Following the blueprint of plasma cells to design a yeast IgG factory

Essi V. Koskela
Following the blueprint of plasma cells to design a yeast IgG factory

Essi V. Koskela

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Chemical Engineering, at a public examination held at the lecture hall KE2 of the school on 10 November 2017 at 12.

Aalto University
School of Chemical Engineering
Department of Bioproducts and Biosystems
Molecular Biotechnology
**Abstract**

IgG antibodies are powerful biotherapeutics that are used in the treatment of several severe diseases, for example cancer and autoimmune diseases. Specialized cells in the human immune system, plasma cells, naturally produce antibodies with high efficiency. However, biotechnological production methods based on mammalian cell cultures remain inadequate and expensive. The expanding market of antibody biotherapeutics has spurred the aspiration to develop alternative production methods. One potential platform for antibody production is the yeast *Saccharomyces cerevisiae*, which has been successfully engineered to produce a range of products by utilizing the versatile genetic toolkit available for modifying this organism. Plasma cell differentiation depicts a comprehensive molecular model of cellular transformation into an efficient antibody factory, and this model provides a blueprint for genetic engineering. In this thesis, we studied whether the key elements from plasma cells would improve IgG secretion in yeast.

First, we modified the yeast ER to mimic the plasma cell ER morphology. Both an increase in ER size and an altered shape, achieved by deletion of *OP1* and shape determinant genes, respectively, increased IgG secretion at least 2.4-fold. In addition, these mutants displayed a reduced stress response related to antibody production. We selected the strain with the *OP1* gene deletion for engineering of protein folding, along with the wild-type production strain. Relying on principles of synthetic biology, we created a modular plasmid library of mammalian folding factors shown to interact with IgG. To aid plasmid library creation, we established a new high-throughput cloning method to complement the available synthetic biology tools. Screening of the plasmid library led us to identify GRP170, BiP and FKB2 as the most potent enhancers of IgG folding and secretion in yeast. We concluded that upregulation of ER-localized PPIase activity is critical for improving IgG titers in yeast.

Additionally, we explored transcriptomics data from plasma cell differentiation to find gene targets which would otherwise be overlooked in engineering approaches. Through this datadriven approach, we selected seven novel genetic modifications to analyze in yeast. Two of these seven modifications, the overexpression of the genes *GOT1* and *IRE1* led to significant improvements in IgG secretion, resulting in a 1.6- and a 3.5-fold increase in specific product yields, respectively. However, in the future the emphasis should be in improving the quality of the secreted antibody.

This thesis demonstrates that plasma cells are useful cellular models for antibody secretion, also when applied to an evolutionary distant species, such as the yeast *S. cerevisiae*. The IgG titers were increased from 40 ng/ml to up to 160 ng/ml, confirming that this yeast is a promising platform for future applications of the biotherapeutics industry.

**Keywords** antibody, plasma cell, yeast, endoplasmic reticulum, protein folding, synthetic biology

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ISSN-L</td>
<td>1799-4934</td>
<td>ISSN (printed)</td>
<td>1799-4934</td>
</tr>
<tr>
<td>Location of publisher</td>
<td>Helsinki</td>
<td>Location of printing</td>
<td>Helsinki</td>
</tr>
</tbody>
</table>
Tiivistelmä


Preface

The work for this thesis was carried out in the research group of Molecular Biotechnology, in the Department of Bioproducts and Biosystems, at the School of Chemical Engineering, Aalto University during the years 2013-2017. I am grateful to the School of Chemical Engineering, Aalto University, for the funded doctoral position I was granted for the years 2014-2017.

Above everyone else, my thanks for enabling this thesis work goes to my supervising professor Dr. Alexander Frey. The interesting topic of my thesis challenged me diversely and after four years of research under your supervision, I feel like I am ready to take on anything. Alex, you gave me unique opportunities, responsibility, and freedom for which I am truly grateful. Regarding the thesis itself, I would like to thank my preliminary examiners Dr. Davis Ng and Dr. Anne Skaja Robinson, who took the time to go through my work and write me such kind feedback.

A significant part of this thesis work I shared with my fellow doctoral student at the time, Jorg de Ruijter. I cherish our fruitful teamwork and I am grateful for your company and support during the long days in the lab. I would also like to thank my other co-authors, Alina Mehl, Mari Piirainen, and Heidi Iivonen. Working with all of you was a great pleasure. Mari, we shared our time in Aalto so far, and I wish you the best of luck in being the next senior doctoral student in the research group. I appreciate also the input of my other collaborators, Dr. András Pásztor and Laura Niemelä, who were centrally involved in determining the direction of my research.

During this thesis work, I had the opportunity to do a research visit in Universität für Bodenkultur, Vienna, Austria. I want to thank Dr. Diethard Mattanovich and Dr. Brigitte Gasser for welcoming me into their group and providing me with an interesting project related to genetic engineering of *Pichia pastoris*. I express my gratitude also to Richard Zahrl, who generously used his time to help and guide me in the lab during my months in
Vienna. I had a great time and I learned a lot from the insightful discussions about science and life. I owe thanks to the Foundation for Aalto University Science and Technology for awarding me a travel grant, which enabled this rewarding research exchange.

There are also other members in the scientific community who aided me in this study. I would like to acknowledge Dr. David Pincus, Dr. Mari Valkonen from the Technical Research Center of Finland, Dr. Linda M. Hendershot, and Dr. Kari Kivirikko for kindly providing materials upon our request. On the same grounds, Dr. Lloyd Ruddock deserves special thanks and additionally, I would like to thank him for the insightful discussions. In addition, I appreciate the time and effort that Dr. Elena Czeizler used to consult me on bioinformatics.

All the people in Aalto and the former and current members of the group of Molecular Biotechnology have made my doctoral studies so delightful. In addition to the above mentioned, German Jürgens, Filip Mollerup, Maciek Spus, Heidi Salminen, Heli Viskari, Markus Räsänen, Janne Wallenius, Edita Jurak, Bart Rooijakkers, Christopher Jonkergouw, Anne Usvalampi, Marcela Ruvalcaba, Riikka Särkelä, etc. made my time in Aalto an eventful journey. I enjoyed the time we had during and outside working hours!

Special thanks goes to my family for their support, especially to my brothers Olli, Vesa, and Pasi. Olli deserves a special acknowledgement for his advices on statistics and for the fact that he took the effort in trying to understand and comment my thesis during the writing process. It has been nice to have a person in the family with whom to share the joys of doctoral studies, as it has been equally enjoyable not to share it with Vesa and Pasi. You were there to make sure I concentrated on the other awesome things in life, like watching our niece, Elsa, grow.

Dear friends, you are the family I got to choose. Although I cannot take all the credit for such excellent choices, you are a great bunch! My lovely ex-wife Inka shared my home for two and a half years during my doctoral work and she was there to cheer me up when work did not. Maria, Emmi, Tiia, Vuokko, Mervi, Vivi, among others, you have accompanied me through these eventful four years, and let there be many more to come. I would also like to thank Helsinki Symphonic Winds, AP, for keeping the music alive during my doctoral studies.
Finally, I saved the best for last. Tim, thank you for being there for me effortlessly despite the fact that most of the time there was 1,500 km between us. Our love hardly fits into words so I refrain from even trying. You know what I mean, and that matters the most to me.

Espoo, August 2017
Essi Koskela

“No hesitation in declaring quite bluntly that the acceptance of a really existing material world as the explanation of the fact, that we all find in the end that we are empirically in the same environment, is mystical and metaphysical.”

-Erwin Schrödinger, 1925
Contents

Preface ..............................................................................................................................1
Contents ............................................................................................................................5
List of publications ..........................................................................................................7
Author’s contributions ...................................................................................................9
List of abbreviations ......................................................................................................11

1. Introduction .............................................................................................................13
   1.1 IgG and therapeutic antibodies ...........................................................................14
       1.1.1 Structure of IgG ........................................................................................................14
       1.1.2 Folding and assembly of the IgG tetramer .................................................................16
       1.1.3 Antibodies as recombinant biopharmaceuticals .........................................................18
       1.1.4 Production systems for antibodies ............................................................................19
   1.2 Nature’s system as the starting point – plasma cells and antibody secretion ..........22
       1.2.1 From B cells to plasma cells ......................................................................................22
       1.2.2 Cellular changes during plasma cell development ......................................................25
       1.2.3 Endoplasmic reticulum folding machinery ..................................................................27
       1.2.4 IgG production characteristics of plasma cells ..........................................................29
   1.3 Saccharomyces cerevisiae as an IgG production platform ........................................31
       1.3.1 S. cerevisiae as a production platform .......................................................................31
       1.3.2 Glycoengineering yeasts ............................................................................................34
       1.3.3 Characteristics of the S. cerevisiae secretory pathway ...............................................37
       1.3.4 Other engineering approaches to yeast antibody production ....................................48
   1.4 Synthetic biology tools for optimizing yeast cell factories ....................................49
       1.4.1 Genetic engineering methods in yeast .........................................................................50
       1.4.2 High-throughput cloning ............................................................................................52
       1.4.3 Integrating systems-level information to synthetic biology ......................................54

2. Aims of the study ......................................................................................................56

3. Materials and methods ............................................................................................57
   3.1 Yeast strains ............................................................................................................57
   3.2 Materials for testing the cloning strategy ...............................................................58
   3.3 Plasmids ..................................................................................................................58
       3.3.1 Plasmid library of mammalian folding factors ..............................................................58
       3.3.2 Other plasmids ............................................................................................................60
   3.4 Measuring specific growth rates .............................................................................61
3.5 Screening procedures, yeast cultures and automated ELISA ......................... 61
3.6 Flow cytometry ............................................................................................... 62
3.7 mRNA extraction and real-time PCR .............................................................. 63
3.8 Western blotting ............................................................................................. 63
3.9 Co-immunoprecipitation .................................................................................. 64
3.10 Data analysis and statistics ............................................................................ 65
3.11 Bioinformatic analysis of plasma cell differentiation ....................................... 65
3.12 IgG production in *Pichia pastoris* ................................................................. 65

4. Results and discussion ....................................................................................... 67
4.1 Reshaping the yeast endoplasmic reticulum ................................................. 67
4.1.1 Modifications to the ER morphology increased specific product yields ........ 67
4.1.2 Expansion of ER sheets diminished the induction of UPR ......................... 70
4.2 Establishing a high-throughput cloning method for plasmid library creation .... 76
4.3 Screening plasmid library of ten mammalian folding factors ......................... 78
4.3.1 Selected folding factors included all functional classes and were expressed at different levels ................................................................. 78
4.3.2 Temperature and induction strength affected the strains similarly ................ 81
4.3.3 FKBP2, P5, mammalian BiP, and its co-chaperone GRP170 improved IgG secretion .... 84
4.3.4 Combinations of several modifications implied that engineering folding alone has limited improvements in secretion ....................................................... 87
4.4 Human BiP interacts with the antibody in yeast cells ................................... 88
4.5 PPIases were the most effective in improving IgG titers ................................. 90
4.6 New genetic targets from exploration of plasma cell differentiation transcriptomics data ........................................................................................................ 94
4.6.1 The data analysis workflow identified several recognized and novel engineering targets .... 94
4.6.2 Two of the identified modifications, overexpression of Got1p and Ire1p, improved IgG secretion significantly ............................................................. 97
4.7 Quality considerations: not all the antibody is in the full-length form ............ 99

5. Conclusions and future prospects ................................................................. 102
References .............................................................................................................. 105
List of publications

This thesis is based on the following four publications, which are referred to by their Roman numerals in the text:


III  Koskela, E.V., de Ruijter, J.C., and Frey, A.D.: Following nature’s roadmap: folding factors from plasma cells led to improvements in antibody secretion in *S. cerevisiae*, *Biotechnology journal* 2017, 12:8


*equal contributions
Author’s contributions

Publication I: Homologous recombinatorial cloning without the creation of single-stranded ends: exonuclease and ligation-independent cloning (ELIC)

The author conceived the study, carried out the initial testing, performed all the laboratory experiments and drafted the manuscript. Alexander Frey participated in the design of experiments and helped drafting the manuscript.

Publication II: Enhancing antibody folding and secretion by tailoring the *Saccharomyces cerevisiae* endoplasmic reticulum

The author designed the screening procedure, created and analyzed the background yeast strains, was responsible of data handling and interpretation, including statistical analyses. The author also contributed to molecular cloning work and screening experiments, which were otherwise mainly performed by Jorg de Ruijter. mRNA measurements and cellular clearance experiments were divided to the author and Jorg de Ruijter, respectively. Jorg de Ruijter and the author drafted the manuscript together, while Alexander Frey helped in the process. Alexander Frey conceived the study and participated in the experimental design.

Publication III: Following nature’s roadmap: folding factors from plasma cells led to improvements in antibody secretion in *S. cerevisiae*

The author designed and conducted the screening, carried out the molecular cloning work and the laboratory experiments. Jorg de Ruijter assisted in screening efforts and in a few laboratory analyses. Data handling, interpretation, and statistical analyses were the author’s responsibility. Jorg de Ruijter and Alexander Frey contributed to molecular cloning work and helped drafting the manuscript. Alexander Frey conceived the study.

Publication IV: Mining data from plasma cell differentiation to find novel genes for engineering of a yeast antibody factory
The author conceived the study, carried out the computational part and performed the strain characterizations. Alina Mehl created the strains and performed the molecular cloning work. Mari Piirainen contributed to experimental design and assisted in the laboratory experiments. Heidi Iivonen contributed to the computational work. Alexander Frey participated in the design of experiments and helped drafting the manuscript.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC</td>
<td>Antibody-drug conjugate</td>
</tr>
<tr>
<td>ADP</td>
<td>Adenosine diphosphate</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>BiP</td>
<td>Immunoglobulin heavy chain binding protein</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>Cas9</td>
<td>CRISPR-associated protein 9</td>
</tr>
<tr>
<td>CHO</td>
<td>Chinese hamster ovary</td>
</tr>
<tr>
<td>COP</td>
<td>Coat protein complex</td>
</tr>
<tr>
<td>CPY</td>
<td>Carboxypeptidase Y</td>
</tr>
<tr>
<td>CRISPR</td>
<td>Clustered regularly interspersed short palindromic repeats</td>
</tr>
<tr>
<td>CypB</td>
<td>Cyclophilin B</td>
</tr>
<tr>
<td>ddH2O</td>
<td>Double-distilled water</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>ELIC</td>
<td>Exonuclease and ligation-independent cloning</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
</tr>
<tr>
<td>ERAD</td>
<td>Endoplasmic reticulum associated degradation</td>
</tr>
<tr>
<td>ERES</td>
<td>Endoplasmic reticulum exit sites</td>
</tr>
<tr>
<td>Fab</td>
<td>Fragment antigen binding domain of the antibody</td>
</tr>
<tr>
<td>Fc</td>
<td>Fragment crystallizable domain of the antibody</td>
</tr>
<tr>
<td>FKBP</td>
<td>FK506-binding proteins</td>
</tr>
<tr>
<td>FSC</td>
<td>Forward scatter</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>GO</td>
<td>Gene ontology</td>
</tr>
<tr>
<td>Hsp</td>
<td>Heat shock protein</td>
</tr>
<tr>
<td>Ig</td>
<td>Immunoglobulin</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin G</td>
</tr>
</tbody>
</table>
kDa  Kilodalton
LB    Luria-Bertani broth
LIC   Ligation-independent cloning
mAb   Monoclonal antibody
OD    Optical density
OST   Oligosaccharyltransferase
PB    Plasmablast
PC    Plasma cell
PCR   Polymerase chain reaction
PDI   Protein disulfide isomerase
PPIase Peptidyl-prolyl cis-trans isomerase
RNA   Ribonucleic acid
scFv  Single-chain fragment variable
SD    Synthetic dropout
SM    Sec1/Munc18-like
SNARE N-ethylmaleimide sensitive factor attachment protein receptor
SRP   Signal recognition particle
SSC   Side scatter
TALEN Transcription activator-like effector nuclease
UDP   Uridine diphosphate
UGGT  UDP-glucose:glycoprotein glucosyltransferase
UPR   Unfolded protein response
Vps   Vacuolar protein sorting
wt    Wild-type production strain
YAC   Yeast artificial chromosome
ZFN   Zinc-finger nuclease
1. Introduction

Antibodies, especially of class IgG, are valuable proteins because of their high specificity and their recognition capability of any desired molecular target. Antibodies raised against pathogens, biomarkers or human hormones are used in the treatment of various diseases and in diagnostics, such as in the common pregnancy test. Because of their high value and multitude of applications, suitable platforms to produce antibody-based products are constantly being researched. Commercial biotherapeutic proteins are mainly produced in living cells, where the target DNA for the desired protein is inserted into the host by genetic manipulation. Using the DNA sequence as a template, the cells produce mRNA and ribosomes translate it to the protein sequence. In eukaryotes, the polypeptides enter the secretory machinery, first from ribosomes into the endoplasmic reticulum (ER), travelling through the cellular postal office, the Golgi apparatus, and finally reaching the plasma membrane and extracellular space. The efficiency with which eukaryotes produce, assemble and secrete antibodies varies greatly between different organisms. This study compares the natural antibody-producer cells, plasma cells, with the microbial eukaryote Saccharomyces cerevisiae in order to identify key modifications that would make the baker’s yeast a competitive platform to produce antibodies for therapeutic applications.

Antibodies are complex molecules that are not easily obtained in sufficient quality with biotechnological means, as described in section 1.1. However, nature has optimized a cellular system to efficiently produce these immune system agents by evolution. Development of plasma cells entails the blueprint for cellular transformation into an antibody factory, which is reviewed in section 1.2. Although baker’s yeast has been successfully utilized for making food and beverages, bioethanol, and with the advent of genetic engineering a whole range of products, it might not intrinsically be suitable for the high-yield production of therapeutic antibodies. Section 1.3 explains these aspects, along with the key engineering targets to increase antibody quality and quantity obtained from S. cerevisiae. In addition, how to realize rational genetic engineering of this yeast with the tools and means of synthetic biology is briefly described in section 1.4. By increasing the amounts
of secreted antibodies obtained from *S. cerevisiae*, this microbe could eventually develop into a commercially viable platform to produce advanced medicines for a wider distribution of patients more cheaply than before.

### 1.1 IgG and therapeutic antibodies

The human adaptive immune system expresses five classes of antibodies named IgA, IgD, IgE, IgG, and IgM, each of which consist of a different assembly of constant regions. From these antibody classes, IgG is the most abundant in human serum and the most common isotype among therapeutic antibodies (Vidarsson et al., 2014; Buss et al., 2012). Four different subclasses have been identified for IgG, and they are referred to as IgG1-4. IgG1 occurs at highest levels of these four, though all subclasses have their designated specialization in the complex network of the immune system (Vidarsson et al., 2014). As antibodies have remarkable potential in tackling diseases by targeting, facilitating, and initiating immune responses, all subclasses of IgG and several engineered variants of IgG are central molecules in biotherapeutics development and research. This section will describe the structure and folding of IgG molecules, review recombinant antibodies and their production and discuss their status as biotechnologically produced pharmaceuticals.

#### 1.1.1 Structure of IgG

Each member of the antibody polypeptide family is constructed from a series of a common structural unit, a domain of around 100 amino acids that displays the immunoglobulin (Ig) fold. The Ig fold has a β-barrel topology composed of two β-sheets, which are formed by several antiparallel β-strands. Each Ig fold has between seven and nine β-strands (Figure 1A). In the hydrophobic core of each domain, a single internal disulfide bond connects the two β-sheets and stabilizes the structure (Huber et al., 1976; Goto and Hamaguchi, 1979). This intradomain disulfide bridge is the yellow bond perpendicular to the β-strands in Figure 1A. Ig domains usually contain several proline residues in the *cis* conformation. Certain proline residues play an important role in the Ig fold structure, because these *cis*-prolines are responsible for the formation of abrupt loops connecting the antiparallel strands (Figure 1A). IgG heavy chain and light chain contain four and two Ig domains, respectively, resulting in 12 Ig domains in total per tetramer (Figure 1B). These domains are named *V*<sub>L</sub> and *C*<sub>L</sub> in the light chain to indicate variable and constant regions, respectively, and similarly *V*<sub>H</sub> and *C*<sub>H1-3</sub> for the variable and constant regions in the heavy chain (Figure 1C).
The IgG tetramer consists of four polypeptides, of which the two light and two heavy chains are identical. The domains in the tetramer associate through lateral interfaces to form the overall Y-shaped structure. The antigen binding sites are located in the two identical Fab regions, in the opposite end from the effector sites in the Fc stem (Figure 1C). The antigen binding sites are formed by hypervariable regions that are located in the interacting loops of the heavy and light chain domains; these areas are also called complementary-determining regions. The hinge region forms a flexible linker between the functional units and allows each Fab fragment to bind an antigen while still enabling the signaling to the Fc. The highly conserved constant region Fc of IgG molecules contains the interaction sites for the downstream effectors of the immunological response. (Voet and Voet, 2011)

Post-translational modifications are essential for proper IgG activity. One intermolecular disulfide bond links each heavy and light chain covalently and in addition, the heavy chains are connected through disulfide bonds in the hinge region (Figure 1C). Disulfide bond structures in IgG are not universally identical and the number of the hinge region -located disulfide bridges varies in the different IgG subclasses, from two in IgG1 to up to 11 in IgG3 (Liu and May, 2012). In addition to disulfide bridges, a central post-translational modification of the IgG molecule is glycosylation. Each heavy chain has one N-linked complex-type biantennary oligosaccharide attached to an asparagine residue in the C1 domain, and these carbohydrates are located in the intermolecular interface (Figure 1B). Approximately 20 % of IgG molecules contain an additional N-glycan within the variable domains, but the significance of the Fab glycans is not fully understood (van de Bovenkamp...
et al., 2016). As the oligosaccharide sequence is modified enzymatically, glycosylation introduces microheterogeneity to antibody molecules which can result in changes to structure and activity (Yamaguchi et al., 2006; Krapp et al., 2003; Jefferis, 2009; van de Bovenkamp et al., 2016). In biotechnologically produced antibodies, several other rare post-translational modifications have been detected, such as trisulfide bonds and terminal amino acid derivatization (Liu and May, 2012; Brorson and Jia, 2014). These unintended modifications result from the various bioprocessing operations and impose a significant immunological risk in the utilization of the therapeutic product (Brorson and Jia, 2014).

1.1.2 Folding and assembly of the IgG tetramer

There is a long way from four linear polypeptides to the quaternary structure of the final IgG heterotetramer. Antibodies and antibody fragments have been popular models to study protein folding, thus the key steps in Ig fold formation and IgG assembly have been elucidated (Feige et al., 2010). Especially, the Ig fold has been regarded as a model to study the general folding pathway of all-β-proteins, although the measured folding rates of Ig domains seem to be slower than reported for other all-β-proteins (Feige et al., 2004). The majority of the secondary and tertiary structures of the domains result from the burial of hydrophobic residues to the inner core of the fold. However, this initial hydrophobic collapse is not sufficient to achieve the final form of the domains. Domain folding and the assembly of IgG is a highly coordinated event, where the folding is assisted by chain-chain associations and folding factors, which catalyze rate-limiting steps (Feige et al., 2010). Individually folded and folding domains interact with each other and with several other molecules in vivo obtaining certain post-translational modifications along the way. The key steps in the folding of antibodies, including the effects of modifications and interactions in the process will be described next.

Although most of the Ig domains can reach their final tertiary structure independently, the tertiary and quaternary structures form more or less simultaneously. The heavy chain associates with the light chain in a partially unfolded state, where CH1 achieves its final form only after interacting with the CL domain of the light chain (Feige et al., 2009). Accordingly, the prolyl-isomerization and disulfide bridge oxidation reactions in CH1 domains occur after the light and heavy chain assembly (Lilie et al., 1995; Feige et al., 2009). The folding of the Fe domains CH2 and CH3 is mainly limited by prolyl-isomerization reactions (Thies et al., 1999; Feige et al., 2004), after which CH3 domains of the heavy chains form a dimer through a protein-protein interface (Huber et al., 1976; Thies et al., 1999). The tendency of CH3
domains to associate with one another guides the heavy chains, or generally the IgG halves, to dimerize (Thies et al., 1999; Feige et al., 2010). Ig domains have a characteristic intrachain disulfide bridge buried in their hydrophobic core (Figure 1A), but it seems that this disulfide bond is not a prerequisite for correct folding, but rather stabilizes the final structure (Thies et al., 2002; McAuley et al., 2008). However, the oxidation reaction of the cysteine residues is considerably slower if it occurs after domain folding (Thies et al., 2002) and thus can slow down the overall folding procedure. Several interchain disulfide bonds form between the chains during folding. Studies utilizing partial reduction suggested that these bonds have only minor effects on antibody conformation, although obtaining correct disulfide patterns might be crucial for antibody activity (Liu and May, 2012).

As several proline residues along the domains of the antibody polypeptides need to rearrange from $trans$ into $cis$ configuration, peptidyl-prolyl isomerization is one of the key features to determine the rate of antibody folding (Feige et al., 2004, 2010). Several Ig domains display faster folding $in$ vitro in the presence of peptidyl-prolyl cis-trans isomerases (PPIases) (Thies et al., 1999; Lilie et al., 1995; Feige et al., 2004, 2009), demonstrating the formation of $cis$-prolines to be a major rate-limiting step in the overall tetramer assembly. In addition to PPIases, several folding factors associate with antibody polypeptides in the ER and $in$ vivo these factors influence the folding rate and patterns substantially (Meunier et al., 2002; Feige et al., 2010). Folding intermediates of antibody domains are commonly found to have partially shielded hydrophobic surfaces (Feige et al., 2004), which explains why interactions of the folding antibody with the folding factor class of holdases are prevalent. Holdases shield the hydrophobic patches and prevent aggregation of folding intermediates. The most notable of the holdases in the ER is the immunoglobulin heavy chain binding protein, BiP, which may directly compete or cooperate with the completely folded light chain domain $\text{CL}$ in associating to the partially folded $\text{CH1}$ heavy chain domain (Lee et al., 1999; Feige et al., 2009). The folding of the $\text{CH1}$ occurs only after quaternary structure assembly, so this domain is a major control point in antibody quality and secretion, while the association with BiP contributes to the control mechanism (Lee et al., 1999; Feige et al., 2009). Overall, the separate steps in IgG folding form a defined sequence with several checkpoints to ensure possibilities for quality control before secretion (Feige et al., 2010).

Interestingly, $\text{CH2}$ domains seem to build an intermolecular interaction only through their attached sugar moieties, and unglycosylated $\text{CH2}$ domains appear as monomers (Feige et al., 2004; Huber et al., 1976). The structure of the N-linked glycan has a direct effect on IgG structure, as the carbohydrate moiety determines the distance between the two $\text{CH2}$ domains.
Although the oligosaccharides are more flexible than the surrounding protein, structures deviating from the common monosaccharide sequence might disrupt stabilizing glycan-protein interactions and lead to small structural differences (Krapp et al., 2003; Yamaguchi et al., 2006). Through these structural effects the changes in glycan sequence induce significant differences in activity between various antibody glycoforms, but the effect of the sugar moiety on folding pathways is so far unknown (Jefferis, 2009; Feige et al., 2010).

