Metabolic Engineering of Recombinant Protein Productions by *Saccharomyces cerevisiae*

Zihe Liu

*Department of Chemical and Biological Engineering*

CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden 2012
Metabolic Engineering of Recombinant Protein Production by

*Saccharomyces cerevisiae*

ZIHE FLORA LIU

Department of Chemical and Biological Engineering

CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden 2012
Metabolic Engineering of Recombinant Protein Production by *Saccharomyces cerevisiae*

Zihe Flora Liu

© *Zihe Flora Liu, 2012*


Doktorsavhandlingar vid Chalmers tekniska högskola

Ny serie nr 3422

ISSN 0346-718X

Department of Chemical and Biological Engineering

Chalmers University of Technology

SE-412 96 Göteborg

Sweden

Telephone + 46 (0) 31-772 1000

Cover: Technology scheme of approaches applied by Zihe Flora Liu for recombinant protein production in this thesis.

Printed by Chalmers Reproservice

Göteborg, Sweden 2012
To my parents…

献给我的爸爸妈妈……

“As a scientist, you must constantly question yourself, your experiment and your data to find out what you might have overlooked, why the experiment did not work or why the results contradict your initial assumptions.”

*Leonardo Almeida-Souza & Jonathan Baets*

花开六月降 天福，偏白城廿年。
人间渐觉香彻骨，廊苑奇葩古今谁与比。
昨朝锦花今更胜， 厚土绿叶伴。
明日再寻花香来，是处繁花满城世间传。

*A poem full of love and compliments from Shuobo to Zihe Flora Liu*

Systems Biology Enabled Engineering for Recombinant Protein Production in *Saccharomyces cerevisiae*

Zihe Flora Liu
The yeast *Saccharomyces cerevisiae* is a widely used cell factory for the production of fuels, chemicals, and it also provides a platform for the production of many heterologous proteins of medical or industrial interest. In this thesis, random and rational approaches, such as vector design, host engineering, fermentation analysis, UV Mutation, coupled with high-throughput systems biology techniques (including whole genomic sequencing, microarray analysis and flux analysis) and integrated analysis (Reporter feature technique), were employed to engineer cellular properties more effectively and purposefully to construct cell factories for protein production. We reported that insulin production mainly depends on the expression level of the gene, whereas amylase tends to achieve higher secretion at lower growth conditions in order to reduce ER stress. Moreover, based on large data generated and systems biology tools, we proposed models to address unknown questions regarding recombinant protein production: i) the futile cycle of protein folding in the ER and the thermodynamic model of non-stoichiometric production of reactive oxygen species explains the oxidative stress that occurred during recombinant protein production, and ii) the final electron acceptor for protein folding and the electron transferring model at anaerobic condition proposed potential electron consuming pathway for protein folding in the ER. Our research provided a set of expression systems that can be used for high-level expression of recombinant proteins in connection with the use of yeast for consolidated bioprocesses, potential targets for future engineering, as well as shed lights for the processing of protein secretory pathway and basic cellular metabolisms.

Keywords: recombinant protein production, α-amylase, insulin precursor, secretory pathway, unfolded protein response, systems biology, UV mutation, anaerobic electron acceptor, growth rate, yeast
PREFACE

This dissertation is submitted for the partial fulfillment of the degree of doctor of philosophy, carried out at the Systems and Synthetic Biology group (Sys²Bio), Department of Chemical and Biological Engineering, Chalmers University of Technology between 2008 and 2012, under supervision of Professor Jens Nielsen. This research was funded by the EU Framework VII project SYSINBIO (grant no. 212766), European Research Council ERC project INSYSBIO (grant no. 247013), the Chalmers Foundation, the Knut and Alice Wallenberg Foundation and the Novo Nordisk Foundation.

Zihe Flora Liu

September 2012
LIST OF PUBLICATIONS

This thesis is based on the following publications:


Other publications related to my thesis:


CONTRIBUTION SUMMARY

A summary of the author’s contribution to the publications on which this thesis is based is provided below:

Paper I
JN conceived the project. DP and JN directed the study. ZL performed the experiments. ZL and KT analyzed the data. ZL wrote the manuscript. All authors have edited and approved of the final manuscript.

Paper II
KT, DP, and JN designed the experiment. DP and JN supervised the research. KT and ZL carried out all cloning, fermentations, and analytical measurements. KT did primary calculations in transcriptomics and metabolic flux data. KT wrote the manuscript. All authors have edited and approved of the final manuscript.

Paper III
JN conceived the project. DP and JN directed the study. ZL carried out all mutagenesis, fermentations and analytical measurements. LL assisted the mutated plasmid evaluation. ZL, TÖ and JH analyzed the genomic sequencing and transcriptomics data. ZL wrote the manuscript. All authors have edited and approved of the final manuscript.

Paper IV
ZL, DP, and JN conceived the project. DP and JN directed the study. ZL performed the experiments. ZL, TÖ and JH analyzed the transcriptomics data. ZL wrote the manuscript. All authors have edited and approved of the final manuscript.

Paper V
ZL, DP, and JN designed the experiment. DP and JN supervised the research. ZL and JM performed fermentations. ZL and JH analyzed the data and wrote the manuscript. All authors have edited and approved of the final manuscript.

Paper VI
DP and JN conceived the project. KT and JH wrote the scaffold part. ZL and JH wrote the biotechnology part. All authors have edited and approved of the final manuscript.
# TABLE OF CONTENTS

ABSTRACT .............................................................................................................................. IV  
PREFACE.................................................................................................................................... V  
LIST OF PUBLICATIONS ........................................................................................................ VI  
TABLE OF CONTENTS ........................................................................................................ VIII  
TABLE OF FIGURES ................................................................................................................ X  
TABLE OF TABLES ................................................................................................................ X  

1.0 RECOMBINANT PROTEIN PRODUCTION ........................................................................ 1  

2.0 RECOMBINANT PROTEIN PRODUCTION IN *S. CEREVISIAE* ...................................... 5  

3.0 THE SECRETORY PATHWAYS IN *S. CEREVISIAE* ...................................................... 7  
  3.1 The Endoplasmic Reticulum Protein Processing ................................................................. 8  
  3.2 Unfolded Protein Response .................................................................................................. 10  

4.0 OPTIMIZATION OF RECOMBINANT PROTEIN PRODUCTION .................................. 12  
  4.1 Expression Design .............................................................................................................. 12  
    4.1.1 Plasmid copy numbers .................................................................................................. 12  
    4.1.2 Marker ...................................................................................................................... 13  
    4.1.3 Promoters .................................................................................................................. 14  
    4.1.4 Leader sequence ........................................................................................................ 15  
  4.2 Host Engineering ............................................................................................................... 16  
    4.2.1 Glycosylation engineering .......................................................................................... 16  
    4.2.2 Disulfide bond formation engineering ........................................................................ 17  
    4.2.3 Engineering protein trafficking .................................................................................. 19  
    4.2.4 Engineering protein degradation pathways .............................................................. 19  
    4.2.5 Random mutagenesis and screening ........................................................................... 20  
  4.3 Fermentation optimization ................................................................................................. 20  
    4.3.1 Temperature and pH ................................................................................................. 21  
    4.3.2 Aeration conditions .................................................................................................... 22  
    4.3.3 Culture Additives ....................................................................................................... 22  
    4.3.4 Cell growth ................................................................................................................ 23  

5.0 SYSTEMS BIOLOGY TOOLS FOR PROTEIN PRODUCTION ................................... 24  
  5.1 Bottom-up Approach: Mathematical Models of the Secretory Pathway ......................... 25
5.2 Top-down Approach: -Omics Analysis ................................................................. 25
6.0 SUMMARY OF RESULTS ...................................................................................... 28
  6.1 Experimental Design .......................................................................................... 28
  6.2 Paper I: Expression systems affect recombinant protein secretion ..................... 31
  6.3 Paper II: Host Engineering-the UPR Dependent and Independent Cell Metabolism .... 34
  6.4 Paper III: Host Engineering-Integrated Analysis of UV Mutation Strains ............... 37
  6.5 Paper IV: Fermentation Analysis-Anaerobic Protein Processing Machinery .......... 43
  6.6 Paper V: Fermentation Analysis-Cell Growth Effects ......................................... 48
CONCLUSIONS AND PERSPECTIVES .................................................................. 52
ACKNOWLEDGEMENTS ...................................................................................... 55
REFERENCES .................................................................................................... 57
TABLE OF FIGURES

Figure 1. Overview of recombinant protein production hosts. ................................................................. 2
Figure 2. The secretory pathway for protein processing............................................................................... 8
Figure 3. Unfolded protein response mechanism.......................................................................................... 11
Figure 4. Illustration of top-down and bottom-up systems biology............................................................... 24
Figure 5. Overall strategies for recombinant protein production................................................................. 28
Figure 6. Recombinant secreted proteins applied in this thesis work.......................................................... 29
Figure 7. Overview of plasmid construction ............................................................................................... 31
Figure 8. Final protein production results for the expression project........................................................... 32
Figure 9. Secretion profiles of IP and α-amylase strains for the expression project......................................... 33
Figure 10. Secretory perturbations and yeast physiology in the UPR project............................................... 35
Figure 11. Cycle thermodynamic model proposes non-stoichiometric ROS produced with incorrect disulfide bond formation........................................................................................................... 37
Figure 12. Experimental design of the random mutagenesis project.............................................................. 38
Figure 13. Amylase production of the UV mutation and reference strains..................................................... 40
Figure 14. Reporter TFs analysis reveals reduced stress responses in mutated strains...................................... 41
Figure 15. Top ten regulated secretory pathway functions in mutated strains............................................. 41
Figure 16. Whole genome sequencing analysis ............................................................................................ 42
Figure 17. Amylase produced more at anaerobic conditions compared to aerobic conditions...................... 44
Figure 18. Common and distinct pathways that were significantly changed at transcriptional level in a RPP-dependent manner under anaerobic and aerobic conditions...................................................... 45
Figure 19. Top 15 reporter metabolites in the three strains when comparing anaerobic to aerobic conditions................................................................................................................................................. 46
Figure 20. Anaerobic electron transfer model predicts fumarate to be the final electron acceptor for protein folding........................................................................................................................................... 47
Figure 21. Fumarate promotes cell growth at anaerobic conditions............................................................... 48
Figure 22. The heterologous protein production in chemostat cultivations................................................ 49
Figure 23. Common and distinct pathways that were regulated at different dilution rates........................... 51

TABLE OF TABLES

Table 1 Advantages and disadvantages of different host systems for protein production............................ 2
Table 2. Benchmarks for recombinant protein produced in different host species....................................... 4
Table 3. Landscape of recombinant protein production in S. cerevisiae ....................................................... 6
Table 4. Effect of ER chaperone over-expression for recombinant protein production.................................... 18
Table 5. Strain information .......................................................................................................................... 30
Table 6. Physiological characterization of mutated strains ............................................................................. 39
Table 7. Genetic changes ............................................................................................................................. 43
Table 8. Genes associated with ER protein processing were regulated among different dilution rates .... 50
1.0 RECOMBINANT PROTEIN PRODUCTION

Proteins, including enzymes and building blocks of life, play crucial roles in cell signaling, immune systems and the cell cycle (Demain and Vaishnav 2009). Many human proteins have important values or great potentials as biopharmaceutical. However, simple extraction of protein from natural sources are often limited for multiple reasons, including very low concentrations which substantially increases the cost for down-stream processing and the risk of infectious contamination during the course of extraction (Porro, Sauer et al. 2005). Since the first recombinant pharmaceutical, recombinant human insulin produced from Escherichia coli, was approved for clinical use (1982), recombinant DNA technology and protein engineering have established an efficient tailor-made industry for protein production. Now there are over 300 biopharmaceuticals proteins and antibodies on the market, with more than $100 billion of sales (Langer 2012), and an annually growth between 15-18% (Schröder 2008). In addition, around 240 monoclonal antibody products and 120 recombinant proteins are in clinical trials (Walsh 2010). In parallel to this, the total market for industrial enzymes has reached $5.1 billion in 2009 and is expected to reach $6.5 billion by 2013 (Freedonia Group 2009).

The increasing demand for recombinant proteins calls for robust production hosts, efficient expression systems and appropriate cultivation conditions. The limitation is often in terms of obtaining upmost quantities at sufficiently low cost to allow for marketing (Werner 2004). Meanwhile, protein quality, stability, yield and productivity are also important factors to be considered. So far, recombinant proteins are produced using a range of different cell factories, including bacteria, yeast, filamentous fungi, insect cells, mammalian cells and cell free systems. As shown in Figure 1, around half of the protein based biopharmaceuticals are produced in microbial systems (~30% in Escherichia coli and ~20% in Saccharomyces cerevisiae), with the rest mainly being produced by mammalian cells (Martínez, Liu et al. 2012). For industrial enzymes, more than half are produced by fungi and 30% in bacteria (Demain and Vaishnav 2009). The common and specific characters for each species used for recombinant protein production are listed in Table 1 (Demain and Vaishnav 2009; Tokmakov, Kurotani et al. 2012).

Escherichia coli is the earliest platform used and also today a work horse for recombinant protein production. The main reasons are the high yield of recombinant protein, reaching up to 50% of total dry cell weight (Porro, Gasser et al. 2011), and the ability for high cell density cultivations (HCDC) reaching up to 100 g biomass per liter (Tripathi 2009). However bacteria suffer from plasmid instability and their limited capacity for post-translational modifications (PTMs) (Porro, Gasser et al. 2011). Additionally, proteins larger than 60 kDa or S-S rich proteins are generally difficult to obtain in soluble correct forms using E. coli (Grauslund, Nordlund et al. 2008).
Figure 1. Overview of recombinant protein production hosts.
(A) Biopharmaceuticals. (B) Industrial enzymes.

Compared to bacteria, the main advantage of yeast expression systems is the similarity of their secretory pathways with mammalian systems and the capacity to perform strict quality control (Hou, Tyo et al. 2012) and post-translational modifications (Tokmakov, Kurotani et al. 2012), including proteolytic processing of signal peptides, disulfide bond formation, subunit assembly, glycosylation, phosphorylation and as well as the ability to secrete proteins in their native forms to facilitate downstream processing (Freigassner, Pichler et al. 2009). Also, yeast systems eliminate contaminations of toxic pyrogens when comparing to E. coli, and do not contain microbial contamination or viral inclusions that were found in mammalian cells (Cregg, Vedvick et al. 1993; Çelik and Çalık 2011). Besides, compared to filamentous fungi yeast secrete much less endogenous proteins, which reduced the cost of the off-line process and chance of proteolytic degradation (Mattanovich, Branduardi et al. 2012). However, yeasts and filamentous fungi suffer from inability to perform correct mammalian PTMs, especially humanized glycosylation, except for a recently developed engineered strain of Pichia pastoris (Hamilton, Davidson et al. 2006; Jacobs, Geysens et al. 2008). Recent advances also make it possible to produce human-type glycosylated proteins in S. cerevisiae in the near future (Amano, Chiba et al. 2008; Chigira, Oka et al. 2008; De Pourcq, De Schutter et al. 2010).

Table 1 Advantages and disadvantages of different host systems for protein production

<table>
<thead>
<tr>
<th></th>
<th>E. coli</th>
<th>S. cerevisiae</th>
<th>Mammalian cells</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantages</strong></td>
<td>High density growth</td>
<td>High yields and cost effective</td>
<td>Produce high quality proteins</td>
</tr>
<tr>
<td></td>
<td>High yields</td>
<td>GRAS strain&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Produce humanized proteins</td>
</tr>
<tr>
<td></td>
<td>Ease of culture and modifications</td>
<td>Ease of culture and modifications</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Whole genome sequence available</td>
<td>High-throughput data available</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cost effective</td>
<td>Stable expression and secretion</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Can perform PTMs</td>
<td></td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td>Produce proteins without key PTM&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Produce hyper-glycosylation proteins</td>
<td>Slow growth and low yield</td>
</tr>
<tr>
<td></td>
<td>Protein produced require refolding</td>
<td>Secretory pathway varies from human</td>
<td>Expensive cultivation</td>
</tr>
<tr>
<td></td>
<td>Intracellular production</td>
<td></td>
<td>Contamination with viruses</td>
</tr>
<tr>
<td></td>
<td>Costly biomass waste treatment</td>
<td></td>
<td>Insufficient for functional studies</td>
</tr>
</tbody>
</table>

<sup>1</sup>Post-translational modifications, <sup>2</sup>Generally regarded as safe.
More than 40 recombinant proteins have been produced by *Saccharomyces cerevisiae* (Hou, Tyo et al. 2012), which is a well-established host system for commercialization of recombinant proteins, due to the deep knowledge of its physiology, the availability of a deletion collection (Giaever, Chu et al. 2002), the extensively reported functional genomics data (Petranovic, Tyo et al. 2010), and a long history of industrial use. Moreover, it has also been developed as probiotics for oral delivery of therapeutic proteins (Blanquet, Marol-Bonnin et al. 2001; Omara, Rash et al. 2010).

Nowadays *P. pastoris* has gained more and more success for recombinant protein production by means of quantity and quality, which makes it impossible to be overlooked, especially after the achievement in genetic engineering to make it able to produce human type N-glycosylated proteins (Choi, Bobrowicz et al. 2003). Compared to *S. cerevisiae*, *P. pastoris* prefers a respiratory mode of growth without accumulation of ethanol and acetate, which enables the ease of high cell density cultures (up to 200 g/L) (Heyland, Fu et al. 2010). Moreover, it is able to grow in methanol containing media which enables avoiding possible contaminations (Demain and Vaishnav 2009). In 2009, the first biopharmaceutical protein, kallikrein inhibitor, produced in *P. pastoris* was approved by the FDA.

Chinese hamster ovary (CHO) cells are the most widely used mammalian cells and they are used for production of around 50% of therapeutic proteins in the market due to its high similarity to human cells. The absolute requirement for glycosylation to be of “human-type” protein is the most important reason for the wide use of mammalian systems for production of biopharmaceutical proteins (Walsh and Jefferis 2006). Proteins produced in mammalian cells are often properly folded and glycosylated (Martínez, Liu et al. 2012) with almost identical pharmaceutical properties to human proteins (Redwan 2007). However, the process is very expensive due to their stringent cultivation requirements and very limited capacities of production and secretion (Demain and Vaishnav 2009). Even so, for some glycoproteins made by CHO cells, off-line modifications are still necessary to achieve their human type quality (Demain and Vaishnav 2009).

Generally, reaching high titers or productivities is as a rule of thumb for industrial enzymes production, while high quality is crucial for biopharmaceuticals production (Porro, Gasser et al. 2011). Bacterial systems are usually applied for production of non-glycosylated proteins with less disulfide bonds and small protein size (Demain and Vaishnav 2009); yeasts are approached when a specific PTM is essential for its function and activity (Ferrer-Miralles, Domingo-Espín et al. 2009). Moreover, if proteins fail to be properly expressed in both microbial systems, higher eukaryotic hosts, such as mammalian cells, insect cells and cell-free systems, will be considered (Freigassner, Pichler et al. 2009). A list of best production of each species is shown in Table 2.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Host</th>
<th>Production</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hirudin</td>
<td><em>S. cerevisiae</em></td>
<td>500 mg/L</td>
<td>(Mendoza-Vega, Hebert et al. 1994)</td>
</tr>
<tr>
<td>Cutinase</td>
<td><em>S. cerevisiae</em></td>
<td>1.6 g/L</td>
<td>(Calado, Ferreira et al. 2004)</td>
</tr>
<tr>
<td>Granulocyte-Colony Stimulating Factor</td>
<td><em>S. cerevisiae</em></td>
<td>1.3 g/L</td>
<td>(Lee, Lee et al. 1999)</td>
</tr>
<tr>
<td>Tetanus Toxin Fragment C</td>
<td><em>P. pastoris</em></td>
<td>12 g/L</td>
<td>(Clare, Rayment et al. 1991)</td>
</tr>
<tr>
<td>Interleukin 2</td>
<td><em>P. pastoris</em></td>
<td>4 g/L</td>
<td>(Cregg, Vedvick et al. 1993)</td>
</tr>
<tr>
<td>Hirudin</td>
<td><em>P. pastoris</em></td>
<td>1.5 g/L</td>
<td>(Demain and Vaishnav 2009)</td>
</tr>
<tr>
<td>scFv</td>
<td><em>P. pastoris</em></td>
<td>4.9 g/L</td>
<td>(Damasceno, Pla et al. 2004)</td>
</tr>
<tr>
<td>Angiostatin</td>
<td><em>P. pastoris</em></td>
<td>108 mg/L</td>
<td>(Xie, Zhang et al. 2003)</td>
</tr>
<tr>
<td>Fab</td>
<td><em>P. pastoris</em></td>
<td>458 mg/L</td>
<td>(Ning, Junjian et al. 2005)</td>
</tr>
<tr>
<td>Insulin Precursor</td>
<td><em>P. pastoris</em></td>
<td>3 g/L</td>
<td>(Gurramkonda, Polez et al. 2010)</td>
</tr>
<tr>
<td>Alpha-amylase</td>
<td><em>P. pastoris</em></td>
<td>340 mg/L</td>
<td>(Lee, Nakano et al. 2003)</td>
</tr>
<tr>
<td>Human Serum Albumin</td>
<td><em>P. pastoris</em></td>
<td>10 g/L</td>
<td>(Kobayashi, Kuwae et al. 2000)</td>
</tr>
<tr>
<td>Human Interferon</td>
<td><em>E. coli</em></td>
<td>42.5 g/L</td>
<td>(Babaeipour, Shojaosadati et al. 2007)</td>
</tr>
<tr>
<td>Human Antithrombin</td>
<td>CHO1 cells</td>
<td>1 g/L</td>
<td>(Kuwae, Ohda et al. 2005)</td>
</tr>
<tr>
<td>Human IgG</td>
<td>CHO cells</td>
<td>130 mg/L</td>
<td>(Wang, Zhang et al. 2005)</td>
</tr>
<tr>
<td>Human tPA</td>
<td>CHO cells</td>
<td>34 mg/L</td>
<td>(Demain and Vaishnav 2009)</td>
</tr>
<tr>
<td>Erythropoietin</td>
<td>CHO cells</td>
<td>121 mg/L</td>
<td>(Yoon, Kim et al. 2006)</td>
</tr>
</tbody>
</table>

1CHO, Chinese hamster ovary.
2.0  RECOMBINANT PROTEIN PRODUCTION IN S. CEREVISIAE

To study recombinant protein production and secretion in S. cerevisiae are generally two-fold: i) it can serve as a simple model to study many human diseases caused by protein misfolding and ER stress, such as Alzheimer’s, Parkinson’s, diabetes mellitus, atherosclerosis ischemia (Yoshida 2007); ii) today around 20% of protein-based biopharmaceuticals on the market are produced by S. cerevisiae (Martínez, Liu et al. 2012), including insulin, hepatitis B surface antigen, urate oxidase, glucagons, granulocyte macrophage colony stimulating factor, hirudin, and platelet-derived growth factor (Demain and Vaishnav 2009).

Due to its model organism status and long history in industry for recombinant protein production of S. cerevisiae, there have been many studies on cell and molecular biology engineering, as well as processes development. Large improvements of the heterologous protein production have been achieved from milligrams to grams per liter. However, many of these attempts have given rather specific conclusions: rational targets have been chosen, but often it was found that the strategy worked successfully only for one (or a few) protein(s), and the same engineered strain could not be used as a general cell factory platform for production of a range of different recombinant proteins. This can be explained by the complexity of protein processing and secretion pathways. Each involved process must be tuned to a specific state based on the secreted protein’s physical properties, e.g. the cell growth needs to be balanced, gene expression needs to be tuned and the endogenous protein processing machinery needs to be modulated. The imbalance of individual proteins in different parts of cellular metabolism, especially RNA and protein synthesis, protein folding, and degradation of misfolded proteins could cause severe cell stress (Schröder 2008). Through detailed understanding of the individual processes and integrated analysis of the interplay between these processes, it should be possible to derive general models for protein secretion that can be used for engineering the secretion pathway and thereby resulting in improved cell factories for recombinant protein production (Graf, Dragosits et al. 2009). Proteins that are produced in S. cerevisiae, either at laboratory or pilot levels, are listed in Table 3.
Table 3. Landscape of recombinant protein production in *S. cerevisiae*.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Production</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Serum Albumin (HSA)</td>
<td>3 g/L</td>
<td>(Chris Finnis 2005)</td>
</tr>
<tr>
<td>Human Transferrin</td>
<td>1.8 g/L</td>
<td>(Chris Finnis 2005)</td>
</tr>
<tr>
<td>Human Insulin-Like Growth Factor (hIGF)</td>
<td>55 mg/L</td>
<td>(Vai, Brambilla et al. 2000)</td>
</tr>
<tr>
<td>Human Platelet Derived Growth Factor (PDGF)</td>
<td>4 mg/L</td>
<td>(Robinson, Hines et al. 1994)</td>
</tr>
<tr>
<td>Glucagon</td>
<td>63.1 mg/L</td>
<td>(Egel-Mitani, Andersen et al. 2000)</td>
</tr>
<tr>
<td>Hirudin</td>
<td>500 mg/L</td>
<td>(Mendoza-Vega, Hebert et al. 1994)</td>
</tr>
<tr>
<td>Bovine Pancreatic Trypsin Inhibitor (BPTI)</td>
<td>180 mg/L</td>
<td>(Parekh and Wittrup 1997)</td>
</tr>
<tr>
<td>Single-chain Antibody (scFv)</td>
<td>3.6 g/L</td>
<td>(Chris Finnis 2005)</td>
</tr>
<tr>
<td>Single-chain T-cell Receptor (scTCR)</td>
<td>7.5 mg/L</td>
<td>(Sagt, Kleizhen et al. 2000)</td>
</tr>
<tr>
<td>Hapatitis Surface Antigen (HBsAg)</td>
<td>19.4 mg/L</td>
<td>(Bulavaite, Sabaliauskaite et al. 2006)</td>
</tr>
<tr>
<td>Llama VHH</td>
<td>100 mg/L</td>
<td>(Frenken, van der Linden et al. 2000)</td>
</tr>
<tr>
<td>Tetanus Toxin Fragment C (TTFC)</td>
<td>1 g/L</td>
<td>(Romanos, Makoff et al. 1991)</td>
</tr>
<tr>
<td>Parvovirus B19 VP2</td>
<td>400 mg/L</td>
<td>(Lowin, Raab et al. 2005)</td>
</tr>
<tr>
<td>Fab</td>
<td>0.2 mg/L</td>
<td>(Edqvist, Ker?nen et al. 1991)</td>
</tr>
<tr>
<td>Interleukin (IL)</td>
<td>30 mg/L</td>
<td>(Guisez, Tison et al. 1991)</td>
</tr>
<tr>
<td>Epidermal Growth Factor (EGF)</td>
<td>5 mg/L</td>
<td>(Chigira, Oka et al. 2008)</td>
</tr>
<tr>
<td>Human Interferon (IFN)</td>
<td>276 mg/L</td>
<td>(Chu, Zhang et al. 2003)</td>
</tr>
<tr>
<td>Insulin Precursor (IP)</td>
<td>90 mg/L</td>
<td>(Liu, Tyo et al. 2012)</td>
</tr>
<tr>
<td>Human Parathyroid Hormone (bPTH)</td>
<td>42 mg/L</td>
<td>(Kang, Kim et al. 1998)</td>
</tr>
<tr>
<td>G-protein Coupled Receptor (GPCR)</td>
<td>84.2 U/mg</td>
<td>(Purvis, Chotai et al. 1991)</td>
</tr>
<tr>
<td><em>A. niger</em> Glucose Oxidase (GO)</td>
<td>9 g/L</td>
<td>(Park, Shin et al. 2000)</td>
</tr>
<tr>
<td><em>A. oryzae</em> α-amylase</td>
<td>4.3 U/ml</td>
<td>(Nieto, Prieto et al. 1999)</td>
</tr>
<tr>
<td>Cutinase</td>
<td>1.6 g/L</td>
<td>(Calado, Ferreira et al. 2004)</td>
</tr>
<tr>
<td>Proteinase A (PrA)</td>
<td>82 mg/L</td>
<td>(Carlsten, Jochumsen et al. 1997)</td>
</tr>
<tr>
<td>Human Lysozyme (h-LZM)</td>
<td>74.5 U/ml</td>
<td>(Choi, Paik et al. 2004)</td>
</tr>
<tr>
<td>Human Adenosine A2a Receptor (A2aR)</td>
<td>28 mg/L</td>
<td>(Wedekind, O'Malley et al. 2006)</td>
</tr>
<tr>
<td><em>A. niger</em> β-galactosidase</td>
<td>5600 U/ml</td>
<td>(Domingues, Lima et al. 2005)</td>
</tr>
<tr>
<td><em>E. coli</em> β-galactosidase</td>
<td>1.15 g/L</td>
<td>(Alberghina, Porro et al. 1991)</td>
</tr>
<tr>
<td>Granulocyte-Colony Stimulating Factor (GCSF)</td>
<td>1.3 g/L</td>
<td>(Lee, Lee et al. 1999)</td>
</tr>
<tr>
<td><em>S. pombe</em> Acid Phosphatase (PHO)</td>
<td>2.5 4455/4600/min</td>
<td>(Robinson, Bockhaus et al. 1996)</td>
</tr>
<tr>
<td><em>P. furiosus</em> β-glucosidase</td>
<td>10 mg/L</td>
<td>(Fabre, Nicaud et al. 1991)</td>
</tr>
<tr>
<td>Pant Thaumatin</td>
<td>0.44 mg/gDCW</td>
<td>(Harmsen, Bruyne et al. 1996)</td>
</tr>
<tr>
<td>Green Fluorescent Protein (GFP)</td>
<td>5682AU</td>
<td>(Štagoj, Comino et al. 2006)</td>
</tr>
</tbody>
</table>
3.0 THE SECRETORY PATHWAYS IN S. CEREVISIAE

For secreted recombinant proteins, there are many steps after translation and before proteins are mature and trafficked to the correct location. A common pathway, called the secretory pathway, is used to complete the protein maturation process. This post-translational protein processing is an extensive pathway with more than 160 proteins responsible for different post-translational processes, where more than 550 proteins (Feizi, Österlund et al. Submitted) pass through several different organelles before they reach their final destinations. The details of the chemical and molecular mechanisms of the secretory pathway processing have been extensively reviewed (Hou, Tyo et al. 2012; Mattanovich, Branduardi et al. 2012). As shown in Figure 2, after translation, the polypeptide get folded and primarily glycosylated in the endoplasmic reticulum (ER), and then the correctly folded proteins are sorted to the Golgi apparatus for further glycosylation and final modifications, whereas misfolded proteins are sorted into the cytosol for degradation. Correctly modified proteins will be targeted to the membrane and secreted to the extracellular region, otherwise they will be sorted to endosome or vacuole for re-cycling or degradation. If a large number of misfolded proteins are retained in the ER for certain amount of time, a transcriptional response, called the unfolded protein response (UPR) will be activated and expression levels of around 400 genes that have close relation with the secretory pathway will be regulated. Besides, the post-translational protein processing has close relations with the oxidative stress response, the general stress response, the general transcription and translation machinery, the amino acid metabolism and the energy metabolism, etc.

The secretory pathway involves several checkpoints where the state of protein folding and its impact on overall cellular stress is monitored. The chaperone capacity, vesicle and cargo proteins, oxidizing equivalents, as well as metabolite requirements, such as ATP, NADH, NADPH, glutathione buffers, glycans, etc, should be well tuned according to the expressed protein and the host system. Protein folding and modification, trafficking, degradation, as well as amino acid metabolism involve many layers of quality control that must be well-coordinated to avoid cellular stress resulting in reduced cell growth and protein secretion (Dürrschmid, Reischer et al. 2008; Nemecek, Marisch et al. 2008) or even apoptosis and cell death (Mattanovich, Gasser et al. 2004). Cells have developed highly regulated networks to balance the proteostasis, including protein degradation processes, e.g. ER associated degradation (ERAD) (Nishikawa, Brodsky et al. 2005), proteasome-ubiquitin system (Ding and Yin 2008), autophagy-lysosome pathway (Yorimitsu, Nair et al. 2006), and several cellular responses can cope with protein misfolding also, such as the heat shock response (HSR) (Westerheide and Morimoto 2005), the unfolded protein response (UPR) (Malhotra and Kaufman 2007) and the environment stress response (ESR) (Perrone, Tan et al. 2008; Schröder 2008). Much of the secretory pathway is managed on the basis of chemical modifications (such as glycosylation and disulfide bond formation) and
protein-protein interactions (such as degradations of misfolded proteins), except for several stress responses (Hou, Tyo et al. 2012).

Figure 2. The secretory pathway for protein processing.

A major problem generally associated with recombinant protein production by mody host species is the accumulation and aggregation of misfolded proteins or polypeptides that causes considerable cell stress (Gasser, Saloheimo et al. 2008). Therefore, a deep understanding of the ER associated protein processing is relevant for both clinical and industrial research.

3.1 The Endoplasmic Reticulum Protein Processing

The protein processing machinery in the ER could be generally dissected into five components (Schröder 2008): (i) ER translocation, (ii) glycosylation, (iii) Disulfide bond formation, (iv) ER associated degradation, and (v) signal transduction pathways.

After ribosomal synthesis begins, a protein bounds for the secretory pathway must be selectively targeted to the ER. The hydrophobicity and amino acid composition of the pre-signal sequence, an N-terminal 15-50 amino acid sequence, determine this step to occur either co-translationally where translation and translation are directly linked (Rapiejko and Gilmore 1997), or post-translationally, which is ribosome-uncoupled (Plath, Mothes et al. 1998; Matlack, Misselwitz et al. 1999). For soluble proteins, the pre-signal is cleaved by the signal peptidase complex (SPC) immediately (YaDeau, Klein et al. 1991). Meanwhile, folding chaperones and cytosolic chaperones (Ssa1p, Ydj1p) bind to the exposed hydrophobic patches to prevent aggregation of the polypeptide (Simons, Ferro-Novick et al. 1995; Willer, Jermy et al. 2003).
Initial glycosylation occurs during translocation. Glycosylation is reported to (a) improve protein folding, (b) protect protein from proteases and (c) act as a step in quality control. Glycosylation occurs in two forms in yeast, N-linked, and O-linked. N-linked glycosylation is the attachment of a sugar molecule to a nitrogen atom in an amino acid residue in a protein; whereas O-linked glycosylation is the attachment of a sugar molecule to an oxygen atom in an amino acid residue in a protein. The initial step of O-linked glycosylation occurs in ER (Strahl-Bolsinger, Gentzsch et al. 1999), together with the N-linked glycosylation (Burda and Aebi 1998; Spiro 2002), and completes later in the Golgi. It is also reported that O-linked glycosylation might occur before N-linked glycosylation, resulting in O-linked glycosylation on the N-linked recognition sequence (Hou, Tyo et al. 2012), which implies that N-linked asparagine glycosylation and O-linked serine/threonine glycosylation may be in competition (Ecker, Mrsa et al. 2003).

Disulfide bond formation must correctly pair cysteines of the polypeptide chain to form and stabilize the protein in its mature conformation. The protein undergoes a series of folding and disulfide bond forming steps. Electrons are transferred from the newly formed disulfide bond to the protein disulfide isomerase (PDI) which in turn passes the electrons to the FAD-bound Oxidoreductin 1 (Ero1p). Finally electrons are passed to the terminal electron acceptor O2 at aerobic conditions (Tu and Weissman 2002). This mechanism forms disulfide bridges at random, and the correct pairings must be found by a trial and error process, involving the repeated oxidation/reduction of cysteines by PDI and its homologues (Wilkinson and Gilbert 2004), with NADPH and glutathione as electron acceptors (Tyo, Liu et al. 2012). Kar2p works on unfolded polypeptides, whereas GRP94 and the lectin chaperones work on partially folded substrates (Schröder 2008). In other words, the unfolded polypeptides firstly bind to Hsp90p or Hsp40p co-chaperones with their unfolded hydrophobic regions, and then to the Hsp70p until the nucleotide exchange triggers the disassociation. The rate of protein folding is dependent upon the expressed proteins, the availability of ATP and chaperones (Tu and Weissman 2004).

Strict quality control sensing determines whether correct structures have been formed before the folded protein is allowed to leave the ER. Protein chaperones assist the polypeptides along the path to correct folding and help to remove proteins from the ER when they have terminally misfolded. Exit from the ER can proceed in two pathways, (a) to the degradation pathway for misfolded and unmodified proteins retained in the ER, and (b) to the Golgi for properly folded and modified proteins. Misfolded proteins are linked to ubiquitin after their re-translocation into the cytosol, and are targeted for degradation by the 26S proteasome (Gasser, Saloheimo et al. 2008). Glycosylation structures of glycoproteins can also traffic proteins to degradation. The UDP-glucose: glycoprotein glucosyltransferase (UGT) is a gate keeper for the degradation of glycoprotein in the ER (Kleizen and Braakman 2004). Kar2p and the Sec61p complex are also involved in the ERAD pathway, with Kar2p-binding acting as a residence-time clock, causing terminally misfolded proteins to be shuttled out of the ER (Plemper, Bohmler et al. 1997;
Brodsky, Werner et al. 1999). Misfolded proteins can also be degraded independent of ERAD by moving them through the Golgi to the vacuole (Hong, Davidson et al. 1996).

