3m*, *CBP2m* and *TFC4m*) and two gene deletion strains *(Δsnf3* and *Δcbp2*) were generated since *TFC4* is an essential gene (Marck *et al.*, 1993). After testing these five strains, only *CBP2m* produced 1.4-fold α-amylase compared to the control, while deletion of *CBP2* severely affected cell growth and significantly reduced α-amylase production (Figure 17).

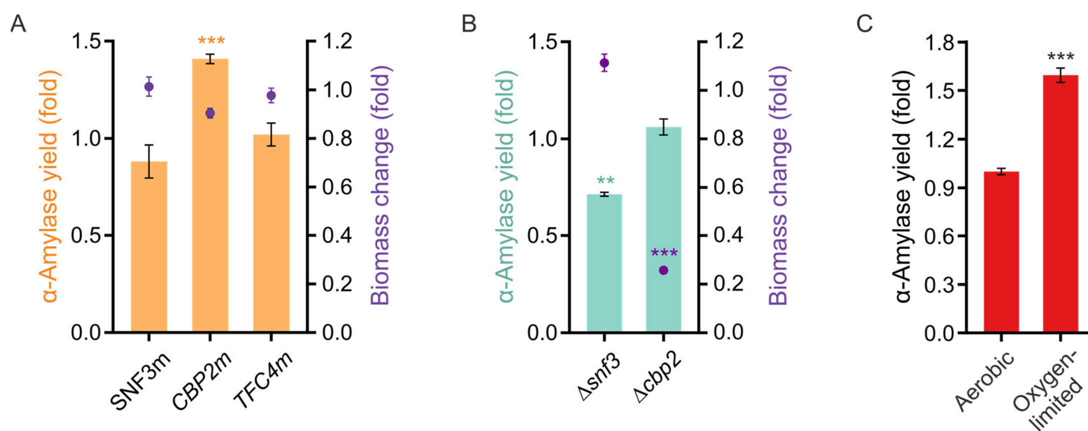

Figure 17. α-Amylase production and biomass change after gene modifications in subset 1. (**A**) Fold changes after point mutations. (**B**) Fold changes after gene deletions. (**C**) α-Amylase production change under aerobic and oxygen-limited conditions. n = 3, error bar = ± SD. Statistical significance $^{**}P < 0.01$, $^{***}P < 0.001$ (Two-tailed Student's t-test). Adapted from **Paper II**.

Cbp2p, a mitochondrial protein that facilitates splicing by binding to RNA, is required for splicing the terminal intron of cytochrome b (COB) pre-mRNA (Gampel *et al.*, 1989, Gampel & Cech, 1991). Cobp is a subunit of the mitochondrial respiratory chain complex III. The lack of *CBP2* leads to the loss of function of Cobp, which therefore blocks the electron transport chain in mitochondria and causes slow growth (Blakely *et al.*, 2005). This is consistent with our results that *Δcbp2* led to lower biomass and reduced protein titer (Figure 17B). Due to the increase in protein production and a slight reduction in biomass of *CBP2m*, I speculated that the point mutation of *CBP2* might only partially impair the enzyme activity of Cobp and therefore reduce respiration. As described in **Chapter 1.3.3**, much work has shown that oxygen-limiting or even anaerobic conditions were beneficial for protein production. To confirm our speculation, I performed aerobic and oxygen-limited cultivation for the control strain K01, respectively. α-Amylase production showed a 0.59-fold increase under the oxygen-limited cultivation condition compared to aerobic conditions (Figure 17C), which supported our hypothesis that the potentially respiration-damaging *CBP2* mutant strain was favorable for α-amylase production.

Subset 2 from M715 to MH34 consists of 25 additional protein-sequence-altering mutations. I actually carried out gene modifications for 24 genes since the gene *BIK1* is positioned on chromosome III, duplicated in mutant strains MH34 and B184 (Huang *et al.*, 2017). As I did not obtain clones for the point mutation of *USO1* after I tested two gRNAs, I finally assessed 23 point mutation strains. On the other hand, since *RSP5* and *USO1* are essential genes (Heo *et al.*, 2020, Sangkaew *et al.*, 2022), 22 gene deletion strains were generated. Compared to the control strain, 8 out of 23 point mutations significantly changed the production of α-amylase, with 5 leading to increased α-amylase production and 3 resulting in decreased α-amylase production (Figure 18A). As deletion of *GOS1* severely impaired cell growth, I did not test this strain. Except for *Δgos1*, out of 21 single gene deletions, 13 significantly changed the production of α-amylase, with 5 improving α-amylase production and 8 reducing α-amylase production (Figure 18B). Ultimately, gene modifications of 6 targets increased protein production more than 0.2-fold, respectively, which were the point mutation of *TIR3* and *GOS1*, both point mutation and gene deletion of *MKT1*, and gene deletion of *TOP1*, *EDE1* and *MAG2*.

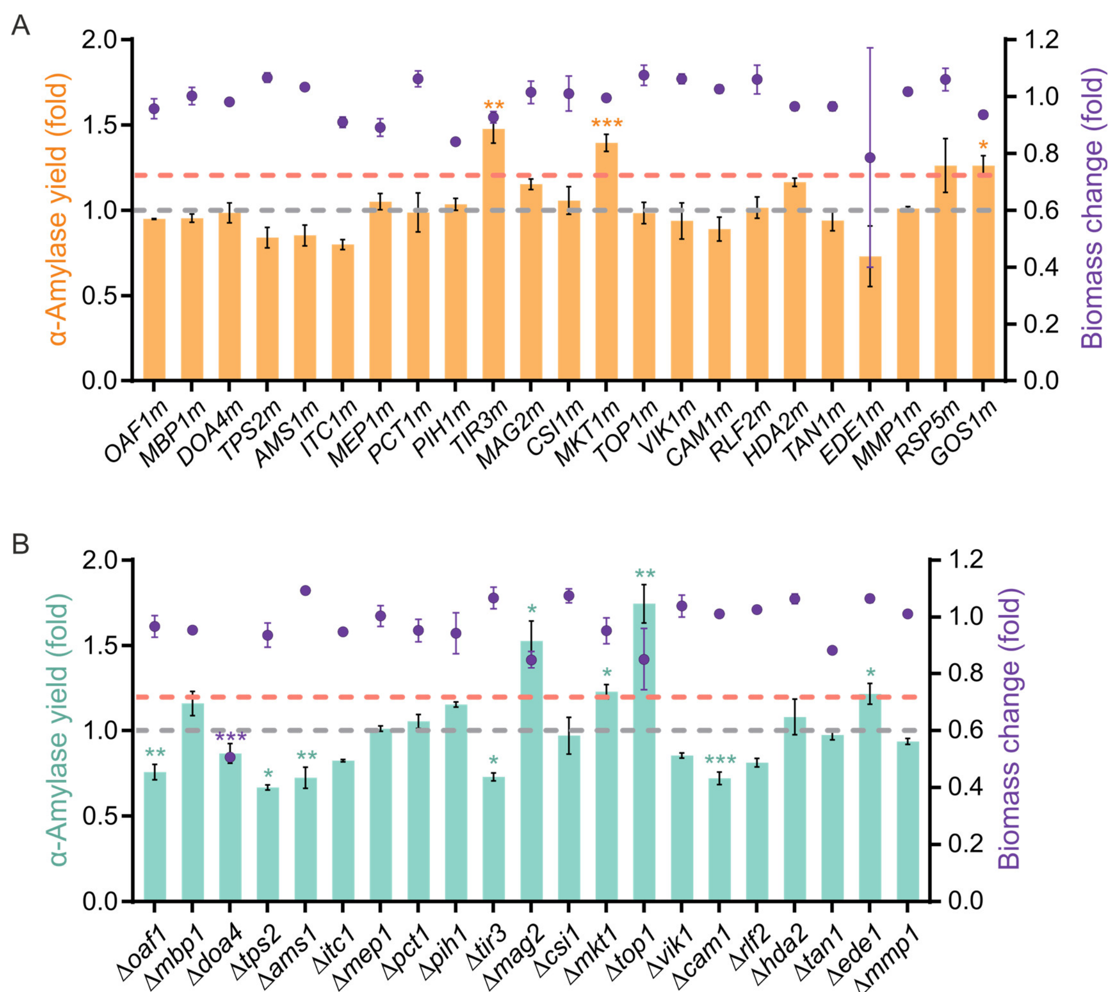

Figure 18. α-Amylase production and biomass change after gene modifications in subset 2. (**A**) Fold changes after point mutations. (**B**) Fold changes after gene deletions. The grey line indicates α-amylase production of the control strain. The red line indicates that the modified strain produces 1.2-fold α-amylase compared to the control strain. n = 3, error bar = ± SD. When the α-amylase yield change was more than 0.2-fold or less than 0.2-fold compared to the control, the statistical significance (**P* < 0.05, ***P* < 0.01, ****P* < 0.001) was determined (Two-tailed Student's t-test). Adapted from **Paper II**.

Tir3p is one of the examples that was investigated in more detail in subset 2. It is a cell wall mannoprotein, which is expressed under anaerobic/oxygen-limiting conditions and required for anaerobic growth (Abe, 2007, Tran Nguyen Hoang *et al.*, 2018). Previous transcriptome analysis demonstrated that the expression levels of *TIR3* and several other genes related to anaerobic metabolism, including *ANB1* encoding the translation elongation factor eIF-5A, *CYC7* encoding isoform 2 of the electron carrier protein cytochrome c, *DAN1* encoding a cell wall mannoprotein, and *AAC3* encoding an ADP/ATP translocator in mitochondrial inner membrane, were considerably up-regulated in the mutant strains (Figure 19), and the dissolved oxygen levels in the medium were approximately 90% when sampling for RNA extraction (Huang *et al.*, 2017). This indicates that characteristics similar to those of anaerobically grown strains shown in the mutant strains may facilitate protein secretion. This phenomenon is similar to the effects caused by the point mutation of *CBP2*. In relation to the significant reduction of α-amylase production by the deletion of *TIR3*, I speculated that the point mutation might increase the activity of Tir3p and therefore improve protein secretion.

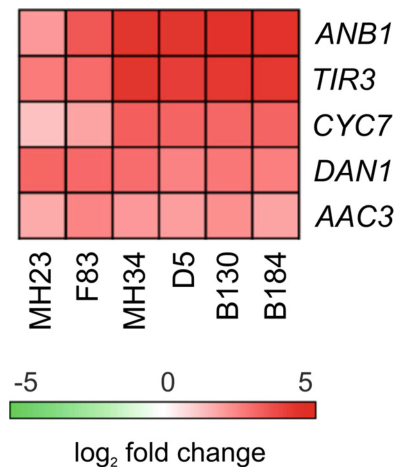

Figure 19. Expression level changes of several genes related to anaerobic metabolism in the mutant strains. Adapted from Huang *et al.* (Huang *et al.*, 2017).

Subset 3 from MH34 to B184 consists of 15 additional protein sequence-altering mutations. I tested 14 genes since I did not obtain mutation and deletion strains for *CWC2* for the same reason as *USO1* (Lu et al., 2012). Out of 14 point mutations, 7 were beneficial for protein production (Figure 20A). And out of 14 gene deletions, 6 significantly changed α-amylase production, with 4 leading to increased and 2 resulting in decreased protein production (Figure 20B). Similar to subset 2, modifications of 7 genes increased protein production more than 0.2-fold, which were both point mutation and gene deletion of *PTC1*, *TOM1* and *ECM3*, and the point mutation of *ATG13*, *FUS2*, *ERV29* and *HEH2*.

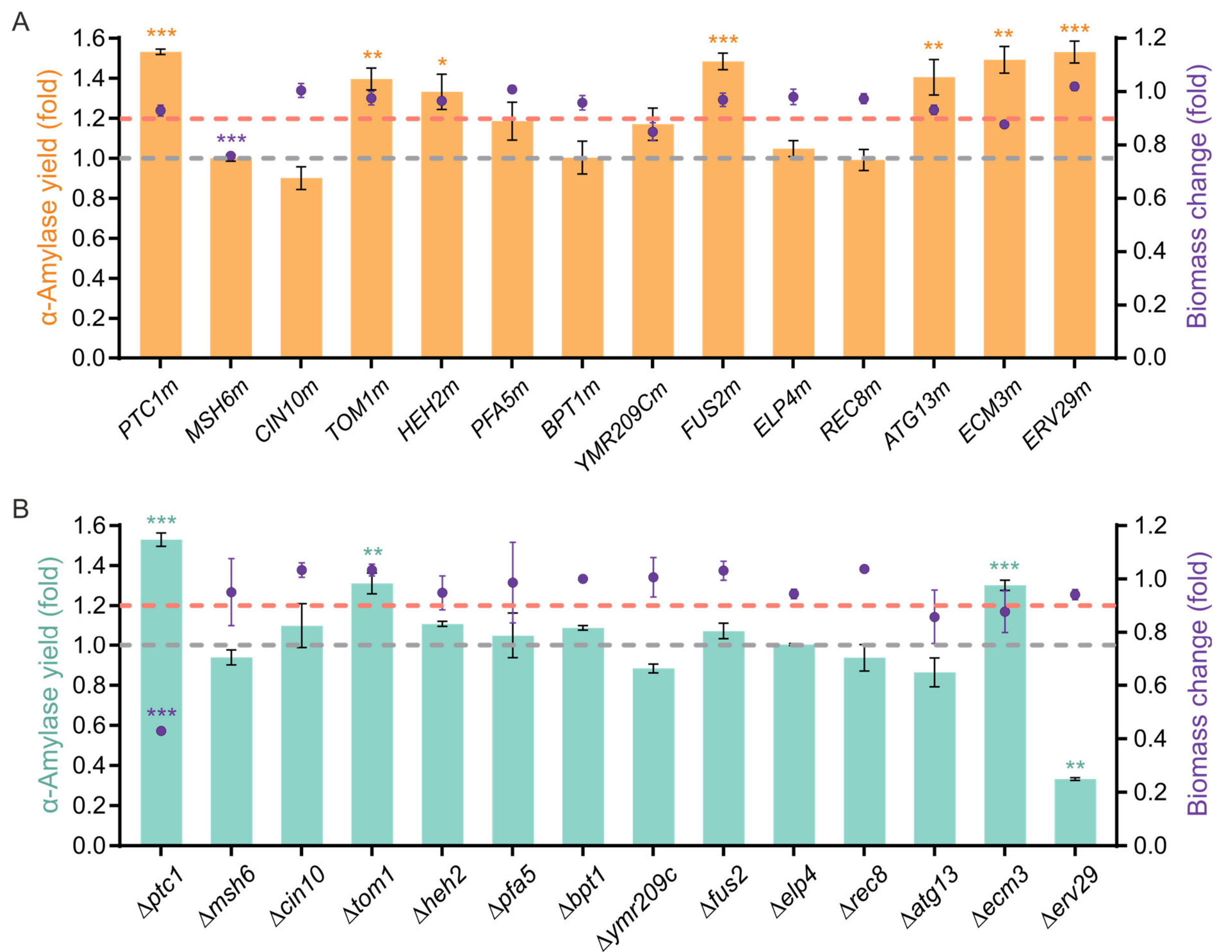

Figure 20. α-Amylase production and biomass change after gene modifications in subset 3. (**A**) Fold changes after point mutations. (**B**) Fold changes after gene deletions. The grey line indicates α-amylase production of the control strain. The red line indicates that the modified strain produces 1.2-fold α-amylase compared to the control strain. n = 3, error bar = ± SD. When the α-amylase yield change was more than 0.2-fold or less than 0.2-fold compared to the control, the statistical significance (*P < 0.05, **P < 0.01, ***P < 0.001) was determined (Two-tailed Student's t-test). Adapted from **Paper II**.