1.1.3 Antibodies as recombinant biopharmaceuticals

The biopharmaceutical industry has bloomed since its emergence in the 1980’s when the first protein products of modern molecular biotechnology appeared on the market. The current biotherapeutics market in the US and EU involves over two hundred approved biopharmaceutical products, of which over one fifth are monoclonal antibody (mAb) products (Walsh, 2014). Antibody biopharmaceuticals are used as treatment for a wide range of conditions, including inflammatory and autoimmune diseases, cancer, and other difficultly cured conditions (Walsh, 2014; Buss et al., 2012; Saeed et al., 2017). Based on the increasing share of antibodies in approved biopharmaceuticals, antibodies and antibody-derived products are a prevalent focus of the pharmaceutical industry. In 2013, antibody products alone generated $75.7 billion of sales, constituting the most profitable product class within biopharmaceuticals (Walsh, 2014). From early murine and chimeric mAbs, humanized and fully human mAbs have started to dominate the market, complemented by emerging innovations of molecular engineering, which greatly expand the market potential of antibody products (Walsh, 2014; Buss et al., 2012). A prominent example of engineered mAbs features antibody-drug conjugates (ADCs), in which the antibody is essentially utilized as a targeting vehicle for a cytotoxic drug (Walsh, 2014; Perez et al., 2014). A few of the approved mAb products are actually ADCs, for example adotrastuzumab emtansine against breast cancer and brentuximab vedotin for the treatment of certain lymphomas (Perez et al., 2014).

In contrast to ADCs extending the antibody molecule and function, several engineering efforts focus on the substructures and functional parts of the antibody molecule. Certain mutations in the Fc sequence or modulation of the N-glycan structure affect the stability and immunological effect of the recombinant antibody, so Fc region engineering is a common approach to optimize mAb performance (Presta, 2008; Buss et al., 2012; Jefferis, 2009). For example, Ferrara et al. were able to increase the response of antibody-dependent cellular cytotoxicity by engineering a non-fucosylated glycovariant of an anti-CD20 antibody (Ferrara
et al., 2006a). Modifications of the variable regions include increasing target affinity or the introduction of bispecificity, which allows the antibody molecule to recognize two target molecules (Presta, 2008; Buss et al., 2012). For several potential applications, there is no requirement to use full-length IgG, but harnessing the specific binding activity of the variable region is sufficient. For such subpart, Fab fragment resulting from proteolytic cleavage in the hinge region is the most traditional molecular design (Lilie et al., 1995; Feige et al., 2010). However, with recombinant technology several other designs have resurrected, such as single-chain fragment variable (scFv), diabodies, minibodies, and other assemblies extending to the level of single variable domain (Holliger and Hudson, 2005; Frenzel et al., 2013). The small fragments are considerably more economical to produce, while the affinity of the antibody can be retained for usage in specific targeting in diagnostics and therapeutic treatment (Holliger and Hudson, 2005; Ahmad et al., 2012). Fc engineered mAbs, bispecific antibodies, and Fab fragments are already among the approved biopharmaceuticals on the market, and the repertoire of engineered mAb therapeutics is anticipated to expand continuously (Presta, 2008; Walsh, 2014).

As the biopharmaceutical industry has matured, patents from the original products expire and pave the way for the newcomers of the industry, biosimilars. As adopted from generic drugs, biosimilars are designed to replicate a reference biological product in structure, efficacy, and usage. The definition of a biosimilar varies slightly between different regulatory frameworks, and the market access of such products depends on national patent protection (Singh, 2013). Roughly a dozen of different biosimilars are approved for the European market, including one antibody, infliximab (Walsh, 2014). Recently, the US Food and Drug Administration also accepted the first biosimilar drug, so biosimilars are slowly gaining regulatory approval and penetrating the US market (Ledford, 2015). Biosimilars may also mark the incentive to develop competing products with alternative and more advantageous production methods. Production platform development is a considerable area of interest both in pharmaceutical industry and in research, thus the cell systems with the potential to produce antibody biopharmaceuticals is explored next.

### 1.1.4 Production systems for antibodies

The ultimate producer of antibodies is the human body. Natural antibody production is obviously important for immune system functioning, but it is not neglected commercially. Nearly 100 tons of plasma-derived intravenous immunoglobulin isolated from blood is consumed in clinical applications every year (Kunert and Reinhart, 2016). In contrast to
blood-derived polyclonal antibodies, which consists of a mixture of several antibody variants, the development of monoclonal antibody technologies enabled the production of a single type of antibody recognizing a pre-defined epitope. The traditional method for monoclonal antibody production relies on cell fusion to create immortal, antibody-secreting hybridomas (Köhler and Milstein, 1975). In this method, laboratory animals are immunized with the antigen, after which differentiated B cells are isolated from the spleen and hybridized with immortal myeloma cells. The hybridomas secreting the desired antibody variant are isolated through screening and selection. Although hybridomas remain a popular platform for production of antibodies for diagnostic purposes (Saeed et al., 2017), recombinant DNA technology and certain mammalian cell lines are the current method of choice for therapeutic antibody production. Most notably, Chinese hamster ovary (CHO) cells are the workhorse of the industry, although several other cell lines, for example baby hamster kidney and human embryo kidney (HEK-293) cells, have been approved for recombinant protein production (Wurm, 2004; Kunert and Reinhart, 2016). Besides CHO cells, the most common choices for commercial antibody production are the murine cell lines NS0 and Sp2/0 (Walsh, 2014).

The selection of mammalian cell lines was largely determined by the adaptability of the cells to grow in suspension. Many cell lines are intrinsically reluctant to proliferate in single-cell cultures, so improving cell viability and growth in the process set-up has been a major focus in system development, in addition to media optimization, notably the development of serum-free media (Wurm, 2004). Through advances in process conditions resulting in improved cellular fitness, product titers reached in mammalian cell systems are more than 100-fold higher today compared to 30 years ago, reaching up to 10 g/l in current CHO-platforms (Wurm, 2004; Kunert and Reinhart, 2016). In the exclusive production scheme, cells secrete the protein product to the culture medium, from which it is purified in downstream processes (Wurm, 2004). Downstream processing constitutes a major proportion of the overall production costs, so considerations of the later phases should be implemented already in making the choice of the cellular production system (Kunert and Reinhart, 2016). In order to reduce downstream processing costs by increasing titers, several cellular engineering strategies have been presented for mammalian cell lines. As process-development focused on increasing cell viability, this aspect has been addressed also with genetic means to circumvent programmed cell death (Fischer et al., 2015). Through expression of anti-apoptotic genes, Figueroa et al. were able to increase operational process time of CHO cultures, which correspondingly increased mAb titers (Figueroa et al., 2007).
Product yields of mAbs benefit also from engineering of secretion and folding processes, such as co-expression of folding factors (Pybus et al., 2014).

Although mammalian expression systems are increasingly common for production of biopharmaceuticals, many of the approved biopharmaceuticals are produced in microbial cells (Walsh, 2014; Sanchez-Garcia et al., 2016). In contrast to mammalian cells, most microbial cells prefer growth in suspension and common organisms already have established culture processes available. In the case of antibodies, the N-glycan moiety in the heavy chain dictates the possible production platforms. Solely mammalian systems can produce human-type glycans, although immunogenic glycan patterns, such as terminal α-1,3-linked galactoses or N-glycolyneuraminic acid residues instead of N-acetylneuraminic acid (sialic acid), from non-human cell lines are of increasing concern (Ghaderi et al., 2012). Although eukaryotic microbes, such as yeasts produce N-glycosylated proteins, the N-glycans differ significantly from the complex-type N-glycans in human antibodies. Optimizing the glycan structures has established the field of glycoengineering, which persists as an active endeavor in all platforms considered for antibody production, from the simple bacterium Escherichia coli to the established mammalian hosts (Fisher et al., 2011; Ghaderi et al., 2012; Laukens et al., 2015; Meuris et al., 2014). Glycan optimization for improved or amended function is likely to remain in focus, since glycoengineered mAb variants have already started to enter the market (Walsh, 2014). Advances in glycoengineering have introduced the potential to use microbial eukaryotes, such as yeasts, as a production platform for therapeutic antibodies. In fact, competitive yields of grams per liter of IgG have been achieved with a glycoengineered Pichia pastoris strain in large-scale fermentations (Ye et al., 2011), demonstrating the commercial potential of the organism. Also the common yeast, S. cerevisiae, could be a future antibody factory, and its potential is reviewed in section 1.3.

Other eukaryotic systems that have been engineered to produce mAbs and antibody variants include plants (Valdés et al., 2003) and insect cells (Hsu et al., 1994), non-conventional yeast hosts (Swennen et al., 2002), certain fungal species (Ward et al., 2004; Nyyssönen et al., 1993), among others (Gasser and Mattanovich, 2007; Durocher and Butler, 2009). In addition, several prokaryotes have been investigated for use as antibody factories. In general, bacterial cells do not offer a proper folding environment for antibody products, but especially E. coli has been studied and engineered for this purpose (Lee and Jeong, 2015). A recent study combined two engineering approaches in E. coli to produce and secrete a correctly folded, active form of an anti-interleukin 1β scFv (Matos et al., 2014). Mimicking the eukaryotic secretion scheme, two main hurdles were surpassed: disulfide bonding inside the
cytoplasm and secretion of the folded product. The researchers used previously engineered factors to induce oxidation in the cytoplasm (Hatahet et al., 2010), after which the correctly folded scFv was exported via the Tat pathway. The Tat pathway in *E. coli* is specialized in folded proteins, and was indeed able to recognize and secrete structurally intact and active scFv (Matos et al., 2014). *E. coli* has been applied mostly for producing small antibody fragments, but recent advancements have pushed this bacterium close to producing full-length antibodies in large-scale (Lee and Jeong, 2015). The future will show how well molecular engineers can employ and design the different cellular systems for antibody production.

### 1.2 Nature’s system as the starting point – plasma cells and antibody secretion

Plasma cells are specialized members of the vertebrate immune system that naturally produce and secrete specific antibodies with an efficiency of up to $10^8$ IgG molecules /cell/h (Hibi and Dosch, 1986). The humoral immune response mediated by the secreted antibodies enables the organism to recognize and selective destroy the invading pathogen carrying the antigenic site. The enormous variety of different antibodies is created and managed in germinal centers of secondary lymphoid organs, where somatic recombination and mutation lead to clonal selection of B cells encoding the antibody in need. The differentiation into plasma cells is triggered by the presence of the antigen and its interaction with the recognizing B cell, which entails the counterpart of that specific antigen. Thus, only B cells able to fight against that specific antigen proliferate. The B cell population is subjected to another round of selection during the differentiation process called affinity maturation, where they specifically interact with T cells, the antigen, and extracellular signaling factors to yield a high-affinity clone which becomes the antibody-secreting plasma cell (Victora and Nussenzweig, 2012). The events leading to an activated B cell destined for plasma cell differentiation consists of a complex molecular interplay, but the details will not be discussed here. Instead, the key molecular determinants and cellular steps leading to the development of B cells into plasma cells will be described, focusing on the equipping of the cell to enable plasma cells to reach their superior capabilities in antibody secretion.

#### 1.2.1 From B cells to plasma cells

The genetic network guiding activated B cells into plasma cells is largely governed by levels of a few master regulatory transcription factors, which appear and dominate differentiation in a timed sequence (Figure 2). The first factors to guide the cells to commit to mature B cell
lineage are Bcl6 and Pax5. Pax5 and Bcl6 act simultaneously on hundreds of gene targets to achieve the necessary early events in mature B cell formation. They affect gene expression widely, mainly as repressors, and they suppress plasma cell activities during the mature B cell developmental phase (Cobaleda et al., 2007; Basso et al., 2010). In addition to downregulating B-cell unrelated signaling pathways, Pax5 induces the recombination events in heavy and light chain genes to create antibody diversity (Cobaleda et al., 2007). As important as their expression is for early differentiation, the disappearance of their activity is required to continue to the next stage of plasma cell development (Tarlinton et al., 2008). Bcl6 is downregulated before Pax5, and while these two gradually decrease in levels, Irf4 is the next key regulator to appear (Cattoretti et al., 2006). Upregulation of Irf4 drives the B cells to develop into antibody-secreting cells, next developmental stage being the plasmablast (Nutt et al., 2011). Irf4 is responsible for class-switch recombination from membrane bound antibody to secreted form and necessary in plasmablast and plasma cell generation (Klein et al., 2006).

The plasmablasts are already capable of secreting antibodies, but among other cellular transformations, they lose their ability to migrate and proliferate during maturation into plasma cells. After Irf4 induction, there is a sudden increase in the expression of Blimp-1, which was recognized as a major promoter of plasma cell differentiation right from its discovery (Turner et al., 1994). Irf4 takes part in activating Blimp-1 and together they act to promote plasma cell fate (Nutt et al., 2011). Blimp-1 represses early factors retaining B cell identity, such as Pax5 and Bcl6 (Figure 2). In fact, Blimp1 and Pax5 have mutually antagonistic effects, so these two determinants are separated in time in the developing cell, and the same has been stated for Blimp-1 and Bcl6 (Cobaleda et al., 2007; Shaffer et al., 2002). Although Irf4 and Blimp-1 are active in the cell at the same time, Blimp-1 has been named as the main orchestrator of the last part of plasma cell differentiation (Tarlinton et al., 2008; Shaffer et al., 2002). Irf4 and Blimp-1 are both required to induce the activation of the most notable of the downstream effectors of Blimp-1, the unfolded protein response (UPR)-related transcription factor Xbp-1 (Klein et al., 2006). Although the other transcription factors act mainly through repression of processes, active upregulation and induction of several genes for fortification of the secretory machinery for antibody production is largely mediated by Xbp-1 (Shaffer et al., 2004). Xbp-1 can be considered as the final player to yield the final phenotype of plasma cells as described below.
Figure 2 The key events in plasma cell development. The appearance and prevalence of major transcription factors in each stage is indicated. These factors influence each other’s expression, activation is shown with an arrow and repression with a blocked arrow. As displayed in the structure of the cell and summarized in lower left corner, the major cellular transformation is the expansion of secretory organelles, the ER (light pink) and Golgi (lavender). The relative volume of the nucleus (pink) in the cell decreases, as cell division arrests.

The exact extra- and intracellular signals required for differentiation vary depending on the antigen, location, and identity of the B cell population; differentiation can also occur outside germinal centers (Oracki et al., 2010). Accordingly, the population of plasma cells is diverse. A division between plasma cells can be made based on their life-span; the majority of plasma cells are short-lived and die after the acute immune response, but approximately 10-20% can become long-lived plasma cells (Tarlinton et al., 2008). Long-lived plasma cells have adapted to the secretory burden by sustaining the upregulation of genes involved in the secretory pathway (Cocco et al., 2012). The selection of which plasma cells become long-lived seems to depend on survival niches in the bone marrow, but the mechanisms are largely unknown (Oracki et al., 2010). Some contribution might come from the persistence in Blimp-1 expression, which has been shown to be a prerequisite for the maintenance of long-lived plasma cells (Shapiro-Shelef et al., 2005; Oracki et al., 2010). B cell selection and development continues throughout the organism’s lifetime in the constant battle against
foreign agents, but stability of some antibodies suggest that long-lived plasma cells can survive decades continuously expressing antibodies and providing life-long resistance to the encountered pathogens (Oracki et al., 2010; Tarlinton et al., 2008).

1.2.2 Cellular changes during plasma cell development

So what happens in the cell during the course of sequential appearance of the transcription factors regulating plasma cell development? After the immunoglobulin gene arrangements and affinity maturation during the rule of Pax5 and Bcl6, cells need to prepare to express and secrete the high-affinity antibody in large quantities. Blimp-1 and Xbp-1 govern the plasmacytic phase of differentiation, where the B cell events of antibody diversification are shut down and the transformation of the cell into an antibody factory is initiated and realized (Figure 2). Several comprehensive studies about transcriptome and proteome changes during terminal B cell differentiation have depicted a detailed view of what kind of cellular changes the key transcription factors push forward through their downstream effectors (Shaffer et al., 2002, 2004; van Anken et al., 2003; Romijn et al., 2005). These studies illustrate how B cells carefully prepare for their role as professional antibody secretors with massive reorganization of cellular structures and changes occurring throughout the proteome. The major events happening on the way to plasma cells under the reign of Blimp-1 and Xbp-1 are described next.

Primary and secondary targets of Blimp-1 include factors that control cell proliferation and growth, and genes involved in DNA synthesis and repair are downregulated. For example, c-myc is repressed by Blimp-1, which causes the terminally differentiating B cell to cease cell cycle progression (Shaffer et al., 2002). Although not considered as a direct consequence of cell cycle arrest, plasma cells are almost double in size and have a lower nucleus to cytoplasm ratio than the preceding differentiation forms (Shaffer et al., 2004; Wiest et al., 1990). Plasma cells are not only bigger but also enhanced in functionality: the overall protein and organelle content of the cell increases, mainly through the activity of Xbp-1 (Shaffer et al., 2004). Lysosomal, mitochondrial and metabolic activities are upregulated linearly with the increase in size to sustain energy and precursor supply for protein production (Shaffer et al., 2004; van Anken et al., 2003; Romijn et al., 2005). It is possible that the increase in organelle biogenesis is actually behind the cellular expansion (Shaffer et al., 2004). In addition to the cessation of division, plasma cells also stop migrating and this is reflected in the decrease in the relative abundance of cytoskeletal proteins (van Anken et al., 2003; Romijn et al., 2005).
Blimp-1 is mainly responsible for the upregulation of immunoglobulin gene expression, amounting to a 8-10-fold increase in antibody coding transcripts and accompanied by a 6-12-fold increase in translation (Shaffer et al., 2004; Gass et al., 2004; Wiest et al., 1990). Cells accommodate this increase with massive expansion of the secretory apparatus initiated by Xbp-1, downstream of Blimp-1 (Shaffer et al., 2004). The morphological changes in late differentiation phases of antibody secreting cells were systematically assessed by Wiest et al. (Wiest et al., 1990). In their study the expansion of the secretory apparatus was quantitatively described: the volumes of rough ER, Golgi and related vesicular structures increased well more than 3-fold, and the authors estimated the actual ER size to be up to 6-fold bigger. Interestingly, the membrane area increased with similar fold change values to the volume, indicating that functional membrane area became enriched in the secretory organelles. In fact, the number of ribosomes and the volume of ER increased to the extent where the ratio of antibody chains to ribosomes and antibody concentration in the ER stayed similar as before the upregulation of antibody synthesis (Wiest et al., 1990). The majority of these changes are actually initiated before the beginning of massive antibody synthesis, while the overall changes in cell size and metabolism occur before the expansion of secretory apparatus (van Anken et al., 2003).

The target genes of Xbp-1 involve practically every stage of the secretory pathway beginning from translation, signal peptide cleavage, and folding, including genes involved in ER-associated degradation, glycosylation, and vesicular traffic (Shaffer et al., 2004). All these processes are upregulated in a manner independent of the UPR, which is the stress response usually initiated due to accumulation of unfolded secretory proteins in the ER. On the protein level, abundance of ER-resident factors display a clear dominance in plasma cells (Romijn et al., 2005; Salonen et al., 2013) and the amount of ER resident proteins steadily increases during differentiation (van Anken et al., 2003; Romijn et al., 2005). ER-resident proteins seem to be upregulated in an antibody-specific manner: for example, the oligosaccharide-binding lectin calreticulin followed the general trend of increasing expression but the soluble homolog calnexin decreased in relative protein levels during differentiation, indicating that the nascent antibodies preferred interaction with the other lectin. Similarly, many other folding factors in the ER became differentially enriched in plasma cells (Romijn et al., 2005). Antibodies clearly have extensive requirements for their folding environment, so the factors in the ER deserve a closer look.
1.2.3 Endoplasmic reticulum folding machinery

The complex folding routes of the IgG tetramer (section 1.1.2) evidently impose the need for plasma cells to upregulate the folding machinery to enhance secretion rates. In effect, the high efficiency of antibody secretion was specifically attributed to the Xbp-1 induced upregulation of ER-resident folding factors which process the nascent immunoglobulin chains (Shaffer et al., 2004). Generally, there are three classes of folding factors: molecular chaperones or holdases, protein disulfide isomerases (PDIs), and PPIases, which catalyze the peptidyl-prolyl isomerization reactions. Mammals express several proteins from each class so they are capable of tailoring the exact composition of the folding machinery according to the client protein (Romijn et al., 2005). The ER-resident folding factors display numerous pairwise interactions resulting in a functional network that assembles into multiprotein complexes on the folding substrate (Jansen et al., 2012; Meunier et al., 2002). A specific subset of folding helpers have been identified to aid antibody folding, and the functioning of these ER components will be discussed next.

The importance of BiP association in antibody folding was already mentioned and in fact, this protein was specifically named after its interaction with the nascent heavy chain (Haas and Wabl, 1983). BiP belongs to the heat shock protein 70 (Hsp70) family of folding factors, which undergo a functional ATP-hydrolysis cycle upon substrate binding. When ATP is attached to the nucleotide-binding domain of BiP, the binding to the unfolded protein is initiated, but it is only stabilized after ATP hydrolysis to ADP. The substrate is released after nucleotide exchange from ADP to ATP, priming BiP for another round of substrate binding (Otero et al., 2010). The activity of BiP is guided by specific co-chaperones that determine the nucleotide status upon substrate binding. ATPase activity is stimulated mainly by ER localized DnaJ proteins, of which ERdj3 is the major player in the case of antibody folding (Meunier et al., 2002; Feige et al., 2010). In addition to ATPase activity control, ERdj3 works in helping to recruit BiP to its substrates and it participates in the process of translocation (Otero et al., 2010; Guo and Snapp, 2013). Nucleotide exchange activity on the other hand is contributed mainly to GRP170 which catalyzes the displacement of ADP with ATP from BiP upon completion of antibody folding (Feige et al., 2010). GRP170 was identified in the heavy chain associated multiprotein complex and it increased in prevalence during plasma cell differentiation (Meunier et al., 2002; Romijn et al., 2005), but its abundance might also be accounted for its ability to function as a holdase (Behnke and Hendershot, 2014). BiP is certainly not the only molecular chaperone associated with IgG,
as it has been reported to act sequentially with another major holdase, GRP94 (Melnick et al., 1994; Feige et al., 2010; Meunier et al., 2002).

One IgG molecule contains between 16 and 25 disulfide bonds (Liu and May, 2012). The formation of these bonds is facilitated, accelerated and enabled by a group of oxidoreductive enzymes from the PDI protein family. PDIs share a common structural element, the thioredoxin fold, and they participate in different stages in oxidoreductive protein folding together with cellular electron-buffering systems such as glutathione (Hatahet and Ruddock, 2009). Several PDI-family members participate in antibody folding and are upregulated during differentiation, including PDI, P5, ERp72, and ERp57 (Romijn et al., 2005; Meunier et al., 2002; Feige et al., 2010). As the founding member, PDI is the most abundant and best characterized of PDIs, whereas the other oxidoreductases might be specialized in certain substrates or localized in specific subcomplexes in the ER, as demonstrated in the case of ERp57 (Hatahet and Ruddock, 2009; Braakman and Hebert, 2013; Oliver et al., 1999). PDIs perform disulfide exchange reactions, where native bonds in the client protein are formed through oxidation, reduction and shuffling of the possible disulfides with a cysteine pair in the PDI. Through these reactions the disulfide bonds in the target protein are correctly formed, false pairings are isomerized to native disulfide bridges, or non-native bonds are reduced back to cysteine residues. However, the overall redox balance is not affected by PDIs, since the enzyme adjusts its oxidation state accordingly. The catalyst can be regenerated in a reaction with glutathione, the major redox buffer in the ER, or by oxidative enzymes, such as Ero1 (Bulleid, 2012; Hatahet and Ruddock, 2009). Both isoforms, Ero1α and Ero1β transfer electrons from PDI to the final electron acceptor oxygen, and especially Ero1α was strongly up-regulated during plasma cell differentiation (Romijn et al., 2005). Ero1 activity can result in the formation of a reactive oxygen species, hydrogen peroxide, which contributes directly to oxidative stress and plasma cell death as described later.

As discussed in section 1.1.2, peptidyl-prolyl isomerizations constitute several rate-limiting steps in antibody folding. These steps can be accelerated significantly with the expression of PPlases, of which the most notable in the ER of antibody-producing cells are cyclophilin B (CypB) and FKBP2 (Meunier et al., 2002; Romijn et al., 2005). These represent two different PPlase families, the cyclophilins and the FK506-binding proteins (FKBP), from the larger family of immunophilins (Göthel and Marahiel, 1999). The exact functioning of PPlases in folding is rather poorly characterized, and thus the specificity and redundancy of the different family members are largely unknown (Braakman and Hebert, 2013). FKBP and cyclophilin family members display some amino acid sequence specificity and they act
Introduction

29

Introduction

differently on Ig domains (Harrison and Stein, 1990; Lilie et al., 1995). In addition to being folding enzymes, PPIase-family members seem to have several other functions. The importance of PPIase-proteins can be highlighted by the fact that they are the targets of the immunosuppressive drugs FK506 and cyclosporin A, hence the name of the proteins families (Galat, 1993).

Antibody heavy chains are subjected to yet another folding system in the mammalian ER that relies upon the N-glycan moiety. The lectin chaperones calnexin and calreticulin are the main actors in this folding cycle, where the N-glycan is used as a folding signal. Indication of the unfolded state is established by UDP-glucose:glycoprotein glucosyltransferase (UGGT), which adds a certain glucose residue to the N-glycan to mark the unfolded or incorrectly folded glycoprotein. Upon folding, the glucose is again removed by ER glucosidases and the deglucosylation releases the protein from the lectin folding complex (Helenius and Aebi, 2004; Williams, 2006). In addition to identifying and binding the oligosaccharide on the folding protein, calnexin/calreticulin recruit folding enzymes ERp57 and CypB to the protein folding complex (Oliver et al., 1999; Kozlov et al., 2010). The heavy chain may undergo several of the N-glycan directed folding cycles, as indicated by the association of the nascent heavy chain to the UGGT (Meunier et al., 2002). This so called calnexin/calreticulin cycle has also a direct link to ER quality control, since prolonged engagement in the cycle exposes the unfolded or misfolded heavy chain for degradation (Williams, 2006).

1.2.4 IgG production characteristics of plasma cells

The antibodies represent 10% of newly synthesized proteins in plasma cells (Wiest et al., 1990). The unsurpassed capability of the cells to produce and secrete antibodies does not come without costs. As mentioned earlier, the majority of the plasma cell population has a short lifespan and programmed cell death has been postulated to be a direct consequence of cellular stress due to antibody production. Plasma cells experience a great metabolic stress when approximately $8 \times 10^6$ amino acid molecules need to be synthesized or obtained per second to continuously replenish the precursor pool for antibody production (Cenci and Sitia, 2007). Metabolic imbalances and toxic byproducts are accompanied by oxidative stress, in which toxic compounds in the form of reactive oxygen species accumulate in the cell (Masciarelli and Sitia, 2008). During differentiation oxidative stress is still balanced with antioxidant responses, but changes in protein redox state coincide with the initiation of antibody production (Hansen et al., 2013). During the antibody production phase the amount
of disulfides to be formed is up to $10^5$ per second (Masciarelli and Sitia, 2008). Proteins involved in redox balance increase linearly during differentiation, but despite this, oxidative stress is one of the several factors contributing to plasma cell death (van Anken et al., 2003; Masciarelli and Sitia, 2008).

Most evident burdens in plasma cells are the stresses caused by protein synthesis and intracellular accumulation of unfolded proteins. UPR and ER stress responses lie at the heart of plasma cell differentiation and function. During B cell differentiation, the central transcription factor Xbp-1 has UPR-independent roles in secretory apparatus expansion (Shaffer et al., 2004). However, some evidence suggest that the ER load also activates the “classical, cargo-induced Xbp-1 mediated UPR” during differentiation; there might be an UPR-related feed-forward loop managing the functioning of the secretory machinery (van Anken et al., 2003). In addition to symptoms of ER stress, plasma cells also display an accumulation of polyubiquitinated proteasome substrates, as the proteolytic capacity is considerably decreased during plasma cell differentiation (Cenci et al., 2006). Proteasomal overload is also considered a driver of apoptosis in plasma cells (Cenci and Sitia, 2007).