### 3.2 Unfolded Protein Response

When ER stress begins to overwhelm the processing machinery, generally caused by (a) aggregation of unfolded and misfolded proteins, (b) elevated levels of long chain fatty acids and (c) imbalance and leaking of ER lumen Ca\(^{2+}\) (Wei, Wang et al. 2006; Pineau and Ferreira 2010), large scale transcriptional alterations become necessary to bring the secretory pathway back into homeostasis. This transcriptional response, called the unfolded protein response, is responsible for detecting protein folding problems in the ER and transferring the information to the nucleus, where expressions of ~400 genes are regulated to adjust secretory resources and increase the capacity of the secretory pathway and bring the secretory pathway back to homeostasis (Patil and Walter 2001).

UPR consists of an upstream sensing mechanism and a downstream activation mechanism, as shown in Figure 3. The upstream mechanism has been studied in great detail and is primarily controlled by two key proteins in *S. cerevisiae*, the ER transmembrane protein Ire1p, and the transcriptional activator Hac1p. Ire1p contains an ER luminal domain that binds to Kar2p and a cytosolic domain that has kinase and endonuclease activity (Okamura, Kimata et al. 2000). Recently, Gardner and Walter reported that the UPR is activated by direct binding of Ire1p to unfolded proteins (Gardner and Walter 2011), which causes Ire1p to dimerize. Then the cytoplasmic portion of Ire1p immediately phosphorylates itself, which in turn, activates its endonuclease domain. This endonuclease activity is specific to an mRNA sequence in *HAC1u*, the transcribed RNA from *HAC1*. Un-activated *HAC1* mRNA is constitutively expressed, but due to the presence of 3’ RNA hairpin, *HAC1u* cannot be translated. Activated Ire1p specifically cleaves *HAC1u* to remove the hairpin, which is followed by Rlg1p mediated ligation, allowing translation to proceed. Hac1p then translocates into the nucleus where it acts as a functional transcriptional activator.

DNA microarray analysis that has identified genes altered by the UPR (Mori, Kawahara et al. 1996; Patil, Li et al. 2004), however, the downstream/implementation part has been limited to identifying promoter sequences that are specific to UPR. The downstream portion of the UPR is characterized by a large, multi-faceted response to bring the secretory pathway back to homeostasis (Travers, Patil et al. 2000). Hac1p is a transcriptional activator that is known to interact with three binding sequences, two of which require Gcn4p for gene activation, to regulate many different activities within the cell in an attempt to correct problems associated with misfolded protein accumulation in the ER (Mori, Kawahara et al. 1996; Travers, Patil et al. 2000; Patil, Li et al. 2004). In all, the expression of approximately 380 genes is altered by the UPR response, although only half of them have Hac1p binding sequences in the promoter.
(Travers, Patil et al. 2000; Kimata, Ishiwata-Kimata et al. 2006). Broadly, the UPR increases (i) the capacity of protein folding and glycosylation processing in the ER, (ii) trafficking components of ER-Golgi vesicles and post-Golgi sorting, (iii) the ERAD and ubiquitin/proteasome system to clear misfolded, unwanted and unnecessary proteins and (iv) up-regulated lipid and inositol synthesis pathways, whereas (v) attenuating general transcriptions, translations and amino acid metabolisms, and (vi) regulating stress response genes (Bernales, Papa et al. 2006; Schröder 2008).

Therefore, as suggested by the mechanism of UPR and demonstrated by many reports, Hac1p plays a significant role in strain modification for enhanced protein production. Disruption of HAC1 in S. cerevisiae reduced production of α-amylase and epidermal growth factor by 75% and 50%, respectively (Valkonen, Penttila et al. 2003), whereas over-expression of endogenous and Trichoderma reesei HAC1 improved α-amylase production by 1.7-fold (Valkonen, Penttila et al. 2003) and 2.4-fold (Higashio and Kohno 2002), respectively. Over-expression of HAC1 or its homologs also increased recombinant protein secretion by Aspergillus niger (Valkonen, Ward et al. 2003) and CHO cells (Tigges and Fussenegger 2006; Ku, Ng et al. 2008).

**Figure 3. Unfolded protein response mechanism.**
4.0 OPTIMIZATION OF RECOMBINANT PROTEIN PRODUCTION

The major challenges facing recombinant protein production are how to decrease the production cost, improve the productivity and titer while maintaining the quality of the products (Chiverton 2010). Even with all the advantages, recombinant protein production (RPP) in *Saccharomyces cerevisiae* is still far from optimal. Therefore, a wide range of studies have been implemented to engineer the yeast secretory pathway and to optimize protein secretion over the past twenty years, as reviewed previously (Freigassner, Pichler et al. 2009; Idiris, Tohda et al. 2010). Generally, smaller proteins (< 60 kDa) tend to be better expressed and a large number of transmembrane helices might decrease expression levels (White, Clark et al. 2007). However, so far, each recombinant protein behaves in an individual, unfortunately unpredictable way in response to overexpression. Thus, no correlation between the production level and protein specific parameters have been reported as guidelines for recombinant protein production (Freigassner, Pichler et al. 2009).

Enhancement of recombinant protein secretion can be achieved by the following factors or combination: (1) engineering DNA sequences and expression systems, (2) engineering the host strains, and (3) optimizing the environmental/cultivation conditions (Homma, Iwahashi et al. 2003), which will be discussed in details in the following. Each of the factors could enhance production levels of protein, and a proper combination will reach an optimal production through repeated trials. For example, a 10-fold increased production of an antibody (OX26 scFv) was reported simply by co-overexpression ER chaperone Kar2p and foldase PDI (Hackel, Huang et al. 2006), while around 50-fold increase of productivity was achieved by tuning the vector system, the chaperone expression and the cultivation condition (Shusta, Raines et al. 1998).

4.1 Expression Design

Vector engineering has been extensively studied for different purposes. The marker type and promoter strength of the expression systems are the key factors that determine the plasmid copy number and the mRNA level of the recombinant protein. Different marker systems (Kuroda, Matsui et al. 2009) and promoter libraries (Fischer, Alper et al. 2006; Partow, Siewers et al. 2010) have been made and evaluated for recombinant protein production.

4.1.1 Plasmid copy numbers

Gene copy number clearly affects transcription rates. As high titer is a key industrial objective, the 2 micron based high-copy plasmids are generally used for heterologous protein production. Whereas in some cases, this strategy could lead to saturation or overloading of the secretory pathway, and low copy plasmids were found to give higher yield for some proteins such as granulocyte-colony stimulating factor and acid phosphatase (Robinson, Bockhaus et al. 1996), bovine pancreatic trypsin inhibitor (Parekh, Forrester et al. 1995) and erythropoietin (Elliott,
Giffin et al. 1989). In some other cases, integrative plasmid also found to be optimal for overexpressing some proteins, e.g. bovine pancreatic trypsin inhibitor (Parekh and Wittrup 1997), single-chain T-cell receptor (Shusta, Holler et al. 2000), and bovine prochymosin (Harmsen, Bruyne et al. 1996).

Multi-copy vectors that do not contain any prokaryotic sequences or antibiotic markers have also been evaluated (Nieto, Prieto et al. 1999), but one problem with this type of vector is that the genes of the vector are in close proximity with each other and can interfere with their expression levels. Therefore, it is recommended to test different arrangements of the transformation modules (Nieto, Prieto et al. 1999) and to find an optimal insertion site, e.g. using the unique SnaBI site in the Delta vectors (Chris Finnis 2005). One type of integration plasmid uses homologues to repetitive DNA sequences in the genome, such as sequence of ribosomal DNA (Lopes, Hakkaart et al. 1991; Nieto, Prieto et al. 1999), to increase the integration efficiencies. The δ-sequences, which carry the bacterial NEO gene to make the integration more stable (Lee and DaSilva 1997) have also been used. Kim et al. (Kim, Han et al. 2003) developed a super δ-integration system which truncated unnecessary bacterial genes and reduced the integration size, and resulted in a 1.3-fold enhancement of hirudin secretion compared with a single δ-system.

4.1.2 Marker

The marker used to maintain stability of heterologous expression and promote copy number plays a significant role in protein secretion. Toxicity genes (Sidorenko, Antoniukas et al. 2008; Agaphonov, Romanova et al. 2010), auxotrophy genes (Chigira, Oka et al. 2008), defective auxotrophy markers (Corrales-Garcia, Possani et al. 2010), and essential genes in the glycolytic pathway (Kjeldsen, Ludvigsen et al. 2002) are commonly used as selective markers.

Auxotrophic markers (LEU2, TRP1, URA3, and HIS3) are often preferred in yeast for protein secretions. In some cases, defective auxotrophic markers, which have truncated regions in their promoter or coding sequence, result in a higher plasmid copy number. This technique has been successfully applied, e.g. LEU2-d (Gabrielsen, Reppe et al. 1990; Romanos, Makoff et al. 1991), TRP1-d (Nieto, Prieto et al. 1999), URA3-d (Mendoza-Vega, Hebert et al. 1994). In order to compare secretion efficiency using different markers, Loison et al. (Loison, Vidal et al. 1989) constructed a plasmid which included both LEU2-d and URA3 markers to express the schistosomal antigen P28-I. The result showed that the product was about 3% of the total cell protein produced using the URA3 marker, whereas it was around 25% of the total cell protein using the LEU2-d marker. Seresht et al. (Kazemi Seresht, Nørgaard et al. 2012) evaluated different truncations of the LEU2-d and URA3-d markers on insulin precursor production, and reported that URA3-d marker led to higher insulin yield whereas the LEU2-d marker caused low plasmid stability. Chen et al. (Chen, Partow et al. 2012) introduced a new approach by applying
the same concept, that a ubiquitin/N-degron tag was fused together with the \textit{URA3}, \textit{URA3-d}, \textit{HXT1} or \textit{KEX2} marker and increased the plasmid copy number and \textit{LacZ} activity by 30-100%.

Other non-traditional essential genes, such as \textit{CDC4}, \textit{PGI1}, \textit{TPI1}, have also been applied as selectable markers for protein secretion, e.g. glucagon (Egel-Mitani, Andersen et al. 2000), insulin precursor (Kjeldsen and Pettersson 2003), human serum albumin (Kjeldsen, Pettersson et al. 1998), and \textit{Aspergillus oryzae} \textalpha-amylase (Liu, Tyo et al. 2012). A typical example is the CPOT plasmid, which use the \textit{POT1} gene from \textit{Schizosaccharomyces pombe} to complement the \textit{tpi1} mutation in the host. Because the foreign \textit{POT1} gene has weaker activity in \textit{S. cerevisiae}, high plasmid copy numbers are necessary to express adequate levels of triosephosphate isomerase activity. This, in turn, increases expression of the recombinant protein of interest. Expression of \textalpha-1-antitrypsin showed a 2-fold increase using \textit{POT1} as the selectable marker while compared to using the \textit{LEU2} marker (Kawasaki 1999).

\textbf{4.1.3 Promoters}

Although transcription factors and plasmid copy numbers can affect transcription, from an engineering standpoint, changing the promoter sequence is the most straightforward method to affect the transcription rate. The strength of different promoters has been evaluated for production of secreted proteins (Smith, Duncan et al. 1985; Alper, Fischer et al. 2005). Promoters that initiate strong and constitutive expression are often chosen for recombinant protein production, such as promoters of \textit{PBRI} (Chris Finnis 2005), \textit{GAPDH1} (Shusta, Raines et al. 1998; Park, Shin et al. 2000), \textit{PGK} (Chu, Zhang et al. 2003), \textit{ACT1} (Nieto, Prieto et al. 1999), \textit{ADH1} (Domingues, Lima et al. 2005) and \textit{TEF1} (Paper I). Specially, the \textit{TPI1} promoter (of strongly expressed glycolytic gene \textit{TPI1} in \textit{S. cerevisiae} coding for triose phosphate isomerase), is widely used and results in high levels of insulin production (Egel-Mitani, Andersen et al. 2000).

Unexpectedly, in some cases weak promoters could result in even higher protein production levels. For example, Ernst (Ernst 1986) reported that up to a two-fold increase in somatomedin-C secretion could be obtained by using the weaker \textit{CYC1} promoter rather than the stronger \textit{ACT1} promoter.

Inducible promoters are also applied to design cell factories and separate cell growing phase with protein expression phase, in order to reduce the metabolic stress caused by recombinant protein production (Mattanovich, Branduardi et al. 2012). \textit{GAL} promoters have been widely applied because they can easily be regulated (Ostergaard, Olsson et al. 2000). \textit{GAL1} (Lowin, Raab et al. 2005), \textit{GAL7} (Calado, Ferreira et al. 2004) and \textit{GAL10} (Kapat, Jung et al. 1998) are among the strongest and most widely used \textit{GAL} promoters. However, \textit{GAL} promoters have certain disadvantages. The inducer, galactose, also serves as a carbon source, and this results in a decrease in the inducer level (Hovland, Flick et al. 1989). Furthermore, the low level of Gal4p,
which is the transcription factor conferring induction, limits induction levels (Johnston and Hopper 1982). In order to solve these problems, Štagoj et al. (Štagoj, Comino et al. 2006) constructed a \textit{GAL1}-\textit{GAL4} promoter by applying an additional copy of the \textit{GAL4} upstream region at the \textit{GAL1} locus and managed to get a higher level production of the recombinant protein. More recently, hybrid promoters have attracted more interest, and examples include \textit{GAL1-GAL10-GAPDH} (Robinson, Hines et al. 1994), \textit{GAL1-GAL10} (Mendoza-Vega, Hebert et al. 1994), \textit{GAL10-PYK1} (Bulavaite, Sabaliauskaite et al. 2006), \textit{ADH2-GAPDH} (Lim, Lee et al. 2002), PGK- \textit{GAL1-10} (Clements, Catlin et al. 1991), \textit{GAPDH/ADH2} (Tottrup and Carlsen 1990) and \textit{CYC1/GAL} promoter (Alberghina, Porro et al. 1991).

\textbf{4.1.4 Leader sequence}

The leader sequence partly determines the trafficking of a secreted protein. The pre-leader is designed to direct the peptide through the translocation step into the ER, and the pro-sequence is responsible for increasing the trafficking efficiency through the inter-organelle transport (Rakestraw, Sazinsky et al. 2009). For most proteins, e.g. human insulin-like growth factor (Romanos, Scorer et al. 1992) and α-globin (Rothblatt, Webb et al. 1987), both the pre- and pro-leader are required in order to achieve an optimal secretion. However, it is reported that the pro region of the alpha factor leader has only a minor effect on secreting aminoglycoside phosphotransferase and granulocyte-colony stimulating factor, and for interleukin, the pre-region decreased Kex2p processing efficiency compared to the case when only the pro-region was applied (Ernst 1988).

The leader sequence can either be a native signal peptide (Bulavaite, Sabaliauskaite et al. 2006), a heterologous secretory peptide (Chigira, Oka et al. 2008) or a synthetic (designed) leader (Hackel, Huang et al. 2006; Rakestraw, Sazinsky et al. 2009). It is obvious that due to evolution, native leaders should possess certain advantages, which is proved by secretion of human serum albumin (Sleep, Belfield et al. 1990), human interferon (Piggott, Watson et al. 1987), human α-amylase (Sato, Uemura et al. 1989) and \textit{Aspergillus niger} glucose oxidase (De Baetseliers, Dohet et al. 1992). However, protein secretion in \textit{S. cerevisiae} usually results in hyper-glycosylation, and leader sequences are often mutated and selected to more efficiently direct proteins through the secretory pathway (Rakestraw, Sazinsky et al. 2009). Highly glycosylated leaders, such as the \textit{S. cerevisiae} alpha factor leader, have been proven to be more efficient for secretion of epidermal growth factor (Chigira, Oka et al. 2008), human platelet derived growth factor (Robinson, Hines et al. 1994), interleukin (Baldari, Murray et al. 1987) and \textit{S. pombe} acid phosphatase (Baldari, Murray et al. 1987). Synthetic leaders are also used to solve secretion problems, such as (i) inefficient processing of leaders, (ii) hyper-glycosylation, and (iii) incorrect trafficking. Examples of synthetic pre-pro leaders include the expression insulin precursor (Kjeldsen 2000), human adenosine A2a receptor (Butz, Niebauer et al. 2003), green fluorescent
protein (Huang and Shusta 2005), hirudin (Mendoza-Vega, Hebert et al. 1994), bovine pancreatic trypsin inhibitor (Parekh and Wittrup 1997), epidermal growth factor (Clements, Catlin et al. 1991), scFv (Shusta, Raines et al. 1998), scTCR (Shusta, Holler et al. 2000), and tetanus toxin fragment C (Romanos, Makoff et al. 1991).

It is difficult to predict which leader is best for efficient secretion of a given protein. It is therefore often required to experimentally evaluate different leaders. For example, various leader sequences including *INU1*, *SUC2*, *PHO5*, and *MEL1* were evaluated, in order to produce either GFP or GFP-hexokinase fusions. In all cases, the majority of the protein accumulated in the vacuole or endosome (Li, Xu et al. 2002). Whereas using a viral leader from the K28 preprotoxin, secretion was improved (Eiden-Plach, Zagorc et al. 2004). Another study showed that *SUC2* signal peptide was correctly cleaved from all secreted human interferon molecules (Parekh and Wittrup 1997), while using the native leader only resulted in 64% of cleavage (Hitzeman, Leung et al. 1983). Whereas in another study, when using the same *SUC2* leader to secrete α-1-antitrypsin, approximately 80% of the produced protein accumulated in the secretory pathway (Moir and Dumais 1987).

**4.2 Host Engineering**

The protein expression optimization is often the simple and initial approach. However, many proteins are only secreted at very limited levels even though their transcription levels are sufficiently optimized, as discussed in 4.1.1 and 4.1.3. This implies that recombinant protein production and secretion also involves the optimization of post-translational modifications (PTMs). PTMs can significantly affect the characteristics of proteins, including charge, hydrophobicity, stability and solubility, etc (Tokmakov, Kurotani et al. 2012). Often, among different PTMs, folding and glycosylation have been identified as rate-limiting steps for heterologous protein production (Tokmakov, Kurotani et al. 2012).

**4.2.1 Glycosylation engineering**

Glycosylation, a posttranslational modification taking place in the ER and Golgi, is the most abundant protein modification in all species (Larkin and Imperiali 2011), and defects in glycosylation profiles of specific proteins have already been recognized as disease markers (Walsh and Jefferis 2006). More than one third of biopharmaceuticals on the market are glycoproteins (Walsh and Jefferis 2006). Glycosylation is also reported to influence the folding process, secretion levels, aggregation, solubility, stability, activity, affinity and selectivity (Çelik and Çalık 2011). Glycosylation has been shown to facilitate protein folding of epidermal growth factor (Parthasarathy, Subramanian et al. 2006; Demain and Vaishnav 2009), immunoglobulin (Rudd, Wormald et al. 1999) and interleukin (Livi, Lillquist et al. 1991).
Glycosylation can be engineered based on the amino acid sequence of the protein or the glycosylation enzymes. Missing one essential glycosylation site of CD47 reduced its surface expression level by more than 90% (Parthasarathy, Subramanian et al. 2006), whereas introducing extra N-glycosylation sites can yield a five-fold increase in secretion of cutinase (Sagt, Kleizen et al. 2000). While on the other hand, glycosylation seems to have no significant effect on the secretion of α-amylase (Nieto, Prieto et al. 1999) and interleukin (Livi, Ferrara et al. 1990). If no glycosylation sites can be engineered in the coding region of the protein, an alternative solution is to engineer the leader sequence (Chen, Pioli et al. 1994). A synthetic leader LA19 with two N-glycosylation sites has been demonstrated to result in optimal production of insulin precursor (Kjeldsen, Hach et al. 1998). Another way to engineer the glycosylation pathway is to develop yeast to build human-type glycans. So far, it has been reported that mutation of yeast-specific hyper-glycosylation genes, especially OCH1 coding for mannosyltransferase combined with disruption of other genes, or overexpression of an α-1,2-mannosidase, or genes involved in the early glycosylation, especially the ALG3 gene have already given positive results in S. cerevisiae (De Pourcq, De Schutter et al. 2010).

4.2.2 Disulfide bond formation engineering

Protein folding in the ER is often considered as the flux controlling step in the secretion pathway (Lim, Lee et al. 2002). The number of disulfide bonds is another factor that affects the protein folding, secretion, stability and function (Hober and Ljung 1999). Correct disulfide bonds stabilize a protein mainly by enclosing hydrophobic regions, making it less favorable for the aggregation and chaperone binding for subsequent degradation (Arolas, Aviles et al. 2006). For example, the expression level and affinity of CD47 (the extracellular immunoglobulin domain of a mammalian membrane protein) decreased by 30% when the core disulfide bond is missing (Parthasarathy, Subramanian et al. 2006); and the expression level of insulin-like growth factor decreased about one-third when removing either Cys23p or Cys96p (Steube, Chaudhuri et al. 1991).

Small proteins could spontaneously fold to their native states in absence of the cellular folding machinery (Arolas, Aviles et al. 2006), whereas more complicated proteins require ER chaperones and protein foldase for disulfide bond formation and isomerization. It was reported that over-expression of recombinant proteins often decreased soluble levels of chaperones, Kar2p and PDI, (Robinson and Wittrup 1995), moreover, decreased levels of Kar2p also in turn resulted in reduction of protein production (Robinson, Bockhaus et al. 1996). Over-expression of chaperones, especially Kar2p and PDI, often allows for improved secretion, as shown in Table 4. Kar2p, a Hsp70p family molecular chaperone, acts as a folding chaperone by binding to exposed hydrophobic stretches of amino acid sequences (Ma, Kearney et al. 1990) and also as an ER detergent functioning in the ERAD process (Robinson, Bockhaus et al. 1996). On the other hand,
over-expression of PDI also improves secretion for proteins that do not contain disulfide-bonds, e.g., *Pyrococcus furiosus* β-glucosidase (Smith and Robinson 2002), suggesting that PDI is not only as a catalyst for disulfide bonds formation and isomerization (Laboissière, Sturley et al. 1995), it may also act in a chaperone-like capacity or cooperate with the folding or degradation mechanisms of non-disulfide containing proteins (Powers and Robinson 2007). In some cases, Kar2p and PDI can work together to further enhance protein production, and it is suggested that Kar2p may maintain the protein in an unfolded state by binding to the protein, and this makes the cysteine residues accessible for PDI activity (Mayer, Kies et al. 2000). However, in other cases, over-expression of ER chaperone yields only a minor increase or even a decrease in protein secretion. Thus, the effect of co-expression of chaperone and foldase also depends on each protein’s unique characteristics, and the fine-tuned over-expression of foreign proteins, as well as the ER chaperone and foldase are highly required to improve the final production (Gasser, Saloheimo et al. 2008).

### Table 4. Effect of ER chaperone over-expression for recombinant protein production.

<table>
<thead>
<tr>
<th>Protein</th>
<th>a.a.</th>
<th>Sulfur Bridge</th>
<th>N-Gly</th>
<th>BiP+</th>
<th>PDI+</th>
<th>BiP+ PDI+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transferrin</td>
<td>679</td>
<td>19</td>
<td>2</td>
<td>-</td>
<td>15</td>
<td>(Chris Finnis 2005)</td>
</tr>
<tr>
<td>Erythropoietin</td>
<td>193</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>-</td>
<td>(Robinson, Hines et al. 1994)</td>
</tr>
<tr>
<td>PDGF-B</td>
<td>109</td>
<td>5</td>
<td>1</td>
<td>-</td>
<td>10</td>
<td>(Robinson, Hines et al. 1994)</td>
</tr>
<tr>
<td>Hirudin</td>
<td>65</td>
<td>3</td>
<td>-</td>
<td>2.5</td>
<td>-</td>
<td>(Kim, Han et al. 2003)</td>
</tr>
<tr>
<td>BPTI</td>
<td>58</td>
<td>3</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>(Robinson, Bockhaus et al. 1996)</td>
</tr>
<tr>
<td>scFv</td>
<td>244</td>
<td>2</td>
<td>1</td>
<td>2.4</td>
<td>2.3</td>
<td>(Shusta, Raines et al. 1998)</td>
</tr>
<tr>
<td>scTCR</td>
<td>240</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>-</td>
<td>(Shusta, Holler et al. 2000)</td>
</tr>
<tr>
<td>HLY</td>
<td>130</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>1.57</td>
<td>(Hayano, Hirose et al. 1995)</td>
</tr>
<tr>
<td>A2aR</td>
<td>412</td>
<td>-</td>
<td>2</td>
<td>1</td>
<td>&lt;1</td>
<td>(Butz, Niebauer et al. 2003)</td>
</tr>
<tr>
<td>GCSF</td>
<td>174</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>(Butz, Niebauer et al. 2003)</td>
</tr>
<tr>
<td>PHO</td>
<td>435</td>
<td>8</td>
<td>9</td>
<td>1</td>
<td>4</td>
<td>(Robinson, Hines et al. 1994)</td>
</tr>
<tr>
<td>β-glucosidase</td>
<td>421</td>
<td>-</td>
<td>-</td>
<td>&lt;1</td>
<td>1</td>
<td>(Smith and Robinson 2002)</td>
</tr>
<tr>
<td>Prochymosin</td>
<td>345</td>
<td>-</td>
<td>2</td>
<td>26</td>
<td>-</td>
<td>(Harmsen, Bruyne et al. 1996)</td>
</tr>
<tr>
<td>Thaumatin</td>
<td>235</td>
<td>8</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>(Harmsen, Bruyne et al. 1996)</td>
</tr>
</tbody>
</table>


Over-expression or mutations of other genes in the ER could also assist with secretion. For example, over-expression of co-chaperones of Kar2p, such as Jem1p, Scj1p, and nucleotide exchange factor Sil1p and Lhs1p, are also reported to increase protein secretion levels, including human albumin, granulocyte–macrophage colony-stimulating factor, and human transferrin (Payne, Finnis et al. 2008). Over-expression of the PDI oxidant Ero1p had enhanced the secretion of scTCR by 5.1-fold (Wentz and Shusta 2007). Overexpression of chaperone holdase, such as calnexin or calreticulin, could also improve protein production (Chung, Lim et al. 2004; Kato, Murata et al. 2005).
4.2.3 Engineering protein trafficking

After correct folding in the ER or being modified in the Golgi, proteins go through membrane-bound trafficking steps among different organelles. However, heterologous proteins are often retained intracellularly at certain steps in the secretion pathway without complete secretion, even though they have folded correctly. For example, α-1-antitrypsin (Moir and Dumais 1987), hepatitis surface antigen (Biemans, Thines et al. 1991) and erythropoietin (Elliott, Giffin et al. 1989) accumulate in the ER, whereas soybean proglycinin is retained in the Golgi (Utsumi, Kanamori et al. 1991). These results point out the importance of genetic optimizations regarding the inefficient trafficking and mis-sorting from the ER to Golgi, internal sorting with Golgi and post-Golgi sorting. Co-overexpression of COG6, COY1, and IMH1, related to Golgi-vesicle transport, enhanced Fab production by 1.2-fold (Gasser, Sauer et al. 2007; Wentz and Shusta 2007). Mutation of PMR1, a Golgi-resident calcium ATPase gene (Rudolph, Antebi et al. 1989), increased the secretion of prochymosin (Harmsen, Bruyne et al. 1996), human single-chain urinary plasminogen activator (Melnick, Turner et al. 1990), propapain (Ramjee, Petithory et al. 1996) and bovine growth hormone (Smith, Duncan et al. 1985). Over-expression of SSO1 and SSO2, which are crucial for vesicle fusion to plasma membrane, increased α-amylase secretion by 2-fold (Larsson, Cassland et al. 2001; Toikkanen, Sundqvist et al. 2004). Mutation of the cell wall protein Gas1p strongly improved the secretion of insulin-like growth factor (Brinkmann, Reiter et al. 1993). Recently we showed that it is also possible to enhance the secretion of insulin precursor and α-amylase in S. cerevisiae by over-expression of SNARE regulating proteins that modulate vesicle transport (Hou, Tyo et al. 2012).

4.2.4 Engineering protein degradation pathways

Incorrectly folded or modified proteins are targeted to degradation either in the cytosol (proteasome-based) or in the vacuole. Delta’s strains with genomic mutations of the UBC4 gene, which encodes the ubiquitin-conjugating enzyme, have been reported to result in extremely high plasmid copy numbers and over-expression of different proteins (Sleep, Finnis et al. 2001). Over-expression of UBC4 is also reported to enhance the secretion level of elafin by 10-fold (Chen, Pioli et al. 1994). Deleting VPS4, VPS8, VPS13, VPS35, VPS36 or PEP4, all encoding vacuolar proteinases, resulted in higher yields of an insulin-containing fusion protein (Zhang, Chang et al. 2001), and disruption of YAP3 alone or together with KEX2 reduced the degradation of HSA (human serum albumin) and HSA-human growth hormone fusion protein (Geisow, Harris et al. 1991). Single deletion of the extracellular protease Ski5p had successfully improved the secretion level of killer toxin (Bussey, Steinmetz et al. 1983).

There are many targets to reduce protein degradations, which are also host and protein specific. Instead, many reports have applied protease-deficient strains to reduce the intercellular and extracellular proteolytic degradations (Newstead, Kim et al. 2007; Li, Hays et al. 2009).
4.2.5 Random mutagenesis and screening

Many attempts have been applied for enhancing protein production in *S. cerevisiae*; however, the production of proteins is highly host and protein specific (Idiris, Tohda et al. 2010). For most proteins, their secretion levels are still 100-, or even 1000-fold lower than their theoretically yield (Schröder 2007). Due to the poor understanding of the protein processing machinery, which involves many tightly cross-reacting factors, molecular engineering to enhance recombinant protein production is sometimes difficult and time-consuming. Recombinant protein production engineering has paved the way to a new era of random mutagenesis accompanied with systems biology analysis, which has been largely promoted by yeast postgenomic technologies and systems biology tools. It is a more convenient way to generate high production strains and fill in the gap in our understanding of how phenotype and genotype are linked.

Various approaches could be applied, including cell-based and protein based, adapted evolution and random mutagenesis, gene shuffling and transcription factors design, etc. Zhang et al. (Zhang, Liu et al. 2003) performed random mutagensis within insulin B-chain, and suggested that the failure of proper disulfide bond formation should contribute to the intracellular trafficking. Kowalski et al. (Kowalski, Parekh et al. 1998) analyzed all possible cysteine mutants within bovine pancreatic trypsin inhibitor and suggested that 5-55 disulfide bond is essential for protein folding and secretion. Payne et al. (Payne, Finnis et al. 2008) identified over-expressed targets involved with the ATPase recycling of Kar2p (*JEM1*, *SIL1*, *LHS1* and *SCJ1*) using chemical mutagenesis, and had enhanced production levels of granulocyte–macrophage colony-stimulating factor, human albumin, and human transferrin. Kanjou et al. (Kanjou, Nagao et al. 2007) found potential deletion targets of vesicle formation by screening the EUROSCARF deletion library, and increased secretion levels of luciferase. Screening surface-displayed cDNA libraries could also help identifying targets for antibody fragments production (Shusta, Kieke et al. 1999). Smith et al. (Smith, Duncan et al. 1985; Wentz and Shusta 2007) found four possible targets by screening mutagenized bovine growth hormone secretion strains and this resulted in a 15-fold increase. Arffman et al. (Arffman, Aho et al. 1990) successfully isolated a strain that could secrete 70-fold more endoglucanase through multiple rounds of mutagenesis and selections.

4.3 Fermentation optimization

Successful improvements of protein production from milligrams to gram per liter have been reported before, and many of which are due to fermentation optimizations (Idiris, Tohda et al. 2010). The environmental conditions and fermentation processes have close correlations with the cell growth, the internal secretory pathway machinery, the secretion levels of proteins and their stability in the medium. It has been reported that a group of genes, the common environmental response (CER) genes, are sensitive to stress conditions (Gasch, Spellman et al. 2000; Causton,
Ren et al. 2001), including protein folding and degradation genes, heat shock responsible genes, energy generation and DNA damage repair genes, etc. Thus, the analysis of the external and internal factors, especially temperature, pH, the aeration condition, nutrient composition and cell growth rate, is also of great importance.

4.3.1 Temperature and pH

Temperature has a profound impact on cellular metabolism, cell wall composition, and regulation of folding-related proteins (Gasser, Saloheimo et al. 2008). Cultivation at a preferred temperature is crucial to obtain optimal protein production and stability. Huang et al. (Huang, Gore et al. 2008) has investigated three different temperatures (20, 30 and 37°C) for green fluorescent protein production in *S. cerevisiae*, and reported that 20°C yielded highest production levels (~5-fold higher than at that 37°C) and longest secretion processes, whereas 30 °C showed the fastest initial secretion rate.

Generally proteins are produced at 30°C, which is the optimal cultivation temperature for yeast cell growth. However, sometimes lowering the temperature to 20–25°C would achieve higher titers of proteins for different reasons. Some argue that it is because the protein’s native host favored a cold environment (Zimmer 2002), for example 20°C for green fluorescent protein (Huang and Shusta 2005). Some claim that un-glycosylated proteins are more easily secreted at lower temperature, such as invertase (Ferro-Novick 1984) and active acid phosphatase could be secreted from tunicamycin treated cells at 20°C or 25°C rather than 30°C (Mizunaga, Izawa et al. 1988). Some report that lower temperature would slow the folding process and hereby alleviate misfolded protein accumulation (Tottrup and Carlsen 1990) and reduce proteolysis (Jahic, Gustavsson et al. 2003), such as 25°C for scTCR (Shusta, Raines et al. 1998), 26°C for human superoxide dismutase-human proinsulin (Tottrup and Carlsen 1990) and 23°C for G-protein coupled receptor (Wedekind, O'Malley et al. 2006). The decrease of growth temperature is also reported to be beneficial for protein production in bacterial and mammalian protein production systems (Dragosits, Frascotti et al. 2011). Interestingly, for some other proteins, a higher temperature is preferred. β-glucosidase could be continuously secreted at high levels in 37°C for over 70 h compared to only 25 h at 30°C (Huang, Gore et al. 2008). At 40°C, yeast could produce β-glucosidase with less ER stress than at 30°C (Smith, Richardson et al. 2005).

Varying pH could also affect secretion levels, protein stability and enzyme activity. A reduced lag phase of cell growth appeared at low pH for arginine kinase expression (Canonaco, Schlattner et al. 2003), and the optimal production took place at pH 5.0 (Canonaco, Schlattner et al. 2002); whereas green fluorescent protein is generally secreted at high pH (pH 7 to 11.5), yeast membrane proteins showed optimal production between neutral and alkaline pH (Sarramegna, Demange et al. 2002; Bonander, Hedfalk et al. 2005) and the fluorescence dropped sharply by about 50% at pH 6 (Laboratories 1999).
4.3.2 Aeration conditions

Investigating effects of aeration conditions and redox processes on the protein production is relevant for both basic and applied research. The disulfide bond formation is achieved through the oxidative folding in the ER resulting in the formation of around 25% of all reactive oxygen species (ROS) generated (Tu and Weissman 2004) and will be increased upon stress conditions (Haynes, Titus et al. 2004). The oxidative stress caused by accumulation of ROS can have multiple consequences, including a leakage in the respiratory pathway and accumulation of misfolded proteins in the ER, etc (Gasser, Saloheimo et al. 2008). Moreover, the property of efficient and rapid secretion of biomass-degrading enzymes under less aerated conditions is also highly required for developing microorganisms in consolidated bioprocessing. It has been reported that, low oxygen level could enhance production of certain proteins, such as glucoamylase (Cha, Choi et al. 1997), 3H6 Fab (Baumann, Maurer et al. 2008) and human trypsinogen (Baumann, Maurer et al. 2008). Similar results have been reported for P. pastoris, i.e. low oxygen supply increased the productivity by around 2.5-fold of a Fab fragment (Baumann, Maurer et al. 2008). Whereas in other cases, a 21-fold increase of carbonylated cutinase was achieved at high oxidative stress conditions (Sagt, Muller et al. 2002).

4.3.3 Culture Additives

Media composition is another important parameter. The optimal culture medium depends on many factors, including host metabolism, potential inhibitory products, target proteins, etc, and the development of the optimal medium is often a trial and error process (Shojaosadati, Kolaei et al. 2008). On one hand, it is desirable from the cost perspective to make the medium as simple as possible by reducing the amount of non-essential carbon and nitrogen components, while on the other hand, supplementing the medium with amino acids and other components has proven to achieve higher secretion levels and reduced protein degradations for different reasons.