Ptc1p is involved in regulating signal transduction by dephosphorylation of Hog1p and negative regulation of the osmosensing mitogen-activated protein kinase (MAPK) cascade (Ariño *et al.*, 2011). When cells sense high osmotic stress, Hog1p will be phosphorylated and imported into the nucleus after the high osmolarity glycerol (HOG) pathway is stimulated (Ferrigno *et al.*, 1998, Ariño *et al.*, 2011). The nuclear-localized Hog1p activates the transcription of approximately 600 genes (Hohmann, 2002), which leads to the production of various small molecules such as glycerol to balance the extracellular hyperosmotic conditions (Westfall *et al.*, 2004). Our results showed that the *Δptc1* strain produced increased glycerol in the supernatant compared to the control strain (Figure 21A), which is consistent with a previous study (Jiang *et al.*, 1995). It has been proved that the accumulation of intracellular glycerol facilitates the enhancement of heterologous protein production (Shi *et al.*, 2003, Chen *et al.*, 2021). The increased α-amylase production might be due to the high glycerol concentration. In addition, the glycerol content in the supernatant of the *PTC1m* strain was also increased to a value between the control strain and the *Δptc1* strain after cultivation for 40 h (Figure 21A). This led us to

speculate that the point mutation weakened the activity of Ptc1p. A reduced growth rate (0.14 h^{-1} in *Δptc1* strain vs. 0.25 h^{-1} in the control strain) and lower final biomass (0.42-fold of control strain) were also observed in the *Δptc1* strain but not in the *PTC1m* strain (Figure 21B and 20B), which proved that weakening rather than deleting *PTC1* was more favorable for protein secretion. This is in agreement with a previous work that a moderate osmolarity level was the most favorable strategy to increase the secretion of heterologous proteins in *Y. lipolytica* (Kubiak *et al.*, 2019).

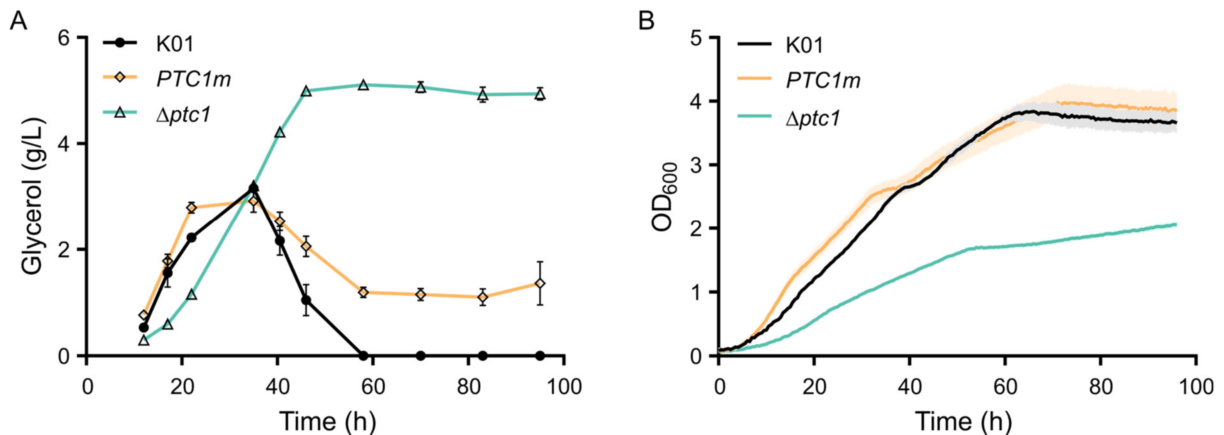

Figure 21. Physiological characterization of the engineered strains *PTC1m* and *Δptc1*. **(A)** Glycerol content in the medium of engineered strains after 96 h cultivations. HPLC was used to measure the glycerol concentration in the supernatants. **(B)** Growth phenotype of engineered strains. Recombinant strains were cultivated in a 96-well microtiter plate. The growth profiler recorded the cell density. n = 3, error bar = ± SD. Adapted from **Paper II**.

Through the detailed exploration of the functions of all beneficial targets, autophagy, cell cycle and the vesicular trafficking pathway between ER and Golgi, are pointed out as playing a vital role in protein secretion. (I) According to GO Slim Term analysis, the targets *EDE1* and *ATG13* are associated with autophagic processes. A previous study demonstrated that attenuated autophagy significantly improved the production of several heterologous proteins in *Kluyveromyces marxianus*, including mannase (Man330), ruminal feruloyl esterase (Est1E), yEGFP, and β-1,4-endoxylanase (XynCDBFV) (Liu *et al.*, 2018). Consistently, reducing the expression levels of some autophagy-related genes increased the secretion of heterologous bovine chymosin (CHY) in *A. oryzae* (Yoon *et al.*, 2013). These studies indicated that rational regulation of autophagic processes is a promising approach for optimizing the production of secreted protein. (II) *TOP1*, *EDE1*, *TOM1* and *FUS2* are related to the regulation of the cell cycle. Much of the work implies a dependence of protein secretion on the cell cycle. For example, the specific production rate of rice α-amylase achieved its maximum value in the M phase in *S. cerevisiae* (Uchiyama *et al.*, 1997). Emerging G1 phase cells did not secrete detectable recombinant human cytokine before reaching cell division (Frykman & Srienc, 2001). Furthermore, a number of genes that are related to various secretory processes, such as vesicle coating, ER and Golgi trafficking and fusion of vesicles to the plasma membrane, are also expressed in a cell cycle-dependent pattern (Spellman *et al.*, 1998). This secretion pattern emphasizes the importance of regulating the cell cycle. (III) *GOS1* and *ERV29* are associated with vesicular trafficking between ER and Golgi. Previous studies showed that engineering the bi-directional vesicular trafficking, including anterograde trafficking by moderately overexpressing *SEC16* (involved in COPII vesicle assembly, described in **Chapter 1.3.2.4**),

and retrograde trafficking by overexpressing *GLO3* (involved in COPI vesicle disassembly, described in **Chapter 1.3.2.5**), enhanced the secretion of several heterologous proteins, including α-amylase, endoglucanase I and GLA (Bao *et al.*, 2017, Bao *et al.*, 2018). These results emphasize the important role of the bi-directional vesicular trafficking pathway in optimizing recombinant protein secretion.

In this chapter, I presented three universally applicable CRISPR/Cas9-mediated strategies to introduce nucleotide substitutions for 42 protein-altering mutations in the strain line to B184. Interestingly, a quite high proportion of the mutations generated during random mutagenesis (14/42) represented beneficial targets. These targets are mainly relevant to stress-related processes, protein degradation, transportation, mRNA processing and export, DNA replication and repair. I also found that potentially effective strategies to further engineer are the oxygen conditions, autophagic processes, cell cycle and bi-directional vesicular trafficking pathway between ER and Golgi for optimizing protein production in the future.

4 Rational optimization of platform strains by investigating novel targets

In the previous chapter, 11 of the 14 beneficial targets were newly identified to be related to recombinant protein production. The novel targets could provide guidelines for designing efficient cell factories. This chapter will mainly summarize diverse engineering strategies applied to investigate potential new genes (**Paper III-IV**). In **Chapter 3**, we saw that the cellular stress response plays an essential role in improving recombinant protein production, and the generation of these stresses is very similar to a previously established yeast AD model with overexpression of Aβ42 (Chen & Petranovic, 2015). Thus, we were wondering: could we identify some effective targets from suppressors of Aβ42 toxicity to alleviate cellular stress and thus increase recombinant protein production? The study is presented in **Paper III** and **IIIb**. During this process, we considered that such engineering strategy is often protein-specific, which could limit the application of identified targets for the improved expression of a wide range of recombinant proteins. We proposed the question: could a more systematic model predict novel targets for the rational design of yeast cells? The study is presented in **Paper IV**.

4.1 Novel targets from Aβ42 yeast model

AD is the most common neurodegenerative disease in aging populations, which is classified as a protein folding disease due to the accumulation of misfolded Aβ peptides (Murphy & LeVine, 2010). Among different Aβ peptides, Aβ42 is more hydrophobic and thus easier to aggregate, being the predominant form in cerebral plaques of AD patients (Selkoe & Wolfe, 2007). Besides, it has been proved that the small oligomers formed by Aβ peptides show increased cytotoxicity (Hartley *et al.*, 1999). As mentioned in **Chapter 1.3**, *S. cerevisiae* is a valuable model for studying the implication of misfolded proteins in human pathologies due to the strong conservation of protein quality control systems among eukaryotic cells. Therefore, to mimic chronic cytotoxicity during AD progression, our laboratory previously developed an Aβ yeast model by constitutively expressing the Aβ42 peptide (Chen & Petranovic, 2015). Compared to the less toxic Aβ40 strain, the expression of Aβ42 peptide in yeast generated more oligomers, induced increased production of ROS and decreased ubiquitin-proteasome activities, and caused a stronger ER stress and UPR (Chen & Petranovic, 2015, Chen *et al.*, 2017). To study gene interactions and thus explore the cellular mechanisms leading to the Aβ42 cytotoxic phenotype, a genome-wide synthetic genetic array (SGA) was applied to the Aβ42 yeast model to screen for gene targets that could alter Aβ42 cytotoxicity when deleted (Chen *et al.*, 2020). Among the subsequently characterized targets, overexpression of the riboflavin kinase gene *FMN1* reduced Aβ42 cytotoxicity by increasing resistance to oxidative stress (Chen *et al.*, 2020). This provided a new direction in that the mutants that suppressed Aβ42 accumulation-related cytotoxicity could be used as potential targets to study cellular stress.

High-level expression of recombinant proteins often causes a protein folding burden and therefore leads to cellular stress, which is very similar to the cytotoxic phenotype caused by the expression of

Aβ42. Thus, we presumed that the suppressors of Aβ42 cytotoxicity could improve recombinant protein production by alleviating ER stress generated by high protein expression.

4.1.1 Effects of single targets on recombinant protein production

The typical SGA screen crosses a query mutation to the gene deletion library, and after forming diploids, sporulation and haploids, the selected double mutants are scored for growth defects (Tong & Boone, 2006). Based on this principle, the previous study constructed two arrays by mating either an Aβ42 expression strain or a control strain without Aβ42 peptides with each deletion strain from the yeast deletion mutant library (Chen *et al.*, 2020) (Figure 22). Scores were obtained by comparing the colony sizes of Aβ42 expression mutants with those of their respective control mutants. Scores > 0 or scores < 0 represent an increase or a decrease of cell growth in Aβ42 expression mutants compared to control strain with mutants, which implies a decreased or an increased Aβ42 cytotoxicity, respectively (Chen *et al.*, 2020).

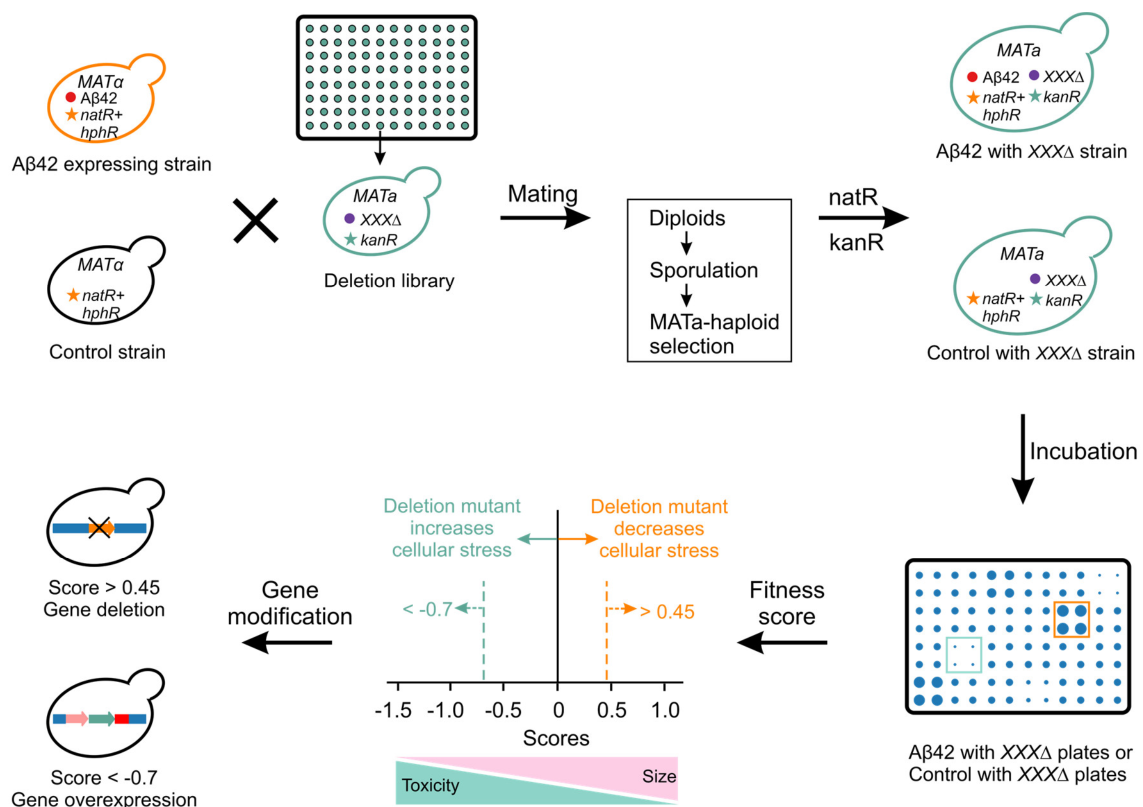

Figure 22. Schematic workflow for screening the suppressors of Aβ42 toxicity by genome-wide SGA technology. Firstly, the Aβ42 expression strain or the control strain with both the nourseothricin resistance (*natR*) and hygromycin B (*hphR*) markers were crossed to the deletion mutants with the kanamycin resistance (*kanR*) marker from the yeast deletion library. The resulting diploids were selected in the medium containing three antibiotics nourseothricin, hygromycin B and G418, and then the diploids were induced to form spores by culturing in reduced carbon and nitrogen medium. The resulting spores were transferred to a synthetic medium lacking histidine to select the *MATa*-haploids, and then the desired strains (double mutants) were selected and scored by culturing these haploids in plates containing three antibiotics nourseothricin, hygromycin B and G418. Score > 0 indicates the increased cell sizes in Aβ42 expression mutants compared to the respective control mutants, while score < 0 indicates the decreased cell sizes.