In combination, the individual stress signals might constitute a timer or counter to control the lifespan of plasma cells by inducing apoptosis when a functional threshold is passed (Cenci and Sitia, 2007). But what about long-lived plasma cells? These cells reside primarily in the bone marrow and persist to secrete antibodies for months, possibly even years (Manz et al., 1997). Long-lived plasma cells have adapted to the secretory burden and they do not seem to experience a sustained UPR, although they retained the capability of inducing an UPR stress response via splicing of Xbp-1 (Cocco et al., 2012). Clearly, the adaptation mechanisms to the proteosynthetic stress induced in long-lived plasma cells differ from that present in their short-lived counterparts. Although this difference in populations has not been elucidated, it could be accounted to enhanced efficiency to cope with cytotoxic and proteasomal stresses (Cenci and Sitia, 2007). The longevity and the nature of stress responses in plasma cells should be kept in mind, when this natural model is utilized in engineering of antibody cell factories.
1.3 Saccharomyces cerevisiae as an IgG production platform

The yeast *S. cerevisiae* has an established place as a model eukaryote and as a widespread biotechnological workhorse. Due to its ability to exist in diploid and haploid states, *S. cerevisiae* was early on geneticists’ favorite in classical experiments and consequently, its genome of 12,000 kilobases divided to 16 chromosomes was the first eukaryotic genome sequence to be fully resolved (Goffeau et al., 1995). Genetic manipulation brings the promise of making any biological product of interest in yeast, and some of the success stories are referred to below. Compared to other platforms for recombinant protein expression and the natural IgG producers, plasma cells, *S. cerevisiae* still lags behind in several aspects. Thus, the remainder of this section discusses the major engineering targets to transform *S. cerevisiae* into a cell factory for biotherapeutics.

1.3.1 *S. cerevisiae* as a production platform

Products derived from *S. cerevisiae* cover the entire range of possible biotechnologically derived substances, varying from simple fuels and bulk chemicals to fine chemicals and complex pharmaceuticals (Hong and Nielsen, 2012), as illustrated in Figure 3. Apart from traditional beverage and food production, bioethanol is by far the most prominent product from *S. cerevisiae*. The annual worldwide production of bioethanol reaches approximately 95 billion liters, of which most is obtained from *S. cerevisiae* fermentations. The bioethanol on the market is mainly from first and second generation bioprocesses, where the generation refers to the complexity and origin of the feedstock (Zabed et al., 2017). Importantly, widening the substrate range of *S. cerevisiae* towards higher generation feedstocks has been a central metabolic engineering target for decades (Hong and Nielsen, 2012; Hahn-Hägerdal et al., 2007). Ultimately, utilization of lignocellulosic raw materials could be achieved by consolidated bioprocessing, in which the *S. cerevisiae* strain producing the ethanol is also engineered to secrete the cellulase enzyme to enable the simultaneous hydrolysis and fermentation (Van Zyl et al., 2016; Zabed et al., 2017). The advances in the yeast bioprocessing and engineering have raised the interest to adapt the current processes for strains producing more advanced fuels, such as biobutanol (Hong and Nielsen, 2012).
Introduction

Figure 3 Examples of products from *S. cerevisiae*. Volumes of the products are indicative and relative. The x-axis shows the product categories arranged by increasing value or complexity of the product. The pie chart depicts the share of the classes of protein pharmaceuticals produced in *S. cerevisiae* as reported in Walsh, 2014.

The extension of the product range is by no means limited to biofuels. In addition, numerous organic chemicals are produced with *S. cerevisiae*, which either provides a biotechnological alternative to petrochemical industry or enables the economical production of complex, non-synthesizable molecules. An example of a commodity chemical present in the *S. cerevisiae* metabolome is succinic acid, a common chemical building block and polymer precursor (Figure 3). The organism natively contains very low amounts of the compound, but through model-based engineering and directed evolution, Otero et al. demonstrated that the accumulation of succinic acid improved up to 30-fold (Otero et al., 2013). Succinic acid is an endogenous product, but most target compounds are not native for *S. cerevisiae* and thus require the introduction of heterologous genes and pathways. A recent triumph in yeast biotechnology was the *S. cerevisiae*-aided synthesis of a potent anti-malarial agent artemisinin (Figure 3), which otherwise has to be isolated from the plant *Artemisia annua* (Paddon et al., 2013). The engineered yeast strain developed by the biotech company Amyris was reported to produce artemisinic acid up to 25 grams per liter, and this precursor was further converted chemically to artemisinin, providing a promise of a reliable supply of an effective antimalarial treatment (Paddon et al., 2013). With the huge range of possible usages from chemical building blocks and food additives to pharmaceuticals, the production of commodity and fine chemicals from *S. cerevisiae* is linked to the ambitions to develop a
general industrial chassis. Most of the yeast-based products derive from the same set of metabolic precursors and cofactors, so Jouhten et al. proposed a modularization of design based on the family of chemical products (Jouhten et al., 2016). By investigating the metabolic pathways of 28 heterologous products, they identified product subgroups based on the direction of the carbon flux in the metabolism. Engineering the carbon flux with tested modifications towards the target product subgroup could reduce the time in the strain engineering workflow and bring *S. cerevisiae* closer to become a general chassis for production of any desired organic compound (Jouhten et al., 2016).

By far the chemically most complex product class from yeast fermentation is recombinant biopharmaceuticals. The first *S. cerevisiae* derived biopharmaceutical was the hepatitis B vaccine Recombivax, which entered the market in the 1980’s (Sanchez-Garcia et al., 2016). The current selection of protein products from *S. cerevisiae* include hormones, notably insulin, other vaccines, growth factors and serum albumin, as summarized in Figure 3 (Ferrer-Miralles et al., 2009; Nielsen, 2013; Walsh, 2014). Around 15% of all recombinant pharmaceuticals is produced in *S. cerevisiae*, making it the third most common production platform for pharmaceutical proteins, after mammalian CHO cells and the bacterium *E. coli* (Baeshen et al., 2014; Ferrer-Miralles et al., 2009). The characteristics supporting recombinant protein production in *S. cerevisiae* are the same that explain its popularity as a production platform in general: it is robust and easy to manipulate, it grows fast in inexpensive media and the organism has been granted the GRAS (generally regarded as safe) status (Nevoigt, 2008; Hong and Nielsen, 2012). However, the intrinsic capability of *S. cerevisiae* to produce and secrete protein is limited as described in section 1.3.3, and clearly inferior to some other yeasts, such as *P. pastoris* (*Komagataella phaffii* in its official classification) (Mattanovich et al., 2014; Porro et al., 2011). In recent years, the biopharmaceutical market has oriented towards treating cancer, but only one growth factor from *S. cerevisiae*, Sagramostim, is among the 20 approved biopharmaceuticals for anticancer treatment (Sanchez-Garcia et al., 2016). Glycoengineering (section 1.3.2) and cellular engineering (section 1.3.3) are exploited in order to broaden the selection of feasible recombinant proteins and to retain *S. cerevisiae* as a competitive platform for recombinant pharmaceuticals.

The advantage of the fact that *S. cerevisiae* is used as a platform for such a wide range of products is that the knowledge gained from each engineered strain and during process development support further applications. Studies improving enzyme secretion for consolidated bioprocessing provide important information on the yeast secretory pathway,
which can be utilized in strain engineering for biopharmaceutical protein production (Hou et al., 2012a; Van Zyl et al., 2016). The genetic tools developed for metabolic engineering, such as promoters, terminators, and vectors, are used invariably throughout the product spectrum. Biotechnological research itself has produced many of these engineering tools: for example, while modifying metabolic pathways, scientists at Amyris company discovered and published a refinement to the new CRISPR-Cas9 technology, which accelerated their strain creation (Horwitz et al., 2015). Transferrable methodologies and accumulating knowledge on cellular processes merge in the idea of synthetic biology, which integrates technological advancements with system-level information for the design of cell factories. The contributions of synthetic biology on \textit{S. cerevisiae} biotechnology are covered in section 1.4.

1.3.2 Glycoengineering yeasts

N-glycosylation is a common post-translational modification reported to be present in over 200 proteins in \textit{S. cerevisiae} (Kung et al., 2009). Early steps of N-glycan assembly are well conserved and the glycan structures in the ER are essentially the same in all eukaryotes. The glycan biosynthesis and subsequent attachment to proteins occur in the ER, although the oligosaccharide assembly is initiated on the cytosolic side of the ER membrane. The first sugar residue, N-acetylglucosamine (GlcNAc), is covalently attached to a lipid carrier, dolichol phosphate, which is embedded in the membrane. Another GlcNAc and five mannose residues are further added on the cytosolic side, before a flippase enzyme upends the glycan-lipid hybrid, the sugar chain now facing to the ER lumen. Growth of the glycan on the lipid-carrier is continued with four mannoses to yield a structure with three branches. Three consecutive glucose residues are added to the first branch, before the structure is recognized by the oligosaccharyltransferase (OST) complex, which transfers the glycan moiety from the lipid to the nascent polypeptide. The glucose residues are utilized as folding signals and are eventually removed along with one mannose from the middle branch to yield the structure shown in Figure 4A, which is the common N-glycan structure in eukaryotes upon ER exit (Helenius and Aebl, 2004). The N-glycan processing pathways in different eukaryotes diverge in the Golgi apparatus. Only the core two GlcNAcs and three mannoses are conserved among species forming a common five-residue core, as the glycan branches gain and lose sugar residues, along with other modifications. Yeasts usually display hypermannosylated glycan structures shown schematically in Figure 4B, extending up to dozens of mannose residues. Three common types of N-glycans have been described in mammals: high-mannose type, complex type and the hybrid type displaying features from
both of the former types (Stanley et al., 2009). N-glycans on antibodies are characteristically of the complex type, depicted in Figure 4C. Thus, the key question in glycoengineering is how to modify the yeast organism to go from hypermannosylated structures in Figure 4B to produce complex-type glycans as shown in Figure 4C.

Humanization of the N-glycan structures has been pursued in several yeast species, but the greatest share of the glycoengineering work is concentrated on the yeasts \textit{P. pastoris} and \textit{S. cerevisiae}. In general, two strategies have been employed to obtain human-like N-glycans (Laukens et al., 2015). The first strategy concentrates on changing the events taking place in the Golgi to prevent hypermannosylation, which was also the strategy chosen and refined by the biopharma company GlycoFi (Hamilton et al., 2003). The steps in this strategy were summarized in the commercially available GlycoSwitch technology (Jacobs et al., 2009). The central modification is to disrupt the \textit{OCH1} glycosyltransferase gene to remove hypermannosyl structures while introducing two heterologous mannosidases to trim down the structure leaving the ER (Figure 4A). The mannosidases remove residues from each branch (Hamilton et al., 2003; Jacobs et al., 2009). The other strategy aims to generate the substrate for mammalian glycosyltransferases by interfering earlier in the N-glycan synthesis pathway: with deletions of ER-localized lipid-linked oligosaccharide synthesis genes \textit{ALG3} and \textit{ALG11}, the glycan synthesis is ceased to the target structure. As the truncated glycan is not a substrate of \textit{OCH1}, hypermannosylation is circumvented. However, in order to transfer the truncated glycan efficiently to protein substrates, an artificial flippase and a protozoan OST needed to be added to the engineered strain (Nasab et al., 2013). From the truncated glycan structure, the two strategies continue similarly by introducing a GlcNAc residue to both branches by expression of mammalian \textit{N}-acetylglucosaminyl transferases (Hamilton et al., 2003; Jacobs et al., 2009; Nasab et al., 2013). Further, expressing galactosyltransferases in the Golgi to obtain the addition of the next residues, galactose, has been described (Jacobs et al., 2009). Also, a combination of the two strategies, involving both \textit{ALG3} and \textit{OCH1} deletions, has been investigated, extending up to the terminal galactoses (Bobrowicz et al., 2004).
Figure 4 N-glycan structures in yeast and mammals. (A) The common N-glycan structure in eukaryotes upon the exit of the protein from the ER. After this step, the N-glycosylation pathways diverse. Yeast produce hypermannosylated structures (B), where n can be up to a hundred residues. N-glycans on human antibodies are of complex type (C), where only the five core residues are similar to yeast structures (two GlcNAc and three mannoses). GlcNAc: N-acetylglucosamines, Neu5Ac; N-acetylneuraminic acid or sialic acid.

Although obtaining humanized glycan structures in yeast is feasible, glycosylation efficiency and heterogeneity remain to be addressed (Piirainen et al., 2014). While testing their glycan synthesis interference strategy, Nasab et al. resolved the reduced glycan site occupancy by optimizing the flipping and glycan transfer processes (Nasab et al., 2013). N-glycan heterogeneity can be addressed by deleting additional Golgi-resident mannosyltransferases to reduce side reactions (Nasab et al., 2013; Piirainen et al., 2016). Extending the N-glycan effectively with the target sugar residues, GlcNAc and galactose, requires that the corresponding substrates, the uridine diphosphate (UDP)-linked sugar monomers, are present in the Golgi lumen in sufficient concentrations. This might require expression of the corresponding transporters (Hamilton et al., 2003; Piirainen et al., 2016), or expressing a Golgi-localized epimerase that converts UDP-glucose into UDP-galactose (Jacobs et al., 2009). In addition to the numerous genetic means to improve glycosylation efficiency (Piirainen et al., 2014), glycoform homogeneity can be influenced by varying growth conditions and process optimization (Jacobs et al., 2009). Another challenge related to the yeast glycoengineering is the diminished growth rate observed for several glycan-modified strains (Nasab et al., 2013; Jacobs et al., 2009).

Engineering of O-linked glycosylation in yeasts is likewise an active field (Laukens et al., 2015). Yeast-specific O-mannosylation pathways seem to recognize antibodies as substrates, so preventing undesired O-glycans may require deletion of specific protein-O-mannosyltransferases from the genome or the use of inhibitory compounds (Nett et al., 2013; Kuroda et al., 2008). Glycoengineering of yeasts is especially targeted to antibodies (Li et al., 2006; Nasab et al., 2013) because of the great interest in developing alternative
production platforms. However, glycoengineering efforts in yeast are still struggling to complete the N-glycan structure shown in Figure 4C. Terminal sialic acids are especially problematic: in addition to the sialic acid transferring and transporting proteins, the process requires the introduction of the biosynthesis pathway for the sugar substrate (Hamilton et al., 2006). Terminal sialic acids on antibodies might not be of interest (Beck and Reichert, 2012), and the absence of the branching fucose residue (see Figure 4C) is potentially a desirable trait (Ferrara et al., 2006a; b). Engineering of antibody glycoforms is an integral part of glycoengineering the production platform, as mentioned in section 1.1.4.

1.3.3 Characteristics of the S. cerevisiae secretory pathway

As eukaryotes, yeast cells contain a similar set of organelles to mammalian cells, although there are several morphological differences between the cells. Yeast cells are smaller in size, and the secretory organelles, the ER and the Golgi apparatus, have a significantly different shape and distribution in the cell compared to regular mammalian cells. The structure of a yeast cell (Figure 5) is far from that represented for an antibody-producing plasma cell (Figure 2) and does not seem to be optimal for protein secretion (Delic et al., 2014, 2013). S. cerevisiae naturally secretes low amounts of proteins such as hydrolases (e.g. invertase) and cell wall components comprising a secretome of over 200 different proteins (Smeekens et al., 2017). However, in the case of high-yield protein secretion, several bottlenecks have been identified in the yeast secretory system, and these engineering targets are highlighted next.
Figure 5 Overview of the S. cerevisiae cell organization and the major events in protein secretion. Protein stages are colored blue and the flow of polypeptide through the cell is indicated by arrows. After translation, secreted protein is translocated to ER, where the folding occurs. Correctly folded proteins gather to ER-exit sites (ERES), from where they are transported in vesicles to the cis-Golgi. During the traverse through the Golgi compartments, the N-glycans are modified before the mature protein is secreted. However, several undesired processes can also be initiated, such as vacuolar sorting from Golgi leading to degradation, or ER-associated degradation (ERAD) leading to proteasomal degradation in the cytosol, in case the protein maturation fails in the ER. If unfolded proteins accumulate in the ER, unfolded proteins response (UPR) is initiated, which signals from the ER to the nucleus to change transcription and adjust secretory processes to restore homeostasis.

The endoplasmic reticulum

Polypeptides are synthesized in the cytosol on ribosomes and targeted to the secretory pathway by transporting them through ER membrane in a process called translocation. In contrast to higher eukaryotes, S. cerevisiae does not seem to prefer post-translational and co-translational translocation to the other but uses both systems in parallel depending on the properties of the N-terminal signal peptide (Ng et al., 1996). In co-translational translocation, the signal-recognition particle (SRP) binds to this signal peptide as soon as the polypeptide part emerges from the ribosome. The SRP guides the polypeptide-ribosome complex to the ER membrane, where an SRP receptor awaits as a docking station. Post-translational translocation is SRP-independent, as the signal peptide is not bound by SRP, but the newly synthesized protein associates to the heptameric SEC-complex, which specifically contains
the Sec62-protein needed for the process. Excluding Sec62p, both translocation events utilize the same set of proteins in the translocon complex, where the transmembrane protein Sec61p forms the pore through which the protein traverses the membrane (Jermy et al., 2006; Delic et al., 2013). Sec61 complex also recruits the signal peptidase complex to the translocation site, assuring that the N-terminal signal peptide is removed immediately after the polypeptide emerges to the ER lumen (Kalies et al., 1998). Additionally, *S. cerevisiae* has another translocon pore, the trimeric Ssh1 complex, which appears to be specific for cotranslational translocation but its biological importance remains to be established (Finke et al., 1996; Delic et al., 2013). As the first step, translocation is the first possible bottleneck in secretion, which Tang et al. demonstrated by improving titers of several heterologous enzymes with overexpression of SRP components (Tang et al., 2015). Cytosolic chaperones are integrally involved in post-translational translocation by maintaining the nascent protein unfolded before ER entry (Ng et al., 1996), which explains some of the beneficial effects of overexpression of cytosolic chaperones in protein secretion (Tang et al., 2015; Gasser et al., 2007).

The polypeptide emerging from the translocon pore is immediately exposed to the folding environment in the ER lumen. The yeast Hsp70-family chaperone Kar2p is associated with the translocon complex on the luminal side and it is the first folding factor to bind the polypeptide. Kar2p has an essential role in both co- and post-translational translocation, as consecutive ATP-driven binding and substrate release cycles help to pull the polypeptide through the translocon pore (Brodsky et al., 1995). As Kar2p is a functional homolog of the mammalian protein BiP, it is similarly regulated by co-chaperones which either stimulate its ATPase function or aid the nucleotide exchange to enable recurrent binding (Normington et al., 1989; Steel et al., 2004). Five ER-resident DnaJ proteins have been characterized in *S. cerevisiae*, three of which have the J domain on the luminal side of a membrane protein (Sec63p, Hlj1p, and Erj5p) and two soluble co-chaperones, Scj1p and Jem1p. In addition to these enablers of the ATPase activity, two nucleotide exchange factors of Kar2p are present in the yeast ER, Sil1p and Lhs1p (Delic et al., 2013). Lhs1p is also a molecular chaperone and a member of the Hsp70 family (Steel et al., 2004; Saris et al., 1997). Sil1p and Lhs1p are not functionally redundant, as specifically Lhs1p is required for translocation (Steel et al., 2004).

Folding of recombinant proteins in yeast has been named as a major bottleneck for secretion (Delic et al., 2014), which specifically holds true for the complex full-length IgG. Although the Ig fold is one of the most widely distributed protein topologies in nature (Feige et al.,
an Ig-like fold for protein domains is relatively rare in yeast, matching only to 19 loci in the yeast genome (EMBL-EBI, 2017). As described in section 1.2.3, antibodies have specific requirements for their folding environment, which are hardly met by the yeast ER. Upregulation of the folding machinery is a common approach to enhance recombinant protein secretion, and the overexpression of BiP/Kar2p has been shown to improve the production of scFv antibodies 2- to 4-fold in *S. cerevisiae* (Xu et al., 2005; Hackel et al., 2006; Shusta et al., 1998). In general, the effects of Kar2p overexpression on recombinant protein production are inconsistent, possibly due to the involvement of the protein in numerous ER processes in addition to folding (Delic et al., 2014). Kar2p is intrinsically highly abundant, present in over 300,000 molecules per cell. In contrast, the amounts of Kar2p regulating proteins are several orders of magnitude lower (Ghaemmaghami et al., 2003), so upregulating the co-chaperones might be a better approach to boost Kar2p activity. Payne et al. showed that overexpression of *LHS1* and *JEM1* had a positive effect on the secretion of three different recombinant proteins (Payne et al., 2008a).

BiP/Kar2p overexpression is often accompanied by the overexpression of PDI. In the same studies that established the benefits of BiP/Kar2p upregulation on scFv production in yeast, also PDI was tested. PDI alone led to improvements of similar extent to BiP, and when co-overexpressed together, combinatorial and synergistic effects were recorded (Xu et al., 2005; Hackel et al., 2006; Shusta et al., 1998). Accordingly, upregulation of PDI was found to enhance secretion of a full-length IgG in *S. cerevisiae* (Rakestraw et al., 2009). Although in some of these studies the used folding factors where from heterologous origin, oxidative protein folding is fully functional in *S. cerevisiae* and realized by the essential Pdi1p with the accessory PDI family members Mpd1p, Mpd2p, Eps1p, Eug1p, and by the oxidase Ero1p (Delic et al., 2013). The number of PDI family members in yeast is considerably lower than in mammals, from which over 20 members of the ER-resident oxidoreductases are known (Braakman and Bulleid, 2011). In addition to Pdi1p, Ero1p was identified as a candidate to improve secretion (Wentz and Shusta, 2007). While PDIs mainly shuffle disulfide bonds, the activity of Ero1p is critical to complete the oxidation and restore the activity of PDIs (Hatahet and Ruddock, 2009). Overexpression of Ero1p was shown to enhance secretion of two different scFv antibodies in *S. cerevisiae* (Wentz and Shusta, 2007), indicating that the functional partner of PDI is as important for efficient secretion as the isomerase itself.

In contrast to PDIs and oxidoreductases, the other class of folding enzymes, PPIases, has been largely overlooked in engineering approaches (Braakman and Hebert, 2013; Delic et al., 2014). In the case of antibody production, the level of PPIase activity might be crucial,
since the isomerization step is often rate-limiting in Ig-fold formation (see section 1.1.2). Although *S. cerevisiae* has five PPIase family members localized to the ER and secretory machinery, their exact biological functions have remained obscure (Arevalo-Rodriguez et al., 2004; Delic et al., 2013). Interestingly, all of these five genes can be removed simultaneously from the genome without a loss of viability, and without the appearance of significant growth defects (Dolinski et al., 1997). The ER PPIases seemingly have a role in protein folding as they are upregulated during ER stress, although there are no reports of their protein substrates in *S. cerevisiae* (Dolinski et al., 1997; Delic et al., 2013).

Once folding is completed, secretory proteins gather to the ER-exit sites (ERES) to be packed in transport vesicles destined to the Golgi apparatus. In *S. cerevisiae*, numerous small spots throughout the ER, typically 30-50 per cell, comprise an ERES and these sites are not very discrete (Rossanese et al., 1999). ERES localize to high-curvature areas in the tubular ER (Okamoto et al., 2012), which is one of the structural subareas of the yeast ER. As indicated in Figure 5, the ER occupies several areas in the cytosol and the ER can be divided into three structurally distinct domains: the central cisternal ER around the nucleus, the plasma-membrane associated ER, and the tubular ER, which extends to the cytosol from the two other structures (West et al., 2011). The shape of the domains is largely determined by the abundance of the reticulon proteins Rtn1p, Rtn2p, and Yop1p, which create and maintain the tubular structures of the ER (Voeltz et al., 2006). However, the functional distinction between the domains is not clear. The traditional differentiation between rough and smooth ER does not seem to hold true, as all the domains are studded with bound ribosomes associated to translocon complexes, although the tubular ER has a considerably lower ribosome density than the other two. Interestingly, in the plasma-membrane associated ER, ribosomes occur solely on the cytosolic side of ER cisternae (West et al., 2011). As indicated by the localization of ERES, the dynamics of tubular and cisternal structures and the cellular localization of the ER domain might determine the function of that area, but this structure-function relationship remains to be elucidated.

*Coping with conformational stress: ER-associated degradation and the unfolded protein response*

Folding and protein assembly is not always successful. As malformed proteins pose a threat to the cells and subsequent organelles of the secretory machinery, they do not gather to ERES but instead are subjected to ER-associated degradation (ERAD). In ERAD, the polypeptide is retrotranslocated from ER lumen back to the cytosol, where it is degraded by the
proteasome, the cytosolic system for controlled protein destruction. The polyubiquitin signal of proteasomal degradation is created at the ER membrane by an ubiquitin-ligase complex, which is composed differently in ERAD subtypes depending on where the unfolded area of the protein is: on a luminal domain (ERAD-L), in a transmembrane domain (ERAD-M) or on the cytosolic side (ERAD-C) (Carvalho et al., 2006). As secretory recombinant proteins are mostly soluble luminal proteins, they are substrates of the ERAD-L, which is characterized by the most elaborate ubiquitin-ligase complex. The ERAD-L complex is centered on the transmembrane ubiquitin ligase Hrd1p, which is also centrally involved in the retrotranslocation of the substrate protein (Carvalho et al., 2006, 2010). Other components of the complex are transmembrane proteins Der1p, Usa1p, Hrd3p, and Ubx2p and the ATPase complex Cdc48p on the cytosolic side, which participates in pulling the substrate through the retrotranslocon with the power of ATP hydrolysis (Carvalho et al., 2006; Stein et al., 2014). On the luminal side of the ERAD-L complex, the soluble Yos9p associates with the luminal domain of Hrd3p to enable substrate delivery to the complex (Denic et al., 2006).

How are ERAD substrates recognized? ERAD is directly linked to protein quality control, where the glycan-directed mechanisms are best characterized. As was described for plasma cells in section 1.2.3, the same glucose residues act as markers for the folding state of the secretory protein in the yeast ER. However, no gene or activity for UGGT has been found in this yeast, although S. cerevisiae has the other components for the calnexin cycle (Fernández et al., 1994; Delic et al., 2013). Thus, the folding cycle is incomplete in S. cerevisiae, and the continuous glycan modifications constitute a trajectory towards ER exit or degradation. After terminal glucose residues are removed, the mannosidase Mns1p removes one of the nine mannoses. Mns1p finalizes the N-glycan structure that is required for ER exit (Figure 4A). However, if the glycoprotein remains unfolded, Pdi1p works together with the mannosidase Htm1p to recognize it as an ERAD-substrate (Gauss et al., 2011). Htm1p removes a mannose residue and the seven-mannose structure left by Htm1p is the recognition signal of Yos9p, which further directs the substrate protein to the ubiquitin-ligase complex (Kim et al., 2005; Carvalho et al., 2006; Gauss et al., 2011). Although modifications of the N-glycan serve as an ERAD-signal, also non-glycosylated proteins undergo quality control. For example, Yos9p/Kar2p and Hrd3p individually and together bind unfolded proteins in a glycan-independent manner and bring the substrate protein in the proximity of the ERAD-L complex (Denic et al., 2006). However, substrate delivery to the Hrd1p complex does not mean definitive degradation, as also Hrd1p can discriminate
between folded and unfolded proteins (Denic et al., 2006; Stein et al., 2014). More studies are required to determine the exact molecular signals and mechanisms of the key players that distinguish mis- and unfolded proteins from folding intermediates.