Feeding sufficient yeast nitrogen base and casamino acids enhanced GFP secretion level while preventing post-secretory losses (Parthasarathy, Subramanian et al. 2006). Adding defined amino acids mixtures (alanine, arginine, asparagine, glutamic acid, glutamine and glycine) enhanced production of xylanase (Goergens, Van Zyl et al. 2005; Huang, Gore et al. 2008) and β-galactosidase (Jin, Ye et al. 1997). Addition of phosphate increased secretion of α-amylase (Faria, Castilho-Valavicius et al. 1989), whereas on the other hand, Seresht et al. (Seresht, Palmqvist et al. 2011) had reported that reduced phosphate concentration yield higher production of insulin precursor. Recently, we have investigated the extracellular protein degradation by comparing different media compositions, and we reported that the specific amino acid composition, the concentration of phosphates and the pH of around 6 are all crucial for efficient protein secretion and reducing protein degradation. The result of microarray analysis suggested that the post-
secretory loss involves the competitive endocytosis of insulin, bovine serum albumin, and yeast extract protein (Paper VIII).

4.3.4 Cell growth

When producing recombinant proteins, it is often found that the optimal growth conditions are not ideal for protein production (Freigassner, Pichler et al. 2009). Expression of recombinant proteins using strong expression systems often result in a metabolic burden resulting in a reduction of cell growth and protein production (Dürrschmid, Reischer et al. 2008). Generally, it could be avoided by slowing down the transcription and translation steps to achieve dynamic equilibrium for translocation and assembly in the secretory pathway (Freigassner, Pichler et al. 2009). In fact, sometimes the recombinant protein production could be promoted by reducing cell growth rates (Lunter and Goodson 2011; Miyawaki 2011). Thus, the investigation of these factors, which influence protein expression and secretion is still of great importance. The production of heterologous proteins has been identified as either growth-dependent or inverse growth associated (Andersen and Krummen 2002), depending on the cell lines, the property of the proteins, promoters and enhancers (Lunter and Goodson 2011). In P. pastoris, the secretion of recombinant proteins was reported to be coupled to the specific growth rate, reaching a plateau at high specific growth rate (Buchetics, Dragosits et al. 2011). Whereas in T. reesei, the protein production rate were negatively correlated with cellular biosynthetic activities (Arvas, Pakula et al. 2011).
5.0 SYSTEMS BIOLOGY TOOLS FOR PROTEIN PRODUCTION

There are many interpretations of systems biology, and in most cases with a common objective to obtain global, quantitative, and hopefully predictive information of the target system (Kitano 2002; Stephanopoulos, Alper et al. 2004; Barrett, Kim et al. 2006; Nielsen and Jewett 2008). Systems biology approaches have been developed as highly valuable tools in *S. cerevisiae* for metabolic engineering and bioprocess development for production of renewable chemicals, biofuels and food ingredients. This is based on the robustness of genome scale metabolic models (GSMM) and together with advances of highly developed integrative approaches (post-genomic analysis), as shown in Figure 4. Top-down approach is based on -omics and high-throughput analysis, which is a data-driven process. On the other hand, bottom-up approaches involves detailed knowledge to generate and reconstruct mathematical models to analyze and simulate the biological system. The bottom-up approach often goes hand in hand with top-down approach to obtain global information, as well as to generate hypothesis for improvement of cell factories (Nielsen and Jewett 2008). Now possibilities have opened for expanding systems biology applications for recombinant protein production, in terms of both gaining deep and systems-level understanding of the secretory pathway and identifying potential targets for further strain and production improvements.

![Figure 4. Illustration of top-down and bottom-up systems biology.](image-url)
5.1 Bottom-up Approach: Mathematical Models of the Secretory Pathway

Mathematical models can assist in gaining understanding of the cell metabolism and to predict cellular responses to different stimuli. Most bottom-up models only focus on specific part of the biological system, since there is limited quantitative and interaction information available. Through basic understanding of the secretory pathway and integrated analysis, it is possible to construct general models for recombinant protein production.

Shelikoff et al. (Shelikoff, Sinskey et al. 1996) reported a structured kinetic model that aims to describe how different co-translational processes affect glycosylations. This model takes glycosylation site occupancy as a black box model and generates the mass balances around this system without detailed biochemical information. Umaña et al. (Umaña and Bailey 1997) described the biochemical pathways responsible for 33 different oligosaccharide modifications and site occupancy in the N-linked glycosylation. Krambeck et al. (Krambeck and Betenbaugh 2005) further extended this model, and evaluated metabolic engineering strategies for overexpressing a target glycoprotein and the effect in glycoprotein distribution. Raden et al. (Raden, Hildebrandt et al. 2005) had developed a mathematical model to describe the early step of the UPR. The model considered the relative concentrations of Ire1p and Kar2p in the ER, combined with expected kinetics, and suggested that only Kar2p is not adequate for UPR activation. Feizi et al. (Feizi, Österlund et al. Submitted) constructed the first genome-scale reconstruction of yeast protein secretion pathway that includes all the known steps of the conventional secretory pathways and presented a decision-matrix that assigns all secretory proteins to a particular secretory class. This study also applied protein abundance data to estimate the activity of each secretory element under exponential growth conditions.

5.2 Top-down Approach: -Omics Analysis

Genome analysis enables identification and analyze of an entire cellular network and this approach is often applied together with evolution and mutagenesis experiments. Transcriptome analysis allows assess of the dynamic patterns of all gene expressions upon stimulus. However, due to the complexity of cellular metabolism, for example, translation, post-translation modifications, chemical alterations, enzyme catalysis, etc, proteomics and metabolomics data are sometimes valuable to identify cellular responses. Fluxomics allows quantitatively characterizing metabolic networks, which could be directly linked to the phenotype data (Kim, Roldão et al. 2012). In order to understand the cellular responses to protein production, -omics analysis has been done for some recombinant strains. Although no general guidelines were identified, factors like chaperones, foldases, cargo proteins, proteases, were often reported to promote protein production (Freigassner, Pichler et al. 2009).
Bonander et al. (Bonander, Darby et al. 2009) analyzed the transcriptomics profile of *S. cerevisiae*, and reported that tuning the ribosomal subunit ratio by adjusting transcript levels of *BMS1* could optimize membrane and soluble protein production. Casagrande et al. (Casagrande, Stern et al. 2000) used transcription data of *S. cerevisiae* to report the crucial role of the UPR pathway on degradation of mouse histocompatibility complex class I heavy chain H-2Kb. Hanlon et al. (Hanlon, Rizzo et al. 2011) performed microarray analysis of *S. cerevisiae* and reported the roles of novel cofactors (Cin5p, Skn7p, Phd1p and Yap6p) on cell responses upon several conditions of stress. Recently, we found that constitutively activation of the heat shock response in *S. cerevisiae* can reduce ER stress in wild type strain, UPR deficient strain, and strain with both UPR and proteasome mutations (Hou, Liu et al. Submitted). Over-expression of *HSF1* increased insulin precursor and α-amylase production by ~50%, and improved endogenous invertase yield by ~100%. Transcriptome analysis revealed that HSR relieved ER stress mainly through up-regulating protein folding genes, whereas repressing the overall transcription and translation (Hou, Österlund et al. Submitted). Sharma et al. (Sharma, Mahalik et al. 2011) analyzed the transcriptome data of *E. coli* strains expressing human interferon, xylanase and GFP, and reported that high level expression triggered the mRNA degradation, osmoprotectant and proteases degradation, as well as repressed aerobic respiration, ATP synthesis, amino acid uptake and biosynthesis pathways. Gasser et al. (Gasser, Sauer et al. 2007) identified potential targets by transcriptome analysis in *P. pastoris*, and yield a 2.5-fold increase of Fab antibody fragment. Kim et al. (Kim, O'Callaghan et al. 2011) used transcriptional analysis to address the CHO cell line instability on recombinant antibody production with the epigenetic-methylation-induced transcriptional silencing and the genetic-progressive loss.

Pandhal et al. (Pandhal, Ow et al. 2011) applied shotgun proteomics of *E. coli*, and suggested that improvement of glycosylation efficiencies could be obtained by enhancing flux through the glyoxylate cycle, and around 3-fold increase of glycosylated proteins was achieved by overexpression of isocitrate lyase. Proteome analysis of *B. megaterium* has identified GroEL and DnaK as important chaperones that assist in protein folding (Wang, Hollmann et al. 2006; Biedendieck, Borgmeier et al. 2011). Quantitative analysis of intracellular amino acids of *Bacillus megaterium* for GFP expression identified the limited levels of tryptophan, aspartate, histidine, glutamine, and lysine, which was supported by that addition of only 5mM of these proteins had increased the GFP yield by 100% (Korneli, Bolten et al. 2012). Sellick et al. (Sellick, Croxford et al. 2011) used metabolite profiling analysis of CHO cells expressing a recombinant IgG4 antibody, and developed a better feed strategy which had increased the cell biomass and antibody titer by 35% and 100%, respectively.

Gonzalez et al. (Gonzalez, Andrews et al. 2003) performed metabolic flux analysis of *S. cerevisiae*, and reported that the flux to pentose phosphate pathway and TCA cycle is lower in
human superoxide dismutase strain comparing to the wildtype strain. Fluxome analysis of *B. megaterium* suggested that pyruvate serves as a more suitable substrate for recombinant protein production (Fürch, Hollmann et al. 2007; Fürch, Wittmann et al. 2007). Metabolic flux analysis was also performed to analysis *P. pastoris* secreting a *Rhizopus oryzae* lipase, and reported increased glycolytic, TCA cycle and NADH regeneration fluxes upon recombinant secretion (Jordà, Jouhten et al. 2012). Driouch et al. (Driouch, Melzer et al. 2011) performed $^{13}$C metabolic flux ratio analysis and in silico elementary flux mode analysis of *Aspergillus niger* expressing the fructofuranosidase, and reported an increased flux through pentosephosphate pathway and mitochondrial malic enzyme for an elevated supply of NADPH, and a reduced flux through the TCA cycle.
6.0 SUMMARY OF RESULTS

6.1 Experimental Design

The work presented in this thesis applies metabolic engineering and systems biology tools to explore ways for recombinant protein over-production and relationship with cellular metabolism and protein production. Enhancing production of recombinant proteins was carried out based on three main approaches: expression design, host engineering and fermentation analysis, as shown in Figure 5. Different examples of recombinant protein production (human insulin precursor and α-amylase), including different tools applied for engineering and analysis, will highlight the parameters and potential mutation target that can be altered for future improvement.

Figure 5. Overall strategies for recombinant protein production.

We already mentioned above that protein production depends on protein size, amino acid demand, disulfide bond number and presence of post-translational sites. When applying the same host strain and the same expression strategy, different proteins will cause different categories and different levels of cellular stresses, and will hence result in different levels of final production. In this thesis work, two recombinant proteins with different properties (size, number of disulfide bonds and glycosylation sites), human insulin precursor and α-amylase from Aspergillus oryzae, were chosen as representatives of simple and multi-domain proteins, as well as glycosylated and un-glycosylated proteins, as shown in Figure 6A. Insulin precursor is one of the first commercialized pharmaceutical protein with a well-studied expression background (Kjeldsen, Brandt et al. 1996; Kjeldsen, Frost Pettersson et al. 1999). It is a single chain peptide that contains a 29-amino acid B chain and a normal 21-amino acid A chain connected by a removable mini-C chain to ensure efficient expression (Kjeldsen, Frost Pettersson et al. 1999). Thus, it is a relatively small and simple protein and it was therefore chosen as one of the model proteins in this study. α-Amylase was selected as the other model protein for our study because it
could degrade starch, which provides chances for consolidated bioprocess analysis and high-throughput put screening for random mutation analysis. Furthermore, α-amylase is a three-domain protein (Randez-Gil and Sanz 1993) with 478 amino acids, 4 disulfide bonds and 1 glycosylation site, and is hence a larger and more complex protein than the insulin precursor. Two secretion leaders, the endogenous alpha factor leader and a synthetic leader Yap3-TA57 (Kjeldsen, Frost Pettersson et al. 1999), are applied as representatives of glycosylated and non-glycosylated leader sequences to make the production of two different proteins, as shown in Figure 6B.

**Figure 6. Recombinant secreted proteins applied in this thesis work.**
(A) Model proteins. (B) Secretion leaders.

Here we applied both an auxotrophy marker and the POT1 expression systems of plasmid to evaluate the production of the two different proteins. A commonly used auxotrophy marker, URA3, was used for plasmid maintenance. For the POT1 expression systems, the host strain has a deletion of TPI1 gene that encodes the glycolytic enzyme triose-phosphate isomerase, which causes a NADH/energy shortage that makes the \( \Delta tpi \) mutant strain unable to grow on glucose as the sole carbon source (Compagno, Brambilla et al. 2001). The POT1 gene from *Schizosaccharomyces pombe* that encodes for the same enzyme and make the TPI1 deletion strain survive growth on glucose was applied as another marker system (Kawasaki 1999).

Two strong promoters, *TEFlp* and *TPI1p*, were applied to generate different mRNA levels of the recombinant protein (Partow, Siewers et al. 2010), which were fused independently with two different leader sequences. Detailed information of plasmids and strains are shown in Figure 7 and Table 5.

Two types of media were applied and evaluated for protein production: i) since the POT1 expression system is designed for complex media, that could yield higher copy numbers, fast cell growth and higher biomass production (Kawasaki 1999), YPD medium was applied in Paper I
and III to generate optimal production, whereas ii) defined medium SD-2xSCAA (Wittrup and Benig 1994) was applied in Paper II, IV and V, in order to have a clean background to investigate cellular metabolism.

**Table 5. Strain information.**

<table>
<thead>
<tr>
<th>Strains</th>
<th>Genotype</th>
<th>Protein</th>
<th>Leader</th>
<th>Promoter</th>
<th>Marker</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC</td>
<td>CEN.PK 113-7D <em>tpi1::kanMX with CPOT1ud</em></td>
<td>Insulin</td>
<td>-</td>
<td>TPI</td>
<td>POT1</td>
<td>Paper I</td>
</tr>
<tr>
<td>AIP</td>
<td>CEN.PK 113-7D <em>tpi1::kanMX with pAlphaInsPOT</em></td>
<td>Insulin</td>
<td>Alpha factor</td>
<td>TEF1</td>
<td>POT1</td>
<td>Paper I</td>
</tr>
<tr>
<td>SIP</td>
<td>CEN.PK 113-7D <em>tpi1::kanMX with pSynInsPOT</em></td>
<td>Amylase</td>
<td>Alpha factor</td>
<td>TEF1</td>
<td>POT1</td>
<td>Paper I</td>
</tr>
<tr>
<td>AAP</td>
<td>CEN.PK 113-7D <em>tpi1::kanMX with pAlphaAmyPOT</em></td>
<td>Amylase</td>
<td>YAP3-5TA7</td>
<td>TEF1</td>
<td>POT1</td>
<td>Paper I</td>
</tr>
<tr>
<td>SAP</td>
<td>CEN.PK 113-7D <em>tpi1::kanMX with pSynAmyPOT</em></td>
<td>Amylase</td>
<td>YAP3-5TA7</td>
<td>TEF1</td>
<td>POT1</td>
<td>Paper I</td>
</tr>
<tr>
<td>AIC</td>
<td>CEN.PK 113-7D <em>tpi1::kanMX with pAlphaInsCPOT</em></td>
<td>Insulin</td>
<td>Alpha factor</td>
<td>TPI</td>
<td>POT1</td>
<td>Paper I</td>
</tr>
<tr>
<td>SIC</td>
<td>CEN.PK 113-7D <em>tpi1::kanMX with pSynInsCPOT</em></td>
<td>Insulin</td>
<td>YAP3-5TA7</td>
<td>TPI</td>
<td>POT1</td>
<td>Paper I</td>
</tr>
<tr>
<td>AAC</td>
<td>CEN.PK 113-7D <em>tpi1::kanMX with pAlphaAmyCPOT</em></td>
<td>Amylase</td>
<td>Alpha factor</td>
<td>TPI</td>
<td>POT1</td>
<td>Paper I</td>
</tr>
<tr>
<td>SAC</td>
<td>CEN.PK 113-7D <em>tpi1::kanMX with pSynAmyCPOT</em></td>
<td>Amylase</td>
<td>YAP3-5TA7</td>
<td>TPI</td>
<td>POT1</td>
<td>Paper I</td>
</tr>
<tr>
<td>WN</td>
<td>CEN.PK 113-7D <em>ura3 with p426GPD</em></td>
<td>-</td>
<td>-</td>
<td>GAPDH1</td>
<td>URA3</td>
<td>Paper II</td>
</tr>
<tr>
<td>WI</td>
<td>CEN.PK 113-7D <em>ura3 with pYapIns</em></td>
<td>Insulin</td>
<td>YAP3-5TA7</td>
<td>GAPDH1</td>
<td>URA3</td>
<td>Paper II</td>
</tr>
<tr>
<td>WA</td>
<td>CEN.PK 113-7D <em>ura3 with pYapAmy</em></td>
<td>Amylase</td>
<td>YAP3-5TA7</td>
<td>GAPDH1</td>
<td>URA3</td>
<td>Paper II</td>
</tr>
<tr>
<td>dN</td>
<td>CEN.PK 113-5D <em>hac1::kanMX with p426GPD</em></td>
<td>-</td>
<td>-</td>
<td>GAPDH1</td>
<td>URA3</td>
<td>Paper II</td>
</tr>
<tr>
<td>dI</td>
<td>CEN.PK 113-5D <em>hac1::kanMX with pYapIns</em></td>
<td>Insulin</td>
<td>YAP3-5TA7</td>
<td>GAPDH1</td>
<td>URA3</td>
<td>Paper II</td>
</tr>
<tr>
<td>dA</td>
<td>CEN.PK 113-5D <em>hac1::kanMX with pYapAmy</em></td>
<td>Amylase</td>
<td>YAP3-5TA7</td>
<td>GAPDH1</td>
<td>URA3</td>
<td>Paper II</td>
</tr>
</tbody>
</table>

Different -omics techniques and integrated analysis were applied in the studies, and a large dataset was generated in order to answer different questions: i) whether there is a correlation between the expression level and final production for both simple and complex proteins; ii) whether the unfolded protein response is important for expression of both simple and complex proteins, what is the UPR-dependent and -independent cellular responses; iii) beside the knowledge of the secretory pathway that is already known, is there other components or pathways that have potential impacts on recombinant protein productions; iv) what is the role of oxygen and oxidative stress on recombinant protein productions; and v) does the protein production profiles correlate with cellular metabolism in the same way for both simple and complex proteins?
6.2 Paper I: Expression systems affect recombinant protein secretion

The eight engineered strains that produce either insulin precursor (IP) or amylase using two different secretion leaders (the alpha factor leader vs. the Yap3-TA57 leader), and expressed from two different promoters (TEF1p and TPI1p) were cultivated in batch fermentations.

The CPOTud strain series showed a notable advantage for production of both insulin precursor and α-amylase, compared with the POTud derived strains, and the advantage was more prominent for the production of insulin precursor than for the production of α-amylase, as shown in Figure 8. IP producing strains with the CPOTud expression system, AIC and SIC, could produce 30 to 50-fold more than strains with the POTud expression system (AIP and SIP), whereas for α-amylase, the CPOTud derived strain AAC could produce 2.7-fold more amylase than the POTud derived strain AAP, but the CPOTud derived strain SAC only produce 8% more amylase than the POTud derived strain SAP.
Figure 8. Final protein production results for the expression project.

Error bars are based on independent duplicate experiments.

Two different leader sequences (alpha factor leader and synthetic leader) resulted in different effects on production of IP and amylase (Figure 8). In all cases, the synthetic leader could direct more IP through the secretory pathway, (a) in the POTud derived strains, SIP produced 1.1-fold more IP than AIP and (b) in the CPOTu derived strains, SIC could produce 19% more IP than AIC. The synthetic leader also showed an advantage for the production of α-amylase but only in the strains secreting a moderate amount (around 15 mg/L in YPD medium) of α-amylase (in the POTud derived strains, SAP produced 1.1-fold more amylase than AAP), whereas in the strains with higher production of amylase, the synthetic leader was less advantageous: in the CPOTu derived strains, the synthetic leader strain SAC only produced 58% of amylase compared with the alpha factor leader strain AAC. The effect of leader sequences on different proteins could be explained by the difference in N-glycosylation sites in the pro-leader sequence. Since secretion of glycosylated proteins in *S. cerevisiae* is often reduced due to hyper-glycosylation and misfolding inside the cell (Srivastava, Piskur et al. 2001), the number of glycosylation sites in the leader sequence is another very important factor to be considered. Kjeldsen et al. (Kjeldsen et al., 1999) reported similar results when secreting insulin precursor under stressed conditions (such as treatment with DTT), and found that TA39 (pro-leader with two glycosylation sites) showed better internal protein trafficking than TA57 (pro-leader with no glycosylation site). In our experiments, when high amount of amylase is produced, the misfolded proteins cause cell stress, possibly in a similar way as low-level DTT induction in the Kjeldsen’s study, and at this condition, the alpha factor pro-leader which possesses three glycosylation sites may provide more stringent guiding for correct fold and consequently, secretion.
In addition to their final titers, the IP and α-amylase also differ in their processing characteristics in the secretory pathway. By plotting the protein production data against dry cell weight to eliminate the effect of the changing cell concentration, it was found that there is a clear shift in the secretion behavior during the diauxic shift (Figure 9). Interestingly, all α-amylase producing strains produced amylase at a higher rate during growth on ethanol, whereas all IP producing strains produced IP at a higher rate in the glucose phase. The shifting patterns of protein production further supported the fact that the rate-controlling step for protein secretion is different for the two proteins. Production of IP is probably mainly limited by expression and for both the used expression promoters (pTPI1 and pTEF1) there is higher expression at fast growth/high glycolytic fluxes. For amylase, which is a larger protein with more diverse modifications, the limitation is likely to be protein processing and folding. We hypothesize that the respiratory conditions prevailing during growth on ethanol may have a beneficial effect on the folding process (compared with the fermentative conditions prevailing in the glucose growth phase), either due to the NADPH/NADP⁺ balance or reduced ER translocated peptides. There may also be a favorable heat shock-like effect induced during growth on ethanol (Piper 1995; Alexandre, Ansanay-Galeote et al. 2001).

**Figure 9. Secretion profiles of IP and α-amylase strains for the expression project.**
Protein production were plotted versus cell growth (expressed as dry cell weight, DCW) to compare single cell producing capacity. (●) protein production (mg/l), (♦) Glucose concentration (g/l) and (▲) Ethanol concentration (g/l). (A) IP production by strain SIC. (B) α-amylase production by strain SAC.
6.3 Paper II: Host Engineering-the UPR Dependent and Independent Cell Metabolism

We already ensured that due to their unique characters, different proteins varied in their expression levels and patterns even if the same host and expression system were applied. We also know the unfolded protein response (UPR) as one of the most important cellular responses that ensure homeostasis of protein synthesis. The next question we want to address is whether UPR was important for different kind of proteins and what are the UPR-independent response.

In order to address this, six strains (Table 5) were constructed with the ability to secret IP or amylase, in wild-type and a Δhac1 genotype. The strains were cultivated in batch fermentations and evaluated for their physiological properties (specific growth rate, carbon utilization efficiency, and recombinant protein secretion), transcriptome analysis and metabolic flux diversion. In producers originated from WT yeast, with the help of uracil auxotrophic plasmid, IP and α-amylase titers were 9 mg/L and 20 mg/L, respectively (Figure 10B), which is around one-tenth of the insulin produced in the CPOTud (POT1 expression system, Paper I), whereas amylase production were comparable between both expression systems. As an essential gene marker, POT1 is reported to yield a plasmid with higher copy number than auxotrophic markers (Kawasaki 1999). Different effects of expression systems on protein production and secretion could be due to the complexity of the expressed protein itself. The rate limiting step for IP secretion is probably not the folding of the protein (Kjeldsen, Frost Pettersson et al. 1999) but rather the IP synthesis (transcription and translation), and thus can be circumvented by increasing transcription. For the structurally more demanding protein, such as the α-amylase, the bottleneck for secretion is likely to be inefficiencies of post-translational processing, for example, translation, folding, and vesicle trafficking or a limitation in metabolic precursors for the protein, etc. Cases with similar opposite effects have been reported before: secretion of human parathyroid hormone, (84 amino acids, 1 disulfide bond and 0 glycosylation sites) (Gabrielsen, Reppe et al. 1990) and granulocyte-colony stimulating factor (174 amino acids, 2 disulfide bonds and 0 glycosylation sites) (Wittrup and Benig 1994) had increased production 17-fold by using a multi-copy plasmid compared to the a single copy plasmid; whereas for secretion of S. pombe acid phosphatase (435 amino acids, 8 disulfide bonds and 9 glycosylation sites), the use of a multi-copy plasmid resulted in a 24% decrease in secretion when compared to a single copy plasmid (Robinson, Hines et al. 1994).

Strains with the Δhac1 genotype had overall lower final cell densities and specific growth rates compared to strains with the WT genotype (Figure 10A), with much less protein production (Figure 10B), which implies the importance of unfolded protein response for efficient secretion (Valkonen, Penttila et al. 2003). In yeast with the WT genotype, IP did not affect growth, whereas the amylase producing strain WA had 25% lower specific growth rate compared to WN,
which implies that amylase is more challenging to be produced than IP. In the Δhac1 background, both IP and amylase production resulted in reduced growth by ~20%. For both strains the reduction in growth was not associated with a change in the specific glucose uptake rate (supplementary data of Paper II), which points out the higher energy requirement in recombinant protein producing strains compared to WT.

Figure 10. Secretory perturbations and yeast physiology in the UPR project.
(A): Specific growth rate on glucose. (B): Final recombinant protein titer. Δhac1 strains were severely inhibited in recombinant secretion. (C): The amount of oxygen consumed per dry cell weight. (D) Specific maintenance ATP consumption. Error bars represents independent triplicate samples.

The maintenance ATP consumption was calculated based on the extracellular fluxes and FBA analysis using a yeast central carbon metabolism model (Förster, Gombert et al. 2002). As shown in Figure 10C, in the WT background, compared to no protein production strain WN, WI did not have a detectable increase in ATP consumption, whereas WA did have a two-fold increase in ATP consumption. In the Δhac1 background, folding efficiency is likely decreased due to ER disfunction and even the smaller, easier to fold IP resulted in ER stress that required increased ATP consumption compared to WT. Despite the increased ATP consumption in dI and dA, little protein was secreted. The oxygen uptake rate was twice as high in the strains that were growth inhibited (for example, WA, dI, dA) than those that were not (Figure 10D). Because the biomass
yields on glucose were also lower in WA, dI, and dA, this increased oxygen uptake might be a result of increased oxidation in connection with formation of disulfide bonds rather than oxidative phosphorylation.

Growth phase transcriptomics measurements to identify the HAC1-dependent and -independent cellular responses associated with recombinant protein production. Cellular adjustments in overall expression level, post-Golgi sorting, oxidative stress, amino acid biosynthesis and savaging, were identified. Both WI and WA presented up-regulated oxidative and osmotic stress responses, compared to WN. Additional oxidative and osmotic stress pathways were activated in WA, as well as a down-regulation in some amino acid synthesis pathways and an overall reduction of transcription. In the Δhac1 background, HAC1 deletion clearly makes the cell more susceptible to recombinant protein secretion, and many of the effects found in WA, were common to both IP and α-amylase producing strains. Some oxidative and osmotic stress pathways also appear to be independent of HAC1. For example, Skn7p and Cin5p were similarly activated in both WT and Δhac1. Oxidative and hypo-osmotic stress, while important for managing the secretory pathway, appears not to be directly managed through the UPR.

Here we put forward a simple thermodynamic model of disulfide bond formation and breaking that explains the increased oxidative stress, oxygen consumption, and the reduced growth observed (Figure 11). Disulfide bond formation has been established to consume oxygen and produce reactive oxygen species (ROS) in stoichiometric quantities with the number of disulfide bonds formed. When non-native disulfide linkages are formed, these linkages must be rearranged until correct disulfide pairings are folded. Disulfide isomerization is reported to be redox neutral, not requiring electron donors or acceptors. Here, we propose that this futile cycle relies on a strong electron affinity gradient to complete an isomerization-like process and requires each disulfide pairing to have a lower electron affinity than the next. The disulfide is formed by the typical oxidation pathway (Figure 11, green) catalyzed by PDI_A, and electrons are shuttled to molecular oxygen, resulting in ROS formation. Instead of isomerization, the incorrect disulfide bond is completely reduced by a different PDI (PDI_B), and electrons are passed from NADPH, through glutathione, to the protein (Figure 11, blue). The difference in electron affinity between the folding protein’s cysteines and a specific PDI’s cysteines can only allow electrons to flow in one direction (toward the higher affinity cysteines). After several cycles, the correct disulfide bond is formed. The net result of the futile cycle is NADPH consumption and ROS production. We further propose that the relative rate of protein folding and disulfide bond formation for nascent peptides has important consequences for oxidative stress. When folding is within similar rate of disulfide bond formation, ROS is produced in near one-to-one amounts with the disulfide bonds formed. However, when folding is slower compared to disulfide bond formation, as is the case when the protein folding machinery gets overloaded, the nascent peptides cycles through the
futile redox cycle producing much more ROS than the final number of disulfide bonds formed. The physiological result following is oxidative damage to a broad range of cellular proteins and consumption of reducing equivalents, GSH and NADPH, that could otherwise be used for anabolism. This model implies that the ROS produced is not stoichiometrically linked to the number of disulfide bonds formed, but varies by the number of futile cycles before the correct bond is formed.

**Figure 11. Cycle thermodynamic model proposes non-stoichiometric ROS produced with incorrect disulfide bond formation.**


### 6.4 Paper III: Host Engineering-Integrated Analysis of UV Mutation Strains

From Paper I and II, we already got first impression that both protein type and host background can affect protein expression, and although UPR is helpful in most cases, there are indeed hidden pathways that could also benefit recombinant protein production. To enhance the production level of a given protein in a specific host strain is often a time consuming process. In order to gain new understanding of the secretory pathway, we performed UV mutagenesis and screening to further push the productions of amylase to the edge, which would be helpful to identify trouble shooting pathways and to unveil potential targets for future engineering. The designed workflow is shown in Figure 12.
Figure 12. Experimental design of the random mutagenesis project.


As shown in Figure 12A, the amylase producing strain AAC was used as starting strain for the UV mutation, and starch plates were applied as a criterion in the first round of selection, with hypothesis that bigger colonies selected have improvements in both cell growth and amylase secretion. The selected 591 strains were further cultivated in falcon tubes and shake flasks and two strains with high amylase production were identified as M715 and M1052 (7 and 10 are the UV dose applied, while 15 and 52 are orders for this two strains at each UV dose, respectively).

In order to test whether the high amylase production of M715 and M1052 is contributed by mutations in the amylase plasmid or the mutated yeast itself, mutated plasmids from both M715
and M1052 were extracted and transformed into normal strains (Figure 12B1). Shake flask cultivations of these two strains holding the mutated plasmids showed no improvement for amylase productions compared with the AAC strain, which suggested that production enhancement were due to the host mutation. Further experiments will be carried out to test whether this mutated yeast is a general better protein producer by replacing mutated amylase plasmid in M715 and M1052 with normal insulin plasmid, Figure 12B2.

Two strains NC and AAC (Paper I) were used as reference strains and four strains (NC, AAC, M715 and M1052) were evaluated under batch cultivations, Figure 12B3. Consistent with what was found in Paper I and Paper II, strains grew slower with the increased amylase production capacities. The glycerol production was similar among the amylase producing strains, although more amylase was produced by the mutated strains, suggesting a more efficient amino acid utilization process in the mutated strains. Less biomass and more ethanol was produced by the mutated strains in the glucose phase, Table 6.

Table 6. Physiological characterization of mutated strains.

<table>
<thead>
<tr>
<th>Strain</th>
<th>$Y_{SX}^a$</th>
<th>$Y_{SG}^b$</th>
<th>$Y_{SE}^c$</th>
<th>$Y_{SA}^d$</th>
<th>$\mu_{max}^e$</th>
<th>Biomass$^f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC</td>
<td>0.20±0.02</td>
<td>0.05±0.01</td>
<td>0.25±0.02</td>
<td>0.033±0.007</td>
<td>0.40±0.01</td>
<td>6.5±0.2</td>
</tr>
<tr>
<td>AAC</td>
<td>0.23±0.02</td>
<td>0.15±0.01</td>
<td>0.20±0.01</td>
<td>0.034±0.001</td>
<td>0.38±0.01</td>
<td>6.7±0.1</td>
</tr>
<tr>
<td>M715</td>
<td>0.18±0.01</td>
<td>0.15±0.01</td>
<td>0.25±0.02</td>
<td>0.036±0.001</td>
<td>0.31±0.01</td>
<td>6.6±0.1</td>
</tr>
<tr>
<td>M1052</td>
<td>0.17±0.01</td>
<td>0.14±0.01</td>
<td>0.28±0.02</td>
<td>0.045±0.005</td>
<td>0.24±0.01</td>
<td>5.6±0.1</td>
</tr>
</tbody>
</table>

Yields (g/g glucose) calculated here only consider the exponential phase and the total consumed substrate. $^a$Biomass, $^b$Glycerol, $^c$Ethanol, $^d$Final succinate production, $^e$Specific growth rate (h$^{-1}$), $^f$Final biomass (g/L). The data represented triplicated biological experiment.

The mutated strains exhibited high amylase producing capacities (yield on cell mass) (Figure 13). For amylase yield on biomass in the exponential phase, M715 strain produced 5.4-fold of amylase and M1052 strain produced 4.9-fold of amylase compared to AAC. When comparing final amylase productions, M715 and M1052 strains produced 2.4-fold and 3.5-fold of amylase, respectively. The mutated strains showed increased amylase production in the glucose phase compared with the whole fermentation phase, suggesting that gene mutations may be related with protein processing in the glucose phase.

Samples were taken for transcriptome analysis of the four strains (NC, AAC, M715 and M1052) during batch cultivations. Through Reporter TFs analysis of the transcriptome data (Patil and Nielsen 2005; Oliveira, Patil et al. 2008), as shown in Figure 14, we identified that most changed genes were related to stress responses. Genes regulated by oxidative stress (Yap1p), osmotic stress (Hog1p), and general stress (Msn2p and Msn4p), were up-regulated in a RPP-dependent manner (comparing all amylase producing strains with the control strain), whereas down-regulated in a mutation-dependent manner (comparing the two UV mutated strains with the non-mutated strain). We also found that genes related to respiration, regulated by Hap2p, Hap3p,
Hap4p and Hap5p, were down-regulated in the mutated strains compared to AAC. Since we identified that amylase is produced at higher levels at anaerobic conditions compared to aerobic conditions (Paper IV), we suggest that the reduced respiration and stress responses might both contribute to the higher production in the mutated strains.

**Figure 13. Amylase production of the UV mutation and reference strains**
(A) Amylase yield on cell mass during the exponential phase. (B) Final Amylase production.

In order to investigate how the improved protein production in the UV mutated strains is related to changes in the protein secretory pathway, we took an integrated data analysis approach. Recently, the first yeast secretory model that covers 170 secretory proteins, classified into 16 secretory classes, was generated (Feizi, Österlund et al. Submitted). Here, the secretory network was further expanded to also involve regulation of the secretory pathway and response to stress. We included genes regulated by UPR (Travers, Patil et al. 2000; Kimata, Ishiwata-Kimata et al. 2006), and secretory genes classified to the transcription factor response to oxidative stress (Yap1p), response to heat shock (Hsf1p) and general stress response (Msn2p and Msn4p). The final list of genes involved in protein secretion was obtained after manually correction based on Saccharomyces Genome Database and literature reading. The Reporter Features algorithm (Patil and Nielsen 2005; Oliveira, Patil et al. 2008) was then used to score the secretory pathway functions by the significant changes of genes expression that belong to the related pathway. As shown in Figure 15, genes belonging to the protein trafficking pathway were up-regulated in both mutated strains compared to AAC. Genes associated with vacuole and amino acid metabolism were specifically regulated in M715 and M1052.
Figure 14. Reporter TFs analysis reveals reduced stress responses in mutated strains.

Figure 15. Top ten regulated secretory pathway functions in mutated strains.
Whole genome sequencing analysis was carried of the mutated strains and genes that have single nucleotide point variations (SNVs), insertions and deletions (INDELs), as well as mutations in the upstream region (0-1000bp) are shown in Figure 16. Silent mutated genes (genes that only changed in nucleotide sequence but not in the amino acid sequence) and genes that changed in the upstream region but not transcriptional regulated were further filtered out. Totally, 17 genes were commonly changed in the two mutated strains.

**Figure 16. Whole genome sequencing analysis.**
(A) Venn diagram of genes that have single nucleotide variations (SNV), insertion and deletions (INDELs) and also changes in the promoter region in mutated strains. (B) Venn diagram of all mutated genes. Upstream region: 0-1000 bp upstream the exon start.

Most mutated genes belong to pathways that are involve in the translation process, and genes that have close relation with protein processing pathway were presented in Table 7, as potential targets for future experiment for enhancing protein production. Similar results showed in genetic changes as in transcriptome analysis, that common mutations in both strains enclosed multiple genes regarding stress response. Though different genes were changed, genes regarding protein degradation, protein trafficking pathways were both mutated also. Genes related to respiration and amino acid metabolism was specifically changed in the M715 strain.