In our study, we first investigated 46 targets with decreased Aβ42 cytotoxicity in the SGA (score > 0.45) by gene deletion and 20 targets with increased Aβ42 cytotoxicity in the SGA (score < -0.7) by gene overexpression to test their individual effects on recombinant protein production in *S. cerevisiae* (Figure 22 and 23). The heterologous protein α-amylase with an α-factor signal peptide was expressed on a high-copy plasmid p423GPD under control of the strong constitutive promoter *TDH3p*. BY4742 (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) was used as the host strain. It was surprising that 26 out of the 46 gene deletion strains dramatically changed α-amylase yield compared to the control strain, of which 20 were beneficial and 6 had negative effects (Figure 23A). To identify the effect of gene overexpression, the targets were expressed on a high-copy plasmid pSPGM1 under control of the strong constitutive promoter *TEF1p*. Here, to avoid the possible plasmid instability caused by the high amount of expression of both α-amylase and overexpression gene from two high-copy plasmid systems, we alternatively used AACD as a host strain. As mentioned in **Chapter 3.1,** AACD expresses α-amylase at high stability and high copy number via the CPOTud plasmid system (Liu *et al.*, 2012). The only difference between AACD and AAC was that AACD has *URA3* deleted. After testing, 4 out of the 20 gene overexpression strains significantly improved α-amylase production compared to the control strain (Figure 23B). These results supported our hypothesis that the suppressors of Aβ42 cytotoxicity could be used to improve the production of recombinant protein.

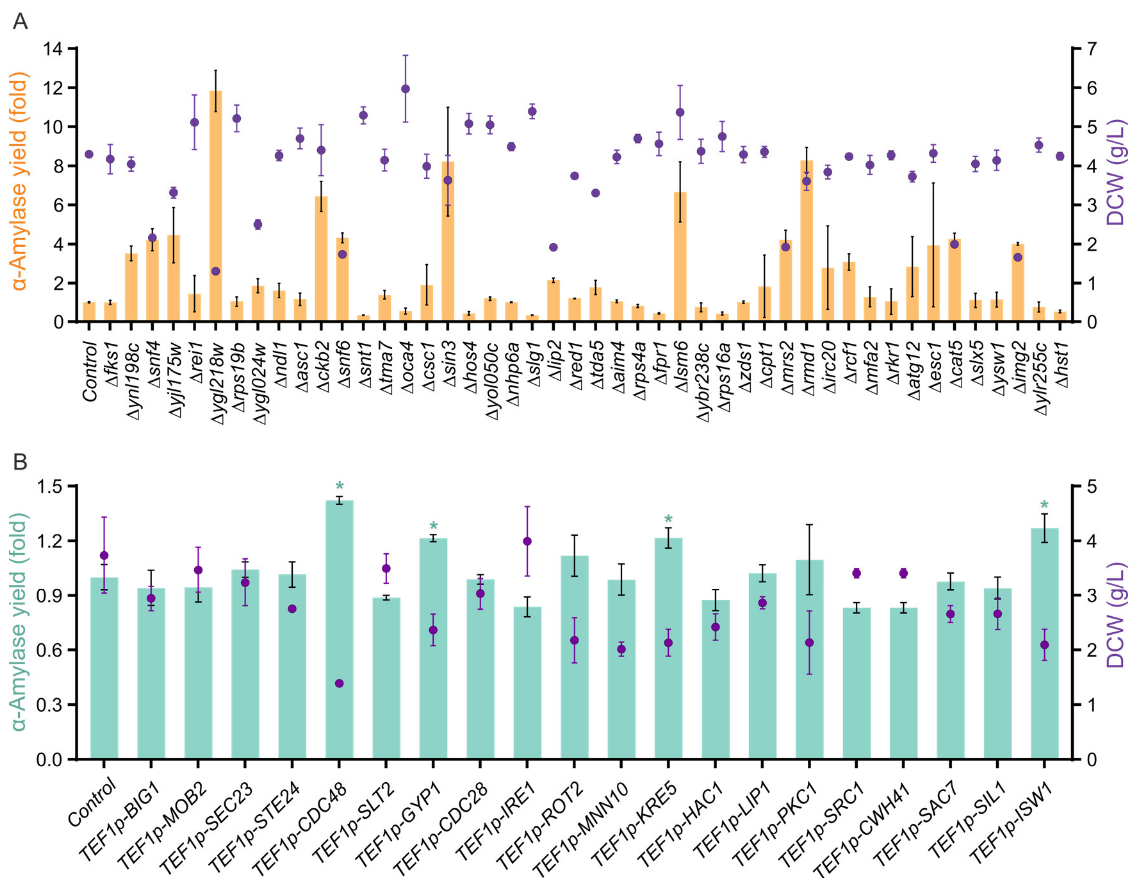

Figure 23. The effect of candidate genes whose deletion decreased or increased Aβ42 toxicity, on α-amylase production. (**A**) α-Amylase yield and dry cell weight (DCW) change by deletion of candidate genes in BY4742. (**B**) α-Amylase yield and DCW change by overexpression of candidate genes in AACD. n = 4, error bar = ± SD. Statistical significance *$P < 0.05$ (Two-tailed Student's t-test). Adapted from **Paper III**.

4.1.2 Effects of combinatorial targets on recombinant protein production

Based on **Chapter 4.1.1**, we selected 6 genes (*YGL218W*, *SIN3*, *RMD1*, *LSM6*, *CKB2*, and *SNF6*) whose deletion resulted in the highest α-amylase production and 4 genes (*CDC48*, *GYP1*, *KRE5*, and *ISW1*) whose overexpression led to the increased production for further combination to test whether these targets had additive effects. According to Figure 23B, we observed that the mutants had a smaller α-amylase yield change (less than 0.5-fold) compared to the control when AACD was used as the host strain. This may be due to the fact that AACD is a preferable strain used to study protein secretion, which has a high secretion capacity compared to other strains (Liu *et al.*, 2012). To make the difference obvious, we turned back to BY4742 as the host strain for the combinatorial characterization. Besides, to ensure stability, genes to be overexpressed were chromosomally integrated.

Since 93% of the open reading frame of *YGL218W* overlaps with the *MDM34* gene, deletion of *YGL218W* is equivalent to inactivating *MDM34*. Given the unknown function of *YGL218W*, we deleted *MDM34* in this study. Compared to the control strain A01 (no deletion), the amounts of α-amylase after 96 h cultivation increased 8.8- and 17.8-fold in the double gene-deletion strains A05 (*Δlsm6* and *Δsin3*) and A06 (*Δlsm6* and *Δmdm34*), respectively (Figure 24A). Both strains showed a significant increase compared to the single gene-deletion strain A02 (*Δlsm6*). We observed a reduction in final dry cell weight when deleting *MDM34* in A06 (*Δlsm6* and *Δmdm34*) compared to A02 (*Δlsm6*). According to the literature, Mdm34p, a mitochondrial component of the ER-mitochondrial encounter structure (ERMES) complex, is involved in the interaction between the ER and mitochondria, the coordination of mitochondrial DNA replication and growth, mitochondrion organization, and phospholipid transport, which is required in maintaining normal intracellular function (Kornmann *et al.*, 2009, Légiot *et al.*, 2019). To reduce the impact on cell growth, instead of deletion, we fine-tuned the expression of *MDM34* using a group of weak promoters (the strength: *GCN4p* < *GLC3p* < *USV1p* < *SDH4p* < *QCR8p*) to replace the native promoter in A05 (*Δlsm6* and *Δsin3*) (Figure 24B). The results showed that all 5 evaluated promoters significantly increased α-amylase production and did not affect the final biomass compared to A05 (*Δlsm6* and *Δsin3*). The strain A08 (*Δlsm6*, *Δsin3* and *GCN4p-MDM34*) with the highest α-amylase production (15.5-fold higher than A01) was selected for additional combinations.

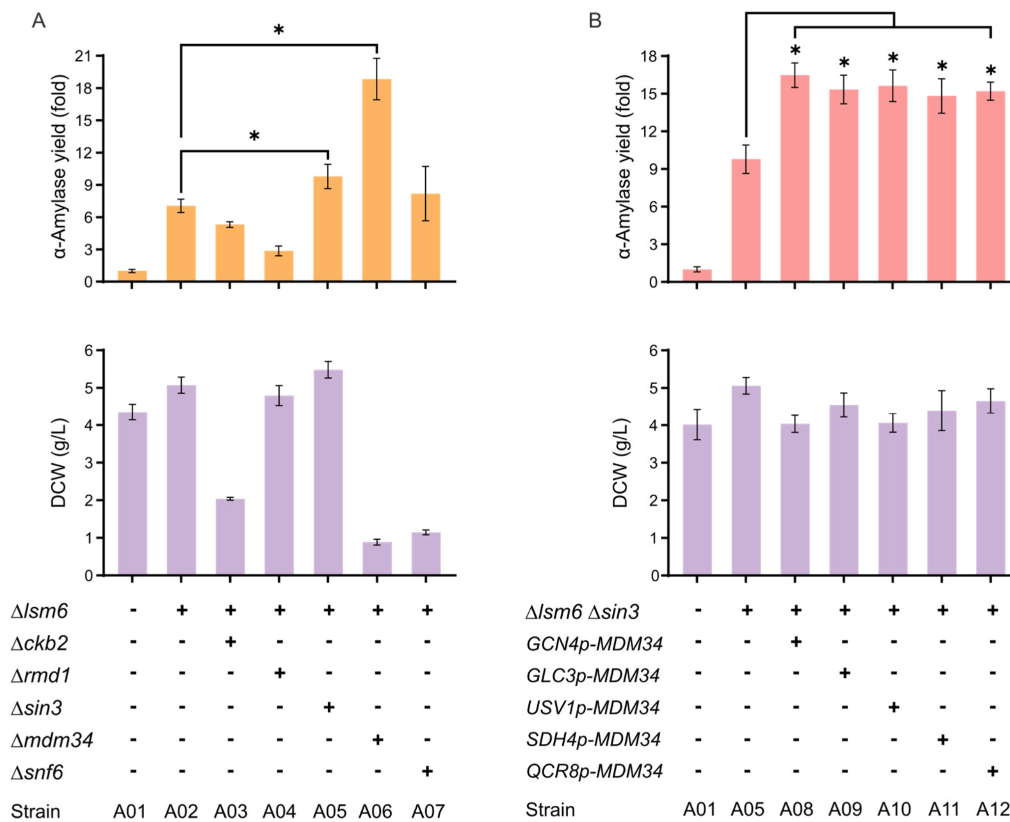

Figure 24. Combinatorial gene deletion of the suppressors of Aβ42 cytotoxicity selected from **Chapter 4.1.1**. (**A**) The effect of combined double gene deletion on α-amylase production and DCW. (**B**) Fine-tuning *MDM34* expression using a group of weak promoters in the A05 (*Δlsm6* and *Δsin3*) background strain. n = 4, error bar = ± SD. Statistical significance **P* < 0.05 (Two-tailed Student's t-test). Adapted from **Paper III**.

For genomic integration, the overexpression cassettes *TEF1p-gene-ADH1t* were integrated into the chromosomal XII-5 locus in the strain A08 (*Δlsm6*, *Δsin3* and *GCN4p-MDM34*), respectively. Out of these four strains, only the overexpression of *CDC48* led to a significant enhancement in α-amylase production compared with A08 (Figure 25). The resulting strain A16 (*Δlsm6*, *Δsin3*, *GCN4p-MDM34* and *TEF1p-CDC48*) produced 18.7-fold α-amylase yield and did not significantly affect the final biomass compared to the control strain A01. None of these four gene targets were previously identified to be relevant to the production of recombinant proteins.

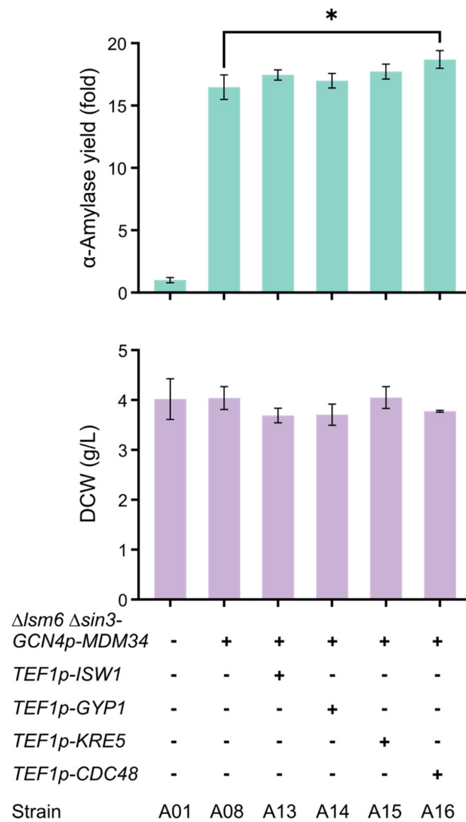

Figure 25. The effect of gene-overexpression targets selected from **Chapter 4.1.1** on α-amylase production and DCW in the A08 (*Δlsm6*, *Δsin3* and *GCN4p-MDM34*) background strain. n = 4, error bar = ± SD. Statistical significance *$P < 0.05$ (Two-tailed Student's t-test). Adapted from **Paper III**.

Lsm6p plays a role in pre-mRNA splicing via the spliceosome (Ingelfinger *et al.*, 2002). The nuclear Lsm6p is involved in RNA processing (Beggs, 2005), while the cytoplasmic Lsm6p is related to mRNA degradation (He & Parker, 2000). Sin3p, a transcriptional corepressor acting via interactions with the histone deacetylase complex, is involved in the negative regulation of the transcription of diverse processes (Grzenda *et al.*, 2009). Our subsequent transcriptomic and proteomic analysis demonstrated that many RNA processing, transcription and translation-associated genes were upregulated in the deletion strains, which supported the fact that deleting *LSM6* and *SIN3* could to some extent alleviate the suppression of transcription and translation and thus increase the production efficiency from DNA to protein. In addition to the function of Mdm34p described above, Cdc48p, a highly conserved AAA ATPase with protein-unfoldase activity, is involved in ERAD. As described in **Chapter 1.3.2.3**, Cdc48p collaborates with its cofactor Ufd1-Npl4p to pull the misfolded proteins out of ER to the cytoplasm by producing unfolded segments in polyubiquitinated proteins that are then targeted to degradation by the proteasome (Olszewski *et al.*, 2019). Based on the SGA screen, the deletion of *CDC48* increased Aβ42 cytotoxicity, implying that Cdc48p might contribute to Aβ42 detoxification by removing the misfolded proteins. Additionally, the increased α-amylase production by overexpressing *CDC48* might be the result of the alleviation of misfolded protein stress and the improvement of protein turnover.