Active ERAD may lead to significant product losses if the substrate recognition machinery fails to dissect the folding forms of the non-native proteins. Several studies have investigated the modulation of the yeast calnexin gene \textit{CNE1}, the chaperone involved in ER quality control, with the rationale of diminishing ERAD losses (Delic et al., 2014). In accordance to the avoidance-of-ERAD-hypothesis, the overexpression of yeast \textit{CNE1} had no effect on recombinant protein production (Čiplys et al., 2011b), while the disruption of the gene increased secretion of a lysozyme variant and \(\alpha_1\)-antitrypsin by over 2-fold (Arima et al., 1998; Parlati et al., 1995). A recent study investigated systematically the effects of removing ERAD components on IgG secretion in \textit{S. cerevisiae}: single gene deletions displayed only minor effects in secreted antibody titers, except for \(\text{A}1\text{yos9}\), which diminished titers significantly (de Ruijter and Frey, 2015). ERAD protects the ER from accumulation of unfolded proteins, and retaining homeostasis effectively during antibody production seems to be necessary for high titers.

Conformational stress in the ER caused by folding intermediates, misfolded, and unfolded proteins can rise to such an extent that a stress response, UPR, is initiated. The purpose of UPR is effectively to lower the concentration of secretory polypeptides in the ER to prevent the formation of toxic aggregates. This is achieved by regulating the processes involved in protein influx and efflux. Protein influx to the ER is essentially reduced by translationally repressing ribosome biogenesis and assembly (Payne et al., 2008b). In contrast to downregulating protein synthesis, efflux processes are upregulated: expression of central ERAD components and proteins involved in vesicle trafficking and ER export is enhanced by the UPR (Travers et al., 2000; Payne et al., 2008b; Higashio and Kohno, 2002). In addition, the UPR causes the proliferation of ER membrane by upregulating lipid metabolism (Schuck et al., 2009; Travers et al., 2000). ER expansion alone has been shown to reduce conformational stress as the concentration of proteins decreased in the ER (Schuck et al., 2009). Central to the UPR, the ER folding machinery and other enzymes involved in protein processing are upregulated to speed up the exit of the proteins from the secretory system (Travers et al., 2000; Payne et al., 2008b). Strikingly, the UPR affects the regulation of over 5\% of the yeast genes (Travers et al., 2000), displaying that restoring ER homeostasis is a comprehensive effort for the yeast cell.
In the center of UPR regulation is the transcription factor Hac1p, which is homologous to the mammalian transcription factor Xbp-1 involved in plasma cell differentiation (Yoshida et al., 2001; Shaffer et al., 2004). An unconventional splicing reaction activates both Xbp-1 and Hac1: the intron-containing mRNA of \( HAC1u \) is constantly present in the cytosol, but the active transcription factor is only translated after the inhibitory intron is removed to create the induced mRNA \( HAC1i \) (Yoshida et al., 2001; Patil and Walter, 2001). A single transmembrane protein Ire1p initiates the splicing in yeast, while the mammalian UPR is a much more elaborate signaling system with three of such UPR sensors (Patil and Walter, 2001; Kimata and Kohno, 2011). Interactions between Ire1p, Kar2p, and unfolded proteins determine a critical level of conformational stress, above which the Ire1p activates (Pincus et al., 2010; Kimata and Kohno, 2011). In order to transmit the signal from the ER lumen to the cytosol, Ire1p dimerizes and undergoes transphosphorylation, which causes the conformational changes needed to activate the endonuclease activity and to induce splicing of \( HAC1u \) into \( HAC1i \) further resulting in the downstream cellular changes (Shamu and Walter, 1996; Welihinda and Kaufman, 1996; Credle et al., 2005; Sidrauski and Walter, 1997). Obviously, recombinant protein production is very likely to induce an UPR, which is an undesired outcome because of the effects of UPR to downregulate translational capacity (Seresht et al., 2013; Payne et al., 2008b). However, prematurely induced UPR can equip the cell with secretory benefits, as Valkonen et al. demonstrated by constitutively overexpressing Hac1p. Constant expression of active Hac1p doubled the amount of secreted \( \alpha \)-amylase and endogenous invertase in yeast cultivations (Valkonen et al., 2003).

The Golgi apparatus

The Golgi apparatus appears as dispersed and individual cisternae throughout the cytoplasm of \( S. \) \( cerevisiae \) (Figure 5) instead of the stacked and ordered morphology found in mammals and other yeast species (Suda and Nakano, 2012). The dispersed Golgi has been suggested to stem from the correspondingly scattered ERES sites (Rossanese et al., 1999), but the functional relevance for this morphological peculiarity is not well characterized. Strong overexpression of a Golgi-localized GDP-mannose transporter, Vig4p, induces mammalian-like stacked Golgi cisternae to appear in \( S. \) \( cerevisiae \) without any other significant effects in cell physiology (Hashimoto et al., 2002). Whether these stacked structures introduce any secretion advantages in recombinant protein production in yeast still remains to be tested. Even though the Golgi parts are separated, \( S. \) \( cerevisiae \) has the common elements of cis-, medial-, and trans-Golgi compartments through which the protein passes consecutively,
being exposed to different sets of modifying enzymes along the path (Suda and Nakano, 2012).

In addition to glycan modifications (section 1.3.2), proteolytic processing is another major protein maturation step occurring in the Golgi apparatus. In addition to the pre-region that is removed at ER entry, some proteins contain an N-terminal leader peptide and this pro-region is removed in the Golgi. Although several protein-specific autocatalytic and endoproteinase-mediated proteolytic reactions have been associated to the Golgi apparatus, mainly the pro-region removal is of major interest in recombinant protein production. Three major proteases contribute to this function in the late-Golgi apparatus: Kex1p, Kex2p and Ste13p. The carboxypeptidase Kex1p and endoprotease Kex2p are serine proteases whereas Ste13p is a dipeptidyl aminopeptidase. These proteases have different substrate specificities, which enable them to recognize a wide range of substrates (Jones, 1991). Although many pharmaceutical proteins are successfully expressed with their native leader peptides (Nielsen, 2013), efficiency of the yeast Golgi to complete the removal of human peptide sequences is varying. Often human leader peptides are exchanged to synthetic or yeast derived sequences, a common choice being the α-factor leader processed by Kex2p (Hou et al., 2012b; Robinson et al., 1994; Rakestraw et al., 2009). In addition to using a different leader peptide, sequence optimization can improve processing efficiency, and further, increase the secreted antibody yields in *S. cerevisiae* (Parekh et al., 1995; Rakestraw et al., 2009). Overexpression of Kex2p has also been studied to improve proteolytic processing, but in the case of bovine pancreatic trypsin inhibitor with a synthetic leader peptide, the overexpression showed only minor effects in protein quality and quantity, possibly because the wild-type activity of the protease is sufficient even in recombinant protein expression (Parekh et al., 1995; Zhang et al., 2001).

Although the ER is considered to include the major quality control point in the secretory system, protein quality assurance continues in the Golgi apparatus. *S. cerevisiae* has molecular mechanisms to target misfolded proteins for degradation in the vacuole; for example, the sorting receptor Vsp10p recognizes unfolded sites on proteins in the Golgi lumen before redirection of the polypeptide to the vacuole (Hong et al., 1996). Vacuolar targeting can be also associated to Kex2p processing, as shown in the case of an insulin-containing fusion protein (Zhang et al., 2001). In addition, the absence of correct lipid associations induced an integral membrane protein Fur4p to be rerouted to the vacuole instead of its target destination, the cell surface (Pineau et al., 2008), highlighting the versatile molecular agents in Golgi-associated quality control. Although N-glycan moieties
function as folding indicators in the ER, the roles of the N-glycans in quality and targeting are less defined in Golgi, despite the fact that most glycan modifications occur in the Golgi cisternae. Certain \( \text{cis} \)-Golgi mannosyltransferases can act on unfolded proteins, which might contribute as a signal to protein retrieval from the Golgi back to the ERAD machinery (Xu and Ng, 2015). Otherwise, N-glycan directed quality control in the Golgi is hardly discussed, despite the obvious importance of the matter in the field of glycoengineering.

**Vesicle trafficking, and the vacuole as the alternative address**

In the \( \text{trans} \)-Golgi network, the synthesized recombinant protein is only a step away from the destination, the plasma membrane. This last vesicular transport step named as exocytosis consists of packaging the soluble protein in heavy density secretory vesicles which later fuse to the plasma membrane releasing the contents to the extracellular space (Harsay and Bretscher, 1995). Although the density and cargo of the vesicles vary, the vesicle formation processes share several common steps in each trafficking step. First, the vesicle buds off from the originating organelle membrane, aided by membrane-curving proteins and the coat proteins assembling on the bud. Coated vesicles travel to the destination membrane, which they recognize via SNARE (N-ethylmaleimide sensitive factor attachment protein receptor) protein pairing. As the vesicle and target membrane SNAREs (v- and t-SNAREs, respectively) twist around each other, membrane fusion occurs, after which the protein complexes are dismantled. In addition to the coat proteins and SNAREs, each vesicle trafficking step in yeast includes an SM (Sec1/Munc18-like) protein facilitating the interaction between the SNAREs and Rab GTPases, which mediate vesicle docking. The identity of each participating protein is specific to the various vesicle transports steps, of which ER-to-Golgi and Golgi-to-plasma membrane routes are the major lines of trafficking in the secretory pathway. The vesicle transport steps can also be distinguished by which coat proteins they use: coat protein complex COPI works in retrograde transport from Golgi and COPII in the ER-to-Golgi step, while clathrin-coated vesicles travel at the plasma membrane (Delic et al., 2013).

With certain recombinant proteins, vesicle trafficking steps form limiting processes for secretion efficiency, as overexpression of vesicle trafficking components has suggested. Factors acting in the exocytosis process, such as the plasma membrane t-SNAREs Sso1p and Sso2p responsible for receiving the Golgi-derived secretory vesicles, and the related SM protein Sec1p, are notable candidates tested. Upregulation of the t-SNAREs improved secretion of heterologous and endogenous enzymes such as cellobiohydrolase I (Xu et al.,
laccase (Larsson et al., 2001), α-amylase and invertase (Ruohonen et al., 1997), while increasing the abundance of Sec1p enhanced the secretion of a variable category of proteins (Hou et al., 2012a). The ER-to-Golgi transport step has also been targeted, and overexpression of the different components of these COPII-coated vesicles successfully increased secretory titers of heterologous enzymes (Van Zyl et al., 2016; Hou et al., 2012a). The limiting trafficking step may reflect the exact compartment to which the protein accumulates, as it can be either the ER or the Golgi depending on the protein (Hou et al., 2012b).

Interestingly, vesicle trafficking targets are largely composed of factors forming and managing the vesicle but not of agents recruiting the cargo inside the vesicle. Secretion has been considered as the default pathway for ER-targeted proteins, but sorting steps are by no means thoroughly characterized. Sorting has a negative tone in recombinant protein production since not all produced proteins reach the plasma membrane. Often a significant portion ends up in the vacuole for degradation and thus, deletions of vacuolar protein sorting (Vps) genes has been a common approach to suppress target protein losses. As Vps10p was characterized as a sorting receptor in the Golgi (Hong et al., 1996), the deletion of this gene is often undertaken (Xu et al., 2014; Zhang et al., 2001; Rakestraw et al., 2009). Removal of several other Vps-proteins, for example Vps13p and Vps8, resulted in increased titers of proteins ranging from insulin-containing protein to scFv (Zhang et al., 2001; Rakestraw et al., 2009). Vps-gene deletions have been associated with enhanced secretion repeatedly in randomized approaches (Zhang et al., 2001; de Ruijter et al., 2017), reflecting the measurable influence of vacuolar sorting to target protein yields. Although the process is sometimes referred to as "mis-sorting" (e.g. in Idiris et al., 2010), targeting to the vacuole can also be due to quality control, as described above.

The route to the vacuole goes through the endosomal system, which is composed of a dynamic assembly of endocytosed and Golgi-arrived vesicles (Pelham, 2002). The importance of this organelle is not well defined or addressed in the field of recombinant protein production, despite the evidence that assembly of high-density exocytic vesicles occurs in endosomes (Harsay and Schekman, 2002). Recombinant proteins not only arrive to endosomes from the Golgi but also through endocytosis. Tyo et al. measured the consumption of protein product in S. cerevisiae cultures, and showed that protein uptake and subsequent catabolism in the vacuole was in the range of 1 g/l/day (Tyo et al., 2014). Vacuolar degradation of recombinant proteins is often addressed through deletion of the major proteases responsible for product digestion, corresponding to genes PRB1 and PEP4.
Deletion of the protease genes is undertaken to stabilize intracellular protein (Hong et al., 1996; Zhang et al., 2001; Rakestraw et al., 2009), and further, protease-deficient strains can exhibit significantly higher product titers (Tomimoto et al., 2013). Product stability is also a matter of process design, which is another approach to improve titers.

### 1.3.4 Other engineering approaches to yeast antibody production

So far, the major genetic targets to improve *S. cerevisiae* as a protein producer were summarized, but other engineering efforts, such as process and expression system design, are equally important. The success of CHO cells is largely attributed to optimization of media composition and growth conditions (Wurm, 2004), which can similarly affect cell growth and protein titers in yeast systems. Although *S. cerevisiae* fermentation processes are fairly established and robust (Nevoigt, 2008), there is a lot of room for manipulation of the feed rate and composition, oxygen, heat and mass transfer efficiencies in different fermentation configurations and scales (Hensing et al., 1995). Cellular physiology and production capacity also change during the course of a production process. Productivity and yield were shown to be growth rate dependent in *S. cerevisiae*, where the produced target protein determined the nature of the exact relationship between the growth rate and product yield (Liu et al., 2013). Cultivation time and temperature have a great influence on the final titer, and in fact, lowering the temperature from 30 ºC to 20 ºC resulted in increased titers of scFv antibodies, although growth rate was reduced (Hackel et al., 2006; Huang et al., 2008).

In addition to process parameters, influence of media composition has been investigated, especially in the perspective of carbon sources. Interestingly, van de Laar et al. reported that specific production of an antibody fragment could be increased 5-fold when ethanol was used as a carbon source instead of glucose (van de Laar et al., 2006). Additionally, carbon source concentration and when applicable, the concentration of an inducer for protein production, have a significant influence on protein product quantity and quality (Kim et al., 2014). Especially in the case of recombinant protein production in *S. cerevisiae*, a media additive to consider is bovine serum albumin (BSA) or other proteinous molecules. Tyo et al. showed that the addition of BSA protected the product protein from uptake and degradation by the yeast cells, highly contributing to the stability of the product in culture medium (Tyo et al., 2014). However, the price of components can impose limitations to media optimization in large scale applications (Hensing et al., 1995).

The molecular configuration of the expression cassette also affects secretory titers. Gene copy number, use of plasmids, or integrative expression cassettes influence protein
expression, also in the context of expression stability (Hensing et al., 1995). Shusta et al. showed that an integrative vector increased production of a scFv 6- to 10-fold compared to a low copy plasmid-based expression (Shusta et al., 1998). However, the authors used two different promoters in their constructs, which might have biased the results, as different promoters result in different expression levels. The promoter strength does not correlate linearly with protein yields, especially with large and complex protein products, which has led to the characterization of a wide range of different promoters (Liu et al., 2012; Partow et al., 2010). In addition to genetic elements such as promoters and terminators (Curran et al., 2015), the target protein and its coding sequence can be subjected to molecular engineering to improve heterologous protein production. These approaches include codon optimization to exclude rare codons and to reduce mRNA secondary structures, in addition to the aforementioned secretion leader engineering. Through codon optimization, Kim et al. were able to increase the solubility and recovery of a viral capsid protein produced in S. cerevisiae (Kim et al., 2013). Secretion leader engineering was discussed in the context of Golgi processing efficiency, but the benefits can also arise from improved protein processing in earlier steps in the secretory pathway or increased protein solubility and stability. These multifaceted effects partly explain the success of synthetic leaders and evolutionary engineering approaches in increasing titers through leader optimization, as naturally occurring leader sequences have evolved for a targeting function and not for high-efficiency protein secretion (Liu et al., 2012; Rakestraw et al., 2009).

Considering the number of process parameters, available choices of genetic elements and cellular engineering targets, it is not surprising that very few have even attempted a holistic approach to optimize yeast recombinant protein production. Finnis et al. presented a combination of engineering target protein, cellular factors and the process to develop a commercial platform for transferrin production. Although the transferrin titers multiplied, their approach was far from comprehensive (Finnis et al., 2010). The complication is, as Liu et al. formulated: “There is no one ultimate method that could work equally well for production of all proteins” (Liu et al., 2012). An integrative approach for holistic optimization for each target protein is simply not feasible. Thus, the framework to predict the necessary components has been envisaged, as presented in the next section.

1.4 Synthetic biology tools for optimizing yeast cell factories

Metabolic engineering and microbial biotechnology are promising industrial fields to produce various high-value products with improved efficiency and environmentally friendly
processes. These approaches have developed into the idea of synthetic biology, which rationalizes all aspects of the development of cell factories. Yeast cell factories have established significance in being the biotechnological alternative for making a variety of valuable chemicals and products, as described in section 1.3.1. Although random approaches in evolutionary engineering have been valuable tools for creating industrially relevant yeast strains (Nevoigt, 2008), these strategies do not fit the deductive design aims of synthetic biology and thus will not be discussed here. Synthetic biology is considered as employing engineering principles in biology, aiming for completely rational and predictable (re)design of biological systems. When the ideas of standardization, abstraction, and decoupling (as described in Endy, 2005) from synthetic biology are combined with metabolic engineering, metabolic pathways are completely reconstructed for medical and industrial uses (Purnick and Weiss, 2009). Although biological complexity is still hindering predictable design, the trends in biotechnology continue to shift towards targeted engineering with the help of accumulating system-level information from various model organisms. The idea of synthetic biology comes directly from the discovery of recombinant DNA technologies (Endy, 2005), thus the most important tools for genome engineering of yeast and constructing recombinant genetic elements will be described next. In addition, this section discusses the contribution of systems biology to synthetic biology engineering approaches.

### 1.4.1 Genetic engineering methods in yeast

The popularity of studying and exploiting *S. cerevisiae* has conceived a large variety of methods for efficient and easy genetic manipulation of the organism. One of the most popular transformation methods, the lithium-acetate method, can introduce any extracellular and foreign DNA into the yeast cells with high efficiency, yielding up to $10^6$ transformants per μg of DNA (Gietz et al., 1992). A convenient way to introduce a new gene to *S. cerevisiae* is to clone it to a yeast shuttle vector, which replicates autonomously inside the host and can be amplified in *E. coli* before transformation. These shuttle vectors are derived from naturally occurring yeast plasmids and are nowadays commonly available with several selection markers, promoters, both in low- and high-copy variants (Mumberg et al., 1995, EUROSCARF-collections).

Another characteristic enabling the genome of *S. cerevisiae* to be easily manipulated is its intrinsic capability to preform homologous recombination with high efficiency. Homologous recombination between two DNA molecules occurs in areas of similar enough sequences that find their complements in the other molecule (depicted in Figure 6). In the case of two
complementary areas for homologous recombination, the area between the sequences is efficiently interchanged between the two molecules (Figure 6). The introduced sequences can be anything from a deletion cassette to additional genes; often the integrative cassette in gene insertions is constructed into a plasmid with the required elements for integration and selection (Taxis and Knop, 2006). Successful integration provides the yeast clone a selection competence, such as antibiotic resistance, so true recombinants can be easily distinguished from the rest of the cell population. To overcome the limitation of selection marker availability, also methods for repeated deletions have been developed. Site-specific recombination is mediated by a defined DNA sequence, and a recombinase that recognizes that sequence. When the integration cassette (resistance marker) is flanked by such sequences, for example loxP-sites (Figure 6), upon expression of the recombinase (Cre in Figure 6), the flanked region is looped out and target genomic area is only left with a single loxP-sequence after the deletion (Figure 6). Thus the same selection marker can be used for another round of deletion (Güldener et al., 1996, 2002) In *S. cerevisiae*, efficient homologous recombination can be achieved already with homologous regions of 30-45 base pairs, so creation of a deletion cassette can be realized fast with a simple PCR-reaction (Güldener et al., 1996; Baudin et al., 1993).

![Figure 6 Homologous recombination and related engineering methods. Grey boxes indicate corresponding sequences which enable the interchange between DNA strands. After the marker is exchanged with the target locus, the marker can be looped out by using the loxP-sites (light grey) and Cre recombinase (brown). Inducing a double-strand break to the target site increases the efficiency of the recombination event. The targeted cut can be achieved with ZFN, TALEN or CRISPR-Cas9 methods (see text).](image)

The efficiency of homologous recombination can be greatly enhanced by inducing a double-strand break to the target locus. Several enzymes are available for creating the double-strand breaks, but the problem lies in the reliable recognition methods of the target sequence. For this purpose, there are two major protein-based systems: zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) (Figure 6). ZFNs and TALENs are designed from modules known to bind certain nucleotides and this part is fused to the nuclease domain. Although ZFNs and TALENs work in pairs and can identify the sequence
on both sides of the cleavage site, there is still significant unpredictability remaining in their construction. Recently, a system enabling base-pair guided DNA cleavage was discovered from studies of bacterial adaptive immunity, and it is called the CRISPR-system (for clustered regularly interspersed short palindromic repeats). In CRISPR, a 20 bp guide-RNA complementary to the target site is produced by CRISPR-associated proteins (usually Cas9), which is also responsible of the target DNA cleavage after the guide RNA pairs with the target DNA to form a hybrid (Carroll, 2012). Although these techniques are not a prerequisite for genome editing of *S. cerevisiae*, for example the CRISPR-system is used to speed up the process, for introducing several modifications simultaneously and possibly with a single marker, even engineering complete metabolic pathways of up to dozens of kilobases in a single step (Horwitz et al., 2015; Mans et al., 2015).

Several hundreds of kilobases of DNA can be introduced to *S. cerevisiae* cells simultaneously also in the form of yeast artificial chromosomes (YAC), which greatly facilitate metabolic engineering efforts (Naesby et al., 2009). Along with the added exogenous DNA, YACs have the chromosomal elements, autonomously replicating sequences, centromeres, and telomeres, so they are maintained like natural chromosomes inside the yeast cell (Burke et al., 1987). As complete chromosomes can be constructed and inserted into the yeast cell, it was also proven that native chromosomes can be replaced with artificially created chromosomes (Annaluru et al., 2014). In the era of synthetic biology, there is an initiative to take this artificial design even further and create a yeast 2.0 system with completely man-interfered genome (Nawy, 2011, www.syntheticyeast.org/). The Synthetic Yeast Genome Project or Sc2.0 has already demonstrated that experimental flexibility can be introduced to a native-like genomic sequences without a significant loss of fitness (Dymond et al., 2011). The Sc2.0 yeast would have a completely redesigned genome, with a reduced and refined sequence to serve as an engineering and experimenting platform for a new generation of synthetic biologists and researchers.

1.4.2 High-throughput cloning

Combining different genetic elements for effective metabolic engineering requires their assembly to larger functional units, such as the plasmids and YACs mentioned above. DNA technologies have developed a long way from the initial discovery of restriction enzymes into a variety of *in vitro* and *in vivo* cloning techniques. However, traditional restriction enzyme cloning or *in vivo* cloning techniques do not meet the criteria for a high-throughput cloning method of sequence-independence, reliability, efficiency, and speed (Marsischky
There are three main strategies of high-throughput cloning methods, from which the most common one is based on site-specific recombination. Univector and Creator™ cloning systems (Clontech) are mediated by the previously described Cre-loxP reaction while Gateway® (Invitrogen™) intermolecular transfers are adapted from sequence-specificity of bacteriophage λ (Liu et al., 1998; Hartley et al., 2000; Marsischky and LaBaer, 2004). Although site-specific recombination is a highly efficient method to create master clones and several expression vectors, the methods have downsides in the high price of the recombinase and their restricted flexibility due to the sequence-requirements of the recombination sites (Marsischky and LaBaer, 2004). To increase the flexibility of expression engineering, modularity has been introduced in the MultiSite Gateway™ system (Invitrogen™), which is also available for *S. cerevisiae* (Nagels Durand et al., 2012). The wide adaption of Gateway® cloning might have spurred from its early commercialization and its usage in early large-scale cloning projects (Hartley et al., 2000; Marsischky and LaBaer, 2004).

A second strategy of high-throughput cloning relies on type II restriction enzymes, familiarly named as Golden Gate assembly (Engler et al., 2008). These enzymes cut outside their recognition sequence, enabling the recognition site to be eliminated from the final construct. This allows simultaneous restriction and ligation and by proper selection of nucleotides in the overhangs, also the assembly and/or shuffling of several fragments in one-pot reaction (Engler et al., 2008, 2009). Although the problem of internal recognition sites of the used enzyme in the insert can be avoided by additional primer design (Engler et al., 2008), Golden Gate assembly is not completely sequence-independent and the method is unsuitable for DNA fragments with unknown sequence. Nevertheless, high efficiency and the possibility of high level of modularization make Golden Gate assembly an attractive tool for synthetic biology applications (Weber et al., 2011).

The third group of high-throughput cloning techniques comprises of various protocols, which rely on homologous annealing of complementary DNA strands. The overall principle is to create single-stranded DNA segments in the end of the molecules, anneal these to homologous sequences in another DNA molecule *in vitro* and transform the reaction mixture to *E. coli* to yield recombinant plasmids. The method of single-strand overhang creation, and the extent to which the plasmid is assembled *in vitro* varies between protocols. Gibson assembly represents the enzymatic *in vitro* completion of a circular DNA molecule (Gibson et al., 2009; Festa et al., 2013). The group of methods which do not complete the plasmid *in vitro* are generally referred to as ligation-independent cloning (LIC), introduced in its first
form over 25 years ago (Aslanidis and de Jong, 1990). Since then, several variations have been described that utilize either exonuclease activity (Li and Elledge, 2007; Jeong et al., 2012; Thieme et al., 2011) or PCR (Tillett and Neilan, 1999; Klock et al., 2008) to create single stranded ends. Also included under the “LIC” term and utilizing homologous stretches of DNA are PCR-mediated strategies such as overlap extension cloning (Bryksin and Matsumura, 2011; Stevenson et al., 2013). The advantages of all these techniques is that they use common, inexpensive enzymes, they are compatible with all vector systems, are highly interconvertible with the same materials and most of them are completely sequence independent (Fasta et al., 2013; Stevenson et al., 2013). However, the efficiency of ligation-independent cloning techniques may be susceptible to insert size and not reliable enough for multiple fragment assembly (Stevenson et al., 2013; de Jong et al., 2006).

These cloning methods are concentrated on creating a variety of expression plasmids, but efforts of synthetic biology are aiming at constructing larger elements of genomes, from artificial chromosomes to completely redefined genomes. Only a few of the above described in vitro techniques have demonstrated the flexibility to be scaled up to dozens of kilobases (Weber et al., 2011; Gibson et al., 2009). Thus, high-throughput cloning is still largely confined in the plasmid-mediated engineering of biological machineries.

1.4.3 Integrating systems-level information to synthetic biology

One question of synthetic biology is, whether living cells are actually too complex to engineer reliably (Endy, 2005). Synthetic biology relies on comprehensive characterization of all the parts in the systems in order to create accurate models and predict interactions between modified parts (Purnick and Weiss, 2009). Several technologies for omics-level analytics have greatly expanded our knowledge on biological systems in a short time period. The discipline dubbed systems biology is the understanding of biological information at a global level which is essential for synthetic biology. Together with metabolic engineering, systems biology is combined in an integrated approach called systems metabolic engineering, which entails both computational and experimental approaches for developing globally optimized cell factories. Systems metabolic engineering has several success stories in the field of small chemicals production, and it is paving the way for synthetic biology approaches to create tailor-made industrial strains (Wittmann and Lee, 2012; Lee and Kim, 2015).