Taken together results of both transcriptome analysis and genome sequencing analysis, we suggested that genetic changes regarding stress response, respiration, protein degradation, protein trafficking and amino acid metabolism might be the reason for the increased amylase production.
### Table 7. Genetic changes.

<table>
<thead>
<tr>
<th>Mutations</th>
<th>Single mutations</th>
<th>INDELs</th>
<th>Promoter region</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SEC7 [Gly92→Val]</td>
<td>CDC27</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HSP82 [Gln135→His]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CO88 [Ser133→Arg]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLR1</td>
<td></td>
</tr>
<tr>
<td>M715</td>
<td>YKR105C [Lys335→Asn]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unique</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA1 [Ser631→Leu]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIR4 [Ser175→Pro]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDH1 [Ser121→Tyr]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILV2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPD1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COQ5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PGM3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KTR2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFU1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1052</td>
<td>RPL11B [Glu168→Glu]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unique</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRS31 [Glu220→His]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WWW1 [Glu198→Asp]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VTA1 [Ser306→Asp]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VPS3 [Ser21→*]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTN2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PUP2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATG23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**6.5 Paper IV: Fermentation Analysis-Anaerobic Protein Processing Machinery**

In Paper I, we found that amylase was produced at lower levels in the glucose phase compared to the ethanol phase. In Paper II, we found that the oxygen consumption rate is much higher in amylase producing strain (WA) compared to insulin producing strain (WI). Is it because that the higher ER translocation caused depletion of chaperones and folding enzymes for amylase production and converted more oxygen to ROS? Did oxidative stress cause the reduction of amylase production? If at aerobic conditions, electrons are transferred from unfolded proteins to oxygen in order to complete the folding process, how does it happen at anaerobic conditions? What is the anaerobic electron acceptor for protein folding?

Three strains with varied amylase production (NC, AAP and AAC from Paper I) were cultivated at aerobic and anaerobic conditions, respectively. For amylase production in the exponential phase, both the AAP and AAC strains presented better amylase production at anaerobic conditions, both in terms of per unit yield of biomass (Figure 17A) and productivity (Figure 17B). These data suggested that anaerobic condition provides a more suitable environment for amylase production, and microarray analysis was therefore performed to identify the molecular mechanisms behind these remarkable differences.
Figure 17. Amylase produced more at anaerobic conditions compared to aerobic conditions
(A) Amylase yields on cell growth. (B) Protein productivities of each strain/condition. (Grey bar): Strain physiologies at aerobic conditions. (Black bar): Strain physiologies at anaerobic conditions. Error bars represented standard errors based on independent triplicate experiments.

Reporter Gene Ontology (GO-terms) and TFs were applied in order to reduce dimensions of the transcriptome data. Figure 18 provides a summary of the common and specific transcriptional changes of genes in secretory pathway functions for the two amylase strains compared to the NC strain at anaerobic and aerobic conditions. We suggest that anaerobic conditions provide better amylase producing environment than aerobic conditions. Cells at normal anaerobic conditions is capable of a low amount of amylase production, between 0.37 mg/g DCW/h (AAP) and 2.6 mg/g DCW/h (AAC), without up-regulation of any secretory helpers (as shown in Figure 18 that, AAP strain did not up-regulate genes related with ER, protein degradation and stress response when compared with NC strain, which is different from aerobic conditions). Whereas when more amylase was produced, key functions in the secretory pathway were activated as at aerobic conditions. Because i) anaerobic cultivations clearly showed a higher amylase production than at aerobic conditions (Figure 17), ii) genes belong to ER functions and stress related responses were up-regulated when comparing AAC strain with NC at both aerobic and anaerobic conditions, iii) even though anAAP produced more amylase than aAAP, genes within key functions in the secretory pathway as well as many stress related pathways were only up-regulated in aAAP but not in anAAP when compared to NC.
In order to identify putative final electron acceptors for the protein folding in the ER at anaerobic conditions, we identified key metabolites around which significant transcriptional changes occurred using the Reporter Metabolite algorithm (Patil and Nielsen 2005; Oliveira, Patil et al. 2008). The top 15 reporter metabolites for each strain when comparing anaerobic and aerobic conditions, were clustered, as shown in Figure 19. It is remarkable that 11 Reporter Metabolites were the same for all three strains, which could be further grouped into two clusters: 1) ATP, ADP, ferricytochrome/ferrocytochrome, orthophosphate and mitochondrial protons, which have close relations with energy metabolism; and 2) fumarate, oxygen, FADH₂, FADH, ubiquinol and ubiquinone-9, which are shown to be even more significant in AAC and AAP than in NC. Here we propose that the second cluster is associated with intercellular electron transfer. We also found that transcriptional levels of: i) FRD1 and OSM1 coding for fumarate reductase and ii) FAD1 coding for FAD synthesis, FLC1 coding for FAD cytosol-ER transporter (Tu and Weissman 2004) and ERV2 coding for ER disulfide bond formation, are all up-regulated at anaerobic conditions compared to aerobic conditions.
Figure 19. Top 15 reporter metabolites in the three strains when comparing anaerobic to aerobic conditions.

Ero1p plays an essential role in catalyzing disulfide formation in folding proteins, whereas over-expression of FMN1 or FAD1 (Tu, Ho-Schleyer et al. 2000) or increasing cellular free FAD levels (Tu and Weissman 2002) could restore cell growth of a temperature-sensitive allele of ERO1 (ero1-1). It was also reported that free FAD was essential for RNase refolding catalyzed by Ero1p and PDI (Tu and Weissman 2002), and therefore it suggested that Ero1p might contain domains that work with free FAD (Tu and Weissman 2004). All this evidence demonstrated the important role of cellular free FAD levels on the protein folding in the ER. It has been reported that under anaerobic conditions Ero1p could directly transfer electrons to free FAD (Gross, Sevier et al. 2006). Here, we suggest that under anaerobic conditions, free FAD could act as the electron carrier who takes parts in the electron transfer during protein folding in the ER.

Single deletion of either OSM1 or FRD1 does not affect the anaerobic cell growth (Camarasa, Faucet et al. 2007), whereas a double deletion is lethal at anaerobic conditions but it has no growth effect at aerobic conditions (Arikawa, Enomoto et al. 1998). It suggests that this essential role of fumarate reductase (Frd1p) is because that it catalyzes the only reaction that could oxidize free FADH2 under anaerobic conditions (Camarasa, Faucet et al. 2007). Many anaerobic species use fumarate as the final electron acceptor, and here we suggest that FADH2 after accepting electrons from the ER protein folding is oxidized by fumarate reductase.
Here we presented a model that demonstrated that at anaerobic conditions, FAD could be the electron transporter and fumarate could be the final electron acceptor for protein folding of *S. cerevisiae*, as shown in Figure 20. There are two electron transferring pathways reported in the ER: in connection with disulfide bridge formation electrons are passed through PDI to either Ero1p or Erv2p (Gross, Sevier et al. 2006). When comparing anaerobic to aerobic conditions, the expression of neither *PDI1* nor *ERO1* were changed, whereas *ERV2* was up-regulated in all three strains, which suggested a more important role at anaerobic conditions (Sevier, Cuozzo et al. 2001). Instead of oxygen, electrons are transferred from the Ero1p bounded FAD to free FAD in two possible routes: i) Since FAD could be transported across the ER membrane (Tu and Weissman 2004), electrons could be transferred to the FAD in the ER lumen and thereafter be exported to the cytosol; or ii) as Ero1p is closely associated with the ER membrane (Pagani, Pilati et al. 2001), electrons could be directly transferred from the membrane spanning part of Ero1p to free FAD in the cytosol. In the cytosol FADH$_2$ could either be consumed when fumarate is converted to succinate by the cytosolic fumarate reductase Frd1p or it could be transferred to the mitochondrion and there get oxidized by the mitochondrial fumarate reductase Osm1p.

**Figure 20. Anaerobic electron transfer model predicts fumarate to be the final electron acceptor for protein folding.**

(Blue box): intracellular metabolites; (Red oval): up-regulated enzymes; (Green oval): down-regulated enzymes; (Grey oval): unchanged enzymes; (Black line): metabolic pathways; (orange line): electron transferring pathways; (dashed line): alternative electron transfer reactions.
From the fermentation, we found that the specific growth rate is much lower in anAAC than anNC, as shown in Figure 21. In Paper II, we found that the oxygen (aerobic electron acceptor for protein folding) consumption rate in amylase producing strain (WA) is twice as high as that in wild type strain (WN). If fumarate is the anaerobic electron acceptor, is the growth defect caused by fumarate starvation? In order to evaluate this hypothesis cell growth was evaluated with 0.5 g/L fumarate added into the SD-2×SCAA media under aerobic and anaerobic conditions, respectively. It is seen that after adding fumarate the growth rate of anAAC increased by about 10%. The effect may partially be explained by fumarate acting as a carbon source, but fumarate addition at aerobic conditions did not show a growth effect, which suggests that fumarate clearly has a positive effect on other aspects. We therefore conclude that the increase in biomass yield is mainly because the electron acceptor role of fumarate.

Figure 21. Fumarate promotes cell growth at anaerobic conditions.
(White bar): Specific growth rate data of the NC strain. (Slash bar): Specific growth rate data of the AAP strain. (Grey bar): Specific growth rate data of the AAC strain. (Black bar): Specific growth rate data of the AAC strain cultured in SD-2×SCAA media with 0.5g/L fumarate. Error bars are based on independent triplicates except for the fumarate fermentations, which are based on independent duplicates.

6.6 Paper V: Fermentation Analysis-Cell Growth Effects

From Paper I, we found that amylase was produced faster in the ethanol phase compared to the glucose phase. Is this because the reduced translocation of growth related proteins allows for improved amylase secretion? Is protein production coupled with the cell growth and metabolism in a protein-dependent or growth rate-dependent manner? To study the effect of growth rate on heterologous protein production, AAC and AIC, amylase and insulin producing strain in Paper I, were compared and evaluated in carbon-limited chemostat cultivations operated at different
dilution rates. The specific growth rate of the recombinant protein production strains in batch cultivation is around 0.25 h\(^{-1}\), and the dilution rate of the chemostat cultivations was therefore controlled as 0.05, 0.1 and 0.2 h\(^{-1}\).

The specific productivity of the two recombinant proteins for the different dilution rates is shown in Figure 22 together with the biomass concentration. We found that although the productivities of both proteins increases with increasing specific growth rates, the yield of IP on biomass resulted in a sharper increase at specific rate of 0.2 h\(^{-1}\), whereas the effect was not so strong on the amylase. On the other hand, the yield of amylase on substrate decreased at higher specific growth rates (with the highest value at specific rate of 0.05 h\(^{-1}\)), whereas the yields of IP on substrates were comparable at different specific growth rates.

![Figure 22. The heterologous protein production in chemostat cultivations](image)

(A) The productivity of amylase and insulin precursor at different growth rates. (Square), insulin producing strain AIC. (Circle), amylase producing strain AAC. (B) Final biomass production of AAC and AIC at different growth rates. (Black bar), 0.05 h\(^{-1}\) dilution rate. (Grey bar), 0.1 h\(^{-1}\) dilution rate. (blank bar), 0.2 h\(^{-1}\) dilution rate.

In order to study general effects of growth rates on heterologous protein production, the transcriptome data of different dilution rates was compared for both strains. Reporter KEGG pathway analysis showed that N- and O-link glycosylations were up-regulated in both RPP strains at higher dilution rate conditions. UPR genes were up-regulated in both strains upon increased dilution rates. And more importantly, the \(HAC1\) gene, which encodes the transcription factor that initiates UPR was also up-regulated. The information in Table 8 indicates that when expressing recombinant proteins (IP and amylase in our case), the UPR level was activated at high specific growth rates, and in the end the so-called super-UPR (besides UPR associated
genes, the HAC1 is also transcriptional up-regulated (Bernales, Papa et al. 2006)) may also be activated when cells were approaching their highest specific growth rate.

**Table 8. Genes associated with ER protein processing were regulated among different dilution rates**

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Unfolded protein response</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common</td>
<td>WSC4, MCD4, LHS1, ERD2, PMT2</td>
<td>SSI1, DGK1, OST1, SIL1</td>
</tr>
<tr>
<td>D0.1/D0.05</td>
<td>HMX1, MNT3, ALG1, ALG5, GPI8, GET1, OST4, OST5, SBI2, VRG4, PMT6, KRE27, PER1, TRS33, SRP21, SNL1, GSF2, ALG12, UBC6, FES1, SSE2, SSAR, RRT12, SSA3</td>
<td></td>
</tr>
<tr>
<td>D0.2/D0.1</td>
<td>CDC48, SEC23, FPR2, KEG1, HLJ1, GET3, CWI41, STT3, ERD1, EMP24, SAR1, KRE11, SRP101, SEC21, SVP26, MSC7, SHE3, SSM4, MIDI, CG2, OST4, BST1, USO1, SEC39, EPS1, ZRG17, HSP26</td>
<td></td>
</tr>
</tbody>
</table>

Bold genes were up-regulated in each comparison.

Reporter TFs showed that genes related to Msn2p and Msn4p were higher expressed at the lowest dilution rate condition (D=0.05 h⁻¹), with the highest protein yields on substrate. On the other hand genes associated with Yap1p were down-regulated when comparing dilution rates of 0.1 to 0.05 h⁻¹, and up-regulated when comparing 0.2 to 0.1 h⁻¹. Reporter KEGG pathway analysis indicated that genes related to the proteasome were expressed in the same manner as Yap1p associated genes: with increasing cell growth rates, they were down-regulated at low specific growth rate conditions (0.05 to 0.1 h⁻¹), whereas they were up-regulated at high specific growth rates (0.1-0.2 h⁻¹). Since a low specific growth rate (0.05 h⁻¹) leads to cell starvation, the up-regulated protein turnover rate can be interpreted as a survival mechanism (Fazio, Jewett et al. 2008), and hence higher amylase production titers are found upon more efficiently utilizing cellular resources and replenishing the pool of free amino acids (Arvas, Pakula et al. 2011).

Here we suggest that growth effects on recombinant protein production mainly rely on ER functions, stress responses and proteasome activities, as summarized in Figure 23A. The growth rate of 0.1 h⁻¹ appeared to be a shifting point between growth effects and RPP effects that take over the main responses. At low specific growth rates (0.05-0.1 h⁻¹), growth effects play the main role based on the nutrient-dependent stress and proteasome responses, whereas at higher specific growth rates (0.1-0.2 h⁻¹), the RPP effects start to play the main role, which were indicated by: i) Genes related to general stress related transcription factors (Msn2p and Msn4p) showed similar expression levels indicating that when growth rates increased above 0.1 h⁻¹, the nutrient starvation was deactivated; ii) Genes related to oxidative stress (Yap1p) were up-regulated when comparing specific growth rates of 0.2 to 0.1 h⁻¹; iii) More importantly, the super-UPR was activated at high specific growth rates, which might positively cause the up-regulated genes associated with proteasome and protein processing in the ER.
In order to further unveil RPP effects, we also performed reporter feature analysis where AAC with AIC were compared at three different dilution rates. As summarized in Figure 24B, since amylase is a bigger protein compared to IP, even though it is produced in a moderate level, the amino acid metabolism and energy metabolism were up-regulated in the amylase producing strains. The RPP effects of both strains also presented a shifting manner around the specific growth rate of 0.1 h\(^{-1}\): i) at the specific growth rate 0.05 h\(^{-1}\), because IP is produced at higher levels, expression of these genes was higher in IP strains; ii) at the specific growth rate 0.1 h\(^{-1}\), the burden of amylase starts to affect the secretory pathway, and causes a similar level of ROS production; and iii) at the specific growth rate 0.2 h\(^{-1}\), even though IP was produced in a much higher level than amylase, the amylase tends to become misfolded, which might result in more ROS and causes oxidative stress, so the expression of Sod1p associated genes was up-regulated in the amylase strain.

**Figure 23. Common and distinct pathways that were regulated at different dilution rates**


Taken together, since amylase is a relatively more complex protein than IP, the post-translational processing in the secretory pathway could be the restrictive step for its over-expression, whereas IP production could be easily manipulated by simply engineering the expression status (Paper I). In other words, \(\alpha\)-amylase needs an increased folding capacity, therefore a lower specific growth rate (including ethanol phase in batch fermentation) may provide a better environment for amylase secretion due to the lower ER burden from endogenous protein secretion and higher protein turnover, which allow a more efficient removal of misfolded proteins.
CONCLUSIONS AND PERSPECTIVES

The work presented in this thesis applies metabolic engineering and systems biology tools to explore ways for recombinant protein over-production and its relationship with cell metabolism and protein production. Careful molecular design (Paper I), fermentation physiological characterization applied with systems biology tools, including genomic sequencing analysis (Paper III), transcriptome analysis (Papers II, III, IV and V) and flux analysis (Papers II and IV), not only enabled explaining specific phenotypes observed, but more importantly, provided insights for second-round engineering for improved recombinant protein production (Paper III). Based on the results and findings presented, we have advanced the understanding of the global regulation by answering key questions listed in the following:

**Question 1: How do the expression factors affect recombinant protein production? Why does the very same approach result in enormous different effects for different proteins?**

We report that although the transcription level of the recombinant gene is important, the final production of the recombinant protein is the result of a combination of effects of transcription and translation levels, protein uniqueness, and leader sequences which influences the secretory pathway processing efficiency. We also report a notable difference in production of IP and α-amylase, and we conclude that this difference is caused by differences in their processing through the secretory pathway. For IP the important step is the synthesis of the protein, whereas for amylase the rate-controlling step for secretion was found to be most likely ER folding and processing.

**Question 2: How do cells manipulate recombinant protein processing in an UPR-dependent and in-dependent manner? Why does heterologous protein production always come along with oxidative stress? What are the hidden factors that could be engineered for host design to achieve high levels of recombinant protein production?**

Host engineering was carried out with focus on effects of the unfolded protein responses (Paper II) and random mutations for understanding the limitations for high protein production (Paper III).

In Paper II, we identified post-Golgi vesicle sorting, high protein degradation rates, repressed overall expression, and oxidative stress in response to +/- UPR strains secreting different size of recombinant proteins. We proposed futile cycling as the dominant disulfide resorting pathway in the ER and used this to explain non-stoichiometric ROS formation observed in our study and elsewhere.

In Paper III, we identified biological mechanisms, which alter the secretory pathway in response to UV random mutagenesis for production of high levels of recombinant proteins, and proposed that genetic changes in stress response pathway, respiration, as well as protein trafficking and...
degradation might contribute to increased amylase production. We also purposed potential targets for enhancing protein production for future engineering.

**Question 3: How does the cultivation condition affect recombinant protein production? Why do some proteins get produced better under anaerobic conditions, some present higher production when cells were grown at a lower specific growth rate, whereas others do not?**

Investigation of interconnections between cell metabolism and recombinant proteins production was carried out using systems biology tools, with a special focus on aeration effects and growth rate effects (Paper IV and V).

In Paper IV, in response to oxygen concentrations in the environment, we have identified cellular adjustments in (a) transcription and translation, (b) amino acid metabolism, (c) protein folding/modification, (d) cytosolic redox control, (e) post-Golgi sorting, and (f) protein degradation. We also provide a model for electron transfers and the final anaerobic electron acceptor for the protein folding in the ER.

In Paper V, our experiments demonstrated that although the specific growth rate may couple to the protein secretion rate, the final effect is strongly correlated with the features of the specific protein. When expressing recombinant proteins, the UPR was activated according with increasing specific growth rates, and the super-UPR might also be activated when approaching the maximum specific growth rate. The impact of specific growth rate was protein specific and fermentation optimization should be based on the properties of proteins.

Taken together, we reported that amylase was produced at a much faster rate in the ethanol phase, whereas the production rate of IP dropped substantially after the diauxic shift (Paper I). IP showed a production increase when we use a higher copy number plasmid, whereas amylase production was comparable between the \textit{POT1} expression system and the \textit{URA3} plasmid (Paper I and Paper II). The amylase yield increased more than 2 fold in fed-batch cultivation (0.08 h\(^{-1}\) feed rate) compared to batch cultivation, whereas the IP yield did not show a clear difference (Paper VII). We also found that amylase was produced more at anaerobic conditions comparing to aerobic conditions (Paper IV). The yield of amylase on substrate decreased at higher specific growth rates (with the highest value at specific rate of 0.05 h\(^{-1}\)), whereas the yields of IP were comparable at different specific growth rates (Paper V). We therefore suggest that IP is produced in a growth associated manner, and that the bottleneck for insulin production remains in the limitation of expression. On the other hand amylase tends to gain higher production levels at lower cell growth conditions, with a reduced rate of ER translocation and ER misfolding. In order to show a broader view, protein production data at three different dilution rates in chemostat cultivation (Paper V) was also compared with production in batch and fed-batch fermentations (Paper VII). Although the cell metabolism is different between batch, fed-batch
and chemostat cultivations, there is still a clear trend which shows that the productivity for IP increased significantly along with increased specific growth rates, whereas for amylase production, the productivity curve remained at moderate levels, which indicated that there should be another key factor that regulates amylase production besides cell growth.

The efficient expression systems, host mutations, fermentation techniques, combined with the advances in systems biology described in this thesis have contributed valuable information to improve recombinant protein production, and also shed lights to a deeper understanding of the secretory pathway. The focus of future engineering might include the following aspects: i) the quantitative understanding of different steps regarding protein production (Graf, Dragosits et al. 2009), ii) whole cell metabolome profiling in order to identify metabolite makers for protein associated cellular stress and mis-functions, iii) genomics-scale model construction that focuses specially and intensively on the secretory pathway, as well as all possible elements that might affect protein production, iv) detailed characterization of recombinant protein production within each phase of the cell cycle, which could gain a deeper insights of the correlation between protein production and cell metabolism, v) combining the state of art systems biology tools to analysis the interreaction between protein and lipid metabolism, especially in the ER.
ACKNOWLEDGEMENTS

When starting my Ph.D. study, I was told that it is a journey fluctuates with joy and pain, and it will possibly reach its utmost decline right before my graduation. Fortunately, there are so many people willing to increase my joy and share my pains, which make it one of the most important periods in my life.

First of all, I am very grateful to my supervisor Prof. Jens Nielsen, who has trusted me and my talent since the very first time we met. I am also thankful for the wonderful research topic he provided me and all those great discussions that contributed substantially to all my scientific discoveries throughout my Ph.D. study. He always believes in me and encourages me to explore my potentials, even during times that I had less belief in myself. Not only being a top-class scientist, he is the best mentor one can ask for, because he also cares to share his wisdoms and support on not only my academic career, but also on my personal life.

I would also like to thank my co-supervisor, Dr. Dina Petranovic. She always looks after me, and translates professional suggestions with respect of my specific circumstances. Dina, your kind support and educations have made me evolve much faster from a student researcher into a young scientist.

A special acknowledgement goes to Dr. Keith Tyo, who I collaborated deeply during my first two years as a Ph.D. student. He taught me generously and continually hand by hand during my new experiments, manuscript writing, new project planning, etc. Keith, I learned so much from you for how to be an independent scientist and also, a supporting friend.

In addition, I am also thankful to all my wonderful collaborators: Dr. Jin Hou, Dr. José L. Martinez, Dr. Mingtao Huang, Ph.D. student Tobias Österlund, Ph.D. student Lifang Liu for their marvelous contributions and excellent scientific discussions on those projects that we worked directly and indirectly together; Dr. Verena Siewers, Dr. Yun Chen and Dr. Siavash Partow for all the valuable suggestions on molecular experiments; Dr. Jie Zhang and Dr. Yun Chen for support and troubleshooting for fermentations; Dr. Intawat Nookaew, Dr. Rahul Kumar and Dr. Pramote Chumanpuen for help with genomic sequencing and transcriptome analysis; Dr. Nathalie Scheers, Dr. William Rodriguez Limas and Dr. Goutham Vemuri for advices on protein measurements; Dr. Liming Liu, Dr. Yu Shen for all the career advices; Erica Dahlin, Martina Butorac, Malin Nordvall, Marie Nordquist and Johanna Zanden for all the social and lab support and in addition all colleagues at Chalmers; Shuobo Shi, Chaoyang Zhu, Yuxin Song, Yuyan Zeng, Yifeng Fu, Bingbing Shi, Huan Zhao, Jianqiang Li, Guowei Lv, Jie Zhang, Lei Yang, Yun Chen, Ruiling Cai, Liming Liu, Jin Hou, Yu Shen, Lifang Liu, Yiming Zhang, Mingtao Huang, Jiufu Qin, Yongjin Zhou, Zhongpeng Guo, Ruifei Wang, Yanming Wang, Jichen Bao, Xiaobing Wu, Haichuan Ji, etc, for all the gathering and joys that made time flies
when I felt lonely and less cheerful. I am very honest to say that I will not make it without your generous and support.

I thank the funding sources, including the EU Framework VII project SYSINBIO (grant no. 212766), European Research Council ERC project INSYSBIO (grant no. 247013), the Chalmers Foundation, the Knut and Alice Wallenberg Foundation, the Novo Nordisk Foundation, Nils Pihlblads resestipendium and FEBS YIF grants for financial support to my research and conference travels.

Finally, I want to thank my parents and families as well as friends in China and Sweden for their love and support. Last but not least, I owe my deepest gratitude to my dear husband. Shuobo, thank you so much for your unconditional love and support. To be with you make me happier than I could ever imagine.
REFERENCES


Feizi, A., T. Österlund, et al. (Submitted). "Genome-scale modeling of the protein Secretory machinery in yeast."


PAPER I

- UPR Effect
- Random Mutation
- Host Engineering
- RPP IP and amylase
- Leader
- Promoter
- Marker
- Expression Design
- Aeration
- Cell growth
- Fermentation Analysis
Different Expression Systems for Production of Recombinant Proteins in Saccharomyces cerevisiae

Zihe Liu, Keith E.J. Tyo, José L. Martínez, Dina Petranovic, Jens Nielsen

Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden; telephone: 46-31-7723804; fax: 46-31-7723801; e-mail: nielsenj@chalmers.se

Received 15 September 2011; revision received 28 November 2011; accepted 8 December 2011
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bit.24409

ABSTRACT: Yeast Saccharomyces cerevisiae has become an attractive cell factory for production of commodity and speciality chemicals and proteins, such as industrial enzymes and pharmaceutical proteins. Here we evaluate most important expression factors for recombinant protein secretion: we chose two different proteins (insulin precursor (IP) and α-amylase), two different expression vectors (POTud plasmid and CPOTud plasmid) and two kinds of leader sequences (the glycosylated alpha factor leader and a synthetic leader with no glycosylation sites). We used IP and α-amylase as representatives of a simple protein and a multi-domain protein, as well as a non-glycosylated protein and a glycosylated protein, respectively. The genes coding for the two recombinant proteins were fused independently with two different leader sequences and were expressed using two different plasmid systems, resulting in eight different strains that were evaluated by batch fermentations. The secretion level (μmol/L) of IP was found to be higher than that of α-amylase for all expression systems and we also found larger variation in IP production for the different vectors. We also found that there is a change in protein production kinetics during the diauxic shift, that is, the IP was produced at higher rate during the glucose uptake phase, whereas amylase was produced at a higher rate in the ethanol uptake phase. For comparison, we also refer to data from another study, (Tyo et al. submitted) in which we used the p426GPD plasmid (standard vector using URA3 as marker gene and pGPD1 as expression promoter). For the IP there is more than 10-fold higher protein production with the CPOTud vector compared with the standard URA3-based vector, and this vector system therefore represent a valuable resource for future studies and optimization of recombinant protein production in yeast.


KEYWORDS: α-amylase; insulin precursor; expression systems; leader sequence; secretory pathway; Saccharomyces cerevisiae

Introduction

Recombinant proteins include important pharmaceuticals for treatment of diseases such as diabetes or cancer, and today there are more than 200 biopharmaceuticals on the market (Walsh, 2010) and new clinical studies show potentials for much wider use of recombinant proteins for treatment of other diseases (Aggarwal, 2010). In order to meet the demand for recombinant proteins, there is a need for efficient expression systems with high productivity. The limitation is often in terms of obtaining sufficient quantities of recombinant proteins for clinical studies or for production at sufficiently low cost to allow for marketing (Werner, 2004). Different host systems have been described, and unicellular microorganisms are often preferred because of their short generation times, high biomass yields, and well-characterized manipulation/modification techniques (Porro et al., 2005).

Saccharomyces cerevisiae is a well-characterized eukaryal model organism for production of heterologous proteins. Contrary to bacterial host systems, S. cerevisiae possess the ability to perform post-translational modifications and secretion, which has dramatically dropped the cost of post-fermentation in vitro purification and modification (Schmidt, 2004). S. cerevisiae is also more tolerant to low pH, high sugar and ethanol concentrations, and high osmotic pressure, which makes it suitable for industrial fermentations (Hahn-Hägerdal et al., 2007).
It has been found that enhancement of recombinant protein secretion can be achieved by the combination of the following factors: (i) engineering of the host strains, for example, over-expressing of genes for folding chaperones (Chigira et al., 2008; Payne et al., 2008); over-expressing of genes for trafficking proteins (Toikkanen et al., 2004) and reducing intracellular and extracellular proteolysis (Zhang et al., 2001); (ii) engineering DNA sequences and expression systems, for example, modifying protein coding sequences (Kim et al., 2006; Zhang et al., 2003) and signal sequences (Li et al., 2002; Rakestraw et al., 2009); optimizing expression systems (increasing plasmid copy numbers (Finnis et al., XXII International Conference on Yeast Genetics & Molecular Biology) and gene expression efficiencies) (Fama et al., 2007; Hackel et al., 2006); and (iii) optimizing the environmental/cultivation conditions (Homma et al., 2003).

Different proteins differ significantly in both their folding behaviors and amino acid demands, which lead to different levels of cell stress, and hence result in different levels of final productions. There is no one ultimate method that could work equally well for production of all proteins. Small and simple proteins could be efficiently folded faster, while multi-domain proteins could need more assistance during folding and require certain chaperones and responses to facilitate the process (Tutar and Tutar, 2010). One well studied and also very successful secretion strategy for one protein (Smith and Robinson, 2002; Smith et al., 2004; Xu et al., 2005), does not always yield a promising production for another protein (Butz et al., 2003; Harmens et al., 1996).

An additional feature should be taken into consideration when devising strategies for efficient protein secretion. The pre–pro leader sequences are very important factors that facilitate secretion of the protein product. The pre-leader is responsible for directing the peptide through the translocation step into the ER, and the pro-leader is designed to facilitate secretion of the protein product. The pre–pro leader sequences are very important factors that are responsible for directing the peptide through the translocation step into the ER, and the pro-leader is designed to facilitate secretion of the protein product. The pre–pro leader sequences are very important factors that are responsible for directing the peptide through the translocation step into the ER, and the pro-leader is designed to facilitate secretion of the protein product.

Vector engineering has also been extensively studied for different purposes. The marker type and promoter strength of the expression systems are key factors that determine the plasmid copy number and the mRNA level of the recombinant protein. Different marker systems (Kuroda et al., 2009) and promoter libraries (Fischer et al., 2006; Partow et al., 2010) have been made and evaluated for recombinant protein production. Toxicity genes (Agaphonov et al., 2010; Sidorenko et al., 2008), auxotrophy genes (Chigira et al., 2008; Stagoj et al., 2006), defective auxotrophy markers (Corrales-Garcia et al., 2011), and essential genes in the glycolytic pathway (Kjeldsen et al., 2002) are commonly used as selective markers. The downside of auxotrophy marker expression systems is that they have to be maintained in the synthetic medium. In contrast the POTI expression systems have the advantage of having high plasmid stability, even when strains are cultivated in rich medium, which can generate higher cell numbers and higher protein production (Kawasaki, 1999, US005871957A). Promoters that initiate strong and constitutive expression are often chosen for recombinant protein production: the widely used TEF1 promoter of S. cerevisiae can drive high gene expression in both high glucose conditions and glucose limited conditions (Partow et al., 2010); and the TPII promoter (of strongly expressed glycolytic gene TPII of S. cerevisiae, coding for triose phosphate isomerase), is also often used for production of recombinant proteins (Egel-Mitani et al., 2000; Kjeldsen et al., 1998b).

In order to further evaluate the process of protein-specific secretion, different types of proteins are often studied and compared using the same strategy (Rakestraw and Wittrup, 2006; Robinson et al., 1996). IP and α-amylase are two widely studied proteins that we also used in our study. IP contains a 29-amino acid B chain and the normal 21-amino acid A chain of insulin connected by a mini-C chain of only three amino acids to ensure efficient expression (Kjeldsen et al., 1999), and it is a single chain peptide with three disulfide bonds and no N-glycosylation sites. α-Amylase from Aspergillus oryzae is a three-domain protein (Randez-Gil and Sanz, 1993) with 478 amino acids, four disulfide bonds, and one glycosylation site.

We report here the construction of eight engineered strains producing two representative recombinant proteins, IP and α-amylase, in batch cultures with diauxic shift. The engineered strains were producing either IP or α-amylase using two different secretion leaders (the native and glycosylated alpha factor leader vs. the synthetic and non-glycosylated leader Yap3-TAS5), using two different promoters (TEF1 promoter and TPII promoter) and using a plasmid that uses the POTI gene (from glycolytic pathway of Schizosaccharomyces pombe) as a marker (in combination with deletion of the corresponding S. cerevisiae gene in the genome). The strain with mutation in the native genomic tpi gene does not grow on glucose and the complementation with the functional copy of the heterologous TPI (in this case the POT gene from S. pombe) results in increasing the
plasmid copy number in the cell, in order to sustain rapid growth on glucose. In order to show the advantage of the POT1 plasmid system, eight different POT1 derived strains were also compared with two strains in which IP and α-amylase were produced using a traditional auxotrophy plasmid-p426GPD, with URA3 marker and the GPD-promoter as expression promoter (Tyo et al. submitted). This study provides insights about the effect of secretion promoter as expression promoter (Tyo et al. submitted).

Materials and Methods

Strains and Media

*Escherichia coli* DH5α (Bethesda Research Laboratories) was used for plasmid constructions. The reference strain *S. cerevisiae* CEN. PK 530-1C (kindly provided by Peter Kötter, University of Frankfurt, Germany) was used as the yeast host for protein secretion. More information about plasmids, strains, and oligonucleotide primers is provided in Table I, Tables S1 and S2, and Figure 1.

YPD media was prepared as follows: 20 g/L D-glucose, 10 g/L yeast extract, 20 g/L peptone, and 1 g/L BSA.

Plasmid Construction

We inserted the KOZAK sequence (aacaaa) (Fujikawa et al., 1986) before the secretion leader to increase the translation efficiency in *S. cerevisiae* (Fig. 1A and Table S2); a Kex2 site (aaaga) (Achstetter and Wolf, 1985) and a spacer (gaagaaggtgaaccaaaa) (Kjeldsen et al., 1996) between the leader and the protein coding sequence were used to increase cleavage efficiencies of the pro-leaders in the late secretory pathway; and a mini-C-peptide (Kjeldsen et al., 2002) between the insulin A-chain and B-chain was used to increase the expression level of IP.

The alpha factor leader and the synthetic leader fused with the insulin cassette, carried by pUC57-NativeInsulin and pUC57-Yap3Insulin plasmid, respectively, were synthesized by GenScript, NJ 08854. The alpha factor leader fused with insulin cassette, the synthetic leader fused with insulin cassette and the synthetic leader fused with amylase cassette were amplified from plasmid pUC57-NativeInsulin, pUC57-Yap3Insulin, and pYapAmy (Tyo et al. submitted) using primers lzh040-lzh045, lzh042-lzh045, and lzh043-lzh044, respectively. The alpha factor leader was amplified from plasmid pUC57-NativeInsulin using primers lzh016-lzh040. The cDNA of α-amylase was amplified from plasmid pYapAmy using primers lzh018-lzh044. The alpha factor fused with the amylase cassette was constructed by fusion PCR of the alpha factor leader with the amplified amylase using primers lzh040-lzh044.

The plasmid POT was constructed by ligation of the *FseI/AscI* digested pSP-G2 (Partow et al., 2010) and the POT1 cassette, which was amplified from the genomic DNA of *S. pombe* (Alao et al., 2009) using primers lzh031-lzh032. Plasmid POTud was derived by ligating the *PstI/Ascl* digested POT vector and the Fl origin, which was amplified from plasmid pSP-G2 using primers lzh046-lzh047. The TPI promoter and TPI terminator were purified from genomic DNA of *S. cerevisiae* CEN.PK 113-7D by primers lzh027-lzh028 and lzh029-lzh030, respectively, and were then ligated together after digested with *NheI*. The CPO-Tud plasmid was derived from POTud by replacing the TEF1 promoter and CYC1 terminator with the TPI promoter and terminator using restriction cisterns of *FseI* and *MluI*. All the IP and amylase cassettes were cloned separately with the *KpnI/NheI* digested POTud and CPO-Tud, resulted in plasmids harboring alpha factor leader insulin (pAlphInsPOT or pAlphInsCPOT), synthetic leader insulin (pSynInsPOT or pSynInsCPOT), alpha factor leader amylase (pAlphAmyPOT or pAlphAmyCPOT), or synthetic leader amylase (pSynAmyPOT or pSynAmyCPOT), respectively.