4.1.3 Evaluating the general applicability of effective targets

To evaluate the capacities of the selected suppressors of Aβ42 cytotoxicity for the production of other proteins, we tested two additional heterologous proteins, GLA from *R. oryzae* and Nanobody (Nan, antibody single V-type domain) from the Camelidae family, in our best engineered strain A16 (*Δlsm6*, *Δsin3*, *GCN4p-MDM34* and *TEF1p-CDC48*) (Figure 26). After shake flask cultures, the concentration of secreted GLA was quantified by measuring the enzyme activity and the protein amount of secreted Nan was analyzed by western blot. Compared with the control strain A01, both proteins showed enhanced production (3.1-fold for GLA and 2.9-fold for Nan) by A16. These results indicated that our findings might generally apply to the production of other recombinant proteins.

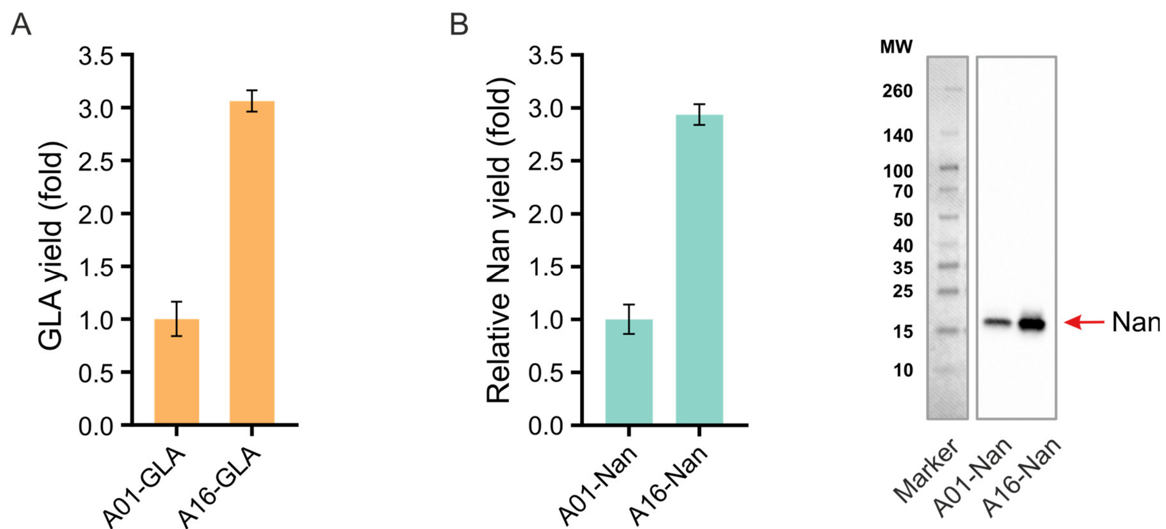

Figure 26. The expression of two heterologous proteins GLA (**A**) and Nan (**B**) in the engineered strain A16. The activity of GLA was measured using the amyloglucosidase assay reagent R-AMGR3 (Megazyme). The relative amount of Nan was quantified by western blot using an anti-6x-His-tag antibody. n = 4, error bar = ± SD. Adapted from **Paper III**.

In this chapter, we investigated the individual effects of suppressors of Aβ42 cytotoxicity from an SGA screening on α-amylase production. Through combinatorial genetic manipulations of several identified positive targets, the best engineered strain A16 (*Δlsm6*, *Δsin3*, *GCN4p-MDM34* and *TEF1p-CDC48*) could produce 18.7-fold α-amylase. Besides, the identified targets can serve as suitable candidates for designing other efficient cell factories and producing other recombinant proteins.

4.2 Novel targets from the pcSecYeast model

As presented in **Chapter 3.2** and **4.1**, the newly identified targets could contribute to optimizing yeast platforms for efficient protein production. Due to the complexity of the protein processing and secretion pathway, the identified positive targets for producing a certain protein sometimes do not have an effect on the production of other proteins. Thus, it is attractive to develop a rational design model for protein secretion and thereby optimize cell factories for different proteins. In this part, we focus on the application of our newly developed pcSecYeast model in guiding *S. cerevisiae* engineering and increasing recombinant protein production.

4.2.1 Brief description of pcSecYeast

As mentioned in **Chapter 1.5.2**, since the first reconstruction of yeast GEM in 2003, there have been multiple rounds of updates. The latest release of the consensus *S. cerevisiae* GEM Yeast8 represents the most comprehensive reconstruction of metabolism in yeast (Lu *et al.*, 2019). Due to the limited size and space of cells, proteome constraints define the limited proteomic resources available within cells, which should be optimally allocated across metabolic pathways and biological processes in response to various environmental conditions (Basan, 2018). The first effort for the integration of proteome constraints into yeast GEMs was enzyme-constrained Yeast7.6, which uses the enzymatic capacity of each metabolic flux calculated from enzyme abundance and turnover number to define its upper limit (Sánchez *et al.*, 2017). Compared to the coarse-grained enzyme-constrained GEMs, the fine-grained approaches that clearly formulate protein expression processes allow for predicting and explaining cell behavior in more detail (Yang *et al.*, 2018). Here, our newly developed pcSecYeast model is one of the extensions of the fine-grained framework of proteome-constrained GEMs, which covers the protein secretory pathway of *S. cerevisiae*.

pcSecYeast first added 92 metabolic reactions to Yeast8 to allow the ability to synthesize precursors required for the secretory pathway, such as glycans and glycosylphosphatidylinositol (GPI) anchor. Subsequently, in addition to adding the reactions of protein expression, translation and degradation, conceptually similar to the earlier published metabolic-expression models (O'Brien *et al.*, 2013, Oftadeh *et al.*, 2021), the model also added the reactions of protein processing and secretion processes including translocation, PTM, folding, misfolding, ERAD and vesicle sorting (Figure 27). Hence, pcSecYeast describes the whole protein secretion process in detail, from nascent polypeptide translocation, protein maturation to trafficking to different final destinations. This model contains 1639 protein-coding genes, representing approximately 70% of total proteome mass based on the PaxDb database, and incorporates 38,020 reactions, 31,824 of which are protein-related, showing the complexity of the secretory pathway.

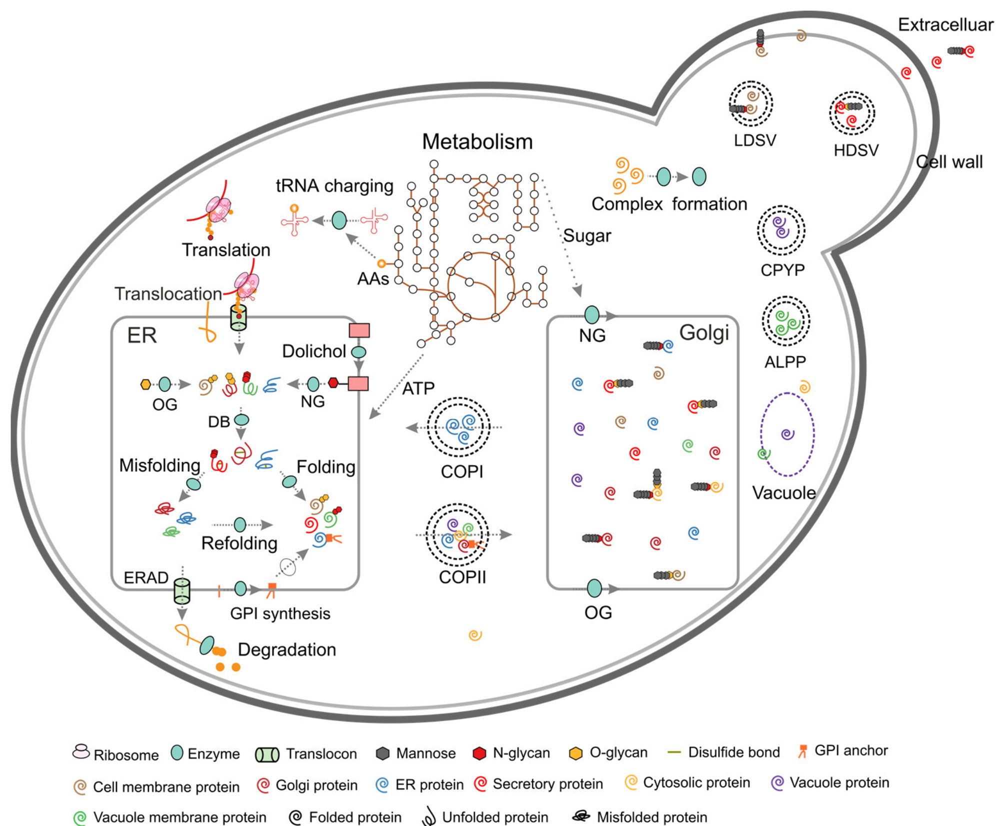

Figure 27. Schematic processes of pcSecYeast. This model includes protein translation, translocation, glycosylation, GPI transfer, ERAD and sorting processes. OG: O-glycosylation; NG: N-glycosylation; DB: Disulfide bond formation; AA: Amino acid; LDSV: Low-density secretory vesicles; HDSV: High-density secretory vesicles; CPYP: carboxypeptidase Y pathway; ALPP: alkaline phosphatase pathway. Adapted from **Paper IV**.

After the protein of interest (including amino acid sequence, disulfide bond number, glycosylation number and GPI number) was added to the model, engineering targets were selected using adapted flux scanning based on enforced objective flux (FSEOF) (Choi *et al.*, 2010) (Figure 28A). FSEOF enforces the flux toward a product formation, which was constrained by a stepwise decrease in the biomass formation rate. The upregulated proteins leading to improved recombinant protein production represented the raw overexpression targets. Then, these predicted targets were ranked with priority scores for further effective selection. Several factors should be considered, including whether this overexpression target is a part of an enzyme complex, contains paralogs, has been significantly changed in abundance, and is abundant in the cell according to PaxDb database (Wang *et al.*, 2015) (Figure 28B).

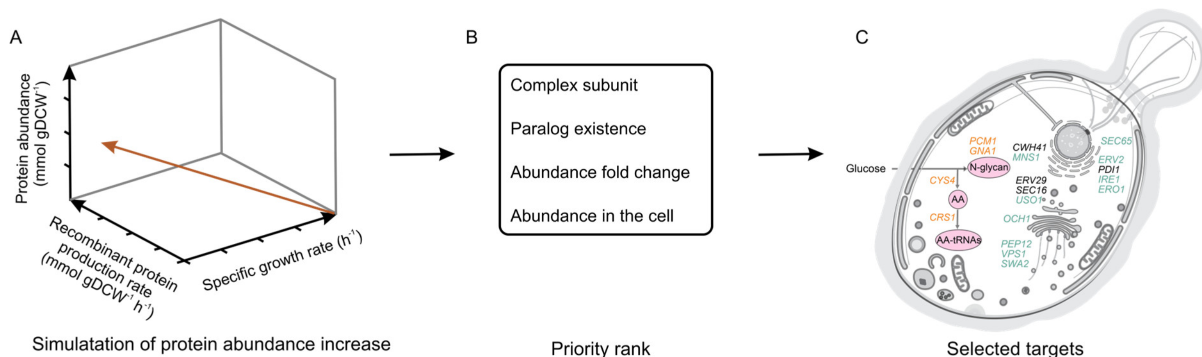

Figure 28. Prediction and selection of overexpression targets for improving α-amylase production. (**A**) An adapted FSEOF method for target prediction. (**B**) Priority rank for further target selection. (**C**) Protein localization of the selected overexpression targets. Yeast compartmentalized figure source: SwissBioPics under CC BY4.0 license. Yellow marked genes indicate metabolic targets, light green marked genes indicate secretory targets, and light green marked genes indicate previously validated secretory targets. Adapted from **Paper IV**.

4.2.2 Verification of the predicted overexpression targets

After selection, 18 targets predicted for improving α-amylase production were experimentally validated. Four metabolic targets are associated with N-glycan synthesis and amino acid synthesis, and 14 secretory targets are involved in translocation, folding, quality control and sorting processes (Figure 28C). Among these targets, overexpression of glucosidase Cwh41p (Qi *et al.*, 2020), COPII vesicle proteins Erv29p (Huang *et al.*, 2018) and Sec16p (Bao *et al.*, 2017), and disulfide isomerase Pdi1p (Huang *et al.*, 2018), respectively, were previously reported to be effective in increasing α-amylase production. Next, we tested the effects of individual gene overexpressions of the remaining 14 targets. In this work, AACD was also used as the host strain, and the selected gene was expressed from the pSPGM1 plasmid under control of the *TEF1* promoter. As shown in Figure 29, the α-amylase production after respective overexpression of *MNS1*, *SEC65*, *ERV2*, *ERO1*, *SWA2* and *CYS4* was significantly increased between 0.32-fold and 1.2-fold compared to the control strain.

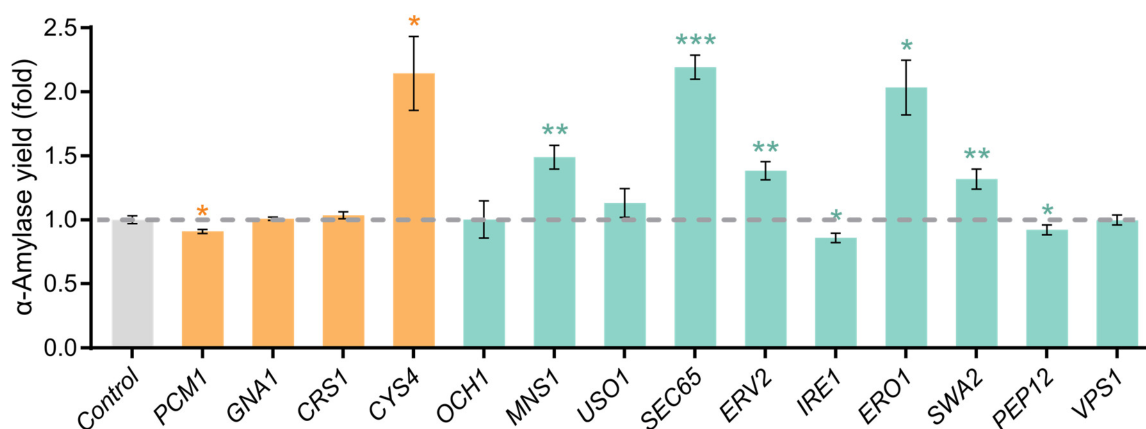

Figure 29. The effect of overexpression of predicted targets on α-amylase production in the AACD background strain. Yellow indicates the overexpression of metabolic targets, while light green indicates the overexpression of secretory targets. The grey line indicates α-amylase production of the control strain. n = 3, error bar = ± SD. Statistical significance *$P < 0.05$, **$P < 0.01$, ***$P < 0.01$ (Two-tailed Student's t-test). Adapted from **Paper IV**.