Oomics-technologies refer to reliable and global quantitative analysis of the selected cell component. They can be considered to have arisen from genomics, the deciphering of full
genomic sequences of the organisms. After genome sequences came available, transcriptomics gained popularity and was followed by several other technologies enabling proteomics, metabolomics, interactomics, fluxomics among others, each named after the cell component they measure. In the case of recombinant protein production, understanding cellular stresses and metabolic bottlenecks through the use of transcriptomics and metabolomics has been applied, but mainly as an analytical effort (Dietmair et al., 2012; Seresht et al., 2013; Tyo et al., 2012). Indeed, instead of gathering data, the main challenge today is how to integrate the data from different omics-technologies for engineering uses and how this data can be used to constructs a comprehensive computational model of the cellular processes with predictive power. As a model organism, \textit{S. cerevisiae} has been under investigation for several of such integrative approaches (Joyce and Palsson, 2006).

Recombinant protein production and secretion in eukaryotes is a tedious effort metabolically, but it is also complicated by compartmentalization and requirement for numerous transport processes across cellular membranes. Although comprehensive genome-scale models for such a complicated phenomenon are not yet feasible, Feizi et al. have provided a genome-scale network model of the yeast secretory machinery describing the transport steps quantitatively (Feizi et al., 2013). In the lack of more sophisticated models, engineering strategies of recombinant protein production might benefit from empirical modeling, as demonstrated with mathematical models of monoclonal antibody production process in CHO-cells (McLeod et al., 2011). In the view of synthetic biology, the ultimate goal is to develop a full-scale genomic model with kinetic information about every component of the cell, which would allow accurate predictions for any type of modification or target process, even recombinant protein secretion. The amount of information and data integration required to obtain such a model is enormous but as data and efforts are accumulating in the case of \textit{S. cerevisiae}, a comprehensive \textit{in silico} model of the organism might become available in the near future.
2. Aims of the study

With the advances in metabolic engineering and synthetic biology, the yeast *S. cerevisiae* is an intriguing platform to be modified for the production of a range of products, including full-length antibodies. As plasma cells are superior antibody-producers, their development provides a blueprint for antibody factory design. In this doctoral work, I studied whether plasma cells could be used as a model to improve *S. cerevisiae* as an IgG factory, especially in the aspect of protein yields. I aimed to identify key modifications in plasma cell development and investigate their transferability to improve IgG production in *S. cerevisiae*. I set out to test several modification that could improve IgG folding and secretion, including

1. Increasing the size of yeast ER to mimic the plasma cell morphology. In addition, the effect of the shape of ER was investigated;
2. Co-overexpressing selected mammalian folding factors to determine modules that improve IgG folding and secretion;
3. Mining transcriptomics data published from plasma cell differentiation to find unexpected gene targets from yeast.

A central theme in this thesis was to utilize methods of synthetic biology, which led us to present a new, streamlined cloning method. Accordingly, I implemented modularity and rational design in the approaches of cellular engineering.
3. Materials and methods

3.1 Yeast strains

Table 1 Saccharomyces cerevisiae strains used in this study.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype or genetic modifications</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>W303a</td>
<td>MATa leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15</td>
<td>Parental strain</td>
</tr>
<tr>
<td>YEK018</td>
<td>his3Δ::natNT2::(HC+LC of anti CD-20 antibody)</td>
<td>Publication II</td>
</tr>
<tr>
<td>YEK019</td>
<td>his3Δ::natNT2::(HC+LC of anti CD-20 antibody) opi1Δ::loxP</td>
<td>Publication II</td>
</tr>
<tr>
<td>YEK052</td>
<td>his3Δ::natNT2::(HC+LC of anti CD-20 antibody) rtn1Δ::loxP tpr1Δ::TRP1::NdegY-GFP</td>
<td>This thesis</td>
</tr>
<tr>
<td>YJR129</td>
<td>his3Δ::natNT2::(HC+LC of anti CD-20 antibody)</td>
<td>Publication II, This thesis</td>
</tr>
<tr>
<td>YEK063</td>
<td>his3Δ::natNT2::(HC+LC of anti CD-20 antibody) opi1Δ::loxP tpr1Δ::TRP1::NdegY-GFP</td>
<td>Publication II, This thesis</td>
</tr>
<tr>
<td>YEK055</td>
<td>his3Δ::natNT2::(HC+LC of anti CD-20 antibody) rtn1Δ::loxP tpr1Δ::TRP1::NdegY-GFP</td>
<td>This thesis</td>
</tr>
<tr>
<td>YMR24</td>
<td>tpr1Δ::TRP1::NdegY-GFP</td>
<td>This thesis</td>
</tr>
<tr>
<td>YEK070</td>
<td>his3Δ::natNT2::(HC+LC of anti CD-20 antibody) ate1Δ::kanMX</td>
<td>Publication IV</td>
</tr>
<tr>
<td>YEK071</td>
<td>his3Δ::natNT2::(HC+LC of anti CD-20 antibody) dph2Δ::kanMX</td>
<td>Publication IV</td>
</tr>
<tr>
<td>YEK072</td>
<td>his3Δ::natNT2::(HC+LC of anti CD-20 antibody) utr4Δ::kanMX</td>
<td>Publication IV</td>
</tr>
</tbody>
</table>

All strains were based on the W303a laboratory strain detailed in Table 1. All gene knock-outs were created with protocols described by Hegemann and Heick (Hegemann and Heick, 2011). In the strains YJR129, YEK063, and YEK055, together with their control strain YMR24, the NdegY-GFP insertion was a UPR-reporter cassette derived from an integrative plasmid pDEP017 provided by Dr. David Pincus and described in Pincus et al., 2010. The integrative plasmid for antibody expression was based on the pRS303N-plasmid (Taxis and Knop, 2006). Cloning of the light chain and heavy chain cDNAs for the monoclonal anti-CD20 antibody C2B8 under the GAL1-promoter was described previously (de Ruijter and Frey, 2015).
3.2 Materials for testing the cloning strategy

Blue-white screening based on the functioning of the lac-operon was used for quick assessment of cloning efficiency. The LacZ gene was amplified from the plasmid pUC19 with primers that contained homologous overhangs to the target plasmid. Overhangs were 5-40 base pairs and the primers are detailed in Publication I. After purification with NucleoSpin Extract kit (Macherey-Nagel), the amplified fragment was introduced to linearized pKQV4. This destination vector was digested either with Smal (FastDigest, Thermo Fisher Scientific) or Cfr9I (XmaI) (Thermo Fisher Scientific) for 16 h at 37 °C to test the influence of blunt (Smal) versus sticky (XmaI) ends. Purified fragments were eluted to ddH2O, mixed in indicated molar ratios in a final volume of 10 μl and reaction mixtures were transformed to Lac-operon deficient XL1-Blue Escherichia coli strain (Stratagene). The LacZ gene was inserted to pKQV4 under the tac-promoter. Positive recombinants were successfully expressing LacZ, and thus turned blue on the used growth medium, LB Agar plates supplemented with ampicillin, isopropyl b-D-1-thiogalactopyranoside, and 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside (Amp IPTG/X-GAL, E. coli FastMedia, Fermentas), while colonies without target plasmid remained white. The number of blue colonies and the ratio of blue to white colonies were used to assess the effects of the length of homologous regions, insert to vector ratio, and the incubation time on cloning efficiency. The testing procedures are described in detail in Publication I. The cloning method was named exonuclease and ligase-independent cloning (ELIC) and was used hereafter for cloning unless otherwise mentioned.

3.3 Plasmids

The yeast shuttle vectors used were based on the pRS41X-plasmid series: plasmids with the marker URA3 were derived from the pRS416 plasmid backbone and LEU2 containing plasmids were pRS415-based (Mumberg et al., 1995). Cloning of the UPR-responsive promoters P_{KAR2} and P_{PDI1} is detailed in Publication II.

3.3.1 Plasmid library of mammalian folding factors

Table 2 lists all the plasmids comprising the mammalian folding factor expression library. Publication III includes the details of the cloning procedures for the library member plasmids. Origin of the cDNAs is as follows: human ERdj3, mouse GRP94, and human GRP170 were obtained from Dr. Linda M. Hendershot, human ERp57 was from Dr. Kari Kivirikko, human PDI was from Dr. Neil Bulleid, Dr. Lloyd Ruddock provided coding
regions for human FKBP2, CypB, P5 L20-L440 and ERp72, and human BiP was obtained from Source Bioscience (IMAGE clone 5020098).

Table 2 The plasmid library consisting of ten mammalian folding factors.

<table>
<thead>
<tr>
<th>Name</th>
<th>Promoter</th>
<th>Marker</th>
<th>Insert</th>
<th>Name</th>
<th>Promoter</th>
<th>Marker</th>
<th>Insert</th>
</tr>
</thead>
<tbody>
<tr>
<td>pEK7</td>
<td>none</td>
<td>LEU2</td>
<td>none (empty marker plasmid)</td>
<td>pEK17</td>
<td>none</td>
<td>URA3</td>
<td>none (empty marker plasmid)</td>
</tr>
<tr>
<td>pAF9</td>
<td>GAL1</td>
<td>LEU2</td>
<td>BiP</td>
<td>pEK16A</td>
<td>GAL1</td>
<td>URA3</td>
<td>GRP170</td>
</tr>
<tr>
<td>pAF7</td>
<td>GPD</td>
<td>LEU2</td>
<td>BiP</td>
<td>pEK16B</td>
<td>GPD</td>
<td>URA3</td>
<td>GRP170</td>
</tr>
<tr>
<td>pAF8</td>
<td>TEF</td>
<td>LEU2</td>
<td>BiP</td>
<td>pEK16C</td>
<td>TEF</td>
<td>URA3</td>
<td>GRP170</td>
</tr>
<tr>
<td>pJR027</td>
<td>PDI1</td>
<td>LEU2</td>
<td>BiP</td>
<td>pEK16D</td>
<td>PDI1</td>
<td>URA3</td>
<td>GRP170</td>
</tr>
<tr>
<td>pJR03I</td>
<td>KAR2</td>
<td>LEU2</td>
<td>BiP</td>
<td>pEK16E</td>
<td>KAR2</td>
<td>URA3</td>
<td>GRP170</td>
</tr>
<tr>
<td>pEK8A</td>
<td>GAL1</td>
<td>URA3</td>
<td>ERdj3</td>
<td>pEK20A</td>
<td>GAL1</td>
<td>LEU2</td>
<td>P5</td>
</tr>
<tr>
<td>pEK8B</td>
<td>GPD</td>
<td>URA3</td>
<td>ERdj3</td>
<td>pEK20B</td>
<td>GPD</td>
<td>LEU2</td>
<td>P5</td>
</tr>
<tr>
<td>pEK8C</td>
<td>TEF</td>
<td>URA3</td>
<td>ERdj3</td>
<td>pEK20C</td>
<td>TEF</td>
<td>LEU2</td>
<td>P5</td>
</tr>
<tr>
<td>pEK8D</td>
<td>PDI1</td>
<td>URA3</td>
<td>ERdj3</td>
<td>pEK20D</td>
<td>PDI1</td>
<td>LEU2</td>
<td>P5</td>
</tr>
<tr>
<td>pEK8E</td>
<td>KAR2</td>
<td>URA3</td>
<td>ERdj3</td>
<td>pEK20E</td>
<td>KAR2</td>
<td>LEU2</td>
<td>P5</td>
</tr>
<tr>
<td>pEK9A</td>
<td>GAL1</td>
<td>LEU2</td>
<td>PDI</td>
<td>pEK25A</td>
<td>GAL1</td>
<td>LEU2</td>
<td>ERp72</td>
</tr>
<tr>
<td>pEK9B</td>
<td>GPD</td>
<td>LEU2</td>
<td>PDI</td>
<td>pEK25B</td>
<td>GPD</td>
<td>LEU2</td>
<td>ERp72</td>
</tr>
<tr>
<td>pEK9C</td>
<td>TEF</td>
<td>LEU2</td>
<td>PDI</td>
<td>pEK25C</td>
<td>TEF</td>
<td>LEU2</td>
<td>ERp72</td>
</tr>
<tr>
<td>pEK9D</td>
<td>PDI1</td>
<td>LEU2</td>
<td>PDI</td>
<td>pEK25D</td>
<td>PDI1</td>
<td>LEU2</td>
<td>ERp72</td>
</tr>
<tr>
<td>pEK9E</td>
<td>KAR2</td>
<td>LEU2</td>
<td>PDI</td>
<td>pEK25E</td>
<td>KAR2</td>
<td>LEU2</td>
<td>ERp72</td>
</tr>
<tr>
<td>pEK10A</td>
<td>GAL1</td>
<td>LEU2</td>
<td>ERp57</td>
<td>pEK33A</td>
<td>GAL1</td>
<td>LEU2</td>
<td>CypB</td>
</tr>
<tr>
<td>pEK10B</td>
<td>GPD</td>
<td>LEU2</td>
<td>ERp57</td>
<td>pEK33B</td>
<td>GPD</td>
<td>LEU2</td>
<td>CypB</td>
</tr>
<tr>
<td>pEK10C</td>
<td>TEF</td>
<td>LEU2</td>
<td>ERp57</td>
<td>pEK33C</td>
<td>TEF</td>
<td>LEU2</td>
<td>CypB</td>
</tr>
<tr>
<td>pEK10D</td>
<td>PDI1</td>
<td>LEU2</td>
<td>ERp57</td>
<td>pEK33D</td>
<td>PDI1</td>
<td>LEU2</td>
<td>CypB</td>
</tr>
<tr>
<td>pEK10E</td>
<td>KAR2</td>
<td>LEU2</td>
<td>ERp57</td>
<td>pEK33E</td>
<td>KAR2</td>
<td>LEU2</td>
<td>CypB</td>
</tr>
<tr>
<td>pEK11A</td>
<td>GAL1</td>
<td>URA3</td>
<td>GRP94</td>
<td>pEK34A</td>
<td>GAL1</td>
<td>URA3</td>
<td>P5</td>
</tr>
<tr>
<td>pEK11B</td>
<td>GPD</td>
<td>URA3</td>
<td>GRP94</td>
<td>pEK34B</td>
<td>GPD</td>
<td>URA3</td>
<td>P5</td>
</tr>
<tr>
<td>pEK11C</td>
<td>TEF</td>
<td>URA3</td>
<td>GRP94</td>
<td>pEK34C</td>
<td>TEF</td>
<td>URA3</td>
<td>P5</td>
</tr>
<tr>
<td>pEK11D</td>
<td>PDI1</td>
<td>URA3</td>
<td>GRP94</td>
<td>pEK34D</td>
<td>PDI1</td>
<td>URA3</td>
<td>P5</td>
</tr>
<tr>
<td>pEK11E</td>
<td>KAR2</td>
<td>URA3</td>
<td>GRP94</td>
<td>pEK34E</td>
<td>KAR2</td>
<td>URA3</td>
<td>P5</td>
</tr>
<tr>
<td>pEK14A</td>
<td>GAL1</td>
<td>URA3</td>
<td>FKBP2</td>
<td>pEK35A</td>
<td>GAL1</td>
<td>LEU2</td>
<td>GRP170</td>
</tr>
<tr>
<td>pEK14B</td>
<td>GPD</td>
<td>URA3</td>
<td>FKBP2</td>
<td>pEK35B</td>
<td>GPD</td>
<td>LEU2</td>
<td>GRP170</td>
</tr>
<tr>
<td>pEK14C</td>
<td>TEF</td>
<td>URA3</td>
<td>FKBP2</td>
<td>pEK35C</td>
<td>TEF</td>
<td>LEU2</td>
<td>GRP170</td>
</tr>
<tr>
<td>pEK14D</td>
<td>PDI1</td>
<td>URA3</td>
<td>FKBP2</td>
<td>pEK35D</td>
<td>PDI1</td>
<td>LEU2</td>
<td>GRP170</td>
</tr>
<tr>
<td>pEK14E</td>
<td>KAR2</td>
<td>URA3</td>
<td>FKBP2</td>
<td>pEK35E</td>
<td>KAR2</td>
<td>LEU2</td>
<td>GRP170</td>
</tr>
<tr>
<td>pEK15A</td>
<td>GAL1</td>
<td>URA3</td>
<td>CypB</td>
<td>pEK15D</td>
<td>PDI1</td>
<td>URA3</td>
<td>CypB</td>
</tr>
<tr>
<td>pEK15B</td>
<td>GPD</td>
<td>URA3</td>
<td>CypB</td>
<td>pEK15E</td>
<td>KAR2</td>
<td>URA3</td>
<td>CypB</td>
</tr>
<tr>
<td>pEK15C</td>
<td>TEF</td>
<td>URA3</td>
<td>CypB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.3.2 Other plasmids

_HAC1_ overexpression plasmid pMS109 and the corresponding control plasmid pKK1 were provided by Dr. Mari Valkonen and they are described in Valkonen et al., 2003. The genes _RTN1, HUT1, IRE1, GOT1, and GSH2_ were amplified from yeast genomic DNA with primers containing 5’ flanks with adaptor sites. These adaptor sites were utilized to create homologous sites to the vector in a second PCR reaction, which enabled ELIC cloning under different promoters as described in Publication III. _RTN1_ with endogenous promoter was amplified with forward primer 5’-AGGGAACAAAAGCTGGAGCTCTACTTACCAGGGTTTGG3’, which recognized the region 400 bp upstream of the gene start, as used by Schuck et al. (Schuck et al., 2009). The forward primer was combined with the same reverse primer as for the _RTN1_ coding region, and the longer _RTN1_ gene fragment was amplified from yeast genomic DNA. _RTN1_ with the endogenous promoter was cloned to SacI/XhoI sites of pRS416-plasmid backbone. Cloning of overexpression plasmid for _HUT1, IRE1, GOT1_, and _GSH2_ is described in Publication IV.

For verifying the expression of folding factors, the sequence for the FLAG-tag (5’-GACTACAAGGACGATGACGACAAG-3’, resulting in oktapeptide DYKDDDDK) was inserted right before the HDEL-sequence of selected folding factor plasmids. A combination of fusion-PCR and ELIC was used to seamlessly insert the sequence inside the cDNA region. Table 3 provides a list of these additional plasmids.

_Table 3 Summary of additional plasmids created and used in this thesis._

<table>
<thead>
<tr>
<th>Name</th>
<th>Promoter</th>
<th>Marker</th>
<th>Insert</th>
<th>Name</th>
<th>Promoter</th>
<th>Marker</th>
<th>Insert</th>
</tr>
</thead>
<tbody>
<tr>
<td>pEK36A</td>
<td>GAL1</td>
<td>URA3</td>
<td><em>RTN1</em></td>
<td>pEK47</td>
<td>GAL1</td>
<td>URA3</td>
<td>GRP94 with FLAG-tag</td>
</tr>
<tr>
<td>pEK36C</td>
<td>TEF</td>
<td>URA3</td>
<td><em>RTN1</em></td>
<td>pEK48</td>
<td>GAL1</td>
<td>LEU2</td>
<td>ERp57 with FLAG-tag</td>
</tr>
<tr>
<td>pEK36F</td>
<td>Endo</td>
<td>URA3</td>
<td><em>RTN1</em></td>
<td>pEK49</td>
<td>GAL1</td>
<td>LEU2</td>
<td>P5 with FLAG-tag</td>
</tr>
<tr>
<td>pEK44</td>
<td>GAL1</td>
<td>URA3</td>
<td>FKPB2 with FLAG-tag</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pEK45</td>
<td>GAL1</td>
<td>URA3</td>
<td>ERdj3 with FLAG-tag</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pHUT1</td>
<td>TEF</td>
<td>LEU2</td>
<td><em>HUT1</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pIRE1</td>
<td>TEF</td>
<td>LEU2</td>
<td><em>IRE1</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pEK50</td>
<td>GAL1</td>
<td>LEU2</td>
<td>PDI with FLAG-tag</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pEK51</td>
<td>GAL1</td>
<td>URA3</td>
<td>CypB with FLAG-tag</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pEK52</td>
<td>TEF</td>
<td>LEU2</td>
<td><em>GOT1</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pEK53</td>
<td>TEF</td>
<td>LEU2</td>
<td><em>GSH2</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Endogenous promoter, 400 bp upstream of coding sequence.*
3.4 Measuring specific growth rates

For measuring specific growth rates, saturated cultures of the strains were inoculated to YPD or YPGal, which contained 1 % yeast extract, 2 % peptone, and 2 % glucose or galactose, respectively. Culture volume was 100 μl per well and starting OD\textsubscript{600} was adjusted to 0.02. Samples were cultivated on round-bottom microtiter plates using continuous orbital shaking (548 rpm, 2 mm) and OD\textsubscript{600}-values were recorded in 15-minute time intervals with an Eon Microplate Spectrophotometer (BioTek, Winooski, USA). Temperature was kept at 30 °C while the experiment was continued for 63.5 hours. Growth rates were determined with R-package grofit (Kahm et al., 2010) from six culture replicates and values describe the average of μ\textsubscript{max}-values from the best model fits.

3.5 Screening procedures, yeast cultures and automated ELISA

We designed a screening procedure for strains based entirely on 96-well deep well plate cultivations as illustrated schematically in Figure 7. In screening experiments, synthetic drop-out (SD) medium (0.67 % yeast nitrogen base without amino acids supplemented with a selection of amino acids and vitamins, and a carbon source) was used. Details are described in Publication II but briefly, selected plasmids were transformed to strains, which were cultivated to saturation once and diluted 1:4 to fresh production media on cultivation plates. After 5.5 hours of culturing at 30 °C, protein expression was induced and plates were moved to production temperature, which was set at 20 °C, 25 °C or 30 °C. Cells were collected after 24 hours and the final OD\textsubscript{600} was measured with an Eon Microplate Spectrophotometer (BioTek, Winooski, USA). Shake flask cultivations are described in Publication III. In analysis of gene targets derived from transcriptomics exploration, a 24-well deep well plate format was used. As described in Publication IV, pre-cultured strains were adjusted to starting OD\textsubscript{600} of 1.0 with inductive SD media containing 2 % galactose as the sole carbon source. After 16 hours of cultivation, OD\textsubscript{600} was measured in cuvettes with eppendorf Biophotometer plus –spectrophotometer (Eppendorf AG, Hamburg, Germany) and samples for IgG determination were collected from supernatants.

Antibody titers were determined from cleared culture supernatants with an enzyme-linked immunosorbent assay (ELISA), which was realized on MICROLAB STAR Liquid Handling workstation (Hamilton, Bonaduz, Switzerland) including an ELx405 Select deep well washer (BioTek Winooski, USA) and a Synergy 2 plate reader (BioTek Winooski, USA). Secretion amounts were routinely measured as described in de Ruijter and Frey, 2015, where both coating (goat anti-human IgG (Fc specific) antibody, Sigma-Aldrich, Finland) and
Materials and methods

detection antibody (goat anti-human IgG (Fc specific) peroxidase-labeled antibody, Sigma-Aldrich, Finland) recognized the antibody heavy chain. For determining the antibody quality, two variants of the ELISA protocol were used to measure the concentrations of full-length (HC-LC) and light chain (LC-LC) fragments. In HC-LC, coating was realized as before, but detection antibody was anti-human kappa light chains (Bound and Free)-peroxidase antibody (Sigma-Aldrich, Finland) diluted 1:4,000. In LC-LC, the detection antibody was used as in HC-LC, but the coating antibody was an anti-human kappa light chain (bound and free) antibody (Sigma-Aldrich, Finland) used in concentration of 4.2 μg/ml. Otherwise protocols were identical, except for substrate incubation, which was increased from eight to ten minutes.

![Figure 7 Schematic summary of the screening procedure. Folding factors in the plasmid library were transformed to production strains and inoculated to selective, non-inductive media in 96-well deep well plates. The inoculum was divided to three plates, which contained fresh media. Each plate was grown at different temperature after addition of galactose. After 24 hours of protein expression, cells were collected by centrifugation, and antibody titers were measured from the culture supernatants with automated high-throughput (HTP) ELISA.](image)

3.6 Flow cytometry

Strains were inoculated to selective SD media with 2 % glucose, or YPD, and grown for 18 hours at 30 ºC shaking. For time course experiments, cells were collected and resuspended to fresh SD media with 0.25% raffinose and diluted to 0.5 OD$_{600}$ in 1 ml culture volume on 96-well deep well plates. Samples were cultured at 30 ºC, 250 rpm for 2 hours before 2 % galactose or 2 % glucose was added at time point zero. One well per sample was inoculated
and collected for each fluorescence measurement at time points 0, 2, 4, 8, and 24 h. Antibody expression was confirmed with ELISA from the samples. Flow cytometry was performed with BD FACS Aria III Cell Sorter (BD biosciences, USA) with the system configuration 70 u 488, 561, 633, 405 (3-5-3-6), using the 70 μm nozzle (BD Biosciences, USA). The measured parameters and set voltages were as follows: forward scatter (FSC) 11, side scatter (SSC) 110 and FITC-channel with 393 for detection of the fluorescence signal. PBS (135 mM NaCl, 2.5 mM KCl, 10 mM Na2HPO4, 1.75 mM KH2PO4) was used as sheet fluid and for resuspension of the sample. Flow rate was 5.0 for samples, in which the cell concentration was adjusted to 0.1-0.5 OD600 per ml. Data was recorded with the BD FACSDiva Software (BD Biosciences, USA) and exported as .fcs-files for further analysis in R software environment complemented with packages flowCore (Hahne et al., 2009) and flowViz (Sarkar et al., 2008).

3.7 mRNA extraction and real-time PCR

Strains were cultivated similarly as described for screening procedure. The growth and induction conditions are detailed in Publication II. Total mRNA was isolated from 1.5 OD600 of cells with RNeasy® Mini kit (Qiagen). Approximately 10 μg of isolated RNA was treated with TURBO DNA-free™ DNAse (Ambion® RNA by Life Technologies), which was inactivated, and the mRNA was converted to cDNA with AffinityScript QPCR cDNA synthesis kit (Agilent Technologies, USA). For qPCR analysis, 100 ng of cDNA from each sample was used for PCR reactions containing commercial PrimePCR™ SYBR® Green Assay CPR5, Yeast primers (qSceCED0004208, Bio-Rad, USA) or primers to determine ACT1 levels as a loading reference (Publication II). Other PCR reagents were from Maxima SYBR green/fluorescein qPCR Master Mix (Thermo Scientific, USA), and the qPCR was performed with CFXConnect thermocycler (Bio-Rad, USA). All kits and reagents were used according to manufacturer’s instructions. Results were analyzed with the Bio-Rad CFX Manager 3.0 software (Bio-Rad, USA).

3.8 Western blotting

Cell lysate preparations were realized with a previously published method (de Ruijter and Frey, 2015) with the exception that lysate volume was 200 μl. Samples were run on SDS-PAGE gels and transferred to nitrocellulose membranes with Trans-Blot® Turbo™ transfer system (Bio-Rad, USA). Blocking and washing were done according to the specifications given by the antibody suppliers. For detection of expressed folding factors, 10 μg/ml of
ANTI-FLAG® M2 (Sigma-Aldrich, Finland) was used with an overnight incubation at 4°C, with an anti-mouse IgG-peroxidase conjugated antibody produced in rabbit (Sigma-Aldrich, Finland) diluted 1:10,000 as the secondary antibody. Chemiluminescence signal was created with SuperSignal® West Pico Chemiluminescent Substrate (Thermo Scientific, Rockford, USA).