Table I. Strains, plasmids, and oligonucleotides.

<table>
<thead>
<tr>
<th>Plasmids and strains</th>
<th>Relevant genotype</th>
<th>Leader</th>
<th>Promoter</th>
<th>Marker</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>pspGM2</td>
<td>TEF1-PGK1 bidirectional promoter (2 μm URA3)</td>
<td>—</td>
<td>—</td>
<td>URA3</td>
<td>Partow et al. (2010)</td>
</tr>
<tr>
<td>pUC57-NativeInsulin</td>
<td>Alpha factor leader insulin synthesized</td>
<td>Alpha factor</td>
<td>—</td>
<td>—</td>
<td>GenScript Co.</td>
</tr>
<tr>
<td>pUC57Yap3Insulin</td>
<td>Synthetic leader insulin synthesized</td>
<td>Yap3-TA57</td>
<td>—</td>
<td>—</td>
<td>GenScript Co.</td>
</tr>
<tr>
<td>CEN.PK 530-1C</td>
<td>MATα URA3HIS3 LEU2 TRP1 SUC2 MAL2-8* tpi1(41-707):loxP-KanMX4-loxP</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>SRD GmbH*</td>
</tr>
<tr>
<td><em>S. pombe</em> L972</td>
<td>h-</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Alao et al. (2009)</td>
</tr>
<tr>
<td>NC</td>
<td>CEN.PK 530-1C with CPO-Tud</td>
<td>—</td>
<td>TPI</td>
<td>POT1</td>
<td>This study</td>
</tr>
<tr>
<td>AIP</td>
<td>CEN.PK 530-1C with pAlphInsPOT</td>
<td>Alpha factor</td>
<td>TEF1</td>
<td>POT1</td>
<td>This study</td>
</tr>
<tr>
<td>SIP</td>
<td>CEN.PK 530-1C with pSynInsPOT</td>
<td>YAP3-TA57</td>
<td>TEF1</td>
<td>POT1</td>
<td>This study</td>
</tr>
<tr>
<td>AAP</td>
<td>CEN.PK 530-1C with pAlphaAmyPOT</td>
<td>Alpha factor</td>
<td>TEF1</td>
<td>POT1</td>
<td>This study</td>
</tr>
<tr>
<td>SAP</td>
<td>CEN.PK 530-1C with pSynAmyPOT</td>
<td>YAP3-TA57</td>
<td>TEF1</td>
<td>POT1</td>
<td>This study</td>
</tr>
<tr>
<td>AIC</td>
<td>CEN.PK 530-1C with pAlphaInsCPOT</td>
<td>Alpha factor</td>
<td>TPI</td>
<td>POT1</td>
<td>This study</td>
</tr>
<tr>
<td>SIC</td>
<td>CEN.PK 530-1C with pSynInsCPOT</td>
<td>YAP3-TA57</td>
<td>TPI</td>
<td>POT1</td>
<td>This study</td>
</tr>
<tr>
<td>AAC</td>
<td>CEN.PK 530-1C with pAlphaAmyCPOT</td>
<td>Alpha factor</td>
<td>TPI</td>
<td>POT1</td>
<td>This study</td>
</tr>
<tr>
<td>SAC</td>
<td>CEN.PK 530-1C with pSynAmyCPOT</td>
<td>YAP3-TA57</td>
<td>TPI</td>
<td>POT1</td>
<td>This study</td>
</tr>
</tbody>
</table>

*Scientific Research and Development GmbH, Oberursel, Germany.
CEN.PK530-1C was transformed separately with the POTud or CPOTud derived plasmids, and resulted in different engineered strains (Fig. 1 and Table I): strain AIP (with pAlphaInsPOT), SIP (with pSynInsPOT), AAP (with pAlphaAmyPOT), SAP (with pSynAmyPOT), AIC (with pAlphaInsCPOT), SIC (with pSynInsCPOT), AAC (with pAlphaAmyCPOT), and SAC (with pSynAmyCPOT). Blank plasmid CPOTud was also transformed to CEN.PK530-1C as the negative control (strain NC). For strains nomenclature see Table I.

Procedures for fermentation and analytics are described in supplementary text S1.

Results and Discussion

Construction of Recombinant *S. cerevisiae* Strains

Three expression systems were evaluated in this study (Fig. 1B): POTud, CPOTud, and P426GPD. POTud and CPOTud are vectors that use the *POT1* gene from *S. pombe* as marker to complement the *tpi1* mutation in the host. *TPI1* is a critical gene in both glycolysis and gluconeogenesis: A *tpi1*Δ strain do not grow on glucose as the sole carbon source (Compagno et al., 2001) and grow very slowly on other carbon sources (Kawasaki, 1999. US005871957A). The *tpi1*ΔA strain containing *POT1 plasmid* therefore allow stable expression in rich media (such as YPD) and also have a very high plasmid stability (Carlsen et al., 1997). In order to show the advantage of the *POT1* plasmid series, we also compared them to our previous studies (Tyo et al. submitted) in which we used the classic auxotrophy plasmid P426GPD, which is a 2µ plasmid carrying the *URA3* marker, the *GPD* promoter, and the *CYC1* terminator. Strain WI produced IP using the p426GPD plasmid and strain WA produced amylase using the p426GPD plasmid.

Overall Strain Characterization

Recombinant protein secretion leads to changes in the cellular metabolism and extracellular fluxes and cell growth parameters were therefore different among the strains (Table S3). The CPOTud strain series grew slightly slower in the glucose phase than the other strains, which suggested significant perturbations to the growth process, but still the final biomass concentration of the different strains were comparable.

In order to demonstrate the specific binding of the insulin antibody used for the Elisa measurement, AIC and SIC strains were cultivated in shake flasks and samples at three different time points (*T*<sub>i</sub>-inoculation, *T*<sub>d</sub>-during diauxic shift, and *T*<sub>f</sub>-final titers) were tested using Western blot. Figure 2 showed that SIC produced higher amount of insulin than AIC. Western blot also showed one additional band that corresponds to a 9 kDa (the IP band corresponds...
to 6 kDa) in the SIC strain (Fig. 2 WB #7). The protein associated with this band was not produced by the NC strain and it was also not present in the culture media (data not shown), and we assume that it is an insulin variant, possibly the un-efficiently cleaved pro-IP that by calculation should be 11.4 kDa (104 amino acids). This result is consistent with the HPLC measurement for another strain using the same leader (WI) (Tyo et al. submitted), and it may be due to the use of a synthetic leader.

**Leader Sequences Affect Recombinant Protein Secretion**

Two different leader sequences (alpha factor leader and synthetic leader) resulted in different effects on IP and amylase production both in glucose phase (Fig. 3) and in final production (Fig. 4). In all cases, the synthetic leader could direct more IP through the secretory pathway throughout the glucose and ethanol phases during the fermentation: (i) in POTud derived strains, SIP produced 55% more IP than AIP in the glucose phase and had a 110% higher final titer, (ii) in CPOTud derived strains, SIC could produce 9% more IP than AIC in the glucose phase and had a 19% higher final titer, and (iii) in p426GPD derived strains, WI produced 15% more IP than AIG (72 h shake flask, data not shown). The synthetic leader showed also an advantage for production of α-amylase but only in the strains secreting a moderate amount (around 15 mg/L in YPD medium) of α-amylase: (i) in the POTud derived strains, SAP produced 36% more amylase than AAP in the glucose phase and had a 110% higher final titer; and (ii) in p426GPD derived strains, strain WA produce 90% more amylase than AAG (strain with alpha factor leader fused with amylase in p426GPD plasmid, 72 h shake flask, data not shown). In the strains with higher production of amylase, the synthetic leader was less advantageous: in the CPOTud derived strains, the synthetic leader strain SAC could only produce 11% more amylase than the alpha factor leader strain AAC in the glucose phase, and additionally it also had a 58% of the final titer.

The effect of leader sequences on different proteins could be explained by the difference of N-glycosylation sites in the pro-leader sequence. Kjeldsen et al. reported that, under stressed conditions (such as treatment with DTT), the fusion of insulin and TA39 (pro-leader with two glycosylation sites) could be transported into late Golgi compartment, while fusion of insulin and TA57 (pro-leader with no glycosylation site) was still retained in the ER (Kjeldsen et al., 1999). They conclude that the lack of N-linked glycosylations of the leader sequence would cause more protein aggregation and precipitation under stressed conditions. In our experiments, amylase is a larger and more complex protein, which may...
cause the protein folding to become the rate-limiting step in the secretion process. When high amount of amylase is produced, the mis-folded proteins would cause cell stress, possibly in a similar way of low-level DTT induction in the Kjeldsen’s study (Kjeldsen et al., 1999). Under this condition, the alpha factor pro-leader which possesses three glycosylation sites provides more stringent guiding for correct fold and consequently, secretion. This may not be the case with folding of IP, which seems to cause only minor ER stress, probably due to its smaller size and simpler folding. In this case the synthetic leader showed its advantage, which is consistent with a previous study (Kjeldsen et al., 1999).

Expression Systems Affect Recombinant Protein Secretion

The CPO Tud strain series showed a notable advantage for production of both IP and α-amylase, compared with the POTud and p426GPD derived strains through different phases during the fermentation. The advantage was more prominent for the production of IP than for the production of α-amylase. IP producing strain with the synthetic leader and CPO Tud expression system, SIC, could produce 26.8-fold more IP than SIP (same construct but with POTud expression system) in the glucose phase and had a 32.5-fold higher final titer. Furthermore, SIC produced 26.6-fold more IP than WI (the synthetic leader fused IP produced with auxotrophy p426GPD system) in the glucose phase and had a 10.7-fold higher final titer. IP producing strain with the alpha factor leader and the CPO Tud expression system, AIC, had a 47.3-fold higher production of IP compared with AIP (same construct but with POTud expression system) in glucose phase and it had a 59.3-fold higher final titer. For the α-amylase producing strains, the results were a bit different. The CPO Tud strain series could still produce more amylase in the glucose phase (Fig. 3B): that is, the synthetic leader strain SAC could produce 3.81-fold more amylase than SAP and 4.79-fold more amylase than WA; and the alpha factor leader strains AAC could produce 6.29-fold more amylase than AAP. However, when it comes to final titers (Fig. 4B), the AAC could produce 2.67-fold more amylase than AAP, but the synthetic-leader-CPO Tud strain series did not possess notable advantages: that is, SAC produce 8% more amylase than SAP, but 3% less amylase than WA.

As an essential gene marker, POT1 is reported to yield a higher copy number than auxotrophic markers (Kawasaki et al., 1999. US005871957A). Different effects of expression systems on protein production and secretion could be due to specific characteristics of the expressed protein itself. The rate limiting step for IP secretion is probably not the folding of the protein (Kjeldsen et al., 1999) but rather the IP synthesis (transcription and translation) and thus can be circumvented by increasing transcription. This is probably the cause of higher production with the CPO Tud system than with p426GPD systems evaluated. For the structurally more demanding protein, such as the α-amylase, the bottleneck for secretion is likely to be post-translational processing, especially folding in the ER, and by increasing the expression with the CPO Tud system more translocated peptides to the ER cause more severe mis-folding stress and more futile cycles of protein generation and degradation which in turn cause increased cell stress, such as induction of ERAD or vacuolar-localized protein degradation (Tyo et al. submitted). As a result protein production is even lower for some conditions. Cases with similar opposite effects have been reported before: both secretion of human parathyroid hormone (hPTH, 84 amino acids, one disulfide bond, and zero glycosylation sites) (Gabrielsen et al., 1990) or granulocyte-colony stimulating factor (GCSF, 174 amino acids, two disulfide bonds, and zero glycosylation sites) (Wittrup et al., 1994) had increased production 17-fold by using a multi-copy plasmid compared to the a single copy plasmid; whereas for secretion of S. pombe acid phosphatase (PHO, 435 amino acids, eight disulfide bonds, and nine glycosylation sites), the use of a multi-copy plasmid resulted
in a 24% decrease in secretion when compared to a single copy plasmid (Robinson et al., 1994).

From these studies as well as our results, we suggest that the limitations are dependent on the molecular weight, and also the complexity of the protein (disulfide bonds, glycosylations, multi-domains, etc.). Since secretion of glycosylated proteins in S. cerevisiae is often reduced due to hyper-glycosylation and mis-folding inside the cell (Srivastava et al., 2001), the number of the glycosylation sites in the leader sequence is another very important factor to consider.

Despite large variations in protein secretion capacity in the strains evaluated here it is interesting to note that the only difference between the POTud and CPOTud plasmids is the promoter that drives the heterologous protein expression. It has been found that the TEF1 promoter is stronger than the TPI1 promoter using lacZ as the reporter gene, both in conditions of glucose excess (1.67-fold compared to TPI promoter) or limitation (5-fold compared to TPI promoter) (Partow et al., 2010). Interestingly, the final protein expression of either IP or amylase from the plasmid including the TEF1 promoter was lower. qPCR assays were therefore performed to compare relative gene expression levels in the yeast strains transformed with the expression systems including either the TPI1 promoter (AIC and AAC strains, respectively) or the TEF1 promoter (AIP and AAP strains), as described in Table I. The relative transcript levels corresponding to both the IP and amylase genes controlled by the TEF1 promoter were indeed higher than those controlled by the TPI1 promoter (Fig. S3). These results, which are consistent with previously reports (Partow et al., 2010) on the relative strength of these two promoters, suggest that the choice of promoter is not directly influencing the final protein titer in the POT1 derived strains. Thus, other events regarding post-transcriptional regulation might be involved and therefore affecting protein production. A follow-up experiment regarding global transcriptional analysis with amylase producing strains (AAP and AAC strains) was performed (data not shown). Using integrated analysis, we found that among the significant reporter GO-terms (FDR < 0.005), many pathways related to the overall transcription and translation were down-regulated in the AAC strain, whereas GO-terms associated with ER protein processing, vacuole degradation, stress response and unfolded protein response were up-regulated. Within the top 10 significant reporter TFs (FDR < 0.005), genes related to all kinds of stress (MSN2 and MSN4 for general stress, HOG1 for osmotic stress and YAP1 for oxidative stress) and heat shock factor which could release ER stress (HSF1) were up-regulated in the AAC strain. All these data suggest that in the amylase producing strains, the high amount of recombinant proteins or peptides are blocking the secretory pathway (possibly inside of the ER) which causes cell stress including the unfolded protein response. The result of this is down-regulation of the general transcription and translation machinery and up-regulation of the ER processing and protein turnover pathways.

In addition to their final titers, the IP and α-amylase also differ in their processing characteristics in the secretory pathway. By plotting the protein production data against dry cell weight to eliminate the effect of the changing cell concentration, it is found that there is a clear shift in the secretion behavior during the diauxic shift (Fig. 5). Interestingly, all α-amylase producing strains produced amylase at a higher rate during growth in ethanol phase, whereas all IP producing strains produced IP at a higher rate in the glucose phase. The shifting patterns of protein productions further supported the fact that the rate-controlling step for protein secretion is different between the two proteins. As mentioned above, production of the IP is probably mainly limited by expression and for all the used expression promoters (pTPI1, pTEF1, and pGPD1) there is higher expression for high growth/high glycolytic fluxes. For amylase, which is a larger protein with more diverse modifications, the limitation is likely to be protein processing and folding. We hypothesize that the respiratory conditions prevailing during growth on ethanol may have a beneficial effect on the folding process (compared with the fermentative conditions prevailing in the glucose growth phase). The conversion of ethanol into acetaldehyde requires NAD(P) as cofactors (Visser et al., 2004), and the hence elevated amount of NAD(P)H could serve as the reducing power either for reduction of ROS generated by the folding stress (Työ et al. submitted) or by converting oxidized glutathione (GSSG) into reduced glutathione (GSH). GSH plays an very important role during the refolding of mis-folded proteins (Tu et al., 2000), and the shortage of GSH could lead to hyper-oxidizing conditions in the ER (Van de Laar et al., 2007), and

Comparison of Insulin Precursor and α-Amylase Secretion

Secretion profiles of IP and α-amylase producing strains were also examined (Fig. 4). Based on this it was found that the trend for production of IP and α-amylase in terms of mg/L is not conserved in the different constructs, whereas in terms of μmol/L the production of IP is always higher. In order to explain the high transcriptional level and the relatively low protein production of amylase producing strain, a follow-up experiment regarding global transcriptional analysis of the amylase producing strain (AAC and NC strains) was performed (data not shown). Using integrated analysis, we found that among the significant Reporter GO-terms (FDR < 0.001), many pathways related to the overall transcription and translation were down-regulated in the AAC strain, whereas GO-terms associated with ER protein processing, vacuole degradation, stress response and unfolded protein response were up-regulated. Within the top 10 significant reporter TFs (FDR < 0.005), genes related to all kinds of stress (MSN2 and MSN4 for general stress, HOG1 for osmotic stress and YAP1 for oxidative stress) and heat shock factor which could release ER stress (HSF1) were up-regulated in the AAC strain. All these data suggest that in the amylase producing strains, the high amount of recombinant proteins or peptides are blocking the secretory pathway (possibly inside of the ER) which causes cell stress including the unfolded protein response. The result of this is down-regulation of the general transcription and translation machinery and up-regulation of the ER processing and protein turnover pathways.

In addition to their final titers, the IP and α-amylase also differ in their processing characteristics in the secretory pathway. By plotting the protein production data against dry cell weight to eliminate the effect of the changing cell concentration, it is found that there is a clear shift in the secretion behavior during the diauxic shift (Fig. 5). Interestingly, all α-amylase producing strains produced amylase at a higher rate during growth in ethanol phase, whereas all IP producing strains produced IP at a higher rate in the glucose phase. The shifting patterns of protein productions further supported the fact that the rate-controlling step for protein secretion is different between the two proteins. As mentioned above, production of the IP is probably mainly limited by expression and for all the used expression promoters (pTPI1, pTEF1, and pGPD1) there is higher expression for high growth/high glycolytic fluxes. For amylase, which is a larger protein with more diverse modifications, the limitation is likely to be protein processing and folding. We hypothesize that the respiratory conditions prevailing during growth on ethanol may have a beneficial effect on the folding process (compared with the fermentative conditions prevailing in the glucose growth phase). The conversion of ethanol into acetaldehyde requires NAD(P) as cofactors (Visser et al., 2004), and the hence elevated amount of NAD(P)H could serve as the reducing power either for reduction of ROS generated by the folding stress (Työ et al. submitted) or by converting oxidized glutathione (GSSG) into reduced glutathione (GSH). GSH plays an very important role during the refolding of mis-folded proteins (Tu et al., 2000), and the shortage of GSH could lead to hyper-oxidizing conditions in the ER (Van de Laar et al., 2007), and

Comparison of Insulin Precursor and α-Amylase Secretion

Secretion profiles of IP and α-amylase producing strains were also examined (Fig. 4). Based on this it was found that the trend for production of IP and α-amylase in terms of mg/L is not conserved in the different constructs, whereas in terms of μmol/L the production of IP is always higher. In order to explain the high transcriptional level and the relatively low protein production of amylase producing strain, a follow-up experiment regarding global transcriptional analysis of the amylase producing strain (AAC and NC strains) was performed (data not shown). Using integrated analysis, we found that among the significant Reporter GO-terms (FDR < 0.001), many pathways related to the overall transcription and translation were down-regulated in the AAC strain, whereas GO-terms associated with ER protein processing, vacuole degradation, stress response and unfolded protein response were up-regulated. Within the top 10 significant reporter TFs (FDR < 0.005), genes related to all kinds of stress (MSN2 and MSN4 for general stress, HOG1 for osmotic stress and YAP1 for oxidative stress) and heat shock factor which could release ER stress (HSF1) were up-regulated in the AAC strain. All these data suggest that in the amylase producing strains, the high amount of recombinant proteins or peptides are blocking the secretory pathway (possibly inside of the ER) which causes cell stress including the unfolded protein response. The result of this is down-regulation of the general transcription and translation machinery and up-regulation of the ER processing and protein turnover pathways.

In addition to their final titers, the IP and α-amylase also differ in their processing characteristics in the secretory pathway. By plotting the protein production data against dry cell weight to eliminate the effect of the changing cell concentration, it is found that there is a clear shift in the secretion behavior during the diauxic shift (Fig. 5). Interestingly, all α-amylase producing strains produced amylase at a higher rate during growth in ethanol phase, whereas all IP producing strains produced IP at a higher rate in the glucose phase. The shifting patterns of protein productions further supported the fact that the rate-controlling step for protein secretion is different between the two proteins. As mentioned above, production of the IP is probably mainly limited by expression and for all the used expression promoters (pTPI1, pTEF1, and pGPD1) there is higher expression for high growth/high glycolytic fluxes. For amylase, which is a larger protein with more diverse modifications, the limitation is likely to be protein processing and folding. We hypothesize that the respiratory conditions prevailing during growth on ethanol may have a beneficial effect on the folding process (compared with the fermentative conditions prevailing in the glucose growth phase). The conversion of ethanol into acetaldehyde requires NAD(P) as cofactors (Visser et al., 2004), and the hence elevated amount of NAD(P)H could serve as the reducing power either for reduction of ROS generated by the folding stress (Työ et al. submitted) or by converting oxidized glutathione (GSSG) into reduced glutathione (GSH). GSH plays an very important role during the refolding of mis-folded proteins (Tu et al., 2000), and the shortage of GSH could lead to hyper-oxidizing conditions in the ER (Van de Laar et al., 2007), and
produce more ROS through futile cycling of the folding process (Nguyen et al., 2011). There may also be a favorable heat shock-like effect induced by ethanol (Alexandre et al., 2001; Piper, 1995).

**Conclusion**

Here, we provide a novel set of expression vectors for recombinant protein production in yeast, and we used these to evaluate the most important expression factors regarding recombinant secretion: protein type, leader sequence, expression system, and promoter. We report that although the transcription level of the recombinant gene is important, the final production of the recombinant protein is the result of a combination of effects of transcription and translation levels, protein uniqueness, and the leader sequences which influences the secretory pathway processing efficiency. We also report a notable difference in production of IP and α-amylase, and we conclude that this difference is caused by differences in their processing through the secretory pathway. For IP the important step is the synthesis of the protein, and this is supported by (i) dramatic IP production changes between the CPOTud and p426GPD systems and (ii) more and faster IP production during growth on glucose. For amylase the rate-controlling step for secretion was found to be most likely ER folding and processing as supported by (i) lower secretion of the α-amylase with a synthetic leader compared to the glycosylated alpha factor leader in the high production strains, (ii) much more amylase produced in AAC compared with AAP, whereas moderate changes of final protein productions, and (iii) the dramatically increased production during growth on ethanol. Our study provides a novel insight into the protein secretion engineering in yeast, and a set of novel expression systems that can be used for high-level expression of recombinant proteins in connection with the use of yeast for consolidated bioprocesses.

We thank Dr. Per Sunnerhagen from Gothenburg University for kindly providing the _S. pombe_ strain and Dr. Peter Kötter from University of Frankfurt for kindly providing the CEN.PK 113-5D and CEN.PK 530-1C strains. This work is financially supported by the EU Framework VII project SYSINBIO (grant no. 212766), European Research Council ERC project INSYSBIO (grant no. 247013), the Chalmers Foundation, the Knut and Alice Wallenberg Foundation, and NIH F32 Kirschstein NRSA fellowship.

**References**


Yu et al.: Factors Affecting Protein Secretion

Biotechnology and Bioengineering
PAPER II
Imbalance of heterologous protein folding and disulfide bond formation rates yields runaway oxidative stress

Keith EJ Tyo1,2, Zihe Liu1, Dina Petranovic1 and Jens Nielsen1*

Abstract

Background: The protein secretory pathway must process a wide assortment of native proteins for eukaryotic cells to function. As well, recombinant protein secretion is used extensively to produce many biologics and industrial enzymes. Therefore, secretory pathway dysfunction can be highly detrimental to the cell and can drastically inhibit product titers in biochemical production. Because the secretory pathway is a highly-integrated, multi-organelle system, dysfunction can happen at many levels and dissecting the root cause can be challenging. In this study, we apply a systems biology approach to analyze secretory pathway dysfunctions resulting from heterologous production of a small protein (insulin precursor) or a larger protein (α-amylase).

Results: HAC1-dependent and independent dysfunctions and cellular responses were apparent across multiple datasets. In particular, processes involving (a) degradation of protein/recycling amino acids, (b) overall transcription/translation repression, and (c) oxidative stress were broadly associated with secretory stress.

Conclusions: Apparent runaway oxidative stress due to radical production observed here and elsewhere can be explained by a futile cycle of disulfide formation and breaking that consumes reduced glutathione and produces reactive oxygen species. The futile cycle is dominating when protein folding rates are low relative to disulfide bond formation rates. While not strictly conclusive with the present data, this insight does provide a molecular interpretation to an, until now, largely empirical understanding of optimizing heterologous protein secretion. This molecular insight has direct implications on engineering a broad range of recombinant proteins for secretion and provides potential hypotheses for the root causes of several secretory-associated diseases.

Keywords: Protein secretion, unfolded protein response, HAC1, protein production, oxidative stress

Background

The protein secretory pathway is an extensive process in eukaryal cells, as it is responsible for processing approximately one-third of all proteins. Substantial cellular resources are therefore utilized to maintain this pathway’s functions, and stressed conditions in the secretory pathway have consequences for the whole cell [1]. Distress in secretory pathway organelles has been implicated as the molecular basis for several diseases, for example, β cell apoptosis in diabetes, cystic fibrosis, and prion-related disease, among others [2]. In biotechnology, efficient secretion of useful recombinant proteins in yeast and fungi is a key industrial objective with applications in enzyme production required for the production of biofuels, detergents, fabrics, food, and biologics, such as immunoglobulins, hormones, and vaccines. Significant effort has gone into engineering yeast for increasing protein secretion [3]. Strategies, such as changing environmental parameters (for example, temperature, media composition) [4] or altering genetics, can increase secretion for some proteins, but they rarely represent generic solutions for improving protein secretion [5,6]. The lack of a single engineering strategy that improves protein secretion across the board implies that there are several possible bottlenecks in the secretory pathway, and different proteins may be constrained in different ways. There is therefore a
requirement for more fundamental insight into this complex pathway that involves a very large number of components.

In yeast, the secretory pathway is a multi-organelle system that is responsible for trafficking proteins to the extracellular space, cell membrane, or vacuole [7]. During this transit, multiple processes must be coordinated, including folding, specific proteolytic cleavage, glycosylation, and disulfide bond formation, all with a layer of quality control at key check points. The pathway requires substantial cellular resources to perform these tasks, such as glycan, electron acceptors, electron donors, and ATP. In the ER, the nascent peptide is folded into its native structure while disulfide bonds are formed. The rate of protein folding is dependent upon the complexity of the protein to be folded, the availability of chaperones to assist folding, and ATP used by the chaperones [1]. Proteins that are slow to fold or terminally misfolded proteins are removed from the ER via the ER-associated degradation (ERAD) pathway [8]. Disulfide bond formation requires the removal of electrons from cysteine thiols via protein disulfide isomerase (PDI) and Ero1p to the final electron acceptor, typically oxygen [9,10]. This process produces reactive oxygen species (ROS) in stoichiometric amounts to the number of disulfide bonds formed [11]. Disulfide bond formation is random, and incorrect bond pairs must be exchanged for native bonds via PDI-based processes [12]. In addition, reduced glutathione (GSH) acts as a buffer for the redox state of the ER [13]. A more detailed description of oxidative protein folding can be found in the reviews by Sevier et al. and Chakravarthi et al. [14,15].

The secretory pathway must adjust the chaperone capacity, oxidizing equivalents, ATP, glycan, and other metabolic requirements, as well as trafficking patterns, based on the portfolio of proteins that need to be expressed at a given time, and the resources required to process that set of proteins. In yeast, the unfolded protein response (UPR) is one transcriptional mechanism that adjusts secretory resources and controls to handle overload of the folding machinery in the ER [16]. In the UPR, accumulation of unfolded proteins in the ER signals a pathway that results in translation of Hac1p, a transcription factor (TF) known to activate or repress over 100 genes, including many ER-associated proteins such as Kar2p, Pdi1p, and Ero1p [17].

In this study, we identified biological mechanisms which alter the secretory pathway in response to secretion of recombinant proteins with different properties (size, number of disulfide bonds, and glycans) in a Hac1p-dependent and independent manner. The secretory pathway was perturbed by secreting a small protein, human insulin precursor (IP), or a comparatively larger protein, α-amylase, in wild-type (WT) and Δhac1 Saccharomyces cerevisiae. These proteins were chosen because the two proteins elicit different behavior in the secretory pathway. These differences will arise because α-amylase is a relatively larger (and likely more difficult to fold), has an odd number of cysteines (which may complicate disulfide isomerization) and has glycosylation, compared to insulin which is small, has even number of cysteines, and is not glycosylated. As well, α-amylase has one more disulfide bond than IP. To identify biological mechanisms, we characterized changes in physiological properties (specific growth rate, carbon utilization efficiency, and recombinant protein secretion), TF activity (as inferred from transcriptome analysis) and metabolic demand (as inferred by changes in metabolic flux diversion). Through this, we identified the following biological processes: amino acids recycling from degraded proteins, trans-Golgi network (TGN) sorting changes, overall expression repression, and oxidative stress. Motivated by secretory-related oxidative stress observations, we present a model for disulfide bond formation and electron transfer in the ER which takes into account thermodynamic irreversibilities caused by differences in electron affinity. The proposed model explains the non-stoichiometric ROS formation that we observed that results from disulfide bond formation and causes oxidative stress under folding-stress conditions. If proven by genetic and biochemical results, the futile cycle model yields insight into a fundamental problem in secretory stress and reveals new avenues to reduce oxidative stress and increase productivity in industrial protein production.

Results
Protein size and Hac1p activity affect protein secretion quantity and cell growth
Yeast strains were constructed that produce and secrete (a) IP or (b) α-amylase and were compared to yeast strains containing (c) an empty vector in both wild-type and hac1 deletion backgrounds. IP and α-amylase were chosen because they are very different types of proteins to secrete. IP is 51 amino acids in length, with six cysteines forming three disulfide bonds, and no glycosylation. α-amylase is 478 amino acid in length, with nine cysteines forming only four disulfide bonds and one glycosylation. The odd number of cysteines in α-amylase complicates disulfide pairing, as the random isomerization process may incorporate the cysteine that should not be incorporated into a disulfide bond. Both proteins were targeted for secretion using a YAP3 pre sequence (21 amino acids, cleaved off in the ER) and a rationally designed pro sequence (TA57, 42 amino acids, no glycosylation or disulfides) were cloned behind a TDH3 promoter in a high copy 2 micron plasmid [18]. α-amylase was expressed using the same plasmid, promoter, and leader sequences. These strains are named WN (WT...
with empty vector), WI (WT secreting IP), WA (WT secreting α-amylase), dN (Δhac1 with empty vector), dI (Δhac1 secreting IP), and dA (Δhac1 secreting α-amylase). Strains were characterized in batch fermentation to understand the effects on cell physiology.

The cellular burden induced by (a) synthesizing and secreting IP and α-amylase and (b) deleting the key TF for the UPR, Hac1p, substantially affected the cells. Protein titers in WT strain were 9 mg/L and 20 mg/L, for IP and α-amylase, respectively (Figure 1a). On a per biomass basis, this is approximately half the insulin produced, and one-third the α-amylase reported for rich media [19,20]. Rich media appears to be favorable for heterologous protein production, but may present complications in downstream separations. Comparing the small and larger proteins, α-amylase was secreted in higher levels on a mass basis, but six-fold more insulin molecules were secreted (1.52 μM IP in WI compared to 0.26 μM α-amylase in WA). Δhac1 strains secreted significantly less protein than WT, confirming that Hac1p is important for efficient secretion (Figure 1a) [5].

Reduced specific growth rates imply impairment of cellular processes (Figure 1b). In WT yeast, IP production did not affect growth; however, α-amylase production reduced growth by 25%. This, combined with the differences in protein titers, implies that α-amylase is more challenging to fold and secrete than IP. In the Δhac1 background, recombinant protein strains dI and dA had approximately 20% lower growth rates compared to dN. This growth reduction occurs despite no change in specific glucose uptake rate (Additional file 1, Tables S1 and S2) pointing toward higher energy requirements to maintain homeostasis in Δhac1 while trying to secrete recombinant proteins. Δhac1 strains had overall lower final cell densities. Δhac1 strains produced more glycerol than WT strains implying impaired oxidative processes in the Δhac1 strains (Additional file 2).

Secretory stress shifts metabolism to increase oxygen and ATP requirements

The physiological changes due to the secretory perturbations affect the distribution of resources through the metabolic network. The glucose uptake and range of products produced were altered by the protein production conditions (Table 1). Changes in the underlying metabolic network were estimated by flux balance analysis (FBA) using a yeast central carbon metabolism model, constrained by measured extracellular fluxes (Additional file 1, Tables S1 and S2, Additional files 3 and 4) [21]. Figure 2a shows a metabolic map of central carbon metabolism for each of the six conditions based on the exchange fluxes in Table 1 and the FBA analysis. The shift in metabolic fluxes were correlated with changes in redox requirements. As expected, the catabolic functions of the TCA cycle was predicted to have very low activity due to glucose repression [22]. Figure 2b shows that the oxygen uptake was twice as high in the strains that were growth inhibited (for example, WA, dI, dA) than those that were not. This increased oxygen uptake was not used for oxidative phosphorylation, as the biomass yields on glucose were lower in WA, dI, and dA, and it may therefore be a result of increased oxidation in connection with formation of disulfide bonds.

Figure 2c shows that the maintenance ATP consumption is increased in WA, dI, and dA according to FBA calculations. In WT background, WI did not consume a detectable increase in ATP, likely because IP is short...
and easily folded, thereby minimally taxing the translation and folding machinery. WA did increase two-fold in ATP consumption, most likely because α-amylase is 10-fold larger and likely more difficult to fold and has more disulfide bond pairing possibilities. In the Δhac1 background, folding efficiency is likely decreased due to ER dysfunction. With native secretion, dN did not require higher ATP maintenance consumption compared to WT. However, even the smaller, easier to fold IP resulted in ER stress that required significant ATP consumption compared to WT. dA, which was already stressed under WT, continued to show high ATP consumption. Despite the increased ATP consumption in dI and dA, little protein was secreted.

Transcription factors controlling oxidative stress, amino acid salvaging, and expression repression are linked to secretory response

Growth phase transcriptomics measurements were carried out to identify cellular processes that were activated under the stresses of HAC1 deletion and recombinant protein production. HAC1 deletion resulted in 339 significantly changed genes in the no recombinant protein case (WN vs. dN). HAC1 deletions in the insulin strain and α-amylase strain resulted in much larger cellular responses of 1628 (WI vs. dI) and 1511 (WA vs. dA) significantly expressed genes, respectively. KAR2 (ER chaperone) expression was significantly reduced upon HAC1 deletion (↓ three-fold dN vs WTN, \( P = 1 \times 10^{-3} \)) and the four yeast protein disulfide isomerases (PDII, EUGI, MPDI, MPD2) reduced an average of 2.9-fold (\( P < 0.05 \)).

The effects of producing IP or α-amylase within a strain background (WT or HAC1) were not as pronounced as the effect of HAC1 deletion, 40 and 194 genes were significantly changed in WI (compared to WN) and WA (compared to WN). Likewise, 74 and 90 genes were significantly changed for dI (compared to dN) and dA (compared to dN).

To reduce the dimensionality of the data and identify putative TFs involved in protein secretion, the Reporter Transcription Factor algorithm was used [23]. TFs were scored by the modulation in expression level of genes that the TFs bind in the upstream region according to ChIP-chip data [24]. Therefore, the score is not indicative of change in the TF expression level itself, but of the genes under its influence. Reporter TF algorithm is useful, because although the statistical significance of an individual gene may not meet an arbitrary threshold, if several genes linked to the same TF have similar behavior, the likelihood of observing the group of genes is low, making TF identification very sensitive. Figure 3 shows significant secretory process TFs shown to be involved in up- and down-regulating different cellular process under their control. Interestingly, different TFs were identified for the two different proteins. This is likely the combined effect of different protein size and number of disulfide bonds. A complete list of significant transcription factors is provided in Additional files 5 and 6.

In WT (Figure 3a), several TFs were activated by protein secretion. Oxidative and osmotic stress pathway up-regulation was common to both proteins. Oxidative stress is likely caused by ROS that is formed when Ero1p shuttles electrons to oxygen in disulfide bond formation [25]. Osmotic stress response, particular hypo-osmotic stress, strengthens the cell wall to counteract internal turgor pressure by changing the cell wall composition. This change in composition requires remodeling the secretory pathway by changing which components are trafficked to the cell wall [26]. Surprisingly, the Reporter TF algorithm found several Hac1p-influenced genes down-regulated. Genes that Hac1p binds from the ChIP-chip data that are significantly down-regulated are KEG1, MCD4, and ERJ5. KEG1 and MCD4 genes are involved in glycan modifications and ERJ5 is a secondary ER chaperone [27-29]. These genes may be influenced by other TFs not included in the ChIP-chip network. Genes known to be regulated by Hac1p (KAR2 and ERO1) were not significantly changed upon secreting recombinant protein, indicating that there is not an actual Hac1p response in the WT.