Mns1p, an α-1,2-mannosidase, is involved in ubiquitin-dependent ERAD and the removal of one mannose residue from a glycosylated protein, which represents the final step in glycoprotein maturation in the ER (Jelinek-Kelly & Herscovics, 1988). An improperly folded protein is released from calnexin and then glycans may be trimmed by Mns1p. These mannose-trimmed glycoproteins are targeted to the ERAD pathway. In addition to the native glycoproteins, α-amylase and its α-factor signal peptide also contain glycosylation sites, which may benefit from the overexpression of *MNS1*. Sec65p, a subunit of SRP that targets nascent peptides from cytosol to ER, is required to associate Srp54p (Hann *et al.*, 1992). Although it is normally involved in co-translational translocation (mentioned in **Chapter 1.3.2.1**), overexpression of *SEC65* seems to benefit post-translational translocation in this case as well. Erv2p and Ero1p are involved in disulfide bond formation (Sevier *et al.*, 2001, Sevier & Kaiser, 2008). As described in **Chapter 1.3.2.3**, overexpression of PDIs and ER oxidase is an effective strategy for improving protein production, which supported our results that overexpression of *ERV2* and *ERO1* was beneficial for α-amylase production. Another gene target, Swa2p, involved in vesicle transport, is required for uncoating clathrin-coated vesicles before fusion of the vesicle with its target membrane (Gall *et al.*, 2000). Cys4p, a cystathionine beta-synthase, is responsible for cysteine biosynthesis (Cherest *et al.*, 1993). Compared to the amino acid composition of *S. cerevisiae* cell protein, a 9-fold higher cysteine amount would be required when producing the same amount α-amylase. This supported the result that overexpression of *CYS4* significantly enhanced the α-amylase production.

In this chapter, we found 9/14 of chosen secretory targets and 1/4 of identified metabolic targets were validated effectively for protein secretion, which indicates a higher accuracy for secretory pathway-related targets. In addition to the selected targets, the model predicted more secretory-related targets, although the metabolic enzyme fraction is higher in this model, which implies that the secretory pathway is probably the rate-limiting step for recombinant protein secretion rather than the metabolic pathways.

5 Conclusion and perspectives

5.1 Conclusion

The work presented in this thesis focused on applying diverse strategies to engineer *S. cerevisiae* platform strains to improve their recombinant protein productivity. Recombinant proteins were chosen because they have a wide range of industrial applications and therapeutic value. Increased protein production will facilitate the profitability of the process and make microbial production of industrial-scale biologicals more attractive compared to traditional production modes. This could be a critical step in improving economic benefits and sustaining long-term development.

The different approaches were described in two parts: 1) how to explore in depth the improved secretory capacity of existing strains and 2) how to rationally optimize the *S. cerevisiae* platform. The first part was presented in **Paper I** and **II** and was based on previously isolated mutant strains. We successfully secreted three biologically active antibody fragments with different configurations using the high-level secretory platform strain HA. We found that the efficient secretion phenotype seen for several heterologous proteins was not valid for all pharmaceuticals. Through RNA-seq analysis, we revealed that both cellular resource allocation and stress response caused by Pex expression might be limiting factors for the low protein production efficiency. We also introduced exact point mutations for 42 protein-sequence-altering mutations from B184 using different CRISPR/Cas9-mediated strategies. We found that 14 targets (point mutations and/or deletions) had positive effects on α-amylase secretion. There were mainly related to stress-related processes, protein degradation, transportation, mRNA processing and export, and DNA replication and repair. Eleven of the 14 beneficial targets were newly identified to be associated with recombinant protein production.

The second part was presented in **Paper III**, **IIIb** and **IV** and investigated potential new target genes using diverse engineering strategies. We first examined the individual effects of suppressors of Aβ42 cytotoxicity from an SGA screening on α-amylase production. We further combined effective targets obtaining an 18.7-fold increase in α-amylase production. We also confirmed that this strategy can be generally applied to the production of other heterologous proteins. On the other hand, we incorporated the protein secretory pathway to the latest release of a *S. cerevisiae* GEM (Yeast8) with proteome-constraints to construct the pcSecYeast model, which allows the simulation of protein synthesis and secretion and prediction of engineering targets for enhancing recombinant protein production. Based on the experimental verification, 10 of 18 predicted targets had positive effects on α-amylase secretion, indicating the high accuracy using our model for predicting and selecting targets for improving recombinant protein production.

The work in this thesis not only increases understanding of the mechanisms associated with efficient protein secretion, but also provides guidelines for constructing novel cell platforms for producing industrial enzymes and pharmaceutical proteins in yeast and even other microorganisms.

5.2 Perspectives

The advances in recombinant DNA technology have opened the possibility of constructing efficient microbial cell factories and producing recombinant proteins in heterologous host organisms. Protein quality, quantification and production costs are essential factors for developing a cell factory for industrial production. Choosing a production platform is only the first step. It is the most challenging and time-consuming journey to understand the protein production machinery and identify rate-limiting factors, and this demands considerable optimizations and modifications in host cells and production pathways. As described above, our work focused on engineering the *S. cerevisiae* platform for improving the production of several recombinant proteins, such as antibody fragments and α-amylase.

Production of mAbs has developed into an extremely important branch of the biopharmaceutical industry. The applications of mAbs are rapidly expanding, especially during the ongoing COVID-19 pandemic. The urgent demands for the upcoming mAbs and therapeutic drugs against the SARS-CoV2 virus promoted a dramatic growth of the biopharmaceutical industry over the past three years. According to **Paper I** and previous work, B184 combining all point mutations improved the production of most recombinant proteins tested, albeit at different extents, which showed its high-secretion capacity (Huang *et al.*, 2015, Huang *et al.*, 2017). However, the structures of these proteins are simple and the molecular sizes are still relatively small. Using B184 to produce more complex full-length antibodies would be very valuable. During this process, glycosylation engineering should be a direction worth considering.

High-level production of recombinant proteins often induces cellular stress, such as from the formation of disulfide bonds, from the large amount of protein accumulation in ER and from misfolded proteins (Martínez *et al.*, 2016). Excessive stress will seriously affect cell growth, reduce protein production and even cause cell apoptosis. In **Paper I**, we demonstrated that increased cellular stress might explain the low efficiency of Pex secretion in the high secretion strain. In **Paper II**, we indicated that some identified beneficial targets were related to various stress processes. In **Paper III** and **IIIb**, we reversely engineered some suppressors of Aβ42 cytotoxicity to alleviate cellular stress. As described in **Chapter 1.3**, many human diseases are also related to the abnormal levels of ER stress. Therefore, future studies can focus on how to efficiently balance stress responses and reduce cellular stress in engineered protein producing strains, which could be of value to the development of protein-based biopharmaceuticals. Interestingly, a previous study reported that adding a moderate amount of antioxidants (such as vitamin C) in the media could protect cells against the harmful effects of ROS accumulation and therefore improve protein production (Bao *et al.*, 2017).

Due to the various possible effects of amino acid substitutions, gene deletion or overexpression are probably not optimal for investigating the association between mutated genes and protein production (Thak *et al.*, 2020). In contrast, point mutations might lead to a fine-tuning of enzyme activity or gene expression level. In **Paper II**, we introduced the exact point mutations from the evolutionary line toward B184 and analyzed the regulatory mechanisms of each beneficial gene target. However, due

to the limited time, we did not perform verification experiments to confirm all of our hypotheses and did not determine the actual enzyme activity or gene expression levels for the effective targets. More studies can further investigate these targets, such as carrying out RNA-seq analysis to explore the effect of the specific mutation on a global level, or conducting protein activity tests and examining protein abundance to study the precise effects of the mutated site on the protein. Besides, by analyzing these targets, we found it should be attractive to engineer the autophagic process, oxygen environment, cell cycle and the bi-directional vesicular trafficking pathway between ER and Golgi for increasing protein production in the future.

A GEM is a valuable knowledge base to systematically understand and rationally engineer an organism's metabolism. In our case, pcSecYeast as a proteome-constrained GEM provides a platform to simulate the protein secretory pathway and its interaction with metabolism and gene expression in yeast (**Paper IV**). Our model currently integrates only the most fundamental processes in the secretory pathway. It will potentially increase the prediction accuracy if UPR, intra Golgi transport and various sorting pathways after Golgi processing are added to the pcSecYeast model. In **Paper III**, we found that combinatorial engineering demonstrates a great advantage for improving protein production. For the moment, our model does not have the capacity to predict combinatorial targets. It would be very important to add this ability to this model in the future. Based on this model, we expect that many similar types of genome-scale secretory models could be built for a broader range of recombinant protein-producing cell factories. In this thesis, we identified a number of beneficial gene targets and investigated their detailed function in *S. cerevisiae* (**Paper II-IV**). Since many essential cellular processes are strongly conserved among eukaryotic cells and other hosts have their own advantages for protein production, these targets can be applied to other organisms, such as filamentous fungi or mammalian cells, to facilitate their protein secretion.

Acknowledgements

It was a long and unforgettable journey during pursuing my PhD degree, and I feel very happy to approach the finish line finally. I have experienced a lot on this journey, and it has been one of hope, happiness, stress, confusion, helplessness, efforts and harvest. Although I am looking forward to this destination, what I cherish most are many amazing people I have met along the way.

My deepest thanks go to my supervisor, Verena. As I start my acknowledgements, the scene from two years ago is vivid in my mind. When I was struggling with my projects and papers, it is you who is willing to help me through those difficult times together. At that time, I felt there was little hope for graduation, but it is you who try to pull me out of a situation like a mud puddle and make it possible. You can always find my problems and find ways to correct them. To improve my ability to deal with different questions, you took your weekly time to talk with me. What impressed me the most was that you taught me that when faced with a question, I should first think from its principle and then dissect it from different directions. You always guide me step by step to think about questions in depth. This was an unforgettable experience and I have learned so much from you about how to think and perform research. If I continue to do research, I do believe that this experience will be an invaluable asset in my future. Thanks for always showing your patience with my constant mistakes and for always very thoroughly reading my various drafts carefully. Thanks a lot for all the efforts and support you put into my PhD destination. Thanks to Xin for all the discussions and good suggestions to my projects. I admire your passion and perseverance for your work, projects and life. Thanks to Ivan and Jens for being my examiners, and Dina and Martin for your supervision. I also learned a lot from you.

I've received many helps from lots of people during my studies. Thanks to Xiaowei for many valuable contributions and constructive suggestions. Thank you for taking the time to inspire me, help me, and keeping patient with my endless questions about my projects. It has been a pleasure to know you as a colleague and as a friend. Thanks to Xin, Feiran and Chunjun for trusting me and allowing me to join your amazing projects. I have learnt a great deal from you, and I am looking forward to future collaborations. I am grateful to Vijay, Ela, Marta, Dany, Veronica, Oliver and Mikael for your time and expert input in helping me prepare for my half-way seminar. This really means so much to me. I am thankful to Lucy, Olena, Quanli, Xiang, Tao, and Tyler for your helpful discussions and feedback regarding my projects.

I've been surrounded by many great people who never hesitate to help with work- and non-work-related problems in the past few years. Thanks to Yi, Quanli and Kate for the fantastic time in the gym. Thanks to Xiaowei, Chunjun and Guokun for the company on the countless lunchtime. I am so fortunate that I could share the office with you, Lucy, and thanks for cheering me up and giving suggestions on my life. Thanks to Feiran, Qi, Zhengming, Yu, Peishun, Jiwei, Xin, Yating, Jichen, Yi, Quanli, Xufeng, Gang, Lei, Boyang, Hao Luo, Yihan for those wonderful afterworks with cooking. Thanks to Elin, Moyra, Jennie and Katarina for your support in helping me get through the stressful period. I would like to thank the research engineers and administrators Emelie, Angelica, Yili, Abder, Marie, Erik, Fredrik, Gunilla, Anne-Lise, Erica and Martina for your professional assistance in the lab

and office. In addition, I want to thank all the people who have made my stay comfortable and colorful. Thanks to Jing, Yun, Yassi, Fei, Xiaozhi, Jian, Le, Mauro, Santosh, Li Han, Yanyan Chen, Jerry, Juan, Lingyun, Xiaofan, Zhiwei, Zongjie, Hongzhong, Liming, Joakim, Patrick, Michi, Shadi, Max, Punchalee, Andrea, John, Mihail, Sara. A very special thanks to my dear friends, Lin, Yilan, and Peng who are always being there for me.

Last of all, my biggest thanks go to my family and my parents. Thank you for your unconditional love and support. I love you all.

References

Abe F (2007) Induction of *DAN/TIR* yeast cell wall mannoprotein genes in response to high hydrostatic pressure and low temperature. *FEBS Lett* **581**: 4993-4998.

Ahn J, Jang MJ, Ang KS, Lee H, Choi ES & Lee DY (2016) Codon optimization of *Saccharomyces cerevisiae* mating factor alpha prepro-leader to improve recombinant protein production in *Pichia pastoris*. *Biotechnol Lett* **38**: 2137-2143.

Al-Hawash AB, Zhang X & Ma F (2017) Strategies of codon optimization for high-level heterologous protein expression in microbial expression systems. *Gene Reports* **9**: 46-53.

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.

Alter TB & Ebert BE (2019) Determination of growth-coupling strategies and their underlying principles. *BMC Bioinformatics* **20**: 447.

Arakel EC & Schwappach B (2018) Formation of COPI-coated vesicles at a glance. *J Cell Sci* **131**(5): jcs209890.

Ariño J, Casamayor A & González A (2011) Type 2C protein phosphatases in fungi. *Eukaryot Cell* **10**: 21-33.

Aza P, Molpeceres G, de Salas F & Camarero S (2021) Design of an improved universal signal peptide based on the α-factor mating secretion signal for enzyme production in yeast. *Cell Mol Life Sci* **78**: 3691-3707.

Backman TWH, Ando D, Singh J, Keasling JD & García Martín H (2018) Constraining genome-scale models to represent the bow tie structure of metabolism for (13)C metabolic flux analysis. *Metabolites* **8**(1): 3.

Baeshen NA, Baeshen MN, Sheikh A, Bora RS, Ahmed MM, Ramadan HA, Saini KS & Redwan EM (2014) Cell factories for insulin production. *Microb Cell Fact* **13**: 141.

Bao J, Huang M, Petranovic D & Nielsen J (2017) Moderate expression of *SEC16* increases protein secretion by *Saccharomyces cerevisiae*. *Appl Environ Microbiol* **83** (14): e03400-16.

Bao J, Huang M, Petranovic D & Nielsen J (2018) Balanced trafficking between the ER and the Golgi apparatus increases protein secretion in yeast. *AMB Express* **8**: 37.

Bao Z, HamediRad M, Xue P, Xiao H, Tasan I, Chao R, Liang J & Zhao H (2018) Genome-scale engineering of *Saccharomyces cerevisiae* with single-nucleotide precision. *Nat Biotechnol* **36**: 505-508.

Basan M (2018) Resource allocation and metabolism: the search for governing principles. *Curr Opin Microbiol* **45**: 77-83.

Beggs JD (2005) Lsm proteins and RNA processing. *Biochem Soc Trans* **33**: 433-438.

Benjamin JJ, Poon PP, Lewis SM, Auger A, Wong TA, Singer RA & Johnston GC (2011) The yeast Arf GTPase-activating protein Age1 is regulated by phospholipase D for post-Golgi vesicular transport. *J Biol Chem* **286**: 5187-5196.