Culture supernatants from *Pichia pastoris* were adjusted to 1x NuPAGE® LDS Sample Buffer (Invitrogen™ by Thermo Scientific, USA) and run onto NuPAGE™ 4-12 % Bis-Tris precast gels (Invitrogen™ by Thermo Scientific, USA). Blotting was realized with Trans-Blot® Turbo™ transfer pack (Bio-Rad, USA) as above. For visualizing light chains, α-κ-light chain-alkaline phosphatase antibody (Sigma-Aldrich, Finland) was diluted 1:10,000. For detection of heavy chains, HRP-conjugated anti-heavy chain antibody (Invitrogen™ by Thermo Scientific, USA) was used, diluted 1:3,000. Alkaline phosphatase conjugate substrate kit (Bio-Rad, USA) was used with the anti-light chain antibody, whereas the substrate for the HRP-conjugated antibody was the same as previously.

### 3.9 Co-immunoprecipitation

Selected strains were grown to saturation and inoculated to 10-20 ml SD-URA-LEU media with 2 % galactose. Starting OD$_{600}$ was 1.5 and cultivation was continued for 16 hours at 30 °C with shaking. Cell pellets were collected, flash-frozen, and subsequently used for preparation of immunoprecipitates. Co-immunoprecipitation protocol was adapted from Gerace and Moazed, 2014, with the following exceptions: cell lysis was realized with Disruptor genie (Scientific Industries, Bohemia, USA) in three 1 min pulses with 5 min incubation on ice in between, the antibody was precipitated with Protein G Sepharose™ 4 Fast Flow –beads (GE Healthcare, UK), and Tween® 20 was used in the lysis buffer instead of NP-40. Cell lysates were adjusted to the same protein concentration as measured with a Bradford assay realized with a commercial protein assay dye reagent concentrate (Bio-Rad, USA). Human BiP was detected with an anti-BiP rabbit monoclonal antibody (Cell Signaling Technology®, USA) and yeast Kar2 with the anti-Kar2 (y-115) rabbit polyclonal antibody (Santa cruz biotechnology, USA). Secondary antibody in both cases was a peroxidase-conjugated anti-rabbit antibody produced in goat (Sigma-Aldrich, Finland). Presence of IgG was confirmed with anti-human kappa light chains (bound and free)-peroxidase antibody produced in goat (Sigma-Aldrich, Finland).
3.10 Data analysis and statistics

Data processing was done in the R software environment using a combination of available packages and the SQL-language for data sorting (sqldf-package was used for language conversion). In screening experiments, replicates were visually analyzed and measurements errors were removed. Statistical relevance of screening results was evaluated with frequency analysis and independent two-group and two-tailed Wilcoxon signed rank test (also known as Mann-Whitney test). In frequency analysis, the cut-off values for 95th percentile was determined with quantile() function, and the count and relative frequency of measurement points inside the percentile was determined for each strain. Statistical tests were performed with wilcox.test() function by comparing the strain measurements to the corresponding values of the respective strain background. The linear model was computed with the lm() function, where the dependency of normalized antibody titers on OD_{600} and normalized specific product yields was visualized as a linear regression plane fitted to all the data points normalized to respective background strain in each condition. Student’s t-test was performed with the t.test() function, adjusting the parameters (unpaired, two-sided) accordingly.

3.11 Bioinformatic analysis of plasma cell differentiation

The transcript levels of selected folding factors were extracted as follows: GEOquery-package (Davis and Meltzer, 2007) was utilized to download the data set with the GEO accession number GSE36975 (Le Gallou et al., 2012) to the R software. The expression data was quantile normalized by using the affyPLM-package (Bolstad, 2004), and the illuminaHumanv4-package (Dunning et al.) was selected for probe annotations. The expression values of selected folding factors were extracted from day 4 after induction of plasma cell differentiation and compared to the unstimulated B-cells. Transcriptomics data analysis workflow for plasma cell differentiation to find gene candidates is described in detail in Publication IV. In this analysis, we utilized the data set GSE 41208 (Cocco et al., 2012) in addition to GSE36975.

3.12 IgG production in Pichia pastoris

Both heavy and light chains of the antibody were cloned under the AOX1-promoter and integrated to the AOX1-locus of parental strain X-33. Light and heavy chains were amplified from the previously cloned plasmids (de Ruijter and Frey, 2015) and introduced to XbaI/XhoI digested pPICZaA (Invitrogen™ by Thermo Scientific, USA) and modified
pPICZαA (Gasser et al., 2006), resulting in plasmids pEK40A and pEK39B. The expression cassette including the AOX1-promoter, cDNA for light chain, and the AOX1-terminator was amplified from pEK40A, with complementary overhangs to the target site on plasmid pEK39B. Heavy chain containing plasmid pEK39B was digested with BamHI and the light chain expression cassette was recombined with the linear pEK39B by using ELIC. The resulting plasmid pEK41B contained expression cassettes for both heavy and light chain, heavy chain being upstream. Plasmid pEK41B was linearized with PmeI and transformed to the selected *P. pastoris* strain by electroporation as described elsewhere (Nett et al., 2005). Transformants were selected based on Zeocin™ (Invitrogen™ by Thermo Scientific, USA) resistance and IgG production was confirmed from true transformants.

The resulting IgG-producing strain was grown on 24-well deep well plates. Colonies were precultured in selective YPG-media (1 % yeast extract, 2 % peptone, 2 % glycerol) for 18-24 hours. Cells were collected and resuspended in 2 ml BM-media (1 % yeast extract, 2 % soy peptone, 0.67 % yeast nitrogen base, 0.4 mg/l biotin, and 0.1 M Na₂PO₄, pH 7) with one glucose FeedBead (diameter 12 mm, Kuhner SHAKER) per well. Culture was started at 8 OD₆₀₀ and grown at 25 °C, 280 rpm (Infors HT orbitron shaker) for 48 hours. To induce the protein production, the cultures were fed with 0.5 % methanol at time point 3 h, and with 1 % methanol at 19, 27, and 43 hours. 1 ml of culture was collected by centrifugation at 16,000 g for 5 min. Supernatants were collected and stored at -20 °C for following analysis on Western blots, as described in section 3.8.
4. Results and discussion

Our model IgG was a chimeric monoclonal anti-CD20 antibody, which is directed against a B-lymphocyte antigen and used as a biopharmaceutical treatment for example against non-Hodgkin’s lymphoma (Buske et al., 1999). The chosen antibody is of isotype IgG1, the most common isotype among therapeutic antibodies (Buss et al., 2012). We cloned the heavy and light chains under the galactose inducible GAL1-promoter, and the cDNAs contained the signal peptide from the yeast PHO5 gene (de Ruijter and Frey, 2015). This IgG expression cassette was integrated into the genomic HIS3-site. The antibody-producing yeast without any further modifications displayed titers between 20-40 ng/ml depending on the conditions. This antibody titer is comparable to previously reported values of secreted full-length IgG from S. cerevisiae (Rakestraw et al., 2009; Horwitz et al., 1988). The wild-type W303α yeast strain with the antibody-production cassette (named from now on as wt) was our starting point for strain engineering, and to this strain, we started implementing the features of plasma cells, concentrating on ER elements.

4.1 Reshaping the yeast endoplasmic reticulum

4.1.1 Modifications to the ER morphology increased specific product yields

The most distinctive feature of plasma cells is their greatly expanded ER, particularly dominated by ribosome-bound rough ER (Wiest et al., 1990). Importantly, the increase in ER capacity was shown to precede the initiation of antibody production (van Anken et al., 2003). ER size seems to be critical for efficient IgG secretion, so we sought for the possibility to affect S. cerevisiae ER morphology. Schuck et al. presented a structural characterization of a yeast OPI1 deletion strain, which was shown to have a 1.5-fold bigger ER than the corresponding parental strain (Schuck et al., 2009). The OPI1 gene encodes a transcriptional repressor, which controls the activity of the heterodimeric Ino2p/Ino4p-complex. The removal of the repression by Opi1p results in constitutively active Ino2p/Ino4p and upregulated lipid synthesis, which can lead to the expansion of the yeast ER (Schuck et al., 2009; Chumnanpuen et al., 2013). We generated an antibody production strain with the
$OPI1$-gene deleted, referred from now on as $Δopi1$, and set out to test whether this ER-size manipulation would have an effect on antibody secretion efficiency.

Throughout the conditions tested, the $Δopi1$ strain proved to be superior in antibody production and secretion (Publication II, Table 4). Although the final cell density and specific growth rates were lower, the $Δopi1$ production strain displayed higher titers of up to 95 ng/ml compared to the wt production strain, which remained around 40 ng/ml at the best screening conditions (Publication II, III, Table 4). Table 4 reports the fold change in specific product yield of the $Δopi1$ strain to be 2.66-fold of the wt strain. Proliferation of ER membranes improved the accumulation of a membrane-localized receptor in the yeast *Yarrowia lipolytica* (Guerfal et al., 2013), indicating that increase in functional area or volume could explain the enhancement of antibody secretion in the $Δopi1$ strain. However, the ER is likely to expand during antibody production also in the wt cells as a consequence of UPR induction. The expansion of ER by deletion of $OPI1$ was shown to be similar in magnitude to the expansion due to protein production stress, but importantly, the ER enlargement in $Δopi1$ occurred independently of UPR and alleviated ER stress itself (Schuck et al., 2009). Modeling studies in *P. pastoris* suggest that the transit through the ER is rate-limiting in protein secretion (Love et al., 2012), thus increasing the ER size without induction of a stress response might equip the cell with enhanced flux of proteins in and out of the ER.

The deletion of the $OPI1$ gene did not only affect the size of the ER but also its shape. The expansion seemed to occur preferably in sheet structures, possibly due to the limited amount of the reticulons Rtn1p and Rtn2p, and their associated protein Yop1p, which are the proteins controlling the shape of the ER (Schuck et al., 2009; Voeltz et al., 2006). As mentioned in section 1.3.3, the yeast ER is organized in sheets and tubules that differ in ribosome density and possibly in other functionalities (West et al., 2011). The sheet structures resemble the structurally less precise classification for this ER appearance, the rough ER, which is the major appearance of ER in plasma cells (Wiest et al., 1990). In order to investigate whether some of the improvements obtained by the $OPI1$ deletion could be accounted for the expansion of sheets, we created a strain with the antibody expression cassette, in which all the three genes for the major ER structure determinants, $RTN1$, $RTN2$, and $YOP1$, were deleted. When all these genes were removed from a yeast strain, the peripheral ER appeared mainly as sheets and growth rate was slightly diminished, although cells otherwise showed no defects in growth or in secretion (Voeltz et al., 2006). Interestingly, when we generated this sheet-enriched production strain (named here as the $Δret$ for reticulon mutant), we were able to recreate the $Δopi1$ secretion phenotype (Table 4). Although the $Δret$ strain was not
as efficient as \( \Delta opi1 \) in secretion, the specific product yield was up to 2.4-fold of the wt production strain (Table 4). We also detected a small decrease in growth rate of \( \Delta ret \), which magnified upon induction of antibody expression and resulted in a lower final cell density, similarly to the \( \Delta opi1 \) strain (Table 4). It is possible that the proliferation of sheets in both the \( \Delta opi1 \) and \( \Delta ret \) production strains creates additional space for high-density ribosome binding and translocation (West et al., 2011), aiding the entrance of the heavy and light chain polypeptides to the ER and secretory pathway.

Table 4 Antibody secretion in strains with ER manipulations compared to the wt production strain.

<table>
<thead>
<tr>
<th>Strain name (number)</th>
<th>Gene deletions</th>
<th>( \mu_{\text{max}} ) (^a)</th>
<th>( \mu_{\text{max (prod.)}} ) (^a)</th>
<th>Antibody titer (ng/ml) (^b)</th>
<th>Final OD(_{600}) b)</th>
<th>Specific product yield (ng/ml/OD(_{600}) b)</th>
<th>Fold change (^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt (YEK018)</td>
<td></td>
<td>0.165±0.029</td>
<td>0.103±0.017</td>
<td>44.46</td>
<td>1.276</td>
<td>35.76±14.93</td>
<td>1</td>
</tr>
<tr>
<td>( \Delta opi1 ) (YEK019)</td>
<td>( \Delta opi1 )</td>
<td>0.123±0.002</td>
<td>0.053±0.002</td>
<td>93.64</td>
<td>0.996</td>
<td>95.00±15.64</td>
<td>2.66</td>
</tr>
<tr>
<td>( \Delta ret ) (YEK052)</td>
<td>( \Delta rt1 \Delta rt2 \Delta yop1 )</td>
<td>0.129±0.005</td>
<td>0.077±0.006</td>
<td>70.25</td>
<td>0.806</td>
<td>86.54±16.90</td>
<td>2.42</td>
</tr>
</tbody>
</table>

\(^a\) Growth was followed on microtiter plates for 63.5 hours, and the values for growth rates were determined from best model fits. The average of six culture replicates is shown, while the error range indicates standard deviation.

\(^b\) Antibody titer, final OD\(_{600}\) and specific product yield (titer normalized to optical density) were recorded in screening conditions at 30 ºC with 0.5 % galactose induction. Error ranges indicate standard deviations. The mean values of at least three biological replicates are shown.

\(^c\) The fold changes were calculated by normalizing the specific product yields with the respective mean value of the wt strain.

Expansion of sheets can also affect the dynamics of ERES, which localize to the high-curvature areas. ERES have been shown to accumulate to fewer distinct sites in \( \Delta rt1 \Delta rt2 \Delta yop1 \) triple deletion strains as these strains displayed fewer tubules and fewer high-curvature areas than the parental strain. The altered localization of ERES can in turn lead to aberrant Golgi cisternae (Okamoto et al., 2012) which might affect secretory capacity. To investigate whether induction of tubules would have any effect on IgG secretion, we overexpressed \( RTNI \) under three different promoters in the wt and \( \Delta opi1 \) strains (Figure 8). Similar constructs have been previously shown to increase the amount of tubular ER (Voeltz et al., 2006; Schuck et al., 2009). However, overexpression of Rtn1p had only minor effects on antibody titers both in the wt (Figure 8A) and in the \( \Delta opi1 \) strain.
Results and discussion

(Figure 8B), implying that the functionalities lying in tubular regions of the ER are not important for increasing IgG secretion in yeast. In fact, forcing the ERES to fewer high-curvature areas by expansion of sheets might actually streamline the downstream secretory pathway, as normally the ERES and correspondingly, the Golgi cisternae are relatively dispersed throughout the cytoplasm and not forming a straight route for secretion (Okamoto et al., 2012).

![Graph](image)

**Figure 8** Effects of increasing the amount of ER tubules. RTN1 was overexpressed under endogenous (P<sub>ENDO</sub>) promoter, GAL1- and TEF-promoters in the wt (A) and Δopi1 (B) production strains. Antibody titers were measured in the screening conditions at 30 °C, both with 0.5 and 2 % induction. The means of three biological replicates, each measured in two technical replicates, are shown, and the error bars indicate standard deviation.

As modifications of the ER morphology had such dramatic effects on specific product yield, the major bottlenecks of IgG secretion must be ER-related processes whether it is folding, translocation or exit from the ER. The model of protein transit through ER in *P. pastoris* suggested that relative rate of secretion decreases with increasing protein complexity (Love et al., 2012). The complexity of the proteins in the model in question was extended only up to glycosylated Fc region of an antibody, highlighting that passing through the yeast ER is a major hindrance for full-length IgG.

### 4.1.2 Expansion of ER sheets diminished the induction of UPR

The small additional increase in specific product yield of Δopi1 compared to Δaret could be explained by the fact that increasing the size of ER has additional benefits in protein folding and UPR-control (Schuck et al., 2009). Physiologically, the UPR is induced in cells upon accumulation of unfolded proteins in the ER. Although stress reactions are undesirable, they can salvage the cell from death and strengthen the secretory system. In addition, the level of
UPR can indicate the fitness of the strain in protein production, as lower level of stress may translate into higher secretion levels. In order to investigate differences in UPR induction, we integrated a UPR-responsive reporter cassette into the genome of the wt, Δopi1 and Δret strains. The cassette is constructed of a cDNA for the green fluorescent protein (GFP) under the control of a modified promoter containing four UPR-response element sequences (Pincus et al., 2010). UPR-induction can thus be measured with the amount of GFP-fluorescence, which we assessed with flow cytometry on a single cell basis.

Flow cytometry analysis revealed that the size distribution of the cells was altered in the three strains. The measured forward scatters (FSC) indicated that the mean size of Δopi1 cells was 9 % bigger than for wt cells, while Δret cells were considerably smaller than the former two (Figure 9A). The side scatter (SSC) is proportional to granularity, or the amount of internal structures, of the cells, and the mean SSC difference between wt and Δopi1 was 24 % (P-value << 0.001, unpaired Student’s t-test). Increased SSC could be a reflection of the increased size of the ER in the Δopi1 cells (Figure 9). SSC of Δret cells might not be comparable to the other strains due to the decreased cell size (Figure 9). The difference in size should be considered in the context of OD 600-values. The high secretion of the Δret strain, as expressed by IgG titer normalized to OD600, might be a result of increased number of cells as the measurement of optical density does not differentiate between an increase in cell number and size.

Figure 9 Flow cytometry analysis of the ER modified strains. The forward scatter (A) and side scatter (B) were recorded for 10,000 events from each strains at induction (0 h) time point after 18 hours of cultivation in non-inductive media. The distribution of measured forward scatter values are shown in (A) while side scatter was transformed to a log-like scale in (B). Relative density reflects the frequency of the measurement in each value point, while the total area (probability) under to curve equals to one. Values on the x-axis are arbitrary, but relate to the measures of forward and side scatter proportionally.
To measure the UPR-response, we set up a time-course analysis to dissect the differences between the strains. Strains were first grown in non-inductive media, and at time point zero, the cultures were supplemented with 2% galactose to initialize the production of heavy and light chain polypeptides (Figure 10, left panel) and corresponding control samples were supplemented with 2% glucose to repress antibody expression (Figure 10, right panel). Already the starting fluorescence was significantly different in the strains, where wt started from mean fluorescence at around 300, \(\Delta \text{o}pil\) at 126, and the \(\Delta \text{aret}\) strain at 23 (Figure 10). The mean fluorescence from control cells with only the UPR-reporter but no cassette for antibody expression was ten at maximum. Wt cells reached the maximum mean fluorescence at eight hours, after which the response decreased (Figure 10A). In contrast, in the \(\Delta \text{o}pil\) strain the maximal response took longer to develop, steadily increasing up to the 24 hour time point (Figure 10B). The delayed maximum and overall low fluorescence levels reflect the alleviated stress response reported for the \(\Delta \text{o}pil\) phenotype (Schuck et al., 2009). Strikingly, no UPR induction seemed to occur in \(\Delta \text{aret}\) cells with the induction of antibody expression (Figure 10C). GFP fluorescence increased in the control \(\Delta \text{aret}\) samples (Figure 10C, right panel), but this can be accounted for by the increased cell size as measured by FSC due to improved growth on glucose (data not shown).

The control samples in the right panel of Figure 10 indicate that cells experience significant protein production stress also when grown in non-inductive media. Indeed, even after cultivation in rich media containing glucose, wt and \(\Delta \text{o}pil\) show measurable GFP fluorescence (Figure 11A). Residual activity of GAL1-promoter on glucose has been reported before (Güldener et al., 2002), but it is surprising that this activity is enough to induce antibody production in such an extent that UPR-signaling is activated. The high level of the signal can be partially explained by the >8 hour half-life of the GFP reporter, causing an accumulation of the fluorescent reporter to the cell even if the UPR does not persist in repressive conditions (Pincus et al., 2010).
Figure 10 Strains with expanded ER sheets displayed less UPR. The UPR-responsive GFP expression cassette was integrated to the genome and GFP fluorescence was followed for 24 hours in the wt (A), Δopi1 (B) and Δret (C) strains. GFP fluorescence was transformed to a log-like scale and the distribution of values from 10,000 cells is depicted for each time point. Left panels show cells grown in inductive, galactose-containing media, and the right panel shows corresponding control samples with added glucose. Boxes display the mean fluorescence for each sample in non-transformed values of linear scale.
Bivariate histograms in Figure 11B summarize the cell populations of different strains at the height of UPR in the wt strain (eight-hour time point). Fluorescence levels in Δopi1 and Δret were visibly lower than in wt, while the Δret strain resembled the control strain as shown in Figure 11B. Since the GFP-fluorescence was unexpectedly low, we tested the performance of the GFP-reporter in the Δret strain by overexpression of active Hac1p (Figure 11C). Hac1p is the transcription factor responsible for the UPR, and the increase in GFP fluorescence upon its overexpression confirmed the proper functioning of the UPR-reporter construct (Figure 11C). Therefore, the results in Figure 11C indicate that UPR sensing is disturbed in the Δret strain, although the downstream signaling pathways function normally.

It is hard to explain exactly why the deletion of RTN1, RTN2, and YOP1 should abolish the UPR. Possibly the absence of these proteins or the expansion of sheets affect the localization of the UPR-sensor Ire1p, preventing its dimerization and activation. Localization of membrane proteins can be affected by lipid rafts, which are segregated functional units within membranes, and these rafts have been implicated to be present in the yeast ER and participate in protein secretion (Bagnat et al., 2000; Helms and Zurzolo, 2004). The membrane-interacting proteins RTN1, RTN2, and YOP1 could participate in the formation of Ire1p containing rafts that act as areas of UPR control. In this case, the absence of the proteins could cause Ire1p to disperse throughout the ER making the local concentrations too low to initiate dimerization and subsequent activation. Another option is that the Δret cells do not experience conformational stress, which seems unlikely, as they secrete significant amounts of antibodies (Table 4). Several aspects of the relationship between structure and function of organelles remain obscure, thus we did not characterize the Δret strain further. Instead, we chose the Δopi1 strain with the enlarged ER to be the platform for reconstructing mammalian folding machinery, in comparison to the wt cells. In order to introduce several expression formats of mammalian folding factors to yeast cells, we needed efficient methods for plasmid library cloning, and optimization of such a method is described next.
Figure 11 Summary of the differences in the UPR induction between the strains. The GFP fluorescence contrasted to side scatter is shown for the strains grown for 18 hours in rich, glucose-containing media (A) and eight hours after induction of antibody expression (B). During antibody production, wt and Δopi1 displayed significant UPR, but Δret resembled the control (ctr) in the left panel in (B). To confirm the proper functioning of the GFP reporter cassette, active Hac1p was overexpressed in Δret (C). HAC1 increased UPR significantly compared to the control (ctr-plasmid), thus lack of UPR in Δret was not due to reporter defect. GFP fluorescence and side scatter values were transformed to a log-like scale. Relative density reflects the distribution of the measured values, while the area under the curve is set to one.
4.2 Establishing a high-throughput cloning method for plasmid library creation

We set out a task to investigate the effects of co-expressing mammalian folding factors with IgG to improve titers which required the creation of yeast vectors containing expression cassettes for selected folding factors under different yeast promoters. As dozens of plasmids needed to be created, we considered a number of high-throughput cloning strategies. First, we decided to test restriction enzyme based method utilizing type II restriction enzyme BsaI. However, the efficiency of the restriction and ligation steps did not meet our requirements. Instead, we used the added sequences containing the restriction enzyme sites as common adaptors to extend the cDNA cassettes with homologous regions to the plasmid (Figure 12). Although our aim was to use the SLIC-protocol for creating the plasmids (Li and Elledge, 2007), we found out that the exonuclease treatment was unnecessary and recombinants could be obtained by mixing purified, double-stranded fragments with homologous ends without any additives to the reaction. We named our method exonuclease and ligation-independent cloning (ELIC) and described it in Publication I. The outline of the cloning procedure is depicted in Figure 12, where also the common elements of the folding factor expression cassettes are shown.

![Diagram](image)

**Figure 12 Summary of the cloning scheme and common plasmid elements.** Yeast ER-retrieval sequence HDEL was inserted or modified to the C-terminus of the selected cDNAs, which are marked with blue. Subsequent PCR amplification introduced common adaptor sequences as shown, before insert was prepared to include the homology regions (HR). The target vector included one of the five different promoters (purple), CYC1-terminator (light blue) and one of the markers (green). The pRS41X-backbone was linearized with restriction enzymes. Insert with corresponding HRs was mixed with the destination vector and simply transformed afterwards to chemicompetent E. coli, where fragments combined to yield the target plasmid. The panel on the right summarizes the workflow for ELIC.

Although our cloning seemed to be robust and efficient with the initial conditions, we wanted to characterize some of the parameters possibly affecting the performance of ELIC. Using blue-white screening for rapid assessment, we tested the effects of the length of homologous regions, insert to vector molar ratio and incubation time. We used LacZ as the insert, and the
pKQV4-plasmid (Strauch et al., 1989) as the destination vector, so true recombinants could be identified by the blue color of the colonies. Length of the homologous regions clearly affected the efficiency of recombination, and the optimal length seemed to settle around 25 bp on both sides (Figure 13A,B). Surprisingly, recombinants were found even with the shortest length of eight bp (Figure 13B). We wondered, if sticky ends in the digested vector might influence the recombination mechanism, so we treated the vector with isoschizomers, XmaI creating blunt ends and SmaI responsible for sticky ends. Treatment of the vector with either of the enzymes displayed similar levels of recombinants, XmaI yielding somewhat higher numbers (Figure 13A,B). The ELIC reaction performed better when vector to insert molar ratio was increased up to 1:4, although recombinants were created in all tested ratios (Figure 13C).

![Figure 13 Performance of the cloning procedure with varying parameters.](image)

After mixing purified insert and destination vector in water, no incubation was needed prior to transformation: the efficiency of ELIC was the same when incubation time was decreased to 0 min (Figure 13D). This was a rather interesting observation, considering that the
fragments were shown to interact already in vitro (Publication I). Although we did not investigate the mechanism of our cloning method further, DNA fragments are able to form stable four-way DNA structures in vitro (Gaillard and Strauss, 1994; Neschastnova et al., 2002), so it is possible that they form a circular structure that *E. coli* cells can take up and replicate. Regardless of the mechanism, ELIC fulfills several of the characteristics required for a high-throughput cloning method listed in Marsischky and LaBaer, 2004. The protocol can be flexibly applied for any vector and insert in a sequence-independent manner, it has very low costs, and it proved to be very reliable for cloning of a large range of insert sizes, from 100 to 4,500 bp (Publication I). Two-fragment assembly was very efficient in terms of correct clones obtained, as in many reactions 100 % of the tested clones were correct, while the average performance was 80 % correct recombination events (Publication I). However, the efficiency of obtaining transformants with low amounts of DNA might be considerably lower with our method than reported for other LIC methods (Publication I, Jeong et al., 2012; Li and Elledge, 2007), so ELIC might not be suitable for small scale. In addition, the number of inserts seemed to be limited to two or three, as the efficiency of obtaining the target plasmid was decreased to 50 % in three-fragment assembly (Publication I). ELIC proved to be very suitable for our applications and it is a simple and easy method to be added to the family of interconvertible ligation-independent cloning techniques with complementary properties (Stevenson et al., 2013).