Clear differences between large and small protein secretion emerge in WT. IP stimulated modification of the TGN through MCM1 and STE12. Overall expression

### Table 1 Physiological parameters of recombinant protein secretion strains

<table>
<thead>
<tr>
<th>Strains</th>
<th>M_{max} ([\text{h}^{-1}])</th>
<th>Y_{SX}</th>
<th>Y_{SE}</th>
<th>Y_{SG}</th>
<th>Y_{SA}</th>
<th>Y_{SCO2}</th>
<th>Carbon balance</th>
</tr>
</thead>
<tbody>
<tr>
<td>WN</td>
<td>0.53 +/- 0.0014</td>
<td>0.14 +/- 0.001</td>
<td>0.32 +/- 0.041</td>
<td>0.067 +/- 0.009</td>
<td>0.048 +/- 0.0005</td>
<td>0.30 +/- 0.019</td>
<td>0.89</td>
</tr>
<tr>
<td>WI</td>
<td>0.40 +/- 0.0012</td>
<td>0.13 +/- 0.002</td>
<td>0.35 +/- 0.029</td>
<td>0.055 +/- 0.005</td>
<td>0.056 +/- 0.0047</td>
<td>0.30 +/- 0.014</td>
<td>0.92</td>
</tr>
<tr>
<td>WA</td>
<td>0.32 +/- 0.007</td>
<td>0.11 +/- 0.003</td>
<td>0.31 +/- 0.006</td>
<td>0.060 +/- 0.007</td>
<td>0.049 +/- 0.0023</td>
<td>0.30 +/- 0.002</td>
<td>0.84</td>
</tr>
<tr>
<td>dN</td>
<td>0.38 +/- 0.005</td>
<td>0.13 +/- 0.0004</td>
<td>0.37 +/- 0.025</td>
<td>0.046 +/- 0.003</td>
<td>0.035 +/- 0.0046</td>
<td>0.29 +/- 0.020</td>
<td>0.91</td>
</tr>
<tr>
<td>dI</td>
<td>0.29 +/- 0.005</td>
<td>0.08 +/- 0.006</td>
<td>0.32 +/- 0.017</td>
<td>0.081 +/- 0.001</td>
<td>0.046 +/- 0.0011</td>
<td>0.31 +/- 0.007</td>
<td>0.84</td>
</tr>
<tr>
<td>dA</td>
<td>0.31 +/- 0.002</td>
<td>0.11 +/- 0.003</td>
<td>0.32 +/- 0.002</td>
<td>0.066 +/- 0.001</td>
<td>0.049 +/- 0.0009</td>
<td>0.30 +/- 0.004</td>
<td>0.85</td>
</tr>
</tbody>
</table>

*aAll yields (Y) are \([\text{g/g}]\). Glucose (S), biomass (X), ethanol (E), glycerol (G), acetate (A), carbon dioxide (CO2).*

*bStrain abbreviations as in Figure 1.*
is reduced by altering mRNA degradation pathways via STO1. \(\alpha\)-amylase had a much larger effect on the cell, as compared to IP, as was implied by physiological parameters of Figure 1 and number of altered genes. Additional oxidative and osmotic stress pathways were activated in WA, as well as a down-regulation in some amino acid synthesis pathways and overall reduction of transcription.
In the Δhac1 background (Figure 3b), many of the effects found in WA, have become common to both IP and α-amylase producing strains. HAC1 deletion clearly makes the cell more susceptible to recombinant secretion overload. Both insulin and α-amylase secretion cause considerable oxidative stress response and down-regulation of amino acid synthesis, including the general amino acid synthesis TF, Gcn4p. In ΔI, translational capacity repression is also employed (via Fhlp/Rap1p) and adjustments in amino acid metabolism. ΔA shows a mix of up- and down-regulation of genes that are controlled by Hac1p. Other TFs appear to be controlling these genes in the absence of HAC1. Some oxidative and osmotic stress pathways appear independent of HAC1. Skn7p and Cin5p were similarly activated in both WT and Δhac1. Oxidative and hypo-osmotic stress, while important for managing the secretory pathway, appears not to be directly managed through the UPR.

Thermodynamic irreversibilities in redox reactions can explain increased oxidative stress in slow protein folding conditions

The increases in oxidative stress, oxygen consumption, and reduced growth observed in the study can be explained by electron transfer in ER redox pathways. Disulfide bond formation has been established to consume oxygen and produce ROS (and thereby consume cellular resources to protect against the ROS) in stoichiometric quantities with the number of disulfide bonds formed [9]. When non-native disulfide linkages are formed, these linkages must be rearranged to the correct disulfide pairings for the native protein to be folded, a process called disulfide isomerization [30].

Disulfide isomerization involves (a) breaking the non-native bond by transferring electrons to the non-native bond creating a cysteine linkage with the PDI, and (b) creating a new disulfide linkage in the nascent protein by transferring the electrons to break the PDI-nascent protein linkage. By random pairing, the native disulfide bonds are found.

Directionality in these redox reactions is determined by thermodynamic favorability through electron affinity of the potential disulfide bonds. Disulfide isomerization is redox neutral, not requiring electron donors or acceptors. However, it does require each disulfide pairing to have a lower electron affinity than the next (non-native disulfide in folding protein < PDI-folding protein disulfide < native disulfide in folding protein) to allow the electrons to transfer. Under slow folding conditions, PDI may hold the disulfide bond (oxidized state) for extended time because a native disulfide cannot be found, resulting in PDI being reduced by other moieties, likely GSH.

Given the observations in our experiments, and the thermodynamic reasoning immediately above, we propose a simple thermodynamic model of disulfide bond formation and breaking that explains increased oxidative stress, oxygen consumption, and reduced growth observed in our experiments. This model expands upon the mechanism by Cuozzo and Kaiser [13]. The thermodynamic model assumes there are PDI disulfide bonds that have electron affinities above and below the nascent
Figure 4 Proposed thermodynamic model predicts non-stoichiometric reactive oxygen species produced with incorrect disulfide bond formation. (a) In the model, forming and breaking an incorrect disulfide bond uses two protein disulfide isomerases (PDIs), one with electron affinity higher (PDI_A) and one lower (PDI_B) than the incorrect disulfide bond. In the formation phase, electrons are shuttled to molecular oxygen, resulting in ROS formation. In the breaking phase, electrons are passed from NADPH, through glutathione, to the protein. In both cases, electrons move along the electron affinity gradient. The net result is a futile cycle that is required to fix incorrect disulfide bonds, but expends redox energy. (b) The thermodynamic model predicts at fast folding rates near stoichiometric ROS is generated per disulfide bond formed. However, when folding rates are slow, the unfolded protein may go through many futile cycles, resulting in excess ROS. Glutathione (GSH), oxidized glutathione (GSSG), disulfide bond formation (DBF), disulfide bond breaking (DBB).
proteins disulfide bonds (Figure 4a). The disulfide is formed by the typical oxidation pathway (Figure 4a, green) catalyzed by high electron affinity PDI (called PDIA here). Instead of isomerization, the incorrect disulfide is reduced by an electron donor with a low electron affinity (most likely a different PDI parologue, called PDIB here) (Figure 4a, blue). The difference in electron affinity between the folding protein's cysteines and a specific PDI's cysteines can only allow the electrons to flow in one direction (toward the higher electron affinity cysteines) (Figure 4a). Therefore, a different PDI is required to form and break the incorrect disulfide bond. This futile cycle relies on a strong electron affinity gradient to complete an isomerization-like process. The net result of the futile cycle is GSH consumption and ROS production. This model implies that the ROS produced is not stoichiometrically linked to the number of disulfide bonds formed, but varies by the number of futile cycles before the correct bond is formed.

The metabolic and transcriptional data supports this model. Upon HAC1 deletion, ER chaperones (KAR2) and PDIs (PD11, FUG1, MPD1, and MPD2) expression is reduced. This downregulation of ER chaperones and PDIs results in suppressed ER folding and disulfide bond formation in the Δhac1 mutants. In the dN case, minimal oxidation stress is seen. However, when there is an increased demand for protein folding and disulfide bond formation, as is the case for dI and dA case, we see high oxygen consumption, ATP requirements, and many oxidative stress pathways being activated transcriptionally. Although both folding and disulfide bond formation is down, an imbalance toward faster disulfide bond formation compared to folding will result in futile cycles. Therefore, this disulfide/folding imbalance acts as a catalyst for drastically increasing ROS production.

Based on this thermodynamic model, the relative rates of protein folding and disulfide bond formation for nascent peptides have important consequences for oxidative stress (Figure 4b). When folding is faster than disulfide bond formation, ROS is produced in near one-to-one amounts with the disulfide bonds formed. Under these conditions, isomerization may be more efficient to resort incorrect disulfide bonds, as native structures with low electron affinity disulfide pairs are favored, and isomerization does not produce ROS. However, when folding is slow compared to disulfide bond formation, as is the case when the protein folding machinery gets overloaded, the nascent peptides cycles through the futile redox cycle producing ROS and consuming GSH in excess to the final number of disulfide bonds formed. The physiological result of a high disulfide bond formation to ER folding rate is oxidative damage to a broad range of cellular proteins and consumption of reducing equivalents that could otherwise be used for anabolism.

**Discussion**

In this study, we have identified biological mechanisms related to protein synthesis and secretion by introducing perturbations to the cell, in the form of HAC1 deletion and different recombinant protein expression, and measuring the system level cellular responses, via transcriptomics and metabolic fluxes. These measurements, combined with data analysis algorithms, Reporter TF algorithm and FBA, were able to identify cellular adjustments in (a) overall expression level, (b) post-Golgi sorting, (c) amino acid biosynthesis and savaging, and (d) oxidative stress. These biological effects are a result of the combined influence of protein synthesis and trafficking through the secretory pathway.

Overall transcription and translation were repressed in response to α-amylase expression (a larger protein) and in the Δhac1 strains with any recombinant protein secretion. Repressing overall expression is a broad spectrum response used to adjust the rates of all other cellular processes to match the reduced folding capacity in the ER. Several mechanisms were used to alter overall expression: repressing mRNA synthesis, increasing mRNA degradation rates, and repressing protein translation rates through reducing ribosome numbers. Specifically, mRNA concentrations are lowered by decreasing RNA polymerase accessibility (HIR2), inhibiting transcriptional elongation (THO2), and controlling RNA degradation (STO1) [31,32]. Ribosome concentration, and thereby translation rates, can be reduced by the TFs Fhl1p and Rap1p which control expression of rRNA and ribosomal proteins [33]. This is seen in IP production in Δhac1 strain, both by the reporter TFs (Figure 3) and by expression of ribosomal proteins (Additional file 7).

In this context, extrachromosomal plasmids offer advantages over chromosomal expression. HIR2, whose mechanism is to silence the chromosome, would not affect extrachromosomal plasmids. Increased recombinant protein secretion would be accomplished by silencing native ER genes, while recombinant, plasmid-born gene would not be affected.

Pronounced adjustments to the TGN were observed in the transcriptome in all conditions. TFs involved in pheromone responses (STE12, MCM1, ASH1), invasive/pseudohyphal growth (STE12, MSN1, PHD1, RIM101), and osmotic stress (CIN5, SKN7, SKO1, YAP6, MSN1) were all identified by the Reporter TF algorithm and point to an underlying set of activities that are required to increase the traffic of secretory vesicles to the membrane. Invasive, pseudohyphal, and filamentous growth morphologies have a high surface to volume ratio and inherently require higher Golgi-to-cell membrane trafficking rates to supply cell membrane and cell wall components for growth. These altered morphologies can be activated through the filamentous and invasive response elements (FREs) [34]
bound by STE12 and used to regulate PHD1 [35]. HAC1 deletion has been shown to cause filamentous growth [36].

Osmotic stress TFs are also responsible for affecting protein secretion, as the external cell wall must be strengthened in response to hypo-osmotic conditions, thereby requiring an efficient secretory pathway to ferry cell wall proteins [26]. MSN1 is known to induce starch degradation, requiring the actions necessary to secrete the appropriate enzymes through filamentous growth activation [37]. SKN7 has a dual role in invasive growth and osmotic stress [38]. Although osmotic stress TFs are commonly associated with the hyper-osmotic glycerol (HOG) pathway, Ypd1p can phosphorylate Skn7p, signaling the hypo-osmotic stress pathway [39]. Because there were no apparent hypo-osmotic conditions in this study, this indicates that these TFs are not directly controlled by osmotic conditions, but possibly through a secondary response to upregulation and increased secretion of cell wall proteins.

TGN TFs and/or the genes they regulate are possible targets for increasing Golgi-to-cell membrane trafficking. In S. cerevisiae, recombinant protein intended for secretion has been found mis-trafficked to the vacuole. This has been shown for insulin and green fluorescent protein secretion in yeast [40,41]. Proteins involved in vesicle trafficking, namely Sly1 and Munc18 have been found to increase recombinant secretory rates in Chinese hamster ovarian (CHO) and several mammalian cell lines [42,43]. It is likely that similar proteins are present in yeast and could be exploited for improving protein secretion.

Significant alterations in amino acid metabolism were observed, particularly in the Δhac1 strains. De novo amino acid synthesis (GCN4, BAS1, MET32, ARG81, RTG3) was suppressed. On the surface, this appears contradictory, as increased amino acid requirements should be observed with recombinant protein production. However, this decrease in de novo amino acid synthesis is accompanied by observed increases in scavenging mechanisms for amino acids (SNT2, CUP9, PUT3). High scavenging rates and decrease synthesis imply high protein degradation rates where the degraded proteins result in available amino acids for scavenging; reducing the need for newly synthesized amino acids. This is consistent with either ERAD, a process where proteins that are stalled in the ER are transported back into the cytoplasm for degradation by the proteosome, or vacuolar-localized protein degradation. In either case, the cell is expending energy on synthesizing proteins that are ultimately degraded. These effects appear in the strains that are the slowest growing with the highest ATP requirements (Figures 1b and 2b). In these cases the ER folding capacity is likely saturated, resulting in ER holdup and ERAD.

Oxidative stress TFs were also found in all conditions. Several were dual oxidative/osmotic stress TFs (CIN5, SKN7, SKO1), and others were dedicated to oxidative stress only (AFT2, YAPI). TFs were found in all three of the major oxidative stress signaling pathways, (a) the Hog1 MAPK pathway (where SKO1 is the DNA binding agent), (b) Sln1 pathway (where SKN7 is the DNA binding agent), and (c) YAP1 and CIN5, which directly sense oxidative stress and bind DNA [44]. The cell's control machinery appears to have hard-wired oxidative stress responses to increased secretory demand, as oxidative/hypo-osmotic pathways have a high degree of overlap, which is appropriate because increased secretion of cell wall proteins will result in higher oxidative stress. In particular, Skn7p, which has already been mentioned for its role in managing secretory pathway directly in an osmotic stress pathway, can also activate oxidative stress response genes [45].

Oxidative stress was pronounced with all secretory perturbations and has been identified in other studies to be associated with secretory stress [1,17]. Futile cycling may be the dominant disulfide resorting pathway when folding is limited. In previous studies, oxidative stress, induced by tunicamycin, a N-linked glycosylation inhibitor, increased with ER stress, despite no increase in the net disulfide bond formation demand [17]. The futile cycle does predict non-stoichiometric ROS formation, while isomerization does not. ROS can be formed at potentially limitless amounts through multiple rounds of disulfide formation and breaking. This will occur under conditions where the rate of folding is slow, a result of proteins that are specifically difficult to fold, or a result of the overall ER folding capacity being saturated. As well, futile cycling will increase as the number of available cysteine residues available for disulfide bonding increase, as is the case for α-amylase, due to the extended amount of isomerization that may be needed to form the correct disulfide bonds.

One implication of the proposed thermodynamic model is that PDI paralogues, or cysteines within a PDI, must exist at different electron affinities that are above and below the electron affinity of the protein to be folded. Although in vivo redox potentials of PDI cysteine pairs were not measured, from first principles it would appear highly likely that these PDIs would need different redox potentials to carry out isomerization. In Figure 4a, we assume that only PDIs interact with the folding protein. This appears the case, as kinetic rates for direct glutathione oxidation/reduction are too slow to be physiologically relevant [9]. Electron affinity (and therefore redox potential) is broadly determined by the proximity of the two cysteines, with the proximity determined by the current structure of the protein [46]. Cysteines that are in the correct orientation will have a low electron
affinity and easily form disulfide bonds, while cysteines that are not in the correct orientation will have a high electron affinity and will have unstable disulfide bonds. Therefore, the electron affinity of a correctly folded/correct disulfide bond would be lower than that of a misfolded or incorrect disulfide bond. This difference in electron affinity may allow PDIs to selectively break disulfides with high electron affinity (incorrect bonds), but not disulfide bonds with low affinity (correct bonds).

The need for different PDIs to form or break disulfide bonds may explain the need for many PDI homologues in the ER, each with different structures, and therefore different electron affinities. These PDIs can only span a finite range of electron affinities, and there may be implications for proteins that have disulfide pairs with electron affinities higher than the highest PDI or lower than the lowest PDI. If no PDI has a lower electron affinity than an incorrect disulfide bond, then the disulfide bond cannot be broken and the protein is terminally misfolded. As well, a protein that has a native disulfide pairing with an electron affinity higher than any PDI cannot form a bond. This may be the case when recombinant proteins are being processed in the ER.

Futile cycling as a large potential ROS source has broad implications on the cell. Tu and Weissman predict Ero1p-produced ROS that is one-to-one with disulfide bond formation could attribute approximately 25% of cellular ROS to the secretory pathway [1]. Therefore, even larger ROS production is likely if the futile cycle is the dominant disulfide resorting pathway under folding stress. This also has implications on GSH and possibly NADPH availability, as it is doubly consumed (a) by the reduction of ROS and (b) directly in the futile cycle. The futile cycle limits reducing equivalents needed for anabolic processes, and may explain the reduced growth rates observed in folding stressed strains (WA, dl, and da).

In all, Figure 4b highlights that the relative rates of two processes, protein folding and disulfide bond formation, must be kept in balance to avoid significant cellular stress. If disulfide bond formation is fast compared to folding, high futile cycle use will result in high ROS formation, NADPH loss, and high protein degradation as a result of ERAD. This scenario is observed in the Δhac1 strains dl and da.

The engineering implications for protein secretion become much clearer with this understanding of protein folding to disulfide bond formation ratio. When overexpressing a recombinant protein, an optimal expression must be found, where transcription is as high as possible without overloading the ER folding capacity and sending the cell into an oxidative stressed state. This optimal expression level will be different for different proteins, as protein folding rates will vary according to the protein size and structure. We see this in comparing IP and α-amylase expression. The concept of an optimal expression has been identified heuristically, in the present study we identify the competing molecular effects that could define these phenomena [47]. This optimal expression ratio extends to recombinant proteins that do not have disulfide bonds. For recombinant proteins without disulfide bonds, recombinant protein folding in the ER will consume folding resources, thus slowing down folding rates. Although the recombinant protein has no disulfide bonds, many native proteins still require disulfide bonds. Because of this, the folding to disulfide bond formation ratio will be disturbed, resulting in similar ROS stress.

To maintain an optimal ratio, either protein folding rates must increase or oxidation rates decrease. Overexpression of chaperones that increase folding capacity has successfully been used to increase protein secretion [6,48]. For particularly large or difficult to fold proteins this may not be adequate. A new approach may be to limit the oxidation rate of Ero1p to slow down the first step of the futile cycle. This would be done in concert with repressing ERAD, as proteins would have long retention times in the ER. In this scenario, recombinant proteins would be slowly folded, albeit without high cellular stresses. This would result in longer overall process times, but may be required for difficult to fold proteins.

Conclusion

In this study, we identified post-Golgi vesicle sorting, high protein degradation rates, repressed overall expression, and oxidative stress in response to +/− UPR strains secreting different sized recombinant protein. These processes were identified through scoring TFs and estimating alteration to the metabolic network. These observations imply our proposed futile cycling is the dominant disulfide resorting pathway in the ER and explains non-stoichiometric ROS formation seen in our study and elsewhere. The futile cycle model, producing ROS and consuming GSH, has a clear thermodynamic driving force compared to disulfide bond isomerization. If correct, futile cycling is likely the dominant mechanism under secretory stress. This interplay between protein folding and futile cycling sheds light on a largely empirical understanding of engineering protein secretion and implies the relative rates of protein folding and disulfide bond formation are critical to maintaining cellular homeostasis. This increased molecular understanding of the secretory pathway should allow for more insightful design of secretory engineering strategies.

Methods

Strains and media

All experiments were performed in the background of CEN.PK 113-5D (MAT α SUC2 MAL2-8α ura3-52, P.
Kötter, Frankfurt, Germany) [49]. Genomic DNA from Y05650 (BY4741; Mat a; his3D1; leu2D0; met15D0; ura3D0; YFL031w::kanMX4, obtained from EUROSCARF) was used as a template for the HAC1 knockout cassette. Standard molecular biology techniques were used [50] and all plasmids were maintained in *Escherichia coli* DH5α in Luria Bertani (LB) broth with 80 mg/L ampicillin. PCR primers are listed in Additional file 8.

**Cloning**

Genomic DNA was purified from Y05650 using Fast DNA Spin Kit for Soil (MP Biomedicals Solon, OH, USA). A 2.6 kb DNA fragment containing the genomic replacement of HAC1 with KanMX and flanking regions was amplified by PCR using primers KT007/KT008 (Additional file 8). The HAC1::kanMX4 fragment was integrated at the HAC1 loci of CEN.PK 113-5D by standard yeast transformation [51] and selected on 200 mg/L G418 to create the Δhac1 strain. Correct integration was confirmed by PCR.

DNA coding for an insulin precursor with a Yap3 pre-leader sequence and the TA57 pro-leader sequence and spacers as described [18] for correct secretory processing was synthesized with optimal codon usage for yeast and delivered on plasmid pUC57-Yap3Insulin (GenScript Co. Piscataway, NJ, USA) (Additional file 9 for sequence). α-Amylase DNA was amplified from *Saccharomyces kluyveri* strain YKM37 [52] using LZH018 and LZH039. The pre-pro-leader was amplified from pUC57-Yap3Insulin using primers LZH015 and LZH039 [53]. Δ-hac1 strains by standard methods [51].

**Fermentor conditions**

Strains were grown in SD-2xSCAA [55], containing 20 g/L glucose, 6.7 g/L yeast nitrogen base minus amino acids (Formedium, Norfolk, UK), 2 g/L KH₂PO₄ (pH = 6 by NaOH), 190 mg/L Arg, 108 mg/L Met, 52 mg/L Tyr, 290 mg/L Ile, 440 mg/L Lys, 200 mg/L Phe, 1260 mg/L Glu, 400 mg/L Asp, 380 mg/L Val, 220 mg/L Thr, 130 mg/L Gly, 400 mg/L Leu, 40 mg/L Trp, 140 mg/L His, 1 g/L bovine serum albumin. Five hundred mL of medium was inoculated in a 1 L bioreactor (DasGip, Jülich, Germany) at 30°C, 600 rpm agitation, 30 standard L/h air flow, pH controlled at 6 by KOH (2 M). Strains were inoculated to an A₆₀₀ = 0.01 from late exponential phase cultures and A₆₀₀ was measured throughout the cultivation. Dry cell weight (DCW) was measured by filtering 5 mL of culture broth through a 0.45 μm nitrocellulose filter and measuring the increased weight of the dry filter. Glucose, ethanol, glycerol, and acetate were measured using a Summit HPLC (Dionex, Thermo Scientific, Waltham, MA, USA) with an Aminex HPX-87H column (Bio-Rad, Hercules, CA, USA). Carbon dioxide and oxygen levels were measured in the off-gas and dissolved oxygen was monitored. Transcriptome samples were taken after 5+ doublings at A₆₀₀ = 1.0-1.4. Triplicate fermentations were carried out for each strain.

**Protein quantification**

Insulin was measured by a modification of the assay by Snell et al. [56]. One mL of cell culture was centrifuged at 4000 × g for 4 min. Eight parts supernatant was added to one part 0.1 N HCl and 5.5 μM sodium azide and stored at 4°C until measurement. Insulin concentration was determined by HPLC using a Luna 5 μ C18(2) (250 mm × 4.6 mm) (Phenomenex, Torrance, CA, USA) column and gradient-based elution. Buffer A contained 68 mM phosphoric acid, 0.2 M sodiumsulphate and 10% (w/v) acetonitrile in water, and Buffer B contained 50% acetonitrile in water. HPLC was run with 25 μL injections at 1 mL/min and 50°C. Gradient protocol: 20% B for 10 min. Linear gradient from 20% B to 60% B over 10 min. Hold at 60% B for 5 min and then to 20% B for 3 min to re-equilibrate for next sample. Insulin standards eluted at 22.6 min and insulin precursor at 20.0 min. HPLC peaks were verified to be the correct protein by SDS-PAGE. Human insulin was used as a standard (Sigma, St. Louis, MO, USA).

α-amylase concentration was calculated from enzyme activity. α-amylase activity was measured using the Cera kit (Megazyme K-CERA, Wicklow, Ireland) using α-amylase from *Aspergillus oryzae* (Sigma, St. Louis, MO, USA) as a standard. This conversion was calculated using a 1.79 U/mg (weight includes salts and purified protein) standard from Sigma using the Protein 80 chip on the Bioanalyzer (Agilent, Santa Clara, CA, USA). By this, α-amylase was found to be 0.0257 g α-amylase/g total α-amylase activity was then converted to mass using 70 U/mg α-amylase protein.

**Transcriptome analysis**

Samples for microarray were taken as described previously and stored at -80°C until processing [57]. RNA was isolated using the RNeasy Minikit (Qiagen, Valencia, CA, USA). Cells were lysed in RNeasy RLT buffer using Lysing Matrix C (MP Biomedicals Solon, OH, USA) in a Fast Prep 24 (MP Biomedicals Solon, OH, USA) as follows: 20 s at speed 6, 1 min at 4°C, 20 s at speed 6. RNA was processed to aRNA using the GeneChip 3’ IVT Express Kit (Affymetrix, Santa Clara, CA, USA) and hybridized/scanned on the Yeast Genome 2.0.
Array (Affymetrix, Santa Clara, CA, USA) following commercial protocols to create CEL files.

Images were analyzed using R 2.10.1 statistical software and the ‘affy’ and ‘limma’ packages as described previously [58]. Briefly, background normalization was carried out using robust multi-array (RMA) average method with perfect match (PM) probes only. Interchip normalization used the qspline algorithm with median polish summary method. Statistical analysis was carried out by comparison of triplicate bioreactor measurements for each strain. Empirical Bayesian statistics were used to moderate standard errors within each gene and Benjamini-Hochberg’s method to adjust for multiple testing. Microarray data was submitted to the GEO database and have accession number GSE27062 (see http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=dpyzfywsoqecbk&acc=GSE27062).

**Reporter transcription factor analysis**

Transcription factor activity was scored using the Reporter Effector algorithm [23]. Transcription factor-DNA interactions were gathered from ChIP-chip with $P < 0.001$ [24]. Significant interactions were found for 176 transcription factors regulating 3,796 genes for a total of 10,849 unique interactions. Gene $P$ values from comparing different strains were used to score transcription factors that were known to bind to the upstream DNA. Transcription factors with $P < 0.05$ of being activated between conditions are reported.

**Flux balance analysis**

Estimates of intracellular reaction rates were performed using measured exchange fluxes of glucose, ethanol, acetate, glycerol, and carbon dioxide. Model-based error correction was used to close carbon and electron balances [59]. Flux balance analysis was carried out using a 85 reaction model of yeast central carbon metabolism and biomass yield were used [21]. Additional file 4 contains the complete results of the analysis which are used to estimate ATP consumption in the different strains.

**Additional material**

<table>
<thead>
<tr>
<th>Additional file</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Measured exchange fluxes in strains. Measured metabolite exchange fluxes for strains used in this study.</td>
</tr>
<tr>
<td>2</td>
<td>Final glycerol concentration of WT and Δhac1 strains. Measured glycerol titers at end of fermentation for strains used in this study.</td>
</tr>
<tr>
<td>3</td>
<td>Estimated exchange fluxes. Metabolite exchange fluxes as estimated by error-correction algorithm for strains in this study.</td>
</tr>
<tr>
<td>4</td>
<td>Intracellular fluxes for metabolic network. Flux balance analysis estimates of internal fluxes for strains in this study.</td>
</tr>
<tr>
<td>5</td>
<td>Reporter TFs for WT protein secretion. Transcription factors activated by recombinant protein secretion in wild-type background.</td>
</tr>
<tr>
<td>6</td>
<td>Reporter TFs for Δhac1 protein secretion. Transcription factors activated by recombinant protein secretion in Δhac1 background.</td>
</tr>
<tr>
<td>7</td>
<td>Expression profiles for ribosomal proteins. mRNA concentrations for yeast ribosomal proteins as determined by DNA microarray.</td>
</tr>
<tr>
<td>8</td>
<td>Oligonucleotides used in this study. PCR primers used for cloning and validation.</td>
</tr>
<tr>
<td>9</td>
<td>Synthesized insulin precursor DNA sequence. DNA sequence for insulin precursor used in this study.</td>
</tr>
</tbody>
</table>

**Abbreviations**

- ATP: adenosine triphosphate; Δ: Δhac1 secreting α-amylase; δ: Δhac1 secreting IP; Δn: Δhac1 with empty vector; mRNA: deoxyribonucleic acids; ER: endoplasmic reticulum; ERAD: ER-associated degradation; FBA: flux balance analysis; FRE: filamentous and invasive responsive elements; GSH: reduced glutathione; HOG: hyper-osmotic glycerol; IP: insulin precursor; mRNA: messenger RNA; NADPH: nicotinamide adenine dinucleotide phosphate; PDI: protein disulfide isomerase; RNA: ribonucleic acid; ROS: reactive oxygen species; TF: transcription factor; TGN: trans-Golgi network; UPR: unfolded protein response; WT: wild-type.

**Acknowledgements**

We thank Dr Intavat Nookaew for assistance with transcriptome analysis and FBA. We thank NIH F32 Kirschstein NRSA fellowship (F32 GM083647), The Knut and Alice Wallenberg Foundation, EU Framework VII project SYNSBIO (Grant no. 212766), European Research Council project INSYSBIO (Grant no. 247013), and the Chalmers Foundation for funding.

**Authors’ contributions**

KT, DP, and JN designed the experiment. KT and ZL carried out all cloning, transcriptomics and metabolic flux data. KT, ZL, DP, and JN analyzed data. KT, DP, and JN wrote the manuscript. KT and JN supervised the research. All authors have read and approved of the final manuscript.

**Competing interests**

The authors declare that they have no competing interests.

**References**

6. Shusta EV, Raines RT, Pluckthun A, Wittrup KD: Increasing the secretory
26. Carla Fama M, Raden D, Zacchi N, Lemos DR, Robinson AS, Silberstein S: Disulfide Formation in the ER and
11. Tu BP, Weissman JS: The FAD- and O-2-dependent reaction cycle of Ero1-
13. Cuozzo JW, Kaiser CA: Intracellular signaling from the endoplasmic reticulum
17. Kimata Y, Ishiwata-Kimata Y, Yamada S, Kohno K: The role of leaders in intracellular
18. Kjeldsen T, Pettersson AF, Hach M: Increasing Yeast Secretion of
43. Peng RW, Guert C, Tigg M, Fussenegger M: The vesicle-trafficking
27. Tao W, Deschesnes RJ, Fassler JS: Intracellular glycerol levels modulate the activity of Sml1p, a Saccharomyces cerevisiae two-component regulator.
...


Cite this article as: Tyo et al.: Imbalance of heterologous protein folding and disulfide bond formation rates yields runaway oxidative stress. BMC Biology 2012 10:16.
PAPER III

- UPR Effect
- Random Mutation
- Host Engineering
- RPP, IP and amylase
- Leader
- Promoter
- Marker
- Expression Design
- Fermentation Analysis
- Aeration
- Cell growth
Systems Biology Analysis of Amylase Producing Yeast Strains

Zihe Liu¹, Tobias Österlund¹, Jin Hou¹,³, Lifang Liu¹, Linn Fagerberg⁴, Dina Petranovic¹, Mathias Uhlén⁴ and Jens Nielsen¹,²,*

¹ Novo Nordisk Foundation Center for Biosustainability, Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-41296 Göteborg, Sweden
² Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Fremtidsvej 3, DK-2970 Hørsholm
³ Current affiliations: State Key Laboratory of Microbial Technology, Shandong University, 27 Shanda Nan Road, Jinan, Shandong, 250100, China
⁴ the School of Biotechnology, AlbaNova, University Center, Royal Institute of Technology, SE-106 91, Stockholm, Sweden

* Corresponding author:
Department of Chemical and Biological Engineering
Chalmers University of Technology, SE-412 96 Gothenburg, Sweden
Telephone: +46 (0)31 772 3804,
Fax: +46 (0)31 772 3801,
E-mail: nielsenj@chalmers.se
Abstract

The increasing demand of industrial enzymes and biopharmaceutical proteins calls for robust production host with high protein yield and productivity. Being one of the best studied model organisms and armed with the ability of performing post-translational modifications; *Saccharomyces cerevisiae* is widely used as a cell factory for recombinant protein production. However, engineering of the secretory pathway has not been exploited in full detail, and many proteins are produced at only 1% or even 0.1% of yeast’s optimal capacity. With the development of next generation sequencing technologies and systems biology analysis it has, however, become possible to use identify novel targets for metabolic engineering of protein production by analyzing mutants generated by random mutagenesis and screening. In this study, we substantially increased yeast’s secretion capacity for amylase production (3-5 fold), and suggested possible ways to further improve the production. Through integrated transcriptome analysis, we identified that most genes related to stress responses were up-regulated in a recombinant protein production (RPP)-dependent manner (comparing all amylase producing strains with a control strain), whereas these genes were down-regulated in a mutation-dependent manner (comparing UV mutated strains with a non-mutated strain). We also found that genes related to respiration, RNA stability, protein trafficking, protein turnover and amino acid metabolism were significantly changed in a strain-specific manner. Furthermore, we identified single nucleotide variations, insertions and deletions as well as mutations in the upstream region (0-1000bp) in the mutated strains from high-throughput sequencing data for the UV mutagenesis strains. Combined with the transcriptome and genome sequencing data, we suggested possible ways that will allow for further improve protein production through metabolic engineering.

**Keywords:** Recombinant protein production, Secretory pathway, *Saccharomyces cerevisiae*
The increasing demand of the recombinant protein industry calls for robust production hosts and efficient expression systems. Today around 20% of protein-based biopharmaceuticals on the market are produced by *Saccharomyces cerevisiae* (Martínez, Liu et al. 2012), including insulin, hepatitis B surface antigen, urate oxidase, glucagon, granulocyte macrophage colony stimulating factor, hirudin, and platelet-derived growth factor (Demain and Vaishnav 2009). *S. cerevisiae* is one of the most well-established host systems for commercialization of biopharmaceutical proteins and industrial enzymes, due to the extensive knowledge of this species genome, metabolism and general physiology (Giaever, Chu et al. 2002; Petranovic, Tyo et al. 2010) and long history of industrial processing. Besides this, the main advantage for yeast expression systems is the similarity of the secretory pathways with mammalian systems and its capacity to perform strict quality control (Hou, Tyo et al. 2012) and post-translational modifications (Tokmakov, Kurotani et al. 2012), including proteolytic processing of signal peptides, disulfide bond formation, subunit assembly, acylation, glycosylation, phosphorylation and as well as the ability to secrete proteins in their native forms to facilitate downstream processing (Freigassner, Pichler et al. 2009).

For secreted proteins, there are many steps after translation before the protein is matured and trafficked to the extracellular region. A common pathway, called the secretory pathway, is used to complete the protein maturation process. This post-translational protein processing is an extensive pathway where proteins pass through several different organelles. The details of the chemical and molecular mechanisms of the secretory pathway processing have been extensively reviewed (Hou, Tyo et al. 2012). After translation, the polypeptides get folded and glycosylated in the endoplasmic reticulum (ER), and the correctly folded proteins are then sorted to the Golgi apparatus for further glycosylation and final modification, whereas misfolded proteins are sorted into the cytosol for degradation. Correctly modified proteins
will be targeted to the membrane and secreted to the extracellular region, otherwise they will be sorted to the endosome or vacuole for re-cycling or degradation. The secretory pathway involves several checkpoints where the state of protein folding and its impact on overall cellular stress is monitored. The chaperone capacity, vesicle and cargo proteins, oxidizing equivalents, as well as metabolite requirements, such as ATP, NADH, NADPH, glutathione buffers, glycans, etc., should be well tuned according to the expressed protein and the host system. Protein folding and modification, trafficking, degradation, as well as amino acid metabolism involves many layers of quality control that must be well-coordinated to avoid cellular stress resulting in reduced cell growth and protein secretion (Dürrschmid, Reischer et al. 2008; Nemecek, Marisch et al. 2008) or even apoptosis and cell death (Mattanovich, Gasser et al. 2004).