Besada-Lombana PB & Da Silva NA (2019) Engineering the early secretory pathway for increased protein secretion in *Saccharomyces cerevisiae*. *Metab Eng* **55**: 142-151.

Björkeroth J, Campbell K, Malina C, Yu R, Di Bartolomeo F & Nielsen J (2020) Proteome reallocation from amino acid biosynthesis to ribosomes enables yeast to grow faster in rich media. *Proc Natl Acad Sci U S A* **117**: 21804-21812.

Blakely EL, Mitchell AL, Fisher N, Meunier B, Nijtmans LG, Schaefer AM, Jackson MJ, Turnbull DM & Taylor RW (2005) A mitochondrial cytochrome b mutation causing severe respiratory chain enzyme deficiency in humans and yeast. *Febs j* **272**: 3583-3592.

Bleichrodt R-J & Read ND (2019) Flow cytometry and FACS applied to filamentous fungi. *Fungal Biology Reviews* **33**: 1-15.

Bricker DK, Taylor EB, Schell JC*, et al.* (2012) A mitochondrial pyruvate carrier required for pyruvate uptake in yeast, *Drosophila*, and humans. *Science* **337**: 96-100.

Brzozowski AM & Davies GJ (1997) Structure of the *Aspergillus oryzae* alpha-amylase complexed with the inhibitor acarbose at 2.0 A resolution. *Biochemistry* **36**: 10837-10845.

Burda P & Aebi M (1999) The dolichol pathway of N-linked glycosylation. *Biochim Biophys Acta* **1426**: 239-257.

Cakar ZP, Turanli-Yildiz B, Alkim C & Yilmaz U (2012) Evolutionary engineering of *Saccharomyces cerevisiae* for improved industrially important properties. *FEMS Yeast Res* **12**: 171-182.

Campbell K, Westholm J, Kasvandik S, Di Bartolomeo F, Mormino M & Nielsen J (2020) Building blocks are synthesized on demand during the yeast cell cycle. *Proc Natl Acad Sci U S A* **117**: 7575-7583.

Cavaillon JM & Legout S (2022) Louis Pasteur: Between Myth and Reality. *Biomolecules* **12**(4): 596.

Cavener DR (1987) Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. *Nucleic Acids Res* **15**: 1353-1361.

Chen X & Petranovic D (2015) Amyloid-β peptide-induced cytotoxicity and mitochondrial dysfunction in yeast. *FEMS Yeast Res* **15**(6): fov061.

Chen X, Li C & Liu H (2021) Enhanced recombinant protein production under special environmental stress. *Front Microbiol* **12**: 630814.

Chen X, Bisschops MMM, Agarwal NR, Ji B, Shanmugavel KP & Petranovic D (2017) Interplay of energetics and ER stress exacerbates Alzheimer's amyloid-β (Aβ) toxicity in yeast. *Front Mol Neurosci* **10**: 232.

Chen X, Ji B, Hao X, Li X, Eisele F, Nyström T & Petranovic D (2020) *FMN* reduces Amyloid-β toxicity in yeast by regulating redox status and cellular metabolism. *Nat Commun* **11**: 867.

Chen Y, Li F & Nielsen J (2022) Genome-scale modeling of yeast metabolism: retrospectives and perspectives. *FEMS Yeast Res* **22**(1): foac003.

Chen Y, Partow S, Scalcinati G, Siewers V & Nielsen J (2012) Enhancing the copy number of episomal plasmids in *Saccharomyces cerevisiae* for improved protein production. *FEMS Yeast Res* **12**: 598-607.

Chen Y, Li F, Mao J, Chen Y & Nielsen J (2021) Yeast optimizes metal utilization based on metabolic network and enzyme kinetics. *Proc Natl Acad Sci U S A* **118**(12): e2020154118.

Cherest H, Thomas D & Surdin-Kerjan Y (1993) Cysteine biosynthesis in *Saccharomyces cerevisiae* occurs through the transsulfuration pathway which has been built up by enzyme recruitment. *J Bacteriol* **175**: 5366-5374.

Chiba Y & Akeboshi H (2009) Glycan engineering and production of 'humanized' glycoprotein in yeast cells. *Biol Pharm Bull* **32**: 786-795.

Chiba Y, Suzuki M, Yoshida S, Yoshida A, Ikenaga H, Takeuchi M, Jigami Y & Ichishima E (1998) Production of human compatible high mannose-type (Man5GlcNAc2) sugar chains in *Saccharomyces cerevisiae*. *J Biol Chem* **273**: 26298-26304.

Chigira Y, Oka T, Okajima T & Jigami Y (2008) Engineering of a mammalian O-glycosylation pathway in the yeast *Saccharomyces cerevisiae*: production of O-fucosylated epidermal growth factor domains. *Glycobiology* **18**: 303-314.

Chng J, Wang T, Nian R, Lau A, Hoi KM, Ho SC, Gagnon P, Bi X & Yang Y (2015) Cleavage efficient 2A peptides for high level monoclonal antibody expression in CHO cells. *MAbs* **7**: 403-412.

Choi BK, Bobrowicz P, Davidson RC, Hamilton SR, Kung DH, Li H, Miele RG, Nett JH, Wildt S & Gerngross TU (2003) Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast *Pichia pastoris*. *Proc Natl Acad Sci U S A* **100**: 5022-5027.

Choi HS, Lee SY, Kim TY & Woo HM (2010) In silico identification of gene amplification targets for improvement of lycopene production. *Appl Environ Microbiol* **76**: 3097-3105.

Coco WM, Levinson WE, Crist MJ, Hektor HJ, Darzins A, Pienkos PT, Squires CH & Monticello DJ (2001) DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat Biotechnol* **19**: 354-359.

Crick F (1970) Central dogma of molecular biology. *Nature* **227**: 561-563.

Curran KA, Karim AS, Gupta A & Alper HS (2013) Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metab Eng* **19**: 88-97.

Davis R, Schooley K, Rasmussen B, Thomas J & Reddy P (2000) Effect of PDI overexpression on recombinant protein secretion in CHO cells. *Biotechnol Prog* **16**: 736-743.

Davy AM, Kildegaard HF & Andersen MR (2017) Cell Factory Engineering. *Cell Syst* **4**: 262-275.

de Ruijter JC & Frey AD (2015) Analysis of antibody production in *Saccharomyces cerevisiae:* effects of ER protein quality control disruption. *Appl Microbiol Biotechnol* **99**: 9061-9071.

de Ruijter JC, Koskela EV & Frey AD (2016) Enhancing antibody folding and secretion by tailoring the *Saccharomyces cerevisiae* endoplasmic reticulum. *Microb Cell Fact* **15**: 87.

Delic M, Valli M, Graf AB, Pfeffer M, Mattanovich D & Gasser B (2013) The secretory pathway: exploring yeast diversity. *FEMS Microbiol Rev* **37**: 872-914.

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

Di Bartolomeo F, Malina C, Campbell K, Mormino M, Fuchs J, Vorontsov E, Gustafsson CM & Nielsen J (2020) Absolute yeast mitochondrial proteome quantification reveals trade-off between biosynthesis and energy generation during diauxic shift. *Proc Natl Acad Sci U S A* **117**: 7524-7535.

Domenzain I, Li F, Kerkhoven EJ & Siewers V (2021) Evaluating accessibility, usability and interoperability of genome-scale metabolic models for diverse yeasts species. *FEMS Yeast Res* **21**(1): foab002.

Egel-Mitani M, Andersen AS, Diers II, 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.

Eguchi Y, Makanae K, Hasunuma T, Ishibashi Y, Kito K & Moriya H (2018) Estimating the protein burden limit of yeast cells by measuring the expression limits of glycolytic proteins. *Elife* **7**: e34595.

Ellgaard L & Helenius A (2003) Quality control in the endoplasmic reticulum. *Nat Rev Mol Cell Biol* **4**: 181-191.

Feizi A, Österlund T, Petranovic D, Bordel S & Nielsen J (2013) Genome-scale modeling of the protein secretory machinery in yeast. *PLoS One* **8**: e63284.

Ferrer-Miralles N, Domingo-Espín J, Corchero JL, Vázquez E & Villaverde A (2009) Microbial factories for recombinant pharmaceuticals. *Microb Cell Fact* **8**: 17.

Ferrigno P, Posas F, Koepp D, Saito H & Silver PA (1998) Regulated nucleo/cytoplasmic exchange of HOG1 MAPK requires the importin beta homologs NMD5 and XPO1. *Embo j* **17**: 5606-5614.

Feyder S, De Craene JO, Bär S, Bertazzi DL & Friant S (2015) Membrane trafficking in the yeast *Saccharomyces cerevisiae* model. *Int J Mol Sci* **16**: 1509-1525.

Fitzgerald I & Glick BS (2014) Secretion of a foreign protein from budding yeasts is enhanced by cotranslational translocation and by suppression of vacuolar targeting. *Microb Cell Fact* **13**: 125.

Frykman S & Srienc F (2001) Cell cycle-dependent protein secretion by *Saccharomyces cerevisiae*. *Biotechnol Bioeng* **76**: 259-268.

Fu X, Zhang Y, Xu Q, Sun X & Meng F (2021) Recent advances on sorting methods of high-throughput droplet-based microfluidics in enzyme directed evolution. *Front Chem* **9**: 666867.

Förster J, Famili I, Fu P, Palsson B & Nielsen J (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* **13**: 244-253.

Gall WE, Higginbotham MA, Chen C, Ingram MF, Cyr DM & Graham TR (2000) The auxilin-like phosphoprotein Swa2p is required for clathrin function in yeast. *Curr Biol* **10**: 1349-1358.

Gampel A & Cech TR (1991) Binding of the CBP2 protein to a yeast mitochondrial group I intron requires the catalytic core of the RNA. *Genes Dev* **5**: 1870-1880.

Gampel A, Nishikimi M & Tzagoloff A (1989) CBP2 protein promotes in vitro excision of a yeast mitochondrial group I intron. *Mol Cell Biol* **9**: 5424-5433.

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.

Gaynor EC, Graham TR & Emr SD (1998) COPI in ER/Golgi and intra-Golgi transport: do yeast COPI mutants point the way? *Biochim Biophys Acta* **1404**: 33-51.

Gething MJ (1999) Role and regulation of the ER chaperone BiP. *Semin Cell Dev Biol* **10**: 465-472.

Gilbert LA, Horlbeck MA, Adamson B*, et al.* (2014) Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**: 647-661.

Goto M (2007) Protein O-glycosylation in fungi: diverse structures and multiple functions. *Biosci Biotechnol Biochem* **71**: 1415-1427.

Gottshall EY, Seebart C, Gatlin JC & Ward NL (2014) Spatially segregated transcription and translation in cells of the endomembrane-containing bacterium *Gemmata obscuriglobus*. *Proc Natl Acad Sci U S A* **111**: 11067-11072.

Grzenda A, Lomberk G, Zhang JS & Urrutia R (2009) Sin3: master scaffold and transcriptional corepressor. *Biochim Biophys Acta* **1789**: 443-450.

Grzeschik J, Yanakieva D, Roth L, Krah S, Hinz SC, Elter A, Zollmann T, Schwall G, Zielonka S & Kolmar H (2019) Yeast surface display in combination with fluorescence-activated cell sorting enables the rapid isolation of antibody fragments derived from immunized chickens. *Biotechnol J* **14**: e1800466.

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

Gutierrez JM, Feizi A, Li S*, et al.* (2020) Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. *Nat Commun* **11**: 68.

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.

Hamilton R, Watanabe CK & de Boer HA (1987) Compilation and comparison of the sequence context around the AUG startcodons in *Saccharomyces cerevisiae* mRNAs. *Nucleic Acids Res* **15**: 3581-3593.

Hamilton SR & Gerngross TU (2007) Glycosylation engineering in yeast: the advent of fully humanized yeast. *Curr Opin Biotechnol* **18**: 387-392.

Hamilton SR, Bobrowicz P, Bobrowicz B*, et al.* (2003) Production of complex human glycoproteins in yeast. *Science* **301**: 1244-1246.

Hamilton SR, Davidson RC, Sethuraman N*, et al.* (2006) Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science* **313**: 1441-1443.

Hann BC, Stirling CJ & Walter P (1992) *SEC65* gene product is a subunit of the yeast signal recognition particle required for its integrity. *Nature* **356**: 532-533.

Hartley DM, Walsh DM, Ye CP, Diehl T, Vasquez S, Vassilev PM, Teplow DB & Selkoe DJ (1999) Protofibrillar intermediates of amyloid beta-protein induce acute electrophysiological changes and progressive neurotoxicity in cortical neurons. *J Neurosci* **19**: 8876-8884.

He W & Parker R (2000) Functions of Lsm proteins in mRNA degradation and splicing. *Curr Opin Cell Biol* **12**: 346-350.

Heo Y, Yoon HJ, Ko H, Jang S & Lee HH (2020) Crystal structures of Uso1 membrane tether reveal an alternative conformation in the globular head domain. *Sci Rep* **10**: 9544.

Herrgård MJ, Lee BS, Portnoy V & Palsson B (2006) Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res* **16**: 627-635.

Hohmann S (2002) Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol Mol Biol Rev* **66**: 300-372.

Holliger P & Hudson PJ (2005) Engineered antibody fragments and the rise of single domains. *Nat Biotechnol* **23**: 1126-1136.

Hong KK, Vongsangnak W, Vemuri GN & Nielsen J (2011) Unravelling evolutionary strategies of yeast for improving galactose utilization through integrated systems level analysis. *Proc Natl Acad Sci U S A* **108**: 12179-12184.

Hou J, Tyo K, Liu Z, Petranovic D & Nielsen J (2012) Engineering of vesicle trafficking improves heterologous protein secretion in *Saccharomyces cerevisiae*. *Metab Eng* **14**: 120-127.

Hou J, Tyo KE, Liu Z, Petranovic D & Nielsen J (2012) Metabolic engineering of recombinant protein secretion by *Saccharomyces cerevisiae*. *FEMS Yeast Res* **12**: 491-510.

Hou J, Osterlund T, Liu Z, Petranovic D & Nielsen J (2013) Heat shock response improves heterologous protein secretion in *Saccharomyces cerevisiae*. *Appl Microbiol Biotechnol* **97**: 3559-3568.

Hsu PD, Lander ES & Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**: 1262-1278.

Huang M, Bao J & Nielsen J (2014) Biopharmaceutical protein production by *Saccharomyces cerevisiae*: current state and future prospects. *Pharmaceutical Bioprocessing* **2**: 167-182.

Huang M, Bao J, Hallström BM, Petranovic D & Nielsen J (2017) Efficient protein production by yeast requires global tuning of metabolism. *Nat Commun* **8**: 1131.

Huang M, Wang G, Qin J, Petranovic D & Nielsen J (2018) Engineering the protein secretory pathway of *Saccharomyces cerevisiae* enables improved protein production. *Proc Natl Acad Sci U S A* **115**: E11025-e11032.

Huang M, Bai Y, Sjostrom SL, Hallström BM, Liu Z, Petranovic D, Uhlén M, Joensson HN, Andersson-Svahn H & Nielsen J (2015) Microfluidic screening and whole-genome sequencing identifies mutations associated with improved protein secretion by yeast. *Proc Natl Acad Sci U S A* **112**: E4689-4696.