4.3 Screening plasmid library of ten mammalian folding factors

4.3.1 Selected folding factors included all functional classes and were expressed at different levels

Studies on IgG folding and plasma cell differentiation have enumerated the specific selection of folding factors required for efficient IgG production. As listed in Table 5, our selection of mammalian folding factors for overexpression studies in yeast consisted of ten different folding factors. Although our subset does not contain all identified components of the plasma cell folding machinery, we kept our selection as diverse and relevant as possible. We included all the major folding enhancers of antibodies as identified in other studies (Feige et al., 2010; Romijn et al., 2005; Meunier et al., 2002) to represent all the folding factor classes. In addition to the molecular chaperones BiP and GRP94, also the co-chaperones of BiP, ERdj3 and GRP170 were included. Notably, GRP170 has been shown to possess the activity of binding unfolded proteins directly and thus to function as a chaperone (Behnke and Hendershot, 2014). Four different PDIs were selected along with two PPIases, which
Results and discussion

represented different subclasses (Table 5). To identify their relevance in plasma cell differentiation, we extracted their expression levels from transcriptomics (Le Gallou et al., 2012) and proteomics data sets (Romijn et al., 2005) which described similar B cell differentiation models. Upregulation of these candidates on mRNA and protein levels corresponded well in these two studies (Table 5), demonstrating that increasing the abundance of these proteins is an integral process in plasma cell differentiation. However, these values describe the upregulation only after four days of B cell development, reaching only the early plasmablast phase (Le Gallou et al., 2012). Therefore, these folding factors might display even higher abundance in mature plasma cells than indicated in Table 5.

Table 5 Selected folding factors, their function and upregulation during plasma cell development.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Full name</th>
<th>Function</th>
<th>Upregulation mRNA</th>
<th>Upregulation protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>BiP</td>
<td>Immunoglobulin heavy chain-binding protein</td>
<td>molecular chaperone</td>
<td>1.11</td>
<td>1.96</td>
</tr>
<tr>
<td>GRP94</td>
<td>94 kDa glucose-regulated protein</td>
<td>molecular chaperone</td>
<td>2.31</td>
<td>1.68</td>
</tr>
<tr>
<td>GRP170</td>
<td>170 kDa glucose-regulated protein</td>
<td>molecular (co-)chaperone</td>
<td>1.68</td>
<td>1.78</td>
</tr>
<tr>
<td>ERdj3</td>
<td>Endoplasmic reticulum DNA J domain-containing protein 3</td>
<td>co-chaperone of BiP</td>
<td>1.22</td>
<td>1.96*</td>
</tr>
<tr>
<td>CypB</td>
<td>Cyclophilin B</td>
<td>peptidyl-prolyl cis-trans isomerase</td>
<td>1.22</td>
<td>1.65</td>
</tr>
<tr>
<td>FKBP2</td>
<td>FK506-binding protein 2</td>
<td>peptidyl-prolyl cis-trans isomerase</td>
<td>1.29</td>
<td>1.6</td>
</tr>
<tr>
<td>PDI</td>
<td>Protein disulfide-isomerase</td>
<td>protein disulfide isomerase</td>
<td>2.09</td>
<td>2.09</td>
</tr>
<tr>
<td>P5</td>
<td>Protein disulfide isomerase P5</td>
<td>protein disulfide isomerase</td>
<td>1.96</td>
<td>2.08</td>
</tr>
<tr>
<td>ERp57</td>
<td>Endoplasmic reticulum resident protein 57</td>
<td>protein disulfide isomerase</td>
<td>-</td>
<td>1.8</td>
</tr>
<tr>
<td>ERp72</td>
<td>Endoplasmic reticulum resident protein 72</td>
<td>protein disulfide isomerase</td>
<td>1.53</td>
<td>1.98</td>
</tr>
</tbody>
</table>

a) Extracted from microarray data (Le Gallou et al., 2012) as described in Materials and Methods.

b) From proteomic study by Romijn et al. (Romijn et al., 2005), values from day 4 after induction of differentiation. *from day 2 as peptides were not detected at day 4.

The experimental testing included only to the folding factors shown in Table 5. We omitted several interesting candidates, for example the glycan-recognizing chaperones calnexin and calreticulin, the oxidoreductase Ero1α and the mammalian UGGT. As S. cerevisiae has no intrinsic UGGT activity (Fernández et al., 1994), so recreating the calnexin cycle in yeast
would have been an interesting engineering effort. Ero1α is significantly upregulated during B cell differentiation (Romijn et al., 2005) and its overexpression could have been beneficial especially in the context of its functional association with PDI. Accordingly, the possibility of pairing calnexin or calreticulin with ERp57 would have been worthy of investigation, as these proteins form a specific complex to aid the folding of glycoproteins such as the antibody heavy chain (Oliver et al., 1999).

The expression level of proteins is tightly regulated in cells and selecting the strength of expression is not a straightforward process. To diversify the amounts of the overexpressed folding factors, we expressed each cDNA under five different promoters. In addition to commonly used GAL1-, GPD-, and TEF-promoters, we selected and cloned two UPR-controlled promoters from the genes KAR2 and PDI1, from which P_{KAR2} was shown to be highly induced by UPR, while P_{PDI1} displayed a moderate induction but has slightly higher basal levels than P_{KAR2} (Mori et al., 1998). We analyzed the strength of each promoter with mRNA analysis of overexpression of the yeast CPR5 gene in the wt production strain at two and ten hours after the induction of antibody production. The result showed that the UPR-controlled promoters complemented the range of expression levels by providing a lower initial overexpression (Figure 14A), and they proved to be useful in protein folding engineering of S. cerevisiae (Publication II). P_{GAL1} displayed a high initial induction as expected, and the two constitutive promoters P_{GPD} and P_{TEF} provided intermediate...
overexpression. Although the activity of $\text{P}_{\text{TEF}}$ has been shown to remain relatively constant (Partow et al., 2010), in our measurements it decreased from two to ten hours (Figure 14A). It should be noted that the strains had the endogenous $\text{CPR5}$ expression, which might influence the mRNA amounts.

By using ELIC, we created a combinatorial plasmid library comprising of the five selected promoters, two auxotrophic selection markers and the ten folding factors yielding all promoter-folding factor pairs and resulting in 65 different plasmids (Table 2). We also created FLAG-tagged versions of some of the folding factors, and showed that these mammalian proteins are successfully expressed in yeast (Figure 14B). Interestingly, the detectable proteins levels varied strongly between the folding factors, even when they were expressed under the same promoter (Figure Expression B). ERdj3, P5 and GRP94 were abundant in cell lysates, but FKBP2 was nearly undetectable (Figure 14B). Although ideally good results should be correlated to optimal strength of folding factor expression, the required amount of each folding factor remains hard to predict. Studies in the fungus $\text{Aspergillus awamori}$ suggested that there is an optimal level of overexpression for folding factors, but this optimum is dependent on the folding factor and the protein substrate in question (Lombraña et al., 2004; Moralejo et al., 2001). In the absence of means to predict this optimum, the possible range of expression strengths was kept wide in this study.

4.3.2 Temperature and induction strength affected the strains similarly

All the individual plasmids and selected combinations of plasmids were transformed to both $\text{wt}$ and $\Delta\text{opi1}$ production strains. The resulting strains were screened for improved secretion at three temperatures (20, 25 and 30 °C) by using two induction levels for the $\text{P}_{\text{GAL1}}$-controlled antibody expression, 0.5 and 2 % galactose. Overall, the lower induction level resulted in higher IgG titers for most strains, as illustrated with all the measurements from $\text{wt}$ control strain at 30 °C in Figure 15A. In addition, the mean final cell density decreased with increasing galactose concentration (Figure 15B). As a result, the specific product yield was slightly lower with 2 % induction compared to 0.5 % induction, although the overall range of measurements within a strain made the differences in induction levels relatively minor (Figure 15). High induction might induce an initial overload to the secretory machinery resulting in a stress response, which would explain the decreased titers. In fact, because of limited capacity of translocation, high-level expression of secretory proteins can result in accumulation of the unfolded protein in the cytosol (Čiplys et al., 2011a), which in turn can reduce fitness and growth (Geiler-Samerotte et al., 2011). The reduction in cell
density with 2% galactose compared to 0.5% induction (Figure 15B) implies that cytosolic antibody aggregates might be the cause for growth reduction, although responses originating from later steps in the secretory pathway cannot be excluded. In any case, lowering the induction level might enable the cell to avoid excess stress.

Figure 15 Influence of the induction strength on antibody secretion. The measurements from wt control samples at 30 °C in the screening format for each galactose concentration are summarized as box plots for antibody titer (A), cell density (B) and specific product yield (C). Specific product yield (SPY) was defined as antibody titer normalized to OD600. The notched box plot representation shows median in the middle, the inner 50% of data in the boxed area, and 25th and 75th percentiles in the lower and upper whiskers, respectively.

The increased secretion phenotype achieved with the Δopi1 deletion remained dominant even with further modifications, as all Δopi1-derived strains displayed consistently higher titers (Figure 16A) and lower final cell densities (Figure 16B) than the wt-based strains at all temperatures. The difference between the strain backgrounds was prominent in the segregated populations of specific product yields (Figure 16C). Expectedly, OD600 increased with temperature, but also IgG titers followed the same trend (Figure 16A,B). Antibody variants have often been expressed at 20 ºC, and lowering the temperature has been shown to result in higher titers in several yeast cultivations (Huang et al., 2008; Shusta et al., 1998; Hackel et al., 2006). Decreasing the expression temperature maintained the secretion capacity for a prolonged period (Huang et al., 2008), but this might not be relevant within our 24-hour time frame. In our set-up, the aeration is not optimal in the 96-well deep well plates, so an increase in temperature might be needed to boost the cell growth (Figure 16B), which also enhanced secretion (Figure 16C). Interestingly, the distributions of specific product yields in wt-based strains are essentially the same at 20 and 25 ºC, while the specific product yields in Δopi1–based strains showed a clear increase from 20 to 25 ºC (Figure 16C). In both strain backgrounds, the highest specific product yields were recorded at 30 ºC, indicating that this cultivation temperature was the best for effective antibody secretion.
Figure 16 Trends and strain populations in the plasmid library screening data. Throughout the conditions and despite of added folding factor elements, the background strains wt (black, solid lines) and Δopi1 (red, dashed lines) remained as distinguishable strain populations as determined by antibody titer (A), optical density (B) and specific product yield (C), which was defined as titer normalized to OD$_{600}$ in each sample. Distributions of measurements is depicted as density or probability plots, where the relative density reflects the amount of occurrence of each value point. Total probability of the measurement to occur under the area defined by the curve is equal to one. X-axis depicts the range of measurements and y-axis is proportional to relative frequency. To evaluate the relationships between the values, the mean values of each strain in each condition were fitted to the same value space by normalizing them to the respective background strain (D). Dependency of antibody titer on optical density and specific product yield was quantified by fitting a linear regression plane (dashed lines) to the data points. Colors in (D) are only for illustrative purposes.
The screening process of the strains was conducted without an initial adjustment of cell densities between the strains. To investigate the possibility that secretion differences between the strains were caused solely by changes in growth phase or cell densities, we fitted a linear model to all the data points from each strain and condition to describe the dependencies between the measured variables (Figure 16D). We normalized all values to the respective strain background to fit all the measurement points in the same value space. Multiple R-squared value for the model was 0.955 and the $F$ test statistics displayed $P$ values well under 0.001, thus the model described our data accurately. The linear model provided us with an estimate of how much the absolute titer was affected by $OD_{600}$ and how much the change in specific product yield contributed. The coefficient for $OD_{600}$ was 0.938, indicating that an increase in cell density was measurably beneficial for the final titer, increasing it linearly in almost 1:1 ratio. As all temperatures were included in the model, the positive effect of expression temperature on both $OD_{600}$ and titer was possibly influencing this measure. The coefficient for specific product yield, displaying the relative secretion of each strain, was 0.911, which implies that although changes in cell density contributed to the change in titers significantly, the inherent differences of the strains in secretion were prominent in our set-up. We conducted a similar screening of a plasmid library of yeast folding factors (Publication II). Compared to what we observed in the overexpression of yeast folding factors, we did not observe a comparable decrease in cell density with improved specific product yields, largely explaining the difference in the values of the linear model coefficients reported in Publication II.

4.3.3 FKBP2, P5, mammalian BiP, and its co-chaperone GRP170 improved IgG secretion

Our goal in the screening was to identify strains, folding factors and modules that displayed high IgG secretion. We took the performance at both induction levels into account and calculated the mean fold change in specific product yield normalized to the respective background strain within a given condition to illustrate the secretion characteristics for each strain (Figure 17). In most cases, the addition of a folding module did not alter specific product yield significantly, as most fold changes were close to one (Figure 17). However, some of the folding factors induced an increase in secretion, namely the holdases BiP and GRP170, PPIases FKBP2 and CypB, and the PDI-family member P5. Interestingly, the UPR-controlled promoters were responsible for the best effects in the $\Delta opi1$ strain background, while in the case of FKBP2, stronger promoters were required in wt strains to induce fold changes above 1.4 (Figure 17).
Results and discussion

Figure 17 The effects of the folding factor constructs on antibody secretion. Heatmaps summarize the effects of each construct on specific product yield in each temperature and background strain. The color and number in each cell of the heatmaps represent the mean fold-change in specific product yield respective to the values in control samples, and the mean from both 0.5 and 2 % galactose induction is shown to reflect the true performance of the construct.

In general, folding factor functionalities or different promoters did not display any distinguishable patterns in secretion phenotypes (Figure 17), possibly because of the varying protein expression levels (Figure 14B). In addition, results were largely dependent on the background strain (Figure 17). Clearly, the expression levels are crucial in engineering of protein folding and unfortunately, we were unable to identify predictability from the data for folding factor expression (Publication II, III). Thus, screening approaches as presented here will be necessary for future work as well. Possibly, the selection of promoters should be widened. Payne et al. used a different set of promoters in their study of co-chaperone
expression in *S. cerevisiae*, and showed that also the transcript levels had measurable differences. Especially, promoters of lower strength, such as the *TDH1* promoter, might be interesting alternatives to avoid any burden caused by folding factor overexpression (Payne et al., 2008a).

BiP and GRP170 increased secretion notably in the Δopi1 strain background, improving their specific product yield up to 1.4- and 1.6-fold, respectively. Similarly to its yeast homolog Lhs1p, GRP170 is both a nucleotide exchange factor and a molecular chaperone (Saris et al., 1997; Behnke and Hendershot, 2014), so the beneficial effects in yeast might be accounted for the latter functionality of GRP170. The Δopi1 strain might have a deficiency in holdase concentration, as the increase in ER space was not accompanied by an increase in Kar2p amounts (Schuck et al., 2009). It seems that this concentration difference was successfully compensated by the low level expression of mammalian holdases (Figure 17), which might explain why BiP and GRP170 mainly increased secretion in the Δopi1 background. As for the other co-chaperones, Guo et al. predicted that ERdj3 overexpression might slow down the rate of secretion in mammalian cells (Guo and Snapp, 2013). Whether or not this mammalian DnaJ-protein has any activity in yeast BiP equivalent, introduction of ERdj3 had no effect on IgG secretion in yeast (Figure 17).

The *PGAL1*-P5 module was performing exceptionally well in the wt background especially at lower temperatures (Figure 17), which is interesting because P5 was the only folding factor expressed with a yeast-derived signal peptide. Other mammalian PDIs had little effect on IgG secretion (Figure 17). According to some reports, disulfide bonds are not essential for obtaining the correct Ig fold (Thies et al., 2002; McAuley et al., 2008), so it is possible that some antibody chains are secreted before sulfhydryl oxidation is complete. Beneficial effects of PDIs might be apparent only in product quality, since disulfides seem to affect mainly antibody activity (Liu and May, 2012). On the contrary, PDIs have been reported to enhance secretion of antibody variants in *S. cerevisiae* in numerous occasions (Hackel et al., 2006; Xu et al., 2005; Rakestraw et al., 2009; Shusta et al., 1998). Interestingly, Hackel et al. showed that overexpression of the native yeast PDI increased scFv secretion to a greater extent than rat PDI (Hackel et al., 2006), indicating that mammalian equivalents might not be as efficient in engineering of protein folding in yeast as their endogenous homologs. However, in our set-up, the upregulation of native *PDI1* displayed mainly negative effects (Publication II) and most heterologous PDIs had only minor effects on secretion, except for *PGAL1*-P5 (Figure 17).
4.3.4 Combinations of several modifications implied that engineering folding alone has limited improvements in secretion

The folding factors do not work alone but interact with each other to the extent that a folding complex forms on the heavy chain in mammalian cells (Jansen et al., 2012; Meunier et al., 2002). After we evaluated the folding modules individually, we paired them with the aim to induce synergistic effects on antibody folding and secretion. The reported primary interactions of our selection of folding factors are shown in Figure 18A, and we tested each of these folding factor pairs. The choice of promoters was based on their performance individually: one or two promoters were selected for each protein, resulting in 35 different combinations (Publication III, Table S5), which were each tested in both strain backgrounds. The measurements from all the combinations are summarized in Figure 18B: in effect, we did not detect any clear synergistic benefits from co-expression of two folding factors. In the \( \Delta opi1 \) strains, which had already displayed high specific product yields, introduction of a second mammalian folding factor only decreased secretion. As a result, \( \Delta opi1 \) strains co-expressing a folding factor pair displayed fold changes that were mostly less than that achieved by \( \Delta opi1 \) deletion alone (Figure 18B). In contrast, titers measured from wt strains showed a small shift towards higher concentrations (Figure 18B). A few folding factor pairs in the wt background induced a fold change more than two, but these benefits could not be accounted for clear combinatorial effects.

As pairs of folding factors did not induce a predictable increase in secretion, we did not proceed to test co-expression of three or more folding factors. The difference between the wt and \( \Delta opi1 \) strain backgrounds in folding factor pair expressions suggested that there is a limit to enhancing secretion by ER modifications. Possibly the increase in ER space by \( \Delta opi1 \) deletion and the addition of one folding factor is sufficient to move the rate-limiting step in secretion elsewhere, whereas the wt might require two folding factors to circumvent the folding bottleneck. Delic et al. pointed out that several targets should be addressed simultaneously to find synergistic benefits for secretion, as concentrating on optimization of one step might not display improvements if the bottleneck is elsewhere in the secretory pathway (Delic et al., 2014). As additional folding factors did not improve secretion, it might be reasonable to target other steps in the secretory pathway to improve secretion further.
Results and discussion

Figure 18 Testing folding factor pairs to improve antibody secretion. (A) The selected folding factors are colored by their molecular function as indicated, and their primary interactions as reported in Jansen et al., 2012 are marked with lines. All pairs connected by lines were tested and the mean fold changes in specific product yield (as compared to the corresponding control) are shown as a function of antibody titer in (B). All conditions are included, and colors mark the cultivation temperature, as indicated in the lower panel. Crosses are strains from the Δopi1 background and dots are wt-based strains. Distribution plots in the lower panel show the range of the IgG titers in each temperature, solid lines for wt and dashed lines for Δopi1 background. The fold changes of the background strain controls are shown as horizontal dashed lines, where wt was set to one and Δopi1 corresponds to an approximately 4-fold improvement alone depending on the conditions. Each data point is the mean of at least three biological replicates in the upper panel of (B).

4.4 Human BiP interacts with the antibody in yeast cells

Our primary effort in overexpressing mammalian chaperones was to increase the efficiency of IgG folding in the yeast ER. However, we cannot exclude other pleiotropic effects that might result in differential secretion. To gain insights to the mechanism of action of the folding factors, we investigated the specific interaction of human BiP with IgG inside the yeast cell by co-immunoprecipitation. Not only did the BiP interact with IgG, but also the expression of BiP seemed to be stabilized with the concomitant expression of IgG (Figure 19A). Hsu et al. also reported co-immunoprecipitation of murine BiP with IgG in a heterologous system, insect cells, demonstrating that the two proteins have a strong specific interaction, which can be recreated in vivo (Hsu et al., 1994). The yeast equivalent of BiP,
endogenous Kar2p, also precipitated in a similar fashion with IgG in its full-length, functional form (Figure 19B), even to a greater extent than the overexpressed human BiP. Kar2p is a highly abundant protein, and the functions of these two proteins are similar (Ghaemmaghami et al., 2003; Normington et al., 1989). Interestingly, when endogenous Kar2p was complemented with overexpression of Kar2p, IgG secretion was not improved (Publication II, Fig. 4). However, human BiP induced 1.5-fold enhancement in secretion in the Δopi1 strain background, while endogenous Kar2p was present (Figure 17). Thus, human BiP seems to contribute to IgG secretion, but it might not interfere with all the functions of Kar2p.

Although mammalian leader peptides have been successful in targeting proteins to the ER and secretory pathway (Hitzeman et al., 1982; Nielsen, 2013), it can be questioned whether the proteins in our constructs localize correctly. ERdj3 appeared in Western blots as two very closely associated bands (Figure 19C), which could be accounted for the size difference due to signal peptide cleavage or due to N-glycosylation, in either case indicating that a
fraction of the protein has entered the ER. Localization to the ER might not be a strict requirement for improving IgG production in yeast, as overexpression of cytosolic chaperones has been found to be beneficial for the secretion of β-glycosidase (Tang et al., 2015) and of a Fab fragment (Gasser et al., 2007). If the antibody chains are translocated post-translationally, chaperones are needed in the cytosol to maintain the polypeptides unfolded (Ng et al., 1996). Furthermore, in case antibodies accumulate in the cytosol due to unavailability of translocation components, additional cytosolic folding factors might help to reduce the formation of aggregates, which could prevent stress responses (Čiplys et al., 2011a; Geiler-Samerotte et al., 2011). We did not investigate the mechanism of action of the folding factor constructs further, although in Publication III we demonstrated that the positive effects also applied for the secretion of another mammalian glycoprotein, human erythropoietin, but not for the endogenous acid phosphatase Pho5p secretion (Publication III, Figure 3).

4.5 PPIases were the most effective in improving IgG titers

By screening the plasmid library, we identified several strains with increased specific product yields. Table 6 summarizes the results from the ten best performing strains in each strain background. The improvements in specific product yields achieved in the screening conditions were 5- to 6-fold compared to the mean secretion of 23.5 ng/ml*OD600 of the wt, resulting in titers of more than 100 ng/ml for several Δopi1-based strains (Table 6). In accordance with earlier statements, folding factors behind the best improvements in the Δopi1 background were expressed under UPR-controlled promoters. While none of the folding factor pairs was among the ten best Δopi1 strains, four such pairs reached the highest specific product yields in wt background (listed in Table 6). As we screened for improvements originating from introduced folding factor constructs, we decided not to minimize biological variation to enable high-throughput assessment. Despite the great variation in measurements, the difference compared to the respective strain background was statistically significant for most of the best performing strains, as determined with low P-value and the high relative frequency of the strains in the highest percentiles of measured specific product yields (Table 6). In conclusion, the screening set-up was successful in identifying interesting strains with improved secretion.
### Table 6 The strains with highest specific product yields in each strain background.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Added elements</th>
<th>Strain background</th>
<th>IgG titer (µg/ml)</th>
<th>OD&lt;sub&gt;600&lt;/sub&gt;</th>
<th>Specific product yield (µg/ml/OD&lt;sub&gt;600&lt;/sub&gt;)&lt;sup&gt;b)&lt;/sup&gt;</th>
<th>Relative frequency&lt;sup&gt;c)&lt;/sup&gt;</th>
<th>P-value&lt;sup&gt;d)&lt;/sup&gt;</th>
<th>Fold change&lt;sup&gt;b)&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GRP170-P&lt;sub&gt;PDI1&lt;/sub&gt;</td>
<td>Δopi1</td>
<td>0.1050</td>
<td>0.7227</td>
<td>0.1482</td>
<td>0.1071</td>
<td>0.0127</td>
<td>6.37</td>
</tr>
<tr>
<td>2</td>
<td>GRP170-P&lt;sub&gt;KAR2&lt;/sub&gt;</td>
<td>Δopi1</td>
<td>0.0966</td>
<td>0.7371</td>
<td>0.1373</td>
<td>0.0714</td>
<td>0.0042</td>
<td>5.89</td>
</tr>
<tr>
<td>3</td>
<td>BiP-P&lt;sub&gt;PDI1&lt;/sub&gt;</td>
<td>Δopi1</td>
<td>0.1031</td>
<td>0.8027</td>
<td>0.1306</td>
<td>0.0714</td>
<td>0.0005</td>
<td>5.58</td>
</tr>
<tr>
<td>4</td>
<td>FKBP2-P&lt;sub&gt;PDI1&lt;/sub&gt;</td>
<td>Δopi1</td>
<td>0.1028</td>
<td>0.8069</td>
<td>0.1295</td>
<td>0.0357</td>
<td>0.0008</td>
<td>5.56</td>
</tr>
<tr>
<td>5</td>
<td>CypB-P&lt;sub&gt;KAR2&lt;/sub&gt;</td>
<td>Δopi1</td>
<td>0.0958</td>
<td>0.7880</td>
<td>0.1209</td>
<td>0.0179</td>
<td>0.0004</td>
<td>5.17</td>
</tr>
<tr>
<td>6</td>
<td>BiP-P&lt;sub&gt;KAR2&lt;/sub&gt;</td>
<td>Δopi1</td>
<td>0.0917</td>
<td>0.7873</td>
<td>0.1203</td>
<td>0.0357</td>
<td>0.0056</td>
<td>5.14</td>
</tr>
<tr>
<td>7</td>
<td>CypB-P&lt;sub&gt;PDI1&lt;/sub&gt;</td>
<td>Δopi1</td>
<td>0.0993</td>
<td>0.8571</td>
<td>0.1197</td>
<td>0.0357</td>
<td>0.0396</td>
<td>5.15</td>
</tr>
<tr>
<td>8</td>
<td>P5-P&lt;sub&gt;PDI1&lt;/sub&gt;</td>
<td>Δopi1</td>
<td>0.0897</td>
<td>0.7807</td>
<td>0.1168</td>
<td>0.0179</td>
<td>0.0051</td>
<td>5.05</td>
</tr>
<tr>
<td>9</td>
<td>ERp72-P&lt;sub&gt;PDI1&lt;/sub&gt;</td>
<td>Δopi1</td>
<td>0.0937</td>
<td>0.8453</td>
<td>0.1153</td>
<td>0.0536</td>
<td>0.0144</td>
<td>5.01</td>
</tr>
<tr>
<td>10</td>
<td>P5-P&lt;sub&gt;KAR2&lt;/sub&gt;</td>
<td>Δopi1</td>
<td>0.0904</td>
<td>0.8013</td>
<td>0.1152</td>
<td>0.0357</td>
<td>0.0274</td>
<td>4.95</td>
</tr>
<tr>
<td>35</td>
<td>none &lt;sup&gt;e)&lt;/sup&gt;</td>
<td>Δopi1</td>
<td>0.0792</td>
<td>0.9123</td>
<td>0.0918</td>
<td>0.1071</td>
<td>-</td>
<td>3.94</td>
</tr>
</tbody>
</table>

<sup>a)</sup> Ranking based on descending specific product yield within a parental strain background.

<sup>b)</sup> The mean from 0.5 and 2% induction at 30°C. Averages were calculated from fold changes that were first compared to the respective wild-type samples. OD<sub>600</sub>-values were measured with a microtiter plate and are not path length corrected: the values are proportional in approximately 1:5.5 ratio to measurements with regular path length of 1 cm.

<sup>c)</sup> Relative frequency in the 95th percentile of measurements within each parental strain at 30°C.

<sup>d)</sup> Calculated with the nonparametric equivalent of t-test, Wilcoxon signed rank test using the parental strain without added elements as the reference.

<sup>e)</sup> Values for the parental strains are added for comparison.