Many attempts have been applied to S. cerevisiae for enhancing protein production, which could be obtained by host design, expression control and environmental optimization. However, the production of proteins is often not improved when transcriptional and translational processes are substantially optimized (Porro, Sauer et al. 2005; Schröder 2007; Liu, Tyo et al. 2012), and optimization of culture conditions are also highly host and protein specific (Idiris, Tohda et al. 2010). There is therefore increased focus on host engineering. However, with all advantages mentioned above, for most proteins produced in S. cerevisiae, their secretion levels are still 100-, or even 1000-fold lower than their theoretically yield (Schröder 2007). Due to the poor understanding of the protein processing machinery, which involves many tightly cross-reacting factors, molecular engineering to enhance recombinant protein production is sometimes difficult and time-consuming. It is therefore interesting to learn from strategies that appear based on selective screening, either following random mutagenesis or adaptive evolution. Using tools from systems biology combined with
When selecting a method for generating random mutations, the question is: should it be a cell-based or protein-based design? If a cell-based engineering is carried out, should it be chemical or physical mutagenesis (Durnev 2008), using cDNA libraries (Shusta, Kieke et al. 1999), mutation strain collections (Kanjou, Nagao et al. 2007), or transcript factor design engineering (Alper, Moxley et al. 2006; Dent, Lau et al. 2007; Bashor, Helman et al. 2008)? What screening method should be used? How should the transcriptional analysis and genomic sequencing be applied in order to identify novel and effective targets (Desai, Rodionov et al. 2009; Vaquerizas, Kummerfeld et al. 2009)? Payne et al. (Payne, Finnis et al. 2008) identified over-expression targets involved with the ATPase recycling of BiP (JEM1, SIL1, LHS1, and SCJ1) using chemical mutagenesis, and this enabled enhanced production of human albumin, granulocyte–macrophage colony-stimulating factor, and human transferrin. Kanjou et al. (Kanjou, Nagao et al. 2007) found potential deletion target of vesicle formation MON2 by screening the EUROSCARF deletion library, and hereby increased the secretion levels of luciferase. Screening surface-displayed cDNA libraries could also help identifying targets for antibody fragments production (Shusta, Kieke et al. 1999; Wentz and Shusta 2007), for example cell wall protein genes (CCW12, CWP2, and SED1) (Wentz and Shusta 2008), the ribosomal subunit gene RPP0 (Wentz and Shusta 2008), and the thiol oxidase gene ERO1 (Gross, Kastner et al. 2004).

In this study, we substantially increased yeasts capacity for amylase production using UV mutagenesis and starch screening. Through whole genome sequencing analysis and transcriptome analysis, biological mechanisms in response to high levels of recombinant proteins were identified, and possible targets to over-explore this potential were suggested.
Results and discussion

Construction of mutation Library

As shown in Figure 1A, the amylase producing strain AAC (Liu, Tyo et al. 2012) was used as the starting strain for UV mutagenesis. The AAC strain was spread on starch plates to obtain single colonies (around 10^6 cells), and exposed to different UV doses (from 4-11 mJ/cm²) at 254 nm using UV cross-linker (Topac Inc., USA) to obtain a mortality rate between 60%-90%. The strains were then cultivated in dark at 30 degrees for six days and 591 clones with large colony size were selected. When performing mutagenesis experiments, trade-offs should always be considered, i.e. improvement of fitness under certain circumstances always comes with some other reduced traits under other conditions (Wenger, Piotrowski et al. 2011).

Enhancing recombinant protein production sometimes comes together with reduced biomass growth. Here the starch plate was applied as a first round of selection trying to minimize this trade-off, namely such that large colonies selected have advances in both cell growth and amylase secretion. The selected 591 strains were further cultivated in falcon tubes and shake flasks and two strains with high amylase production were identified, named M715 and M1052 (7 and 10 are the UV dose applied, while 15 and 52 are identification numbers for these two strains at each UV dose, respectively).

Characterizations of mutation strains

In order to test whether this high amylase production is because of mutations in the amylase plasmid or the mutated yeast itself, mutated plasmids from both strains were extracted and the sequencing result showed that there is no mutation site in the POT1 marker region, pTPI1 promoter region or the alpha factor leader region, however, there was a common single site mutation in the amylase region (Thr331→Gln). The mutated plasmids were then transformed into normal strain (CEN. PK 530-1C), and shake flask cultivations of these two strains
showed no improvement for amylase production compared with the AAC strain, Figure S1, which suggested that it is host mutations that caused production enhancement. Further experiments will be carried out to test whether the mutated yeast is a general better protein producer by replacing the mutated amylase plasmid with normal insulin plasmid, Figure 1B2. The strain NC (with no amylase production) and AAC (with amylase production without UV mutation) were used as reference strains (Liu, Tyo et al. 2012) and the four strains (NC, AAC, M715 and M1052) were evaluated under batch cultivations, Figure 1B3. Strains grew slower when amylase production increased. The glycerol production was similar among the amylase producing strains, although more amylase was produced in the mutated strains, suggesting a more efficient amino acid utilization process in the mutated strains. Less biomass and more ethanol was produced in the mutated strains during the glucose phase. Detailed strain information, extracellular product yield and cell growth parameters are listed in Table I and II. The mutation strains exhibited high amylase producing capacities (yield on cell mass) (Figure 2). For amylase production in the exponential phase, the M715 strain produced 5.4-fold of amylase and the M1052 strain produced 4.9-fold of amylase compared to AAC. When comparing final amylase production, M715 and M1052 strains produced 2.4-fold and 3.5-fold more amylase than AAC. The mutation strains showed a relatively larger improvement in productivity in the glucose phase compared with the overall productivity, suggesting that the gene mutations may be related with protein processing in the glucose phase.
Common and specific transcriptional responses for UV mutation strains

Genome-wide transcription analysis of the mutated strains (M715 and M1052) and the control strains (AAC and NC) was carried out during the exponential growth phase (glucose phase). The expression of 63 genes change significantly in expression (FDR<0.001) when comparing M715 with AAC, and 1452 genes change significantly (FDR<0.001) when comparing M1052 with AAC, Figure S2.

In order to identify key transcription factors that could play an important role in the regulation of RPP we applied the reporter TF algorithm (Patil and Nielsen 2005; Oliveira, Patil et al. 2008) to the microarray data, using the TF-DNA interaction network presented in Harbison et al. 2004. As shown in Figure 3, we found that most genes related to stress responses, such as genes regulated by oxidative stress (transcriptionally regulated by Yap1p), osmotic stress (regulated by Hog1p), and general stress (regulated by Msn2p and Msn4p), were up-regulated in a RPP-dependent manner (comparing all amylase producing strains with the control strain, Figure 3A), whereas down-regulated in a mutation-dependent manner (comparing all UV mutagenesis strains with the non-mutated strain, Figure 3B). We also found that genes related to respiration, regulated by Hap2p, Hap3p, Hap4p and Hap5p, were down-regulated in mutated strains compared to AAC (Figure 3B). Since we identified that amylase is produced at higher levels at anaerobic conditions compared to aerobic conditions (Liu, Österlund et al. Submitted), we suggest that both the reduced respiration and reduced stress responses might contribute to the higher production in the mutated strains.

Reporter secretory pathway analysis

In order to investigate how the improved protein production in the UV mutated strains is related to changes in the protein secretory pathway, involving protein processing, sorting and other ER- and Golgi functions we took an integrated data analysis approach. Recently, the
first yeast secretory model that covers 170 secretory proteins, classified into 16 secretory
classes, was generated (Feizi, Österlund et al. Submitted). Here, the secretory network was
further expanded to also involve regulation of the secretory pathway and response to stress.
We also included genes regulated by UPR (Travers, Patil et al. 2000; Kimata, Ishiwata-
Kimata et al. 2006), and secretory genes classified to the transcription factor response to
oxidative stress (Yap1p), response to heat shock (Hsf1p) and general stress response (Msn2p
and Msn4p). The final list of genes involved in protein secretion was obtained after manually
correction based on Saccharomyces Genome Database and literature reading. The Reporter
Features algorithm was then used to score the secretory pathway functions by the significant
changes of genes expression that belong to the related pathway. As shown in Figure 4, genes
belonging to the protein trafficking pathway were up-regulated in both mutated strains
compared to AAC, whereas genes belonging to the proteasome associated degradation
pathways were down-regulated in both strains. Genes associated with vacuole and amino acid
metabolism were specifically regulated in M715 and M1052.

Changes in genotype

The genomes of the UV mutation strains (M715 and M1052) and the wild type CEN.PK 113-
7D strain were sequenced (Materials and Methods, table III), aligned and compared to the
reference sequence of CEN.PK 113-7D. In total 1713 putative mutations were identified in
any of the two protein producing strains. Mutations that were present (same position and
same variant) in both UV mutation strains and in the wild type strain were considered to be
due to genetic differences between our background strain and the strain that was used to
obtain the reference sequence, and were therefore filtered out. 496 mutations were unique for
any of the two UV mutation strains and out of these were 328 single nucleotide point
mutations and 84 were INDELs. The majority of mutations were found in the non-coding
regions. All mutations identified in coding and upstream regions in all strains are presented in Additional dataset S1-S4.

Furthermore, genes that have single nucleotide variations (SNVs), insertion and deletions (INDELs), as well as mutations in the upstream region (0-1000bp) are shown in Figure 5. Genes with only synonymous mutations (genes that only changed in nucleotide sequence but not in the amino acid sequence) and genes that changed in the upstream region but not transcriptionally regulated (FDR<0.05) were further filter out (Figure 5A). This resulted in the identification of 17 genes that were commonly changed in their genomic sequence in both mutated strains, whereas 40 and 41 genes were changed uniquely in either M715 or M1052, respectively.

Mutated genes involved in the translation process and the protein secretion pathway are presented in Table IV. There is consistency with the found genetic changes and the transcriptome analysis, i.e. there are mutations in both strains in several genes related with cellular stress response. Though different genes were changed, genes regarding protein degradation, protein trafficking pathways were also found to be mutated. Genes related to respiration and amino acid metabolism were specifically changed in the M715 strain. Detailed information of all mutated genes is shown in additional files.

Taken together results of both transcriptome analysis and genome sequencing analysis, we suggested that genetic changes regarding stress response, respiration, protein degradation, protein trafficking and amino acid metabolism might be the reason for the increased amylase production. Molecular experiment will be carried out to validate key genes that benefit for amylase production.
Conclusion

In this study, we have identified biological mechanisms related to the protein processing machinery by introducing perturbations to the cell, in the form of high levels of α-amylase production and UV mutagenesis, and measuring cellular responses at the systems level via whole genome sequencing and transcriptome analysis. These measurements, combined with Reporter TFs and Reporter Secretory Pathway algorithm, were able to identify cellular adjustments in (a) amino acid metabolism, (b) protein degradation, (c) protein trafficking, (d) respiration, and (e) stress responses, which could provide a clue about why amylase production is higher in the mutated strains. Potential targets for enhancing protein production were also suggested.

Materials and Methods

Media and cultivation conditions

YPD media was prepared as follows: 20 g/L D-glucose, 10 g/L yeast extract, 20 g/L peptone, and 1 g/L BSA. Starch plate was prepared as follows: 0.04 g/L D-glucose, 10 g/L starch, 6.7 g/L YNB (Yeast Nitrogen Base) without amino acids, 20 g/L agar.
Seed cultures were grown over-night, and inoculated into the fermenter at an initial OD (A600) of 0.01. All fermentations were performed in DasGip 1.0-liter stirrer-pro vessels (Drescher Arnold Schneider, Germany) with a working volume of 500ml YPD media, at 30°C, 600 rpm agitation. Aerobic conditions were controlled by keeping flowing 1 vvm (volume of flow per working volume per minute) of air during fermentation. One drop of antifoam was added to each fermenter. Dissolved oxygen was measured using a polarographic oxygen electrode (Mettler Toledo, Switzerland). The pH was maintained at 6.0 by the pH sensor (Mettler Toledo, Switzerland) using 2 M KOH. All fermentations were done in biological triplicates.

**Analytical methods**

1 ml of the culture was centrifuged at 4000g for 5 min, and 800 μl of the culture supernatant was mixed with 5.5 mM NaN₃ final concentrations, and stored at 4°C until measurement. Concentrations of glucose, glycerol, ethanol, and acetate were analyzed by the Dionex Ultimate 3000 HPLC (Dionex Softron GmbH, Germany) with an Aminex HPX-87H column (BIORAD, USA) at 65°C using 5 mM H₂SO₄ as the mobile phase at a flow rate of 0.6 ml/min. The activity of α-amylase was measured using the Ceralpha kit (Megazyme, Ireland) using α-amylase from A. oryzae (Sigma, USA) as standard. The dry cell weight (DCW) was acquired by filtering the cell culture through a 0.45 μm filter (Sartorius Stedim, Germany) and measuring the increased weight.

**Transcriptome analysis**

RNA for microarray was isolated using the RNeasy Minikit (Qiagen) and processed to aRNA using the Genechip 3’ IVT Express Kit (Affymetrix) and hybridized/scanned on the Yeast Genome 2.0 Array (Affymetrix) to create CEL files. The moderated t-statistic and reporter
analysis was performed using the Platform for Integrated Analysis of Omics-data (PIANO) package for R (Varemo, Nielsen et al. submitted).

**Illumina genome sequencing**

Whole genome sequencing of the two UV mutagenesis strains (M715 and M1052) and the wild type CEN.PK 113-7D strain was performed using Illumina hiseq 2000. The reads were checked for quality and reads with an average quality score less than 28 was filtered out. The reads were aligned to the recently published genome sequence of CEN.PK 113-7D (Nijkamp, van den Broek et al. 2012) using Stampy version 1.0.17 (Lunter and Goodson 2011). On average 83% of the reads could be mapped to the reference sequence. A summary of the sequencing and reads mapping is presented in Table III. Further, single nucleotide point mutations in each of the sequenced genomes as compared to the reference sequence were identified using Atlas-SNP2 version 1.0 (Shen, Wan et al. 2010) and INDELs were detected using Atlas-Indel2 version 1.0 (Challis, Yu et al. 2012). A single nucleotide point mutation was considered as highly confident by the Atlas-SNP2 algorithm if it was found on both strands, if the calculated posterior probability was greater than 0.95 and if the coverage at that position (number of aligned reads) was at least 8. INDELs were detected using the default parameters of Atlas-Indel2 and INDELs that were located in regions where the reference sequence contained at least one unknown bases (N) was filtered out.

Single nucleotide variations and INDELs which were detected as having the same variant in all three strains were considered as genetic differences between the background strain and the published reference sequence, and therefore not further investigated. Mutations and INDELS detected in exons and upstream regions (0-1000 bp upstream the exon start) in the M715 and M1052 strains could be beneficial mutations for improved protein production, and were therefore investigated further.
Acknowledgments

This work is financially supported by the European Research Council ERC project INSYSBIO (Grant no. 247013), the Novo Nordisk Foundation, the Chalmers Foundation, and the Knut and Alice Wallenberg Foundation.
Reference


Feizi, A., T. Österlund, et al. (Submitted). "Genome-scale modeling of the protein Secretory machinery in yeast."


Liu, Z., T. Österlund, et al. (Submitted). "Transcriptome analysis reveals the anaerobic protein processing machinery and the final electron acceptor in Saccharomyces cerevisiae."


### Table I. Strain information.

<table>
<thead>
<tr>
<th>Strains</th>
<th>Relevant Genotype</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC</td>
<td>CEN.PK 530-1C with TPII promoter (2 μm POT1)</td>
<td>[17]</td>
</tr>
<tr>
<td>AAC</td>
<td>CEN.PK 530-1C with pTPII-alpha factor leader amylase (2 μm POT1)</td>
<td>[17]</td>
</tr>
<tr>
<td>M715</td>
<td>UV mutated AAC strain under 7 mJ/cm²</td>
<td>This study</td>
</tr>
<tr>
<td>M1052</td>
<td>UV mutated AAC strain under 10 mJ/cm²</td>
<td>This study</td>
</tr>
<tr>
<td>M715n</td>
<td>CEN.PK 530-1C with mutated plasmid of M715</td>
<td>This study</td>
</tr>
<tr>
<td>M1052n</td>
<td>CEN.PK 530-1C with mutated plasmid of M1052</td>
<td>This study</td>
</tr>
</tbody>
</table>
Table II. Physiological characterization of mutated strains.

<table>
<thead>
<tr>
<th>Strain</th>
<th>$Y_{SX}^{a}$</th>
<th>$Y_{SG}^{b}$</th>
<th>$Y_{SE}^{c}$</th>
<th>$Y_{SA}^{d}$</th>
<th>$\mu_{\text{max}}^{e}$</th>
<th>Biomass$^{f}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC</td>
<td>0.20±0.02</td>
<td>0.05±0.01</td>
<td>0.25±0.02</td>
<td>0.033±0.007</td>
<td>0.40±0.01</td>
<td>6.5±0.2</td>
</tr>
<tr>
<td>AAC</td>
<td>0.23±0.02</td>
<td>0.15±0.01</td>
<td>0.20±0.01</td>
<td>0.034±0.001</td>
<td>0.38±0.01</td>
<td>6.7±0.1</td>
</tr>
<tr>
<td>M715</td>
<td>0.18±0.01</td>
<td>0.15±0.01</td>
<td>0.25±0.02</td>
<td>0.036±0.001</td>
<td>0.31±0.01</td>
<td>6.6±0.1</td>
</tr>
<tr>
<td>M1052</td>
<td>0.17±0.01</td>
<td>0.14±0.01</td>
<td>0.28±0.02</td>
<td>0.045±0.005</td>
<td>0.24±0.01</td>
<td>5.6±0.1</td>
</tr>
</tbody>
</table>

Yield (g/g glucose) calculated here only consider the exponential phase and the total consumed substrate. $^{a}$Biomass, $^{b}$Glycerol, $^{c}$Ethanol, $^{d}$Final succinate production, $^{e}$Specific growth rate (h$^{-1}$), $^{f}$Final biomass (g/L). The data represented triplicated biological experiment.
Table III. Overall statistics of the Illumina sequencing results

<table>
<thead>
<tr>
<th></th>
<th>M715</th>
<th>M1052</th>
<th>WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total reads</td>
<td>53846120</td>
<td>27382708</td>
<td>18413972</td>
</tr>
<tr>
<td>Coverage</td>
<td>443x</td>
<td>225x</td>
<td>151x</td>
</tr>
<tr>
<td>Reads mapping to genome</td>
<td>42388376</td>
<td>22603996</td>
<td>16067934</td>
</tr>
<tr>
<td>Mapped reads (%)</td>
<td>79%</td>
<td>83%</td>
<td>87%</td>
</tr>
<tr>
<td>SNVs</td>
<td>INDELs</td>
<td>Upstream</td>
<td>Function</td>
</tr>
<tr>
<td>-------------</td>
<td>--------</td>
<td>----------</td>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>Common</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mutations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEC7 [Gly&lt;sup&gt;92&lt;/sup&gt;→Val]</td>
<td>CDC27</td>
<td>Ubiquitin-protein ligase</td>
<td>ER-Golgi and inter-Golgi transport</td>
</tr>
<tr>
<td>HSP82 [Gln&lt;sup&gt;135&lt;/sup&gt;→His]</td>
<td></td>
<td>Heat shock response</td>
<td></td>
</tr>
<tr>
<td>COS8 [Ser&lt;sup&gt;193&lt;/sup&gt;→Arg]</td>
<td></td>
<td>Unfolded protein response</td>
<td></td>
</tr>
<tr>
<td>HLR1</td>
<td></td>
<td></td>
<td>Response to osmotic stress</td>
</tr>
<tr>
<td>IMD2</td>
<td></td>
<td></td>
<td>Resistance to the drug</td>
</tr>
<tr>
<td>M715</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YKR150C [Lys&lt;sup&gt;115&lt;/sup&gt;→Asn]</td>
<td></td>
<td>Amino acid permease</td>
<td></td>
</tr>
<tr>
<td>Unique</td>
<td>PCA1 [Ser&lt;sup&gt;411&lt;/sup&gt;→Leu]</td>
<td></td>
<td>ubiquitination</td>
</tr>
<tr>
<td>TIR4 [Ser&lt;sup&gt;778&lt;/sup&gt;→Pro]</td>
<td></td>
<td>expressed under anaerobic conditions</td>
<td></td>
</tr>
<tr>
<td>SDHI [Ser&lt;sup&gt;121&lt;/sup&gt;→Tyr]</td>
<td></td>
<td></td>
<td>Respiration</td>
</tr>
<tr>
<td>LYS2</td>
<td></td>
<td>Glycerol synthesis, essential for growth under osmotic stress</td>
<td></td>
</tr>
<tr>
<td>HK25</td>
<td></td>
<td>Response to stress</td>
<td></td>
</tr>
<tr>
<td>ETR2</td>
<td></td>
<td>N-linked protein glycosylation</td>
<td></td>
</tr>
<tr>
<td>BFU1</td>
<td></td>
<td>Ubiquitin homeostasis</td>
<td></td>
</tr>
<tr>
<td>M1052</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YLR11B [Asp&lt;sup&gt;168&lt;/sup&gt;→Glu]</td>
<td></td>
<td>Depletion causes protein degradation</td>
<td></td>
</tr>
<tr>
<td>Unique</td>
<td>TRS31 [Glu&lt;sup&gt;210&lt;/sup&gt;→His]</td>
<td></td>
<td>ER-Golgi and inter-Golgi transport</td>
</tr>
<tr>
<td>WWJ1 [Glu&lt;sup&gt;193&lt;/sup&gt;→Asp]</td>
<td></td>
<td>Regulates H&lt;sub&gt;2&lt;/sub&gt;O&lt;sub&gt;2&lt;/sub&gt;-induced apoptosis</td>
<td></td>
</tr>
<tr>
<td>VTA1 [Ser&lt;sup&gt;196&lt;/sup&gt;→Asn]</td>
<td></td>
<td>Endosomal protein sorting</td>
<td></td>
</tr>
<tr>
<td>VPS3 [Ser&lt;sup&gt;521&lt;/sup&gt;→*]</td>
<td></td>
<td>Vacuolar protein sorting</td>
<td></td>
</tr>
<tr>
<td>BTN2</td>
<td>Protein retrieval from a late endosome to Golgi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PUP2</td>
<td>Ubiquitin-dependent catabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATG23</td>
<td>Cytoplasm-to-vacuole targeting pathway and efficient macroautophagy</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure Legends

**Figure 1.** Experimental design of the random mutagenesis project. (A) Mutant construction. (B1) Evaluation of the mutated plasmid in normal strains. (B2) Evaluation of the mutated yeast for insulin production. (B3) Evaluation of the mutated amylase producing strains. (C) Integrated analysis.

**Figure 2.** Protein yield during batch fermentations. (A) Amylase yield on cell mass during the exponential phase. (B) Final Amylase production.

**Figure 3.** Reporter TFs analysis reveals reduced stress responses in mutated strains. (A) Many stress related transcription factors showed up as key transcription factors when comparing amylase producing strains with NC (reporter p-value<0.000001). (B) The reporter TF results when comparing UV mutation strains with AAC. Red color indicates that the genes regulated by this transcription factor are upregulated in the amylase producing strains as compared to NC. Blue color indicates down-regulated genes. Reporter p-value<0.000001.

**Figure 4.** Top ten regulated secretory pathway functions in mutated strains.

**Figure 5.** Whole genome sequencing analysis. (A) Venn diagram of genes that have single nucleotide variations (SNVs), insertion and deletions (INDELs) and also changes in the promoter region in mutated strains. (B) Venn diagram of all mutated genes. Upstream region: 0-1000 bp upstream the exon start.
Figure 1

A

B1

B2

B3

C

Figure 2

A

B

Glucose phase U Amy/g DCW

Final amylase production (U Amy/g DCW)
Figure 3

A

M1052-NC (UP)
M1052-NC (DOWN)
M715-NC (UP)
M715-NC (DOWN)
AAC-NC (UP)
AAC-NC (DOWN)

B

M1052-AAC (UP)
M1052-AAC (DOWN)
M715-AAC (UP)
M715-AAC (DOWN)

-RAS2
-YAP1
-STE7
-HOG1
-HSF1
-MSN2
-MSN4
-GCN4

-RAS2
-PDR1
-ROX3
-ARG81
-CCR4
-RCS1
-MSN4
-MSN2

-GCN4
-HOG1
-HAP2
-HAP3
-HAP5
-HAP4
Figure 4

Amino acid metabolism
Translocation
Vesicle trafficking-ER to Golgi
Retrograde transport, endosome to Golgi
Vacuole
Dolichol pathway
Protein folding
Heat shock response
Oxidative stress response
General stress response
A  

**SNVs**

<table>
<thead>
<tr>
<th>M715</th>
<th>WT</th>
<th>M1052</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>16</td>
<td>31</td>
</tr>
</tbody>
</table>

Non-synonymous mutations

<table>
<thead>
<tr>
<th>M715</th>
<th>WT</th>
<th>M1052</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>7</td>
<td>17</td>
</tr>
</tbody>
</table>

B  

**INDELs**

<table>
<thead>
<tr>
<th>M715</th>
<th>WT</th>
<th>M1052</th>
</tr>
</thead>
<tbody>
<tr>
<td>54</td>
<td>137</td>
<td>97</td>
</tr>
</tbody>
</table>

Transcriptional changed

<table>
<thead>
<tr>
<th>M715</th>
<th>WT</th>
<th>M1052</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>5</td>
<td>19</td>
</tr>
</tbody>
</table>

Upstream region

<table>
<thead>
<tr>
<th>M715</th>
<th>WT</th>
<th>M1052</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>17</td>
<td>41</td>
</tr>
</tbody>
</table>
PAPER IV
Anaerobic α-amylase production and secretion with fumarate as the final electron acceptor

in Saccharomyces cerevisiae

Running Title: Anaerobic α-amylase production by yeast

Zihe Liu\textsuperscript{1}, Tobias Österlund\textsuperscript{1}, Jin Hou\textsuperscript{1,3}, Dina Petranović\textsuperscript{1}, Jens Nielsen\textsuperscript{1,2} \textsuperscript{*}

\textsuperscript{1} Novo Nordisk Foundation Center for Biosustainability, Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-41296 Göteborg, Sweden

\textsuperscript{2} Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Fremtidsvej 3, DK-2970 Hørsholm

\textsuperscript{3} Current affiliations: State Key Laboratory of Microbial Technology, Shandong University, 27 Shanda Nan Road, Jinan, Shandong, 250100, China

\textsuperscript{*} Corresponding author

Telephone: +46 31 772 3804

Fax: +46 31 772 3801

E-mail: nielsenj@chalmers.se
Abstract

In this study we focus on two *Saccharomyces cerevisiae* strains with varying production of heterologous α-amylase and we compare the metabolic fluxes and transcriptional regulation at aerobic and anaerobic conditions, in particular with the objective to identify the final electron acceptor for protein folding at anaerobic conditions. We found that anaerobic conditions promote amylase production when comparing to aerobic conditions and genome-scale transcriptional analysis show that genes related to the endoplasmic reticulum (ER), lipid synthesis and stress responses were up-regulated at anaerobic conditions. Based on our integrative analysis we proposed a model for the electron transfer from ER to the final electron acceptor, fumarate under anaerobic conditions. This was supported by findings that the addition of fumarate under anaerobic conditions improves cell growth in α-amylase producing strain. Our findings provide a model for the molecular mechanism of anaerobic protein secretion using fumarate as a final electron acceptor, which may allow for further engineering of yeast for improved protein secretion at anaerobic growth conditions.

Keywords: anaerobic fermentation/ consolidated bioprocesses/ recombinant protein production/ protein folding
Introduction

Post-translational modifications in eukaryal cells comprise an extensive pathway where the proteins pass through several different organelles, compartments, and vesicles before they reach their final destinations. Protein folding and modifications, trafficking, degradation, as well as amino acid metabolism involve many layers of quality control that must be well-coordinated to avoid cellular stress resulting in reduced cell growth and protein secretion (1,2) or even apoptosis and cell death (3,4). Understanding of the molecular processes in the secretory pathway may have implications for the development of therapies for human proteostasis diseases such as Alzheimer's (5) and Parkinson's disease (6), as well as it may find applications for improving microbial based production of pharmaceutical proteins and industrial enzymes, for example, vaccines (7) and α-amylase (8). The yeast secretory pathway is responsible for processing proteins (peptides) through the ER, Golgi, trans-Golgi network and either to the extracellular space, plasma membrane, endosome or vacuole (9). The protein secretory pathway involves several checkpoints where the state of protein folding and its impact on overall cellular stress is monitored. Misfolded proteins are detected and removed via the ER-associated degradation (ERAD) pathway (10), the ubiquitin–proteasome system (UPS) (10), the autophagy pathway (11), or the unfolded protein response (UPR) pathway (12). Although many studies have provided much insight into the protein secretory pathway, most of them focus on regulations during aerobic growth.

Investigating how yeast handles folding and secretion of recombinant proteins under anaerobic conditions is relevant for both basic and applied research. There is growing interest in developing consolidated bioprocesses, with an increasing call for the cell factory to be able to secrete enzymes needed for the degradation and utilization of complex substrates (such as
amylases for the degradation of starch). Considering both industrial processes for production of
pharmaceutical proteins (13) and consolidated bioprocesses to be carried out at anaerobic
conditions there is therefore a requirement for a more fundamental insight into how recombinant
protein production is affected by the oxygen supply. It has also been reported that low oxygen
levels could enhance production of certain proteins, such as glucoamylase (14), 3H6 Fab (15)
and human trypsinogen (15). In order to investigate whether it is because there is less oxidative
stress at anaerobic conditions, it is interesting to study the impact of oxygen supply on the
secretory pathway for production of recombinant proteins. Considering the complexity of the
protein secretion pathway it is necessary to apply a systems approach, and we therefore
combined macroscopic flux analysis with genome-wide transcription analysis of several different
strains producing and secreting a heterologous α-amylase at both aerobic and anaerobic
conditions.

Additionally, for the produced protein to be active it has to fold correctly and this involves the
formation of cysteine bridges in the ER which requires transferring enormous amount of
electrons to an electron acceptor, in order to match the flux of translocations for both secreted
and cell mass proteins. In aerobic conditions, electrons removed from cysteine thiols for
disulfide bond formation are transferred to oxygen as the final electron acceptor (16,17),
resulting in the production of reactive oxygen species (ROS). In a previous study, we reported
that the oxygen uptake and ATP consumption were twice as high in the amylase producing strain
than the control strain, which we suggested to be a result of the increased oxidation in connection
with the electron transfer in ER redox pathways (18). However, it has stayed unclear what is
used as the final electron acceptor for protein folding under anaerobic conditions. *In vitro*
experiments suggest that under anaerobic conditions Ero1p of yeast *S. cerevisiae* could transfer
electrons to different types of exogenous acceptors, such as free FAD, yeast cytochrome b5 and bacterial azurin (19). Other species of either bacteria or eukarya that can live under anaerobic conditions use several alternative electron acceptors, as summarized in Table 1. For species that live in both aerobic and hypoxic conditions like the mussel *Geukensia demissa* and the lugworm *Arenicola marina*, it has been shown that they respire oxygen under aerobic conditions and switch to fumarate respiration when oxygen is limited (20,21).

In this study, we identified biological mechanisms in response to secretion of recombinant proteins at aerobic and anaerobic conditions. We identified common and specific cellular processes responding to increased loading of the protein secretory pathway, and concluded that anaerobic condition is more suitable for the overall protein processing. Combined with Reporter Metabolite analysis, quantification of overall carbon fluxes and physiological characterization allowed us to propose fumarate as the final electron acceptor at anaerobic conditions.

**Materials and methods**

**Strains and Media**

The reference strain NC, the amylase producing strain AAP and AAC were constructed, described and characterized in our previous study (22).
SD-2×SCAA media was prepared as previous (23): 20 g/L D-glucose, 6.7 g/L YNB (Yeast Nitrogen Base) without amino acids, 2 g/L KH$_2$PO$_4$ (pH 6.0 by NaOH), 1 g/L BSA (Bovine Serum Albumin), containing filter sterilized SCAA solution (190 mg/L Arginine, 108 mg/L Methionine, 52 mg/L Tyrosine, 290 mg/L Isoleucine, 440 mg/L Lysine, 200 mg/L Phenylalanine, 1260 mg/L Glutamic acid, 400 mg/L Aspartic acid, 380 mg/L Valine, 220 mg/L Threonine, 130 mg/L Glycine, 400 mg/L Leucine, 40 mg/L Tryptophan, 140 mg/L Histidine). During anaerobic conditions, the anaerobic growth factors (10 mg/L ergosterol and 420 mg/L Tween-80) were added into the medium (24).

**Fermentations**

Seed cultures were grown in shake flasks for 24h at 30 °C, 180 rpm, and inoculated into the fermentor at an initial OD ($A_{600}$) of 0.01. All fermentations were performed in DasGip 1.0-liter stirrer-pro vessels (Drescher Arnold Schneider, Germany) with a working volume of 500 ml of SD-2×SCAA media, at 30°C, 600 rpm agitation. Aerobic and anaerobic conditions were controlled by keeping the gas flow at 1 vvm (volume of flow per working volume per minute) with either air or nitrogen throughout the fermentations. In order to keep the cultivation fully anaerobic, 1 vvm of nitrogen was flushed through the fermentor overnight before inoculation. One drop of antifoam (Sigma, USA) was added to each fermentor. Dissolved oxygen was measured using a polarographic oxygen electrode (Mettler Toledo, Switzerland). The pH was maintained at 6.0 by the pH sensor (Mettler Toledo, Switzerland) using 2 M KOH. All fermentations were done in biological triplicates.

**Analytical methods**

One ml of the culture medium was centrifuged at 4000g for 5 min, and 800 μl of the culture supernatant was mixed with 100μl 0.1 M HCl and 5.5 mM NaN$_3$ final concentrations, and stored
at 4°C until measurement. Concentrations of glucose, fumarate, succinate, glycerol, ethanol, and acetate were analyzed by the Dionex Ultimate 3000 HPLC (Dionex Softron GmbH, Germany) with an Aminex HPX-87H column (BIORAD, USA) at 65 °C using 5 mM H₂SO₄ as the mobile phase at a flow rate of 0.6 ml/min. The activity of α-amylase was measured using the Ceralpha kit (Megazyme, Ireland) using α-amylase from A. oryzae (Sigma, USA) as standard, the activity of the standard amylase was calculated to be 69.6 U/mg amylase (22). For calculating the amylase yield on biomass, we plotted the amylase production against the biomass concentration for all the time points in the log phase, and calculated the yield as the slope of the data. The dry cell weight (DCW) was determined by filtering the cell culture through a 0.45 µm filter (Sartorius Stedim, Germany) and measuring the weight increase.

**Transcriptome analysis**

Samples for microarray analysis were taken as described previously and stored at -80 °C until processing (25). RNA was isolated using the RNeasy Minikit (Qiagen) and processed to cRNA using the Genechip 3’ IVT Express Kit (Affymetrix) and hybridized/ scanned on the Yeast Genome 2.0 Array (Affymetrix) to create CEL files. Images were analyzed using R 2.10.1 software and Bioconductor packages. Briefly, data normalization was carried out using the method of Probe Logarithmic Intensity Error (PLIER) with perfect match probe only (PM-only). The moderated t-statistic was used to identify differentially expressed genes between two conditions, comparing the two RPP strains (AAP and AAC) with the NC strain separately, in anaerobic and aerobic conditions respectively. One-way ANOVA was carried out to identify transcriptional responses between the anaerobic and aerobic conditions regardless of amylase productions. Benjamini–Hochberg’s method was used to adjust the p-values for multiple testing (FDR). PCA and hierarchical clustering analysis were performed to identify the general
transcriptional patterns among different conditions. Microarray data was submitted to the NCBI’s Gene Expression Omnibus database [Accession number: GSE38848]. Reviewer Access Link is also provided:


The FDR from the statistical analysis was used as input to the Reporter Features algorithm (26,27) to identify key biological features (Reporter GO-terms, Reporter Transcription Factors and Reporter Metabolites), whose neighboring genes in the corresponding biological network were significantly changed between two conditions. The algorithm was also run with a subset of the original biological network as input, containing the up-regulated or down-regulated genes only in order to identify the influence of the transcriptional changes on the different features in one direction (up- or down-regulated).