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.

Ikemura T (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J Mol Biol* **146**: 1-21.

Ingelfinger D, Arndt-Jovin DJ, Lührmann R & Achsel T (2002) The human LSm1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrnl in distinct cytoplasmic foci. *Rna* **8**: 1489-1501.

Irani ZA, Kerkhoven EJ, Shojaosadati SA & Nielsen J (2016) Genome-scale metabolic model of *Pichia pastoris* with native and humanized glycosylation of recombinant proteins. *Biotechnol Bioeng* **113**: 961-969.

Ishchuk OP, Domenzain I, Sánchez BJ, Muñiz-Paredes F, Martínez JL, Nielsen J & Petranovic D (2022) Genome-scale modeling drives 70-fold improvement of intracellular heme production in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **119**: e2108245119.

Ishino Y, Shinagawa H, Makino K, Amemura M & Nakata A (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* **169**: 5429-5433.

Itakura K, Hirose T, Crea R, Riggs AD, Heyneker HL, Bolivar F & Boyer HW (1977) Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science* **198**: 1056-1063.

Ito Y, Yamanishi M, Ikeuchi A, Imamura C, Tokuhiro K, Kitagawa T & Matsuyama T (2013) Characterization of five terminator regions that increase the protein yield of a transgene in *Saccharomyces cerevisiae*. *J Biotechnol* **168**: 486-492.

Jelinek-Kelly S & Herscovics A (1988) Glycoprotein biosynthesis in *Saccharomyces cerevisiae*. Purification of the alpha-mannosidase which removes one specific mannose residue from Man9GlcNAc. *J Biol Chem* **263**: 14757-14763.

Jiang B, Ram AF, Sheraton J, Klis FM & Bussey H (1995) Regulation of cell wall beta-glucan assembly: *PTC1* negatively affects *PBS2* action in a pathway that includes modulation of *EXG1* transcription. *Mol Gen Genet* **248**: 260-269.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA & Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816-821.

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

Kalies KU & Hartmann E (1998) Protein translocation into the endoplasmic reticulum (ER)--two similar routes with different modes. *Eur J Biochem* **254**: 1-5.

Kesik-Brodacka M (2018) Progress in biopharmaceutical development. *Biotechnol Appl Biochem* **65**: 306-322.

Kintaka R, Makanae K & Moriya H (2016) Cellular growth defects triggered by an overload of protein localization processes. *Sci Rep* **6**: 31774.

Kjeldsen T, Pettersson AF & Hach M (1999) The role of leaders in intracellular transport and secretion of the insulin precursor in the yeast *Saccharomyces cerevisiae*. *J Biotechnol* **75**: 195-208.

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.

Kornmann B, Currie E, Collins SR, Schuldiner M, Nunnari J, Weissman JS & Walter P (2009) An ER-mitochondria tethering complex revealed by a synthetic biology screen. *Science* **325**: 477-481.

Korpys-Woźniak P, Kubiak P & Celińska E (2021) Secretory helpers for enhanced production of heterologous proteins in *Yarrowia lipolytica*. *Biotechnol Rep (Amst)* **32**: e00669.

Kozak M (1986) Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proceedings of the National Academy of Sciences* **83**: 2850-2854.

Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**: 283-292.

Krambeck FJ & Betenbaugh MJ (2005) A mathematical model of N-linked glycosylation. *Biotechnol Bioeng* **92**: 711-728.

Kubiak M, Borkowska M, Białas W, Korpys P & Celińska E (2019) Feeding strategy impacts heterologous protein production in *Yarrowia lipolytica* fed-batch cultures-Insight into the role of osmolarity. *Yeast* **36**: 305-318.

Kuersten S & Goodwin EB (2003) The power of the 3' UTR: translational control and development. *Nat Rev Genet* **4**: 626-637.

Lanza AM, Curran KA, Rey LG & Alper HS (2014) A condition-specific codon optimization approach for improved heterologous gene expression in *Saccharomyces cerevisiae*. *BMC Syst Biol* **8**: 33.

Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS & Qi LS (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc* **8**: 2180-2196.

Le Marquer M, San Clemente H, Roux C, Savelli B & Frei Dit Frey N (2019) Identification of new signalling peptides through a genome-wide survey of 250 fungal secretomes. *BMC Genomics* **20**: 64.

Lee J, Moir RD, McIntosh KB & Willis IM (2012) TOR signaling regulates ribosome and tRNA synthesis via LAMMER/Clk and GSK-3 family kinases. *Mol Cell* **45**: 836-843.

Lee T-H, Bae Y-H, Kim M-D & Seo J-H (2012) Overexpression of *HAC1* gene increased levels of both intracellular and secreted human kringle fragment in *Saccharomyces cerevisiae*. *Process Biochemistry* **47**: 2300-2305.

Légiot A, Céré C, Dupoiron T, Kaabouni M, Camougrand N & Manon S (2019) Mitochondria-Associated Membranes (MAMs) are involved in Bax mitochondrial localization and cytochrome c release. *Microb Cell* **6**: 257-266.

Levi T, Sloutskin A, Kalifa R, Juven-Gershon T & Gerlitz O (2020) Efficient *in vivo* introduction of point mutations using ssODN and a co-CRISPR approach. *Biol Proced Online* **22**: 14.

Li D, Wu Q & Reetz MT (2020) Focused rational iterative site-specific mutagenesis (FRISM). *Methods Enzymol* **643**: 225-242.

Li J, Zhang MM, Wan C, Den Haan R, Bai FW & Zhao XQ (2020) Improved cellulase production in recombinant *Saccharomyces cerevisiae* by disrupting the cell wall protein-encoding gene *CWP2*. *J Biosci Bioeng* **129**: 165-171.

Li Z & Rinas U (2020) Recombinant protein production associated growth inhibition results mainly from transcription and not from translation. *Microb Cell Fact* **19**: 83.

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

Liu L, Zhang Y, Liu Z, Petranovic D & Nielsen J (2015) Improving heterologous protein secretion at aerobic conditions by activating hypoxia-induced genes in *Saccharomyces cerevisiae*. *FEMS Yeast Res* **15**(7): fov070.

Liu Q, Liu Y, Li G, Savolainen O, Chen Y & Nielsen J (2021) *De novo* biosynthesis of bioactive isoflavonoids by engineered yeast cell factories. *Nat Commun* **12**: 6085.

Liu SH, Chou WI, Lin SC, Sheu CC & Chang MD (2005) Molecular genetic manipulation of *Pichia pastoris SEC4* governs cell growth and glucoamylase secretion. *Biochem Biophys Res Commun* **336**: 1172-1180.

Liu Y, Mo WJ, Shi TF, Wang MZ, Zhou JG, Yu Y, Yew WS & Lu H (2018) Mutational Mtc6p attenuates autophagy and improves secretory expression of heterologous proteins in *Kluyveromyces marxianus*. *Microb Cell Fact* **17**: 144.

Liu Z, Tyo KE, Martínez JL, Petranovic D & Nielsen J (2012) Different expression systems for production of recombinant proteins in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* **109**: 1259-1268.

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.

Liu Z, Österlund T, Hou J, Petranovic D & Nielsen J (2013) Anaerobic α-amylase production and secretion with fumarate as the final electron acceptor in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* **79**: 2962-2967.

Liu Z, Liu L, Österlund 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.

Lu H, Li F, Sánchez BJ*, et al.* (2019) A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat Commun* **10**: 3586.

Lu P, Lu G, Yan C, Wang L, Li W & Yin P (2012) Structure of the mRNA splicing complex component Cwc2: insights into RNA recognition. *Biochem J* **441**: 591-597.

Madhavan A, Arun KB, Sindhu R, Krishnamoorthy J, Reshmy R, Sirohi R, Pugazhendi A, Awasthi MK, Szakacs G & Binod P (2021) Customized yeast cell factories for biopharmaceuticals: from cell engineering to process scale up. *Microb Cell Fact* **20**: 124.

Malina C, Yu R, Björkeroth J, Kerkhoven EJ & Nielsen J (2021) Adaptations in metabolism and protein translation give rise to the Crabtree effect in yeast. *Proc Natl Acad Sci U S A* **118**(51): e2112836118.

Mans R, van Rossum HM, Wijsman M, Backx A, Kuijpers NG, van den Broek M, Daran-Lapujade P, Pronk JT, van Maris AJ & Daran JM (2015) CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res* **15**(2): fov004.

Marck C, Lefebvre O, Carles C, Riva M, Chaussivert N, Ruet A & Sentenac A (1993) The TFIIIB-assembling subunit of yeast transcription factor TFIIIC has both tetratricopeptide repeats and basic helix-loop-helix motifs. *Proc Natl Acad Sci U S A* **90**: 4027-4031.

Martin DF, Maguire MG, Ying G, Grunwald JE, Fine SL & Jaffe GJ (2011) Ranibizumab and bevacizumab for neovascular age-related macular degeneration. *The New England journal of medicine* **364**: 1897-1908.

Martínez JL, Liu L, Petranovic D & Nielsen J (2012) Pharmaceutical protein production by yeast: towards production of human blood proteins by microbial fermentation. *Curr Opin Biotechnol* **23**: 965-971.

Martínez JL, Liu L, Petranovic D & Nielsen J (2015) Engineering the oxygen sensing regulation results in an enhanced recombinant human hemoglobin production by *Saccharomyces cerevisiae*. *Biotechnol Bioeng* **112**: 181-188.

Martínez JL, Meza E, Petranovic D & Nielsen J (2016) The impact of respiration and oxidative stress response on recombinant α-amylase production by *Saccharomyces cerevisiae*. *Metab Eng Commun* **3**: 205-210.

Marx H, Sauer M, Resina D, Vai M, Porro D, Valero F, Ferrer P & Mattanovich D (2006) Cloning, disruption and protein secretory phenotype of the GAS1 homologue of *Pichia pastoris*. *FEMS Microbiol Lett* **264**: 40-47.

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.

Mattanovich D, Branduardi P, Dato L, Gasser B, Sauer M & Porro D (2012) Recombinant protein production in yeasts. *Methods Mol Biol* **824**: 329-358.

Mehnert M, Sommer T & Jarosch E (2014) Der1 promotes movement of misfolded proteins through the endoplasmic reticulum membrane. *Nat Cell Biol* **16**: 77-86.

Meusser B, Hirsch C, Jarosch E & Sommer T (2005) ERAD: the long road to destruction. *Nat Cell Biol* **7**: 766-772.

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

Murphy MP & LeVine H, 3rd (2010) Alzheimer's disease and the amyloid-beta peptide. *J Alzheimers Dis* **19**: 311-323.

Muyldermans S, Baral TN, Retamozzo VC*, et al.* (2009) Camelid immunoglobulins and nanobody technology. *Vet Immunol Immunopathol* **128**: 178-183.

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M & Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**: 1344-1349.

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.

Nelson AL (2010) Antibody fragments: hope and hype. *MAbs* **2**: 77-83.

Nielsen J (2019) Yeast systems biology: model organism and cell factory. *Biotechnol J* **14**: e1800421.

Nielsen J & Jewett MC (2008) Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*. *FEMS Yeast Res* **8**: 122-131.

Nocon J, Steiger MG, Pfeffer M*, et al.* (2014) Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production. *Metab Eng* **24**: 129-138.

Nookaew I, Jewett MC, Meechai A, Thammarongtham C, Laoteng K, Cheevadhanarak S, Nielsen J & Bhumiratana S (2008) The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Syst Biol* **2**: 71.

O'Brien EJ, Lerman JA, Chang RL, Hyduke DR & Palsson B (2013) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol* **9**: 693.

Obst U, Lu TK & Sieber V (2017) A modular toolkit for generating *Pichia pastoris* secretion libraries. *ACS Synth Biol* **6**: 1016-1025.

Oftadeh O, Salvy P, Masid M, Curvat M, Miskovic L & Hatzimanikatis V (2021) A genome-scale metabolic model of *Saccharomyces cerevisiae* that integrates expression constraints and reaction thermodynamics. *Nat Commun* **12**: 4790.

Oka T & Jigami Y (2006) Reconstruction of de novo pathway for synthesis of UDP-glucuronic acid and UDP-xylose from intrinsic UDP-glucose in *Saccharomyces cerevisiae*. *Febs j* **273**: 2645-2657.

Olszewski MM, Williams C, Dong KC & Martin A (2019) The Cdc48 unfoldase prepares well-folded protein substrates for degradation by the 26S proteasome. *Commun Biol* **2**: 29.

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

Overton TW (2014) Recombinant protein production in bacterial hosts. *Drug Discov Today* **19**: 590-601.

Packer MS & Liu DR (2015) Methods for the directed evolution of proteins. *Nat Rev Genet* **16**: 379-394.

Parthasarathy R, Subramanian S, Boder ET & Discher DE (2006) Post-translational regulation of expression and conformation of an immunoglobulin domain in yeast surface display. *Biotechnol Bioeng* **93**: 159-168.

Partow S, Siewers V, Bjørn S, Nielsen J & Maury J (2010) Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. *Yeast* **27**: 955-964.

Passolunghi S, Riboldi L, Dato L, Porro D & Branduardi P (2010) Cloning of the *Zygosaccharomyces bailii* GAS1 homologue and effect of cell wall engineering on protein secretory phenotype. *Microb Cell Fact* **9**: 7.

Patil C & Walter P (2001) Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals. *Curr Opin Cell Biol* **13**: 349-355.

Payne T, Finnis C, Evans LR, Mead DJ, Avery SV, 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.

Petersen SD, Zhang J, Lee JS, Jakociunas T, Grav LM, Kildegaard HF, Keasling JD & Jensen MK (2018) Modular 5'-UTR hexamers for context-independent tuning of protein expression in eukaryotes. *Nucleic Acids Res* **46**: e127.

Peyroche A, Paris S & Jackson CL (1996) Nucleotide exchange on ARF mediated by yeast Gea1 protein. *Nature* **384**: 479-481.

Pirkalkhoran S, Grabowska WR, Kashkoli HH, Mirhassani R, Guiliano D, Dolphin C & Khalili H (2023) Bioengineering of antibody fragments: challenges and opportunities. *Bioengineering (Basel)* **10**(2): 122.

Plath K & Rapoport TA (2000) Spontaneous release of cytosolic proteins from posttranslational substrates before their transport into the endoplasmic reticulum. *J Cell Biol* **151**: 167-178.

Plath K, Wilkinson BM, Stirling CJ & Rapoport TA (2004) Interactions between Sec complex and prepro-alpha-factor during posttranslational protein transport into the endoplasmic reticulum. *Mol Biol Cell* **15**: 1-10.

Poon PP, Cassel D, Spang A, Rotman M, Pick E, Singer RA & Johnston GC (1999) Retrograde transport from the yeast Golgi is mediated by two ARF GAP proteins with overlapping function. *Embo j* **18**: 555-564.