We aimed to characterize the best strains further so we cultivated a selection of strains in shake flasks and determined antibody titers at different time points. We chose the Δopi1-based strains expressing GRP170-P<sub>PDI1</sub>, BiP-P<sub>PDI1</sub> or FKBP2-P<sub>PDI1</sub>, which had displayed the highest titers, and additionally, we included the strains expressing P5-P<sub>GAL1</sub> or the GRP94-P<sub>PDI1/ERp72-P<sub>TEF</sub></sup> pair from the wt background. The differences in secretion between the strains and the absolute titers were similar in shake flasks to the screening measurements, but the different time points revealed interesting details of the secretion characteristics.
Evidently, measuring the endpoint titer did not reveal the pattern of how or when the folding factor improved IgG secretion. In the wt background, the folding factor co-expression actually decreased IgG production during the first six hours, after which the cells seemed to adapt to production and the titers increased above the control strain at 12 hours (Figure 20A).

The initial decrease in modified wt strains implies that the benefits in secretion resulted from various effects in ER transit and homeostasis due to the expression of several heterologous proteins. Furthermore, degradation of the product from culture supernatant seemed prevalent as the decrease in titers within strains from 12 hours to 24 hours suggested (Figure 20). The disappearance of the antibody product was more prominent in the wt strains than in the Δopi1 strains (Figure 20), possibly because of the decreased cell density of Δopi1-derived strains (Publication III, Fig. 4). Yeast cells can consume and degrade the target protein (Tyo et al., 2014), which proved to be a problem despite the supplementation of BSA in the media. Maintaining the titers in high levels might benefit from genetically blocking the endocytic internalization mechanisms leading to target protein degradation as demonstrated by Rodriguez-Limas et al. (Rodriguez-Limas et al., 2015). In addition, deleting endogenous proteases from the production strain can improve titers, as rupturing cells can release these protein degrading enzymes to the culture medium (Tomimoto et al., 2013; Idiris et al., 2010).

Figure 20 Enhancement in antibody titers occurred differentially in the engineered strains. (A) Antibody titers in four time points for strains of wt background, where maximum titers were reached already at 12 hours. (B) Titers for strains in the Δopi1 background. Constructs in the strains are as follows: G+E; GRP94-PDII and ERp72-PDII pair expression, GRP; GRP170-PDII, FKBP; FKBP2-PDII and CPR5 is the yeast PPIase Cpr5p expressed under PGPD. BiP and P5 were expressed with constructs BiP-PDII and P5-PGAL1, respectively. Points represent the mean of three biological replicates, each measured in at least three technical replicates.
As pointed out for human BiP and Kar2p in section 4.4, and shown elsewhere for human and yeast calnexins (Ćiplys et al., 2011b), and yeast and rat PDI (Hackel et al., 2006), the functional homologs are not equivalent in their effects on protein production. For most of the folding factors, we measured variable effects for both yeast and mammalian variants, but in both studies, PPIases struck out as improving IgG secretion significantly (Publication II, III). To gain insights into the secretion behavior of homologous PPIases, the Δopi1-based strain expressing the yeast ER localized PPIase construct P_{GPD}-CPR5 (Publication II) was included in the shake flask experiments. Cpr5p seemed superior to mammalian folding factors as the strain recorded the highest titer of 159±28ng/ml after 24 hours of cultivation in shake flasks (Figure 20B). The final titer from the FKBP2 expressing strain was 135±30ng/ml, which was effectively reached already after 12 hours (Figure 20B). Although Cpr5p increased specific product yield to a greater extent than FKBP2, the difference in final titers was relatively small and displayed a certain degree of statistical uncertainty (P-value was 0.1019 with unpaired Student’s t-test). However, the specific product yield is dependent on the final cell density, and throughout the experiments we noticed that high titers were accompanied with a decrease in OD600 values (Table 6; Publications II, III). The systematic decrease in cell density associated with high specific product yields is potentially problematic when the industrial adaptability of the engineered strains is evaluated. As Cpr5p had a stronger impact on cell density, and correspondingly reached final peak titer slower, the strain with Cpr5p might not outperform FKBP2 in large scale cultivations. In general, the antibody titers reached here are modest compared to commercially viable levels, where titers from mammalian cell lines and for example, a glycoengineered P. pastoris strain, are orders of magnitude higher (Kunert and Reinhart, 2016; Ye et al., 2011). Nevertheless, our screening resulted in strains with approximately 3-4 times the titers of the wt strain (Figure 20), which is a significant improvement and can be largely contributed to the enhancement of PPIase activity in the yeast ER.

Pybus et al. demonstrated that the PPIase CypB increased mAb titers specifically through improved CHO culture viability (Pybus et al., 2014). It seems unlikely that mammalian folding factor expression would improve viability in yeast, and equally unlikely for the yeast PPIase, which led to a decrease in final cell density. Considering folding enhancement, Cpr5p might have expanded substrate specificity, unlike FKBP2 and CypB, which may catalyze only certain peptidyl-prolyl isomerizations on antibodies (Lilie et al., 1995). Every antibody domain undergoes proline isomerizations, and these reactions have been identified as a rate-limiting step in several domains (Lilie et al., 1995; Thies et al., 1999; Feige et al.,
Differential substrate specificity would explain the difference in titers between the PPIase overexpressions, if Cpr5p acted on more prolines or had higher activity on the rate-limiting one than the mammalian equivalents. The substrate specificity of yeast PPIases is not characterized and generally, the significance of these enzymes in the yeast ER is questionable (Dolinski et al., 1997; Arevalo-Rodriguez et al., 2004). Based on our results, we postulate that overexpression of ER-localized PPIase is a prerequisite to increase IgG titers in yeast. Considering the importance of the proline isomerization reaction in antibody folding, this engineering approach should be extended to other similar expression platforms, such as the yeast *P. pastoris*.

### 4.6 New genetic targets from exploration of plasma cell differentiation transcriptomics data

B cell differentiation has been characterized comprehensively with data from several omics-platforms (Le Gallou et al., 2012; Romijn et al., 2005; Salonen et al., 2006; Shaffer et al., 2002). Effectively, all the molecular changes needed for creating an efficient antibody-secreting cell are hidden in the wealth of these extensive data sets. As not all the genetic changes are obvious, we aimed to re-analyze some of the available data sets to identify novel genetic targets, which might be relevant in *S. cerevisiae*.

#### 4.6.1 The data analysis workflow identified several recognized and novel engineering targets

For data exploration, we selected two transcriptomics data sets that were realized on an illumina-platform. The analyzed differentiation points covered several distinct cellular stages between mature B-cells and long-lived plasma cells (Le Gallou et al., 2012; Cocco et al., 2012). To extract the essential modifications in each phase of differentiation, we divided the analysis into two phases: from mature B cells to plasmablast phenotype (PB), and the plasmacytic phase from plasmablasts to plasma cells (PC). After we determined the differentially expressed genes in each phase, we divided them roughly to up- and downregulated genes based on the mean fold changes in their expression. Before GO-term enrichment analysis, we converted the identified genes to their *S. cerevisiae* homologs, whenever applicable. From all genes annotated to the microarray probes, a total of 1364 yeast homologs were identified, covering approximately 25 % of the yeast genes. In the end, we obtained a list of GO-terms relevant in yeast, and we could explore the *S. cerevisiae*-specific targets based on their annotated biological processes. Our workflow enabled us to explore new genetic targets easily through their functional significance.
In the early PB phase, we obtained a list of 45 and 94 yeast homologs in the up- and
downregulated groups, respectively. The associated GO-terms were dominated by processes
involved in DNA metabolism and cell division (Figure 21A), reflecting the recombination
and proliferation events occurring during the affinity maturation in B cell differentiation
(Victora and Nussenzweig, 2012). Since we decided to exclude genes associated with
essential biological processes, such as cell organization and division, central metabolism,
DNA, RNA and cytoskeleton related terms, from the modifications of yeast cells, the PB
analysis did not yield many interesting gene candidates for laboratory experiments.
However, from the group of downregulated genes we identified \textit{DPH2}, which participates
in modifying translation initiation factors in \textit{S. cerevisiae} (see Table 7 for details). In order
to determine whether \textit{DPH2} could affect IgG production by regulating translation, we
decided to test the deletion of this gene.

In the PC analysis, we obtained a larger number of yeast homologs in the up- and
downregulated groups, 255 and 219 genes, respectively. A major portion of the cellular
transformation occurs in the plasmacytic phase of differentiation covered by the PC analysis,
and accordingly, several interesting biological processes were enriched in the GO-terms
which have a clear connection to polypeptide modification and secretion, and thus to IgG
production (Figure 21B). The genes linked to the GO-terms depicted in Figure 21B included
several genetic candidates that have already been analyzed in recombinant protein
production in yeast, indicating that the human-specific genetic events showed some level of
transferability to the yeast platform. For example, \textit{KAR2} and \textit{SIL1} appeared in the
upregulated-group. Accordingly, overexpression of these yeast genes has been a rational
approach in engineering of protein folding, also in the case of antibodies (Hackel et al., 2006;
Payne et al., 2008a; Delic et al., 2014). Interestingly, deletion of several different vacuolar
sorting proteins have been linked to high-secretion phenotypes in random screening
experiments (Zhang et al., 2001; de Ruijter et al., 2017) and accordingly, at least Vps41p
and Vps45p appeared in the category of downregulated genes in the PC analysis. Thus, at
least the differentiation phases in PC analysis were successful in listing several genetic
targets that affect the protein and antibody secretion in \textit{S. cerevisiae}. 

Results and discussion

[371x686]
Figure 21 Enriched GO-terms identified from yeast gene targets after plasma cell differentiation analysis and homology conversion. A selection of the yeast-relevant enriched GO-terms for biological processes are shown from early phases of plasma cell differentiation from mature B cells to plasmablasts (A) and for terminal differentiation from plasmablasts to long-lived plasma cells (B). Genes were roughly divided into up- and down-regulated groups. The color of the nodes reflect in which group the majority of genes in that GO-term were, red for upregulated and blue for downregulated. Size of the node indicates the number of genes linked to the process in question and the color intensity reflects the statistical reliability of the GO-term enrichment. The networks in (A) and (B) were created with different setting of GO-levels and network connectivity and thus are not directly comparable. GO-analysis was conducted with Cytoscape (Shannon et al., 2003) extended with the ClueGO-app (Bindea et al., 2013, 2009).
4.6.2 Two of the identified modifications, overexpression of Got1p and Ire1p, improved IgG secretion significantly

In addition to the already mentioned deletion of DPH2, we selected additional candidates from the PC analysis for laboratory experiments. We aimed to have a large variety of untested targets, in which well-characterized genes were accompanied by genes with no clear function or relevance to the antibody production process. The tested genetic targets and their secretion phenotypes are summarized in Table 7. Considering our efforts to select unexpected modifications, it is not surprising that many of them did not have a significant effect on antibody secretion (Table 7). However, we identified two new genes that induced a significant increase in IgG secretion upon overexpression: GOT1 and IRE1. Although the deletion of DPH2 induced a significant increase in specific product yield, this was only due to a decrease in final cell density, as the titer was comparable to the control strain (Publication IV). ATE1 was actually identified from the group of upregulated genes, but we decided to analyze the deletion phenotype because of the reported connection of ATE1 overexpression with increased rates of apoptosis (Kumar et al., 2016). Considering this, it is interesting that inducing a modification to the opposite direction as suggested by the analysis correspondingly led to a minor decrease in secretion (Table 7).

Table 7: The selected gene candidates and their effects on antibody secretion.

<table>
<thead>
<tr>
<th>Modification</th>
<th>Gene function</th>
<th>Titer (ng/ml)</th>
<th>Final OD600</th>
<th>Fold change</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Δate1</td>
<td>Arginyl-tRNA-protein transferase; N-end rule degradation pathway</td>
<td>18.17±3.25</td>
<td>11.70</td>
<td>0.83</td>
<td>0.126</td>
</tr>
<tr>
<td>Δdph2</td>
<td>Protein required for synthesis of diphthamide (a modified histidine residue of translation elongation factor 2 (Eft1p or Eft2p))</td>
<td>23.67±4.18</td>
<td>8.73</td>
<td>1.48</td>
<td>0.011</td>
</tr>
<tr>
<td>Δutr4</td>
<td>involved in methionine salvage</td>
<td>24.83±7.52</td>
<td>10.28</td>
<td>1.30</td>
<td>0.108</td>
</tr>
<tr>
<td>ΔGOT1</td>
<td>COPII vesicles; cycles between the ER and Golgi; involved in secretory transport</td>
<td>35.33±6.25</td>
<td>11.42</td>
<td>1.63</td>
<td>0.001</td>
</tr>
<tr>
<td>ΔGSH2</td>
<td>Glutathione synthetase</td>
<td>22.50±3.94</td>
<td>11.45</td>
<td>1.10</td>
<td>0.577</td>
</tr>
<tr>
<td>ΔHUT1</td>
<td>Function unclear: involved in sugar nucleotide transport</td>
<td>24.50±4.89</td>
<td>11.39</td>
<td>1.18</td>
<td>0.274</td>
</tr>
<tr>
<td>ΔIRE1</td>
<td>ER-resident sensor for unfolded proteins: initiates UPR</td>
<td>69.83±17.27</td>
<td>10.77</td>
<td>3.55</td>
<td>0.003</td>
</tr>
</tbody>
</table>

a) Both gene deletions (Δ and small characters) and overexpressions (capital letters, italics) were investigated based on the available literature and the direction of the change in the B cell differentiation analysis. In overexpressions, the gene was expressed under the TEF-promoter.

b) As reported in Saccharomyces genome database (www.yeastgenome.org)

c) Means of three biological replicates are shown. Error values indicate standard deviation. Fold change was calculated by normalizing the specific product yield with the respective value of the control strain.

d) P-values were determined with unpaired Student’s t-test.
Although modulating the UPR-response has been demonstrated to enhance recombinant protein production, the modification has always occurred downstream of \textit{IRE1}, targeting the expression of \textit{HAC1} \cite{Gasser2006, Valkonen2003}. As the UPR is centrally involved in coping with protein production, it is hardly surprising that targeting the signaling pathway affected secretion markedly. Ire1p forms a dimer upon activation and UPR signal initiation \cite{Credle2005, Shamu1996}, so increasing the concentration of Ire1p is likely to increase the rate of dimerization effectively sensitizing the cell for unfolded proteins. However, the cells putatively retain the possibility to silence the UPR signaling, which is not possible upon active Hac1p overexpression. Our results suggested that overexpression of Ire1p is a better strategy to modulate UPR, as \textit{IRE1} improved secretion up to 3.6-fold (Table 7), considerably more than the overexpression of Hac1p \cite{PublicationIV, Figure3}.

In contrast to \textit{IRE1}, \textit{GOT1} appeared as one of the nondeducible targets to improve secretion. There are two, partially contradictory reports for the function of Got1p. Conchon et al. showed that deletion of \textit{GOT1} reduces transport efficiency substantially but Lorente-Rodríguez et al. linked this phenotype to the other genetic changes present in the original \textit{Δgot1} strain from the Research Genetics collection \cite{Conchon1999, Lorente-Rodriguez2009}. Actually, Lorente-Rodríguez et al. reported that strong overexpression of Got1p reduced secretion, but their conclusion was not based on direct evidence but on phenotypic comparison and on the increased amount of carboxypeptidase Y (CPY) precursor form. They ignored the fact that the amount of the mature form of CPY also seemed to increase in their experiments \cite{Lorente-Rodriguez2009}. \textit{GOT1} is not an essential gene, but appears to play a role in ER-to-Golgi transport. It is packaged into COPII-vesicles and Conchon et al. suggested that it mediates certain vesicle fusion events on the target membrane \cite{Conchon1999, Lorente-Rodriguez2009}. Interestingly, overexpression of Got1p induces morphological changes in the ER and Golgi, increasing the prevalence of ER membranes and creating diffuse Golgi patterns \cite{Lorente-Rodriguez2009}. It is possible that actually these structural changes in the secretory organelles induce the increase in secretion, as the overexpression of another factor, Sly1p, which functions in the ER-to-Golgi transport, did not display any effects on IgG secretion \cite{PublicationIV}.

Although we drafted this model-guided approach, the evolutionary distance between yeast and humans hinders us from considering the engineering strategy as a rational design. Candidates might occur at random and through to statistical chance, but as we evaluated each gene individually, vigorous accuracy and data integration in the analysis phase was not
considered necessary. Obviously, the parameters in the analysis influence the exact list of genes obtained in the end. Effectively, our approach resembled more of a random screen, but with the experimental control of a targeted approach. In contrast to random approaches (de Ruijter et al., 2017), none of the modifications actually induced a significant decrease in IgG secretion (Table 7). In the end, we were able to identify two new modifications, overexpression of *IRE1* and *GOT1*, which improved IgG secretion 3.6- and 1.6-fold, respectively. Such data exploration provides an interesting engineering approach for the future, to be used in case of other production platforms or with other target proteins.

### 4.7 Quality considerations: not all the antibody is in the full-length form

Throughout this study, the IgG concentrations presented were measured with ELISA, in which both coating and detection antibodies target the heavy chain. Based on the reports that the heavy chain is not secreted alone and that heavy chain availability limits antibody secretion (Gasser et al., 2006; McLeod et al., 2011), we reasoned that the heavy chain concentration in the supernatant would be the most descriptive estimate for the protein secretion capacity of the strains. However, there is a question about product quality. The antibody tetramer is not native to *S. cerevisiae*, thus there is no guarantee for complete assembly of IgG in this organism. Light chains can be secreted separately (Gasser et al., 2006; McLeod et al., 2011), and various dimers and degradation fragments are possible. Unfortunately, our titers are too low for direct visualization of the antibody product on SDS-PAGE or Western blot, and different concentration methods hardly retain the original assembly of the complex antibody product. To gain more insight to product quality, we assessed the antibody structures from *S. cerevisiae* with variations of the ELISA measurements. In addition, we created a *P. pastoris* strain, which secreted the same antibody in considerably higher amounts than *S. cerevisiae*. With this other yeast, titers in the culture supernatants were high enough for discrete bands to appear on Western blot, depicting what kind of fragments are secreted and which are the dominant variants.

Optimally, if the antibody was in uniform format, all ELISA variations would give the same final concentrations. As shown in Figure 22A, this is hardly the case. Although light and heavy chains appear in nearly equimolar amounts, the concentration of fragments with complete heavy chains attached to a light chain is strikingly low. We assessed the amount of full-length fragments by using a light-chain specific antibody for detection, while the coating antibody was heavy-chain specific. As these antibodies are polyclonal, is hard to assess which part or domain they are specific for. It seems counterintuitive that light and heavy
chains would appear in comparable amounts, but that they would not be associated with each other. Another explanation for the low concentration is that they occur in forms not expanding the complete heavy chain length as a tetramer. This seems possible as illustrated by the fragments secreted by *P. pastoris* (Figure 22B,C). In addition to full-length tetramer, the antibody was secreted as a dimer, and further, several additional small fragments were visible in Western blots probed with anti-heavy chain (Figure 22B) and anti-light chain (Figure 22C) detection antibodies. Thus, the antibody product from yeast displayed significant heterogeneity which could not be fully assessed or explained by the present methods.

Clearly, significant efforts need to be targeted to the improvement of product quality. Interestingly, Figure 22B shows that the majority of the secreted antibody from *P. pastoris* is actually in a form of a dimer. The tetramer interface between heavy chain C\(_2\) domains is constructed solely from the carbohydrate moiety (Huber et al., 1976), implying that not all forms of N-glycans can be accommodated in the antibody structure. Yeasts have considerably larger N-glycans with hypermannosylation, which differ from the mammalian complex N-glycans natively found on antibodies (Figure 4). Hypermannosylation in the Golgi complex might be extensive enough to pry the antibody tetramer into dimers, as suggested by the prevalence of the dimers in Figure 22B. Further, unspecific O-mannosylation occurring in yeast might affect the performance of the detection methods.

---

**Figure 22** Quality of the secreted antibody from yeast is not optimal. (A) The IgG concentration from culture supernatants of *S. cerevisiae* by using antibodies with different specificity in the measurements. In HC, both coating and detection antibodies target the heavy chain, and in LC, both recognize the light chain. HC-LC is a combination of heavy-chain specific coating antibody and light chain specific detection antibody. Western blot analysis of *Pichia pastoris* culture supernatant after IgG expression, probed with an anti-heavy chain (B) and anti-light chain detection antibody (C) revealed, that several forms of the antibody fragments are present. Estimations of the configurations of fragments in some of the bands are shown in the schematics, where heavy chains are colored black and light chain domains are white rectangles. The calculated sizes for the heavy and light chains in the expression constructs are 49.2 kDa and 23.1 kDa, respectively.
(Nett et al., 2013). Reduction of O-glycan occupancy has been shown to increase antibody titers, which can be a result from improved recognition by the antibodies used in the measurements (Kuroda et al., 2008). Our heavy and light chains may be differently O-glycosylated, which can affect the concentrations obtained from ELISA-measurements. These structural implications give more weight to the importance of glycoengineering, which evidently should be implemented in all phases of strain engineering in the case of full-length antibodies.
5. Conclusions and future prospects

In this thesis work, we investigated the use of plasma cells as a blueprint to improve IgG secretion in yeast. In the end, the approach enabled us to create a yeast strain with a more than 6-fold increase in specific product yield, and finally, to increase the initial titer of 40 ng/ml up to 160 ng/ml. The first evident question is whether using the blueprint created by evolution provided an advantage over other cellular engineering strategies. Although in the end the yeast PPIase Cpr5p was responsible for the highest improvements, we showed that the effects achieved with mammalian folding factors were comparable to the enhancements reported for yeast folding factor overexpressions. Notably, following the example of plasma cells led us to discover several novel modifications, which improved IgG secretion, such as the modifications to ER size and shape, and the gene candidates identified from the transcriptomics data exploration. As discussed by Dinnis and James (Dinnis and James, 2005), the use of plasma cells as a model is intriguing for mammalian cell systems, but as we demonstrated here, the model can be extended to evolutionary distant expression platforms, such as S. cerevisiae. However, plasma cell differentiation might provide only a limited model for growing cell cultures since plasma cells are non-dividing. Furthermore, the majority of plasma cells are short-lived (Manz et al., 1997), so following the example of B cell differentiation obediently does not necessarily result in a sustainable process. This is more of a problem for mammalian systems, in which maintaining viability and growth is already a major engineering target (Fischer et al., 2015), but in yeast such a linear use of the model would probably not be feasible. Here, we adopted selected features of plasma cell morphology and antibody folding to yeast, leading to changes in antibody secretion. Additionally, Swers et al. demonstrated that the B cell differentiation model can be used in S. cerevisiae for other purposes, in their case for introducing diversity to antibody molecules for improved and altered affinity (Swers et al., 2011). Clearly, nature has provided a comprehensive collection of molecular clues that can be exploited in a number of ways in antibody production and engineering.
Especially the modifications to the organelle size and shape were effective in improving antibody secretion. The increase in ER size by the OPI1 deletion created a promising yeast chassis for further antibody cell factory design. An interesting additional modification to this strains would be the corresponding expansion of the Golgi apparatus, similarly to the plasma cell model (Wiest et al., 1990; Shaffer et al., 2004). Notably, also the shape of the Golgi apparatus should be considered, as the yeast S. cerevisiae natively exhibits dispersed Golgi structures. It would be interesting to test whether the mammalian stacked Golgi morphology provides a secretory advantage in S. cerevisiae, in reflection to the importance of ER shape demonstrated in this study. Interestingly, one of our novel gene targets, the secretory protein Got1p, also induced altered ER and Golgi morphology (Lorente-Rodríguez et al., 2009), which could explain the benefits of GOT1 overexpression on antibody production. Much remains unknown from the relationship of structure and function of cellular organelles, which should be a focus of a number of fundamental and applied studies in the future.

Synthetic biology provides a comprehensive framework for re-designing biological systems for useful purposes, and the methods and ideas of synthetic biology bring new aspects to metabolic engineering and recombinant protein production. In this study, we contributed to the toolkit of synthetic biology by describing a new version of high-throughput cloning. In addition, we implemented the principles of synthetic biology to the folding factor library design in attempt to include modularity in the expression cassettes and combinations. However, we did not manage to improve predictability of engineering of protein folding, possibly due to nonlinearity in the expression levels (Figure 14). Effects of folding factor overexpression are not only specific for a particular protein type but can differ within the same IgG subclass (Pybus et al., 2014). The use of screening approaches, such as presented here, to optimize the expression platform for every different protein product is laborious and expensive. Thorough characterization of the expression cassettes of folding factors and their cellular environment could contribute to the construction of a predictive folding model, which could reduce the efforts needed to optimize folding and secretion. Our model, plasma cells, was able to provide us with a multitude of engineering targets, also through omics-level analysis. The plasma cell derived information on the identified targets could be integrated in a full-scale model of the yeast secretory system, such as the one described by Feizi et al. (Feizi et al., 2013), to predict expression levels and effects of the modification on antibody production. With the efforts of synthetic biology, the usability and predictive power of such integrative models are increasing continuously. With further elucidation of expression and system dynamics, our results of the folding factor library screening might
provide experimental data for constructing a model for engineering of protein folding to streamline future synthetic biology approaches for recombinant protein production.

Lastly, the results in this work provided new matter to the evaluation of the potential of yeasts as antibody factories. Significantly, antibody secretion was multiplied through genetic engineering, although implementation of several ER modifications into one strain had limited effects. The final peak titer in *S. cerevisiae*, 160 ng/ml is still modest compared to commercial mammalian systems (Kunert and Reinhart, 2016) and lagged behind the other yeast used here, *P. pastoris*. Major issues remain in the assessment and improvement of antibody quality, which is prominently far from the strict standards required for therapeutic applications. However, the value of using such a flexible system as *S. cerevisiae* should not be neglected. The screening approach of folding factors or many of the modifications could not have been so easily tested with any other systems due to lack of experimental methods and system characterization. As we implemented the model of plasma cells to the design of *S. cerevisiae*, our results could be applied to other, commercially relevant systems. Additionally, future applications and innovations often arise unexpectedly and unpredictably. Considering response time, yeasts are clearly superior to mammalian cells. Generating a stable CHO cell line can take up to 25 weeks, and further several weeks before fermentation batches are completed. With yeast strains, the whole workflow has a usual duration of 8-12 weeks (Fujifilm Diosynth biotechnologies, personal communication). In the era centered on personalized medicine, enabling fast production of tailored biologicals and biosimilars for diagnostics and therapy seems like an intriguing vision for the future. Although the realization of personalized biopharmaceuticals is still a regulatory impossibility, it is easy to picture yeast platforms in a central role in the development of the industry.
References


Chumnanpuen, P., I. Nookaew, and J. Nielsen. 2013. Integrated analysis, transcriptome-lipidome, reveals the effects of 


Dunning, M., A. Lynch, and M. Eldridge. illuminaHumanv4.db: Illumina HumanHT12v4 annotation data (chip illuminaHumanv4).


Fischer, S., R. Handrick, and K. Otte. 2015. The art of CHO cell engineering: A comprehensive retrospect and


Goto, Y., and K. Hamaguchi. 1979. The Role of the intrachain disulfide bond in the conformation and stability


References


essential protein homologous in sequence and function to mammalian BiP. Cell. 30:1223–1236.


References


Biopharmaceutical antibodies such as IgGs are proteins that can be used as medicine to combat serious illnesses, including cancer and autoimmune diseases. Protein drugs are produced biotechnologically with living organisms, but the current commercial methods to produce antibodies are expensive and inadequate. The yeast *Saccharomyces cerevisiae* has the potential to become an IgG factory through applying the versatile genetic toolkit to engineer this microbe up for the task. In this thesis, we exploited these molecular methods to improve IgG production and secretion in yeast. We used human body’s natural IgG producer cells, plasma cells, as a blueprint to guide our engineering strategy.