Porro D, Gasser B, Fossati T, Maurer M, Branduardi P, Sauer M & Mattanovich D (2011) Production of recombinant proteins and metabolites in yeasts: when are these systems better than bacterial production systems? *Appl Microbiol Biotechnol* **89**: 939-948.

Prelich G (2012) Gene overexpression: uses, mechanisms, and interpretation. *Genetics* **190**: 841-854.

Puetz J & Wurm FM (2019) Recombinant proteins for industrial versus pharmaceutical purposes: a review of process and pricing. *Processes* **7**: 476.

Qi Q, Li F, Yu R, Engqvist MKM, Siewers V, Fuchs J & Nielsen J (2020) Different routes of protein folding contribute to improved protein production in *Saccharomyces cerevisiae*. *mBio* **11**(6): e02743-20.

Qin X, Qian J, Yao G, Zhuang Y, Zhang S & Chu J (2011) *GAP* promoter library for fine-tuning of gene expression in *Pichia pastoris*. *Appl Environ Microbiol* **77**: 3600-3608.

Rader RA (2008) Expression systems for process and product improvement. *BioProcess Int* **6**: 4-9.

Rajkumar AS & Maerkl SJ (2012) Rapid synthesis of defined eukaryotic promoter libraries. *ACS Synth Biol* **1**: 483-490.

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.

Randez-Gil F & Sanz P (1993) Expression of *Aspergillus oryzae* alpha-amylase gene in *Saccharomyces cerevisiae*. *FEMS Microbiol Lett* **112**: 119-123.

Rodríguez-Limas WA, Tannenbaum V & Tyo KE (2015) Blocking endocytotic mechanisms to improve heterologous protein titers in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* **112**: 376-385.

Romualdi C & Lanfranchi G (2009) Statistical tools for gene expression analysis and systems biology and related web resources. *Bioinformatics for Systems Biology,*p.^pp. 181-205. Springer.

Rosano GL & Ceccarelli EA (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol* **5**: 172.

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.

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.

Sánchez-Rivera FJ & Jacks T (2015) Applications of the CRISPR-Cas9 system in cancer biology. *Nat Rev Cancer* **15**: 387-395.

Sánchez BJ, Zhang C, Nilsson A, Lahtvee PJ, Kerkhoven EJ & Nielsen J (2017) Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol Syst Biol* **13**: 935.

Sangkaew A, Kojornna T, Tanahashi R, Takagi H & Yompakdee C (2022) A novel yeast-based screening system for potential compounds that can alleviate human α-synuclein toxicity. *J Appl Microbiol* **132**: 1409-1421.

Schuck S, Prinz WA, Thorn KS, Voss C & Walter P (2009) Membrane expansion alleviates endoplasmic reticulum stress independently of the unfolded protein response. *J Cell Biol* **187**: 525-536.

Selkoe DJ & Wolfe MS (2007) Presenilin: running with scissors in the membrane. *Cell* **131**: 215-221.

Sevier CS & Kaiser CA (2008) Ero1 and redox homeostasis in the endoplasmic reticulum. *Biochim Biophys Acta* **1783**: 549-556.

Sevier CS, Cuozzo JW, Vala A, Aslund F & Kaiser CA (2001) A flavoprotein oxidase defines a new endoplasmic reticulum pathway for biosynthetic disulphide bond formation. *Nat Cell Biol* **3**: 874-882.

Shariati SA, Dominguez A, Xie S, Wernig M, Qi LS & Skotheim JM (2019) Reversible disruption of specific transcription factor-DNA interactions using CRISPR/Cas9. *Mol Cell* **74**: 622-633.e624.

Sharon E, Kalma Y, Sharp A, Raveh-Sadka T, Levo M, Zeevi D, Keren L, Yakhini Z, Weinberger A & Segal E (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat Biotechnol* **30**: 521-530.

Shi X, Karkut T, Chamankhah M, Alting-Mees M, Hemmingsen SM & Hegedus D (2003) Optimal conditions for the expression of a single-chain antibody (scFv) gene in *Pichia pastoris*. *Protein Expr Purif* **28**: 321-330.

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

Si T, Chao R, Min Y, Wu Y, Ren W & Zhao H (2017) Automated multiplex genome-scale engineering in yeast. *Nat Commun* **8**: 15187.

Sicard D & Legras JL (2011) Bread, beer and wine: yeast domestication in the *Saccharomyces sensu stricto* complex. *C R Biol* **334**: 229-236.

Siewers V (2014) An overview on selection marker genes for transformation of *Saccharomyces cerevisiae*. *Methods Mol Biol* **1152**: 3-15.

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.

Sjostrom SL, Bai Y, Huang M, Liu Z, Nielsen J, Joensson HN & Andersson Svahn H (2014) High-throughput screening for industrial enzyme production hosts by droplet microfluidics. *Lab Chip* **14**: 806-813.

Soares MB, Bonaldo MF, Jelene P, Su L, Lawton L & Efstratiadis A (1994) Construction and characterization of a normalized cDNA library. *Proc Natl Acad Sci U S A* **91**: 9228-9232.

Solis-Escalante D, Kuijpers NG, Bongaerts N, Bolat I, Bosman L, Pronk JT, Daran JM & Daran-Lapujade P (2013) *amdSYM*, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS Yeast Res* **13**: 126-139.

Solow SP, Sengbusch J & Laird MW (2005) Heterologous protein production from the inducible *MET25* promoter in *Saccharomyces cerevisiae*. *Biotechnol Prog* **21**: 617-620.

Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D & Futcher B (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* **9**: 3273-3297.

Sreenivas S, Krishnaiah SM, Govindappa N, Basavaraju Y, Kanojia K, Mallikarjun N, Natarajan J, Chatterjee A & Sastry KN (2015) Enhancement in production of recombinant two-chain Insulin Glargine by over-expression of Kex2 protease in *Pichia pastoris*. *Appl Microbiol Biotechnol* **99**: 327-336.

Stephanopoulos G, Alper H & Moxley J (2004) Exploiting biological complexity for strain improvement through systems biology. *Nat Biotechnol* **22**: 1261-1267.

Stern DL (2014) Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test. *Trends Genet* **30**: 547-554.

Supek F, Madden DT, Hamamoto S, Orci L & Schekman R (2002) Sec16p potentiates the action of COPII proteins to bud transport vesicles. *J Cell Biol* **158**: 1029-1038.

Szul T & Sztul E (2011) COPII and COPI traffic at the ER-Golgi interface. *Physiology (Bethesda)* **26**: 348-364.

Tang H, Hou J, Shen Y, Xu L, Yang H, Fang X & Bao X (2013) High β-glucosidase secretion in *Saccharomyces cerevisiae* improves the efficiency of cellulase hydrolysis and ethanol production in simultaneous saccharification and fermentation. *J Microbiol Biotechnol* **23**: 1577-1585.

Tang H, Bao X, Shen Y, Song M, Wang S, Wang C & Hou J (2015) Engineering protein folding and translocation improves heterologous protein secretion in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* **112**: 1872-1882.

Tang H, Song M, He Y, Wang J, Wang S, Shen Y, Hou J & Bao X (2017) Engineering vesicle trafficking improves the extracellular activity and surface display efficiency of cellulases in *Saccharomyces cerevisiae*. *Biotechnol Biofuels* **10**: 53.

Tavassoly I, Goldfarb J & Iyengar R (2018) Systems biology primer: the basic methods and approaches. *Essays Biochem* **62**: 487-500.

Testa L, Van Gaal WJ, Bhindi R, Biondi-Zoccai GG, Abbate A, Agostoni P, Porto I, Andreotti F, Crea F & Banning AP (2008) Pexelizumab in ischemic heart disease: a systematic review and meta-analysis on 15,196 patients. *J Thorac Cardiovasc Surg* **136**: 884-893.

Thak EJ, Yoo SJ, Moon HY & Kang HA (2020) Yeast synthetic biology for designed cell factories producing secretory recombinant proteins. *FEMS Yeast Res* **20**(2): foaa009.

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.

Tong AH & Boone C (2006) Synthetic genetic array analysis in *Saccharomyces cerevisiae*. *Methods Mol Biol* **313**: 171-192.

Tran Nguyen Hoang P, Ko JK, Gong G, Um Y & Lee SM (2018) Genomic and phenotypic characterization of a refactored xylose-utilizing *Saccharomyces cerevisiae* strain for lignocellulosic biofuel production. *Biotechnol Biofuels* **11**: 268.

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.

Tu BP & Weissman JS (2004) Oxidative protein folding in eukaryotes: mechanisms and consequences. *J Cell Biol* **164**: 341-346.

Tyo KE, Ajikumar PK & Stephanopoulos G (2009) Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nat Biotechnol* **27**: 760-765.

Tyo KE, Liu Z, Magnusson Y, Petranovic D & Nielsen J (2014) Impact of protein uptake and degradation on recombinant protein secretion in yeast. *Appl Microbiol Biotechnol* **98**: 7149-7159.

Uchiyama K, Morimoto M, Yokoyama Y & Shioya S (1997) Cell cycle dependency of rice alpha-amylase production in a recombinant yeast. *Biotechnol Bioeng* **54**: 262-271.

Vai M, Brambilla L, Orlandi I, Rota N, Ranzi BM, Alberghina L & Porro D (2000) Improved secretion of native human insulin-like growth factor 1 from *gas1* mutant *Saccharomyces cerevisiae* cells. *Appl Environ Microbiol* **66**: 5477-5479.

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 Anken E, Romijn EP, Maggioni C, Mezghrani A, Sitia R, Braakman I & Heck AJ (2003) Sequential waves of functionally related proteins are expressed when B cells prepare for antibody secretion. *Immunity* **18**: 243-253.

van der Heide M, Hollenberg CP, van der Klei IJ & 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 Der Kelen K, Beyaert R, Inzé D & De Veylder L (2009) Translational control of eukaryotic gene expression. *Crit Rev Biochem Mol Biol* **44**: 143-168.

Van der Weken H, Cox E & Devriendt B (2019) Rapid production of a chimeric antibody-antigen fusion protein based on 2A-peptide cleavage and green fluorescent protein expression in CHO cells. *MAbs* **11**: 559-568.

Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE, Jr., Hieter P, Vogelstein B & Kinzler KW (1997) Characterization of the yeast transcriptome. *Cell* **88**: 243-251.

Vieira Gomes AM, Souza Carmo T, Silva Carvalho L, Mendonça Bahia F & Parachin NS (2018) Comparison of yeasts as hosts for recombinant protein production. *Microorganisms* **6**(2): 38.

Villalobos A, Ness JE, Gustafsson C, Minshull J & Govindarajan S (2006) Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics* **7**: 285.

Walsh G (2018) Biopharmaceutical benchmarks 2018. *Nat Biotechnol* **36**: 1136-1145.

Wang G, Huang M & Nielsen J (2017) Exploring the potential of *Saccharomyces cerevisiae* for biopharmaceutical protein production. *Curr Opin Biotechnol* **48**: 77-84.

Wang G, Björk SM, Huang M, Liu Q, Campbell K, Nielsen J, Joensson HN & Petranovic D (2019) RNAi expression tuning, microfluidic screening, and genome recombineering for improved protein production in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **116**: 9324-9332.

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR & Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**: 894-898.

Wang M, Herrmann CJ, Simonovic M, Szklarczyk D & von Mering C (2015) Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* **15**: 3163-3168.

Wang X, Rong L, Wang M, Pan Y, Zhao Y & Tao F (2017) Improving the activity of endoglucanase I (EGI) from *Saccharomyces cerevisiae* by DNA shuffling. *RSC advances* **7**: 46246-46256.

Wang Y, Li X, Chen X, Nielsen J, Petranovic D & Siewers V (2021) Expression of antibody fragments in *Saccharomyces cerevisiae* strains evolved for enhanced protein secretion. *Microb Cell Fact* **20**: 134.

Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LB & Gill RT (2010) Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat Biotechnol* **28**: 856-862.

Wells JA, Vasser M & Powers DB (1985) Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites. *Gene* **34**: 315-323.

Wendeler MW, Paccaud JP & Hauri HP (2007) Role of Sec24 isoforms in selective export of membrane proteins from the endoplasmic reticulum. *EMBO Rep* **8**: 258-264.

Westfall PJ, Ballon DR & Thorner J (2004) When the stress of your environment makes you go HOG wild. *Science* **306**: 1511-1512.

Wilson DS & Keefe AD (2001) Random mutagenesis by PCR. *Curr Protoc Mol Biol* **Chapter 8**: Unit8.3.

Wingler LM & Cornish VW (2011) Reiterative Recombination for the in vivo assembly of libraries of multigene pathways. *Proc Natl Acad Sci U S A* **108**: 15135-15140.

Wittrup K & Benig V (1994) Optimization of amino acid supplements for heterologous protein secretion in *Saccharomyces cerevisiae*. *Biotechnology techniques* **8**: 161-166.

Wu X & Rapoport TA (2018) Mechanistic insights into ER-associated protein degradation. *Curr Opin Cell Biol* **53**: 22-28.

Yamanishi M, Katahira S & Matsuyama T (2011) *TPS1* terminator increases mRNA and protein yield in a *Saccharomyces cerevisiae* expression system. *Biosci Biotechnol Biochem* **75**: 2234-2236.

Yamanishi M, Ito Y, Kintaka R, Imamura C, Katahira S, Ikeuchi A, Moriya H & Matsuyama T (2013) A genome-wide activity assessment of terminator regions in *Saccharomyces cerevisiae* provides a ”terminatome” toolbox. *ACS Synth Biol* **2**: 337-347.

Yang L, Yurkovich JT, King ZA & Palsson BO (2018) Modeling the multi-scale mechanisms of macromolecular resource allocation. *Curr Opin Microbiol* **45**: 8-15.

Yoon J, Kikuma T, Maruyama J & Kitamoto K (2013) Enhanced production of bovine chymosin by autophagy deficiency in the filamentous fungus *Aspergillus oryzae*. *PLoS One* **8**: e62512.

Yoshida H (2007) ER stress and diseases. *Febs j* **274**: 630-658.

Yu R, Campbell K, Pereira R, Björkeroth J, Qi Q, Vorontsov E, Sihlbom C & Nielsen J (2020) Nitrogen limitation reveals large reserves in metabolic and translational capacities of yeast. *Nat Commun* **11**: 1881.

Zahrl RJ, Prielhofer R, Ata Ö, Baumann K, Mattanovich D & Gasser B (2022) Pushing and pulling proteins into the yeast secretory pathway enhances recombinant protein secretion. *Metab Eng* **74**: 36-48.

Zhang W, Zhao HL, Xue C, Xiong XH, Yao XQ, Li XY, Chen HP & Liu ZM (2006) Enhanced secretion of heterologous proteins in *Pichia pastoris* following overexpression of *Saccharomyces cerevisiae* chaperone proteins. *Biotechnol Prog* **22**: 1090-1095.

Zheng N & Gierasch LM (1996) Signal sequences: the same yet different. *Cell* **86**: 849-852.

Zimmermann R, Eyrisch S, Ahmad M & Helms V (2011) Protein translocation across the ER membrane. *Biochim Biophys Acta* **1808**: 912-924.