Results and discussion

Expression levels of α-amylase and oxygen levels affect protein secretion and cell growth

Three yeast strains producing different levels of amylase (22) were compared and evaluated under aerobic and anaerobic batch cultivations. These strains were named as follows: NC (negative control: S. cerevisiae CEN. PK 530-1C transformed with empty vector), AAP (CEN. PK 530-1C with amylase expression under TEF1 promoter and CYCI terminator) and AAC (CEN. PK 530-1C with amylase expression under TPII promoter and TPII terminator). As described previously, the strain CEN.PK530-1C has a deletion in the TPII gene that encodes for the glycolytic enzyme triose-phosphate isomerase, and all the vectors contain the POTI gene from Schizosaccharomyces pombe that encodes for the same enzyme (22). To ensure efficient secretion of amylase we used the alpha factor leader sequence, which has been found to result in more efficient secretion than other leader sequences (22). The physiological parameters are listed
in Supplementary Table S1 for all the strains grown at the two different conditions. The data suggests that there is a trade-off between amylase production and cell growth, as well as glycerol and ethanol production. In other words, strains with higher amylase production grew more slowly and produced more glycerol and less ethanol.

When comparing aerobic and anaerobic conditions, both AAP and AAC strains presented higher amylase production at anaerobic conditions (Fig. 1A): per unit of biomass the AAP strain produced 85% more amylase than under aerobic conditions and the AAC strain produced 3.3-fold more amylase than under aerobic conditions. Amylase production is also more efficient at anaerobic conditions (Fig. 1B): the productivity of AAP and AAC is 1-fold and 2-fold higher at anaerobic conditions compared to aerobic conditions, respectively. At anaerobic growth the AAC strain produces about 28 mg amylase/g biomass, and as the typical protein content of yeast is approximately 500 mg protein/g biomass this means that about 5.6% of all cellular protein produced is amylase. These data suggested that at anaerobic conditions the cellular regulations are more suitable for amylase production. To study transcriptional regulations occurring in anaerobic conditions while producing α-amylase, global transcriptome analysis was therefore performed to identify the possible molecular mechanisms.
Transcriptional responses in protein-producing strains in aerobic and anaerobic conditions

Microarray analysis of the amylase-secreting strains (AAC and AAP) and the control strain (NC) was carried out during both aerobic and anaerobic conditions. After data normalization, principal component analysis (PCA) was performed (Fig. S1). The data suggests that the global transcriptional response changed in the same direction at both conditions, which suggested that many pathways were altered regardless of the growth condition. In order to reduce the dimensionality of the data and to filter out global biological responses related to production of a heterologous protein, integrated analysis was performed for both aerobic and anaerobic conditions (details showed in Fig. S2 and S3).

Fig. 2 provides a summary of the common and specific changes in secretory pathway functions for the two production strains compared to the NC strain at anaerobic and aerobic growth conditions. Based on this analysis we found that anaerobic conditions provide a better RPP-producing environment than aerobic conditions, and the advantage level is suggested to be the ability to tolerate certain amount of RPP perturbation between 0.58 mg/g DCW/h (anAAP) and 5.5 mg/g DCW/h (anAAC), because i) anaerobic cultivations clearly showed a higher amylase production than at aerobic conditions (Fig. 1), ii) genes belong to ER functions and stress related responses were up-regulated when comparing AAC strain with NC at both aerobic and anaerobic conditions, iii) even though anAAP produced more amylase than aAAP, genes within key functions in the secretory pathway as well as many stress related pathways were only up-regulated in aAAP but not in anAAP when compared to the control strain.

Recombinant protein production at anaerobic conditions

To analyze whether anaerobic conditions provide a better α-amylase-producing environment than aerobic conditions in a protein production-independent manner, we also performed one-way
ANOVA to filter out global significant pathways between anaerobic and aerobic conditions regardless of production levels (Supplementary Fig. S4). Among the significant reporter GO-terms with a very high significance (p-value < 1E-06), we found that genes belong to ER and protein trafficking functions were up-regulated under anaerobic conditions, which suggested that it is the anaerobic condition rather than some protein-production-dependent transcriptional responses that provides up-regulations that allow for more efficient synthesis and secretion of α-amylase. Table 2 collects all significantly changed genes (FDR < 0.05) from the protein secretion related Reporter GO-terms. We found that some of these genes were also up-regulated when comparing both production strains to NC under aerobic conditions, for example, genes encoding proteins involved in protein folding (KAR2, SCJ1, EUG1 and ERV2) and ER associated degradation (UBC7). Some genes were also up-regulated when comparing both production strains to NC under anaerobic conditions, for example, genes encoding proteins involved in protein folding (SWP1 and DPM1) and protein export (SSO2 and SEC61). It has been reported that over-expression of some of these genes could increase recombinant protein production in yeast, for example, KAR2 (28), SCJ1 (29), EUG1 (30), DPM1 (31), SSO2 (32) or SEC61 (33). Our study increases the list of possible valuable metabolic engineering targets that were not found previously, such as the protein folding pathway gene ERV2 (34), the ERAD pathway gene UBC7 (35), the essential glycosylation gene SWP1 (36).

Genes of stress responses and cell redox homeostasis were also up-regulated under anaerobic conditions suggesting that the reason for the more efficient secretory pathway at anaerobic conditions might be a combination of several factors: i) at anaerobic conditions there is increased cell stress and there is therefore up-regulation of stress response genes and this allows for improved processing of secretory proteins; and ii) the anaerobic redox condition provides a
better environment for protein processing. We also found that besides the protein processing pathways, genes belonging to many lipid associated GO-terms were also significantly up-regulated at anaerobic conditions. Previous work has reported similar result when they compared anaerobic and aerobic effects under four different carbon-limited conditions, and demonstrated that lipid metabolism was up-regulated at anaerobic conditions in all conditions (37). Similar responses were also reported in Pichia pastoris where protein folding and trafficking pathways, lipid metabolism and stress responses were up-regulated at low oxygen conditions (38).

The three major functions of the ER are: i) storing and regulating the level of intracellular \( \text{Ca}^{2+} \), ii) protein folding and modifications and iii) lipid biosynthesis (39,40). Perturbations of the homeostatic state of lipids, especially saturated fatty acids can cause ER stress (41,42), and many UPR targets are genes in the lipid catabolism (43,44). It was also reported that regulation of glutathione can reduce cell stress caused by perturbation of either fatty acid oxidation or protein folding (45). Here, we performed hierarchical clustering of all significant genes belonging to protein processing in the ER, lipid synthesis and glutathione metabolism when comparing anaerobic to aerobic conditions (Fig. 3). We found that genes belong to these three functions were well-distributed into four different clusters, and also interesting to see that in each of these 4 clusters, there are genes related to glutathione metabolism, that has important roles for protein folding in the ER (18) and the cellular redox balance (46). Here, we suggest that there is close correlation between protein processing in the ER and lipid metabolism, possibly through the metabolism of glutathione. Since both protein secretion and lipid biosynthesis involve the ER, lipid associated stress might be one of the processes that caused anaerobic stress.
Electron acceptors for protein folding in anaerobic conditions

In order to propose a putative final electron acceptor for the protein folding in the ER at anaerobic conditions, we overlaid our data onto a genome-scale metabolic model using the Reporter Metabolite algorithm (27,26) and identified the key metabolites around which significant transcriptional changes occurred. The top 15 Reporter Metabolites for each strain when comparing anaerobic and aerobic conditions were clustered in Fig. 4. It is remarkable that the 11 common Reporter Metabolites for all three strains could be grouped into two clusters: 1) ATP, ADP, ferricytochrome/ferrocytochrome, orthophosphate and mitochondrial protons, which have close relations with the energy metabolism; and 2) fumarate, oxygen, FADH₂, FADH, ubiquinol and ubiquinone-9 which are related to electron transport, details shown in Table 3.

More precisely, FADI that is involved in FAD synthesis, FLCI that is responsible for FAD ER transport, and ERV2 that codes for flavin-bound thiol oxidase (34) for disulfide bond formation, were all up-regulated at anaerobic conditions. Under aerobic conditions the level of free FAD and total FAD were estimated to be ~3 µM and 15 µM, respectively, in wild type strains (47) and our data suggested that at anaerobic conditions the FAD synthesis was further up-regulated, which suggested that FAD might have important functions in the anaerobic metabolism. Indeed, it has been reported that all sulfhydryl oxidases and most disulfide reductases have flavin as essential cofactors (48,49). Depletion of riboflavin, the precursor of flavins, resulted in a severe defect in oxidative folding (16), whereas increasing cellular free FAD levels (50) could restore cell growth of the ero1 mutant. It was also reported that free FAD was essential for RNaseA refolding catalyzed by Ero1 and PDI (50), and therefore suggested that Ero1p might contain domains that work with free FAD (51). All these evidence demonstrated the important role of cellular free FAD levels on the protein folding in the ER. It has been reported that under
anaerobic conditions Ero1p could directly transfer electrons to free FAD (19). Here, we suggest that under anaerobic conditions, free FAD could act as the electron carrier that takes part in the electron transfer from Ero1p to the final electron acceptor during protein folding in the ER.

We further found that the fumarate reductase encoding OSM1 and FRD1, were up-regulated at anaerobic conditions. It was reported that a single deletion of either OSM1 or FRD1 does not affect the anaerobic cell growth (52), whereas a double deletion is lethal at anaerobic conditions but it has no growth effect at aerobic conditions (53). It is suggested that this essential role of fumarate reductase is because it catalyzes the only reaction that could oxidize free FADH$_2$ under anaerobic conditions (52). Here we suggest that the FAD after accepting electrons from the ER protein folding is then oxidized by the fumarate reductase. A model for electron transfer from the ER to fumarate is presented in Fig. 5.

There are two electron transferring pathways reported in the ER (Fig. 5): for disulfide bridge formation electrons pass through PDI to either Ero1p or Erv2p, that both can reduce free flavins (19). It has been further shown that over-expression of Erv2p can restore cell growth in an ero1 mutant both under aerobic (34) and anaerobic conditions (50). When comparing anaerobic to aerobic conditions, the expression of neither PDII nor ERO1 was changed, whereas ERV2 was up-regulated in all three strains (34). Instead of oxygen electrons are further transferred to free FAD, possibly in the following two routes: i) Since FAD could be transported across the ER membrane (51), electrons could be transferred to the free FAD in the ER lumen directly by the Ero1p bounded FADH$_2$ (19) and thereafter be exported to the cytosol; or ii) as Ero1p is closely associated with the ER membrane (54,55), electrons could be directly transferred from the membrane spanning part of Ero1p to free FAD in the cytosol. In the cytosol FADH$_2$ could either be oxidized when fumarate is converted to succinate by the cytosolic fumarate reductase Frd1p.
or it could be translocated to the mitochondrion and there get oxidized by the mitochondrial fumarate reductase Osm1p.

**Fumarate as the final electron acceptor in *S. cerevisiae***

Fumarate is reported to be the electron acceptor for the dihydroorotate oxidase Ura1p catalyzing reaction in the pyrimidine synthesis pathway in *S. cerevisiae* (56). This reaction converts dihydroorotate to orotate and at the same time ubiquinone is converted to ubiquinol. Interestingly, the genes *COQ5, COQ6* and *COQ9* that are related to the mitochondrial synthesis of ubiquinol were significantly up-regulated under anaerobic conditions (Fig. 5), which points to ubiquinol as the possible electron donor for fumarate.

In order to evaluate our hypothesis that fumarate may act as the final electron acceptor for both pyrimidine biosynthesis and for protein folding, the number of electrons generated and consumed at anaerobic conditions was calculated based on our experimental data, details are showed in Supplementary Table S1 and Fig. S5. If we assume that all electrons formed by the disulfide bridge formation and pyrimidine biosynthesis have fumarate as the final acceptor, the total amount of succinate formed would be about 0.11 mmol/g biomass. In our anaerobic experiments, the succinate production amount to about 0.22-0.41 mmol/g biomass, which corresponds to 2-4 fold of the theoretical calculation. In this context it is quite interesting to notice that the higher amylase producing strain AAC produces more succinate than AAP and NC, even though it has a lower biomass production. This could be explained by that high levels of heterologous protein production generate more cell stress, possibly higher ER stress (this could also be the reason why AAC grew slower at both aerobic and anaerobic conditions), and hence resulting in futile cycling of disulfide bond formation, and hence more electrons need to be consumed by the fumarate reductase.
Addition of fumarate promotes cell growth in anaerobic conditions

In a previous study, we demonstrated that when yeast produces amylase, the protein folding machinery easily get overloaded, protein folding cycles then go through futile redox cycles, which consume potentially limitless amounts of oxygen, the final electron acceptor in aerobic conditions (18). From the fermentation experiments, we found that the specific growth rate ($\mu$) is much lower at anaerobic condition for the high producing strain (AAC), Fig. 6. Above we hypothesize that fumarate could act as the electron acceptor, and since futile protein folding cycles might occur at anaerobic conditions, limited fumarate levels could explain this growth limitation. In order to test this hypothesis cell growth was assessed with the addition of 0.5 g/L fumarate and evaluated in aerobic and anaerobic conditions, and we found that indeed the growth of AAC increased by about 10% only at anaerobic conditions (Fig. 6). It is possible that fumarate is used as a carbon source, but fumarate addition at aerobic conditions did not show increased growth (aAAC$_f$, Fig. 6). Also, fumarate addition did not promote cell growth in the NC strain (with no amylase production), which further supported that hypothesis that only when the futile cycle of protein folding carries a high flux, does fumarate become limited for cell growth as a role of the electron acceptor.

However, the $\alpha$-amylase titer did not increase by fumarate addition, suggesting that addition of fumarate might mainly helped by recycling folding associated intermediates, which could help to produce amylase in a more efficient way, i.e. produce the same amount of amylase faster. There was still fumarate left by the end of the fermentation, which suggested that there are other limiting steps for anaerobic amylase productions besides electron transferring.
Conclusions

In this study, we investigated the mechanisms underlying increased α-amylase production and secretion in anaerobic conditions in yeast *S. cerevisiae*. By measuring and comparing the transcriptional responses, using Reporter GO-terms, Reporter Transcription Factors, Reporter Metabolites and flux analysis we conclude that several cellular pathways were regulated as results of protein production and secretion: (a) transcription and translation, (b) amino acid metabolism, (c) protein folding/ modification, (d) cytosolic redox control, (e) post-Golgi sorting, and (f) protein degradation. More specifically, genes related to the endoplasmic reticulum, lipid synthesis and stress responses were generally up-regulated at anaerobic conditions when comparing to aerobic conditions. Based on the integrative analysis we also provide a model for electron transfer and we propose the final electron acceptor to be fumarate in anaerobic condition. Our genome wide transcription data points to the significant up-regulation of FAD synthesis, mitochondrial ubiquinol synthesis and fumarate reductase at anaerobic conditions. Indeed, addition of small amounts of fumarate provided a significant growth improvement at anaerobic conditions, but not at aerobic conditions. In conclusion, we propose the use of integrated data analysis to generate new hypotheses for engineering that will further improve the design of cell factories for protein production and secretion.

Acknowledgments

This work is financially supported by the European Research Council ERC project INSYSBIO (Grant no. 247013), the Novo Nordisk Foundation, the Chalmers Foundation, and the Knut and Alice Wallenberg Foundation.
Reference


43. Tyra HM, Spitz DR, Rutkowski DT. 2012. Inhibition of fatty acid oxidation enhances oxidative protein folding and protects hepatocytes from endoplasmic reticulum stress. Molecular Biology of the Cell 23:811-819


Fig. 1. Amylase production during the glucose log phase of the fermentation.

(Grey bar) Aerobic conditions. (Black bar) Anaerobic conditions. (A) Amylase yield (B) Protein productivity. NC stands for the reference strain, AAP and AAC stand for amylase producing strains. Error bars are based on independent triplicates.
<table>
<thead>
<tr>
<th>Function</th>
<th>GO-terms</th>
<th>TFs</th>
<th>aAAC /aNC</th>
<th>aAAP /aNC</th>
<th>anAAC /anNC</th>
<th>anAAP /anNC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER processing</td>
<td>ER lumen</td>
<td>Hsf1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PDI activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein degradation</td>
<td>Vacuolar protein catabolic process</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stress</td>
<td>Response to Stress</td>
<td>Msn2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Man4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hog1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yap1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transcription/translation</td>
<td>Ribosome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Processome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nucleolus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rRNA processing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Translation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amino acid metabolism</td>
<td></td>
<td>Gcn4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 2.** Pathways that were significantly changed in α-amylase producing strains (AAC and AAP).

Under anaerobic (anAAC, anAAP or anNC) and aerobic conditions (aAAC, aAAP or aNC).

Reporter Gene Ontology (GO-terms) or Reporter Transcription Factors (TFs) were selected from figure S2 and figure S3. **(Red):** up-regulated. **(Green):** down-regulated.
**Fig. 3. Heatmap of significant genes when comparing anaerobic and aerobic conditions.**

Genes (one-way ANOVA p-value < 0.01 and logarithmic fold change > 0.5) belonging to protein processing in the ER, lipid metabolism and glutathione metabolism when comparing anaerobic and aerobic conditions. We found that most of the genes in each cluster related to stress response. For example, in cluster 1, *HSP26, ALE1* and *RNR3* for copper response (57), *ARE1, MPD1, HSP26* and *DER1* for UPR (44, 58); in cluster 2, genes involved in ethanol tolerance, *FAA1, ELO1, MRP1, ERG5* and *SOD2* (59-61); in cluster 3, *EUG1, PHS1, FAA4, ADH4, ERG1, ERG24* for ethanol tolerance (60, 62, 63), *RNR2* and *ERG1* for DNA damage response (64); and in cluster 4, *SOD1, ERG6, ERG25* and *IDP1* for DNA damage response (65-67).
Fig. 4. Top 15 reporter metabolites in the three strains when comparing anaerobic to aerobic conditions.

11 metabolites were commonly presented in all three comparisons (metabolites around which the most significant transcriptional changes occur). The upper cluster including Fumarate, Oxygen, FADH$_2$, FAD, Ubiquinol and Ubiquinone-9 is shown to be even more significant anaerobic/aerobic reporter metabolites in the two production strains (AAC and AAP) than in the wild type strain (NC), under anaerobic (anAAC, anAAP or anNC) and aerobic conditions (aAAC, aAAP or aNC).
Fig. 5. Proposed model for anaerobic electron transfer with fumarate as the final electron acceptor for protein folding.

For the disulfide bridge formation electrons pass through PDI to either Ero1 or Erv2. Then instead of oxygen, electrons are transferred through free FAD to the final electron acceptor, fumarate, either in the cytosol or in the mitochondrion. (Blue box): intracellular proteins and metabolites; (Red oval): up-regulated enzymes; (Green oval): down-regulated enzymes; (Grey oval): un-regulated enzymes; (Black line): metabolic pathways; (orange line): electron transferring pathways; (dashed line): alternative electron transfer reactions.
Fig. 6. Fumarate promotes cell growth in anaerobic conditions.

(White bar): Specific growth rate data of the NC strain. (Dot bar): Specific growth rate data of the NC strain cultured in SD-2×SCAA media with 0.5g/L fumarate. (Slash bar): Specific growth rate data of the AAP strain. (Grey bar): Specific growth rate data of the AAC strain. (Black bar): Specific growth rate data of the AAC strain cultured in SD-2×SCAA media with 0.5g/L fumarate. Error bars are based on independent triplicates except for the fumarate fermentations, which are based on independent duplicates.
Table 1. Alternative final electron acceptors in different species that grow anaerobically

<table>
<thead>
<tr>
<th>Species</th>
<th>Acceptor</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>Fumarate</td>
<td>(68)</td>
</tr>
<tr>
<td><em>Veillonella parvula</em></td>
<td>Fumarate</td>
<td>(69)</td>
</tr>
<tr>
<td><em>Wolinella succinogenes</em></td>
<td>Fumarate, Nitrate</td>
<td>(69,70)</td>
</tr>
<tr>
<td>Sulfate reducing bacteria</td>
<td>Sulfite, Sulfur</td>
<td>(71)</td>
</tr>
<tr>
<td><em>Sporomusa acidovorans</em></td>
<td>CO₂</td>
<td>(72)</td>
</tr>
<tr>
<td><em>Rhodopseudomonas capsulata</em></td>
<td>DMSO</td>
<td>(73)</td>
</tr>
<tr>
<td>Eukarya</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Geukensia demissa</em></td>
<td>Fumarate</td>
<td>(21)</td>
</tr>
<tr>
<td><em>Arenicola. marina</em></td>
<td>Fumarate</td>
<td>(20)</td>
</tr>
<tr>
<td>Ciliates</td>
<td>Nitrite, Nitrate</td>
<td>(74)</td>
</tr>
<tr>
<td>Fungi</td>
<td>Nitrite, Nitrate</td>
<td>(75)</td>
</tr>
</tbody>
</table>
Table 2. Significantly regulated genes in all strains as a function of anaerobic /aerobic conditions

<table>
<thead>
<tr>
<th>GO term</th>
<th>function</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER</td>
<td>Translocation</td>
<td>KAR2 SEC62 SEC66 YPT10 WSC4 SBH1 SSA3 (SEC61)</td>
</tr>
<tr>
<td>Folding</td>
<td></td>
<td>KAR2 SCJ1 LHS1 HSP26 SSA3 EUG1 ERV2 SSE2 (YDJ1)</td>
</tr>
<tr>
<td>Glycosylation</td>
<td></td>
<td>SWP1 DPM1 OST3 PMT2 MNN1 MNT2 PMT3 PMT5 KTR1 KRE2 (OST4 DIE2 RFT1 SEC20)</td>
</tr>
<tr>
<td>ERAD</td>
<td></td>
<td>UBC7 DER1 DFMI CUE1</td>
</tr>
<tr>
<td>Anchored to membrane</td>
<td>General</td>
<td>SPO20 SYN8</td>
</tr>
<tr>
<td>ER to Golgi</td>
<td>SEC23 SEC20 BOS1 SEC20 BOS1 SEC22 TRX1 (USE1)</td>
<td></td>
</tr>
<tr>
<td>Protein export</td>
<td>KAR2 SSO2 SEC61 SRP102 SBH1 SPC3 SPC2 SPC1 SEC9 SEC11 SEC62 (SNC2 OXA1)</td>
<td></td>
</tr>
</tbody>
</table>

All genes were significantly up-regulated (value) in anaerobic conditions, except for genes in the brackets. Genes with FDR lower than 0.05 were selected.
Table 3. Reporter Metabolites: Significantly changed genes (FDR<0.05) as a function of anaerobic/aerobic conditions

<table>
<thead>
<tr>
<th>Reporter Metabolite</th>
<th>Genes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fumarate</td>
<td><strong>FRD1</strong> Fumarate reductase</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>OSM1</strong> Fumarate reductase</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>FUM1</strong> Fumarase</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>SFC1</strong> Mitochondrial succinate-fumarate transporter</td>
<td></td>
</tr>
<tr>
<td>FAD/ FADH$_2$</td>
<td><strong>FAD1</strong> FAD synthesis</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>FLC1</strong> FAD ER transporter</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>ERV2</strong> Disulfide bond formation</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>SDH3</strong> succinate dehydrogenase</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>FLX1</strong> FAD transporter</td>
<td></td>
</tr>
<tr>
<td>Ubiquinol/</td>
<td><strong>URA1</strong> Pyrimidine synthesis</td>
<td></td>
</tr>
<tr>
<td>Ubiquinone</td>
<td><strong>COQ1</strong> Ubiquinone synthesis</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>COQ2</strong> Ubiquinone synthesis</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>COQ3</strong> Ubiquinone synthesis</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>COQ4</strong> Ubiquinone synthesis</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>COQ5</strong> Ubiquinone synthesis</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>COQ6</strong> Ubiquinone synthesis</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>COQ9</strong> Ubiquinone synthesis</td>
<td></td>
</tr>
</tbody>
</table>

Bold genes were significantly up-regulated (value) in anaerobic conditions
PAPER V
Correlation of cell growth and heterologous protein production in

*Saccharomyces cerevisiae*

Zihe Liu¹, Jin Hou¹,², José L. Martínez¹, Dina Petranovic¹, Jens Nielsen¹,³, *

¹ Novo Nordisk Foundation Center for Biosustainability,
Department of Chemical and Biological Engineering,
Chalmers University of Technology, SE-41296 Göteborg, Sweden

² Present address: State Key Laboratory of Microbial Technology, Shandong University, 27 Shanda Nan Road, Jinan, Shandong, 250100, China

³ Novo Nordisk Foundation Center for Biosustainability,
Technical University of Denmark, Fremtidsvej 3, DK-2970 Hørsholm

* Corresponding author:

Telephone: +46 (0)31 772 3804,
Fax: +46 (0)31 772 3801,
E-mail: nielsenj@chalmers.se
Abstract

With the increasing demand for biopharmaceutical proteins and industrial enzymes it is necessary to optimize the production by microbial fermentation or cell cultures. Yeasts are well established for the production of a wide range of recombinant proteins but there are also some limitations, e.g metabolic and cellular stresses have a strong impact on recombinant protein production. In this work we investigated the effect of the specific growth rate on production of two different recombinant proteins. Our results show that human insulin precursor is produced in a growth associated manner whereas α-amylase tends to gain higher production levels at low specific growth rates. Based on transcriptional analysis, we found that the difference in production of the two proteins as function of the specific growth rate is mainly due to differences in ER processing, protein turnover, cell cycle, and global stress response. We also found that there is a shift at a specific growth rate of 0.1 h⁻¹ that influences protein production. Thus, for lower specific growth rates the α-amylase and insulin precursor producing strains present similar cell responses and phenotypes, whereas for higher specific growth rates the two strains respond differently to changes in the specific growth rate.

Keywords: Recombinant protein production, Chemostat, Insulin Precursor, Amylase, Yeast
Introduction

The increasing demand for industrial enzymes and biopharmaceutical proteins calls for robust production hosts ensuring high yield and productivity. There are now over 300 biopharmaceuticals proteins and antibodies on the market with annual sales exceeding USD100 billion (Langer 2012) and with an annual growth of about 19% (Schröder 2008).

Yeasts are well established for the production of a wide range of recombinant proteins, due to their rapid growth, robustness to industrial conditions, and more importantly their ability to perform post-translational modifications including proteolytic processing of signal peptides, disulfide bond formation, subunit assembly, acetylation, glycosylation, phosphorylation and secretion of heterologous proteins in their native forms (Hou et al. 2012b; Freigassner et al. 2009). However, despite all these advantages, recombinant protein production (RPP) in *Saccharomyces cerevisiae* is not optimal. The main limitations include ER misfolding, hyper-glycosylation, and inefficient trafficking, etc. Therefore, a wide range of studies have been implemented to engineer the host, and engineering of the yeast secretory pathway is typically done in these studies (Idiris et al. 2010; Freigassner et al. 2009).

The investigation of recombinant protein production at varied specific growth rates is relevant for both basic and applied biology. Expression of recombinant proteins often triggers many physiological changes, resulting in metabolic burden and reduction of cell growth and protein production (Dürrschmid et al. 2008). Generally, this can be avoided by slowing down the transcription and translation, to balance these with translocation and assembly in the secretory pathway (Freigassner et al. 2009). It was reported that in *S. cerevisiae* the highest yield of β-galactosidase was obtained at the lowest specific growth rate (Hardjito et al. 1993), and that the specific activity of cutinase decreases with increasing specific growth rates (Ferreira et al. 2003; Verripsab et al. 2000). Similar results have been reported in *Escherichia coli* also, namely that the yields of glyceraldehyde-3-phosphate dehydrogenase (Nancib and
Boudrant 1992) and β-lactamase (Fu et al. 1993) were higher at low dilution rates. The production of heterologous proteins has been reported as either growth associated (reaching a plateau at high specific growth rate) or inversely associated with growth (Andersen and Krummen 2002), depending on the cell lines, the properties of the proteins and expression (promoters and enhancers) (Lunter and Goodson 2011). In *Pichia pastoris*, the secretion of recombinant proteins was reported to be coupled to the specific growth rate (Buchetics et al. 2011), whereas in *Trichoderma reesei*, the protein production rate was negatively correlated with growth (Arvas et al. 2011). Our previous study revealed that the heterologous α-amylase yield on biomass increased 2-fold in fed-batch cultivations (0.08 h\(^{-1}\) feed rate) compared with batch cultivation, whereas that of insulin precursor (IP) did not increase in fed-batch cultivations compared with batch cultivations. Therefore, production of hterologous proteins could be coupled differently with cell growth even when the same expression system and host strain is used.

The general approach to study the kinetics of protein production in response to changes in the specific growth rate is through the use of chemostat cultivations (Arvas et al. 2011). Chemostat cultures generally represent industrial processes better, as they are operated with a low glucose concentration and reduced specific growth rate, which represent more similar conditions than found in the typical operation of a fed-batch process, by far the preferred industrial operation, than found in a batch culture (Seresht et al. 2011). Furthermore, the kinetics of protein production obtained from chemostat cultures can be used directly for design of a fed-batch process. Comprehensive high-throughput analyses have been applied to analyze the effect of cell growth on the overall cellular physiology (Pincus et al. 2010; Nurse 2003; Regenberg et al. 2006), and this has revealed a large influence of the specific growth rate on yeast metabolism. Castrillo et al. (Castrillo et al. 2007) analyzed the impact of the specific growth rate at the transcriptome, proteome and metabolome levels and identified
around 900 genes that are regulated by growth. Fazio et al. (Fazio et al. 2008) carried out a three factor design study to identify growth rate dependent genes, and showed that the specific growth rate had a positive correlation with ATP producing and consuming pathways, cell cycle regulation, protein biosynthesis, ribosome biogenesis and translation process, whereas the response to stress, proteins involved in ER-associated protein catabolism, and ubiquitin and proteasome activity were down-regulated with increasing growth rate. To date, few studies have focused on using systems biology (Kim et al. 2012) to study heterologous protein secretion related to cell growth and metabolism.

In this study, we performed carbon-limited chemostat cultivations by culturing two recombinant protein producing strains, in the same host and with the same expression system: a strain producing IP and a strain producing α-amylase. The specific growth rate was controlled at 0.05, 0.1 and 0.2 h⁻¹ and through integrated analysis (Oliveira et al. 2007), we identified protein specific and growth rate specific processes that affect heterologous protein secretion.

**Materials and Methods**

*Strains and Media*

Strains used in this study were named as follows: AAC (CEN. PK 530-1C with amylase expression) (Liu et al. 2012), AIC (CEN. PK 530-1C with insulin expression) (Liu et al. 2012). The strain CEN.PK530-1C has a deletion in the *TPII* gene that encodes for the glycolytic enzyme triose-phosphate isomerase, and all the vectors contain the *POTI* gene that in *Schizosaccharomyces pombe* encode for the same enzyme (Liu et al. 2012). To ensure efficient secretion of amylase and IP, we used the alpha factor leader sequence and the *TPII* promoter, which has been found to result in an high level protein secretion (Liu et al. 2012).
SD-2×SCAA medium was prepared as previously (Hackel et al. 2006; Tyo et al. 2012): 10 g/L glucose, 6.7 g/L Yeast Nitrogen Base without amino acids, 2 g/L KH2PO4 (pH 6.0 by NaOH), 1 g/L BSA (Bovine Serum Albumin), containing filter sterilized SCAA solution (190 mg/L Arginine, 108 mg/L Methionine, 52 mg/L Tyrosine, 290 mg/L Isoleucine, 440 mg/L Lysine, 200 mg/L Phenylalanine, 1260 mg/L Glutamic acid, 400 mg/L Aspartic acid, 380 mg/L Valine, 220 mg/L Threonine, 130 mg/L Glycine, 400 mg/L Leucine, 40 mg/L Tryptophan, 140 mg/L Histidine).

Fermentations

Seed cultures were grown in 100 ml shake flasks containing 20 ml SD-2×SCAA medium for 24h at 30°C, 180 rpm, and inoculated into the fermentor at an initial OD (A_{600}) of 0.01. All fermentations were performed in DasGip 1.0-liter stirrer-pro vessels (Drescher Arnold Schneider, Germany) with a constant working volume of 400ml, at 30°C, 600 rpm agitation and dilution rates of 0.05 and 0.1 h^{-1}. Aerobic conditions were controlled by keeping flowing 1 vvm (volume of flow per working volume per minute) of air, and the concentration of dissolved oxygen was measured using a polarographic oxygen electrode (Mettler Toledo, Switzerland). The pH was maintained at 6.0 by the pH sensor (Mettler Toledo, Switzerland) using 2 M KOH. Stable chemostat cultivation was reached when at least five residence times had passed since starting the continuous cultivation. All fermentations were done in biological triplicates.

Analytical methods

1 ml of the culture broth was centrifuged at 4000g for 5 min, and 800 μl of the culture supernatant was mixed with 100 μl 0.1 M HCl and 5.5 mM NaN₃ final concentrations for insulin measurement, and with 100 μl 5.5 mM NaN₃ final concentrations for amylase measurement. Concentrations of glucose, glycerol, ethanol, and acetate were analyzed using a
Dionex Ultimate 3000 HPLC (Dionex Softron GmbH, Germany) system with an Aminex HPX-87H column (BIORAD, USA). The dry cell weight (DCW) was acquired by filtering the cell culture through a 0.45 µm filter (Sartorius Stedim, Germany). The activity of α-amylase was measured using the Ceralpha kit (Megazyme, Ireland) using α-amylase from *A. oryzae* (Sigma, USA) as standard (Liu et al. 2012). The insulin was measured by HPLC using a Luna 5μ C18(2) (250mm x 4.6 mm) (Phenomenex) column and gradient-based elution as described in (Tyo et al. 2012), human insulin was used as a standard (Sigma).

**Transcriptome analysis**

Samples for microarray were carried out as described previously (Tyo et al. 2012). Images were analyzed using R 2.10.1 software. The microarray data was submitted to the NCBI GEO database (accession number: GSE40934) and Reviewer access link: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=xxonpkomccwoexs&acc=GSE40934. The FDR from the statistical analysis was used as input to the Reporter Features algorithm (Oliveira et al. 2008; Patil and Nielsen 2005) to identify key biological features (Reporter KEGG pathway and Reporter TFs). The algorithm was run with a subset of the original biological network as input, containing the significant changed genes in order to identify the influence of the transcriptional changes on the different features in one direction.

**Relative transcript levels determination by qPCR**

cDNA was synthesized by adding 100 ng of total RNA to a final RT reaction volume of 20 µl, and 2 µl of the cDNA were used as template with the Brilliant III Ultra fast SYBRGreen QPCR Master mix (Stratagene) in a Mx3005P QPCR System (Agilent Technologies). Cycle thresholds (Ct) were normalized and gene expression calculated relative to *S. cerevisiae ACT1* expression levels.
Results

Recombinant protein production is different at different dilution rates

To study the effect of specific growth rate on the kinetics of heterologous protein production, the previously constructed α-amylase producing strain (AAC) and the insulin producing strain (AIC) (Liu et al. 2012) were compared and evaluated using carbon-limited chemostat cultivations operated at different dilution rates. The specific growth rate of the recombinant protein production strains in batch cultivations was around 0.25 h\(^{-1}\) (Liu et al. submitted for publication), and the dilution rate of the chemostat cultivations was therefore controlled as 0.05, 0.1 and 0.2 h\(^{-1}\). Samples from the chemostat cultures were analyzed for the concentration of glucose, ethanol, glycerol, acetate, biomass and recombinant protein. Based on these measurements the specific glucose uptake rate and the yield coefficients for different metabolites were calculated and the results are collected in Table 1 for both strains grown at the different conditions. The biomass yield on glucose decreased significantly at higher specific growth rates, indicating high energy requirement for protein production.

The specific productivity of the two recombinant proteins for the different dilution rates is shown in Figure 1 together with the biomass concentration. We found that although the productivities of both proteins increases with increasing specific growth rates, the yield of IP on biomass resulted in a sharper increase at specific rate of 0.2 h\(^{-1}\), whereas the effect was not so strong on the amylase. On the other hand, the yield of amylase on substrate decreased at higher specific growth rates (with the highest value at specific rate of 0.05 h\(^{-1}\)), whereas the yields of IP on substrates were comparable at different specific growth rates. In order to determine whether these results were based on changes in the transcription of the recombinant gene, qPCR experiments were carried out for both the α-amylase and insulin precursor genes. The relative transcript levels of IP were higher than those of amylase, and whereas the transcript level for the amylase encoding gene did not change significantly with
specific growth rate there was a small increase in the transcript level of the IP encoding gene (Figure 2).

**Genome-wide Transcription analysis of protein producing strains, at different dilution rates**

Genome wide transcription analysis was performed in order to further dissect growth effects (at 0.05, 0.1 and 0.2 h\(^{-1}\)) and the effects of recombinant protein production (RPP). The expression of many genes changed between different dilution rates, clearly showing that the growth effect had a much bigger effect than the protein production effect. The expression levels of 1716 and 1331 genes commonly changed in both protein producing strains when comparing different dilution rates, whereas only 60 genes transcriptionally changed when comparing IP and amylase producing strains at all three different dilution rates (Figure 3).

**Cell cycle regulation and ER functions have a positive correlation with both the growth rate and protein productivity**

In order to study the effects of the specific growth rate on heterologous protein production, transcriptomes of different dilution rates were compared for both strains. Reporter KEGG pathway analysis showed that N- and O-link glycosylation processes were up-regulated in both production strains, at higher dilution rate conditions (Figure S1) and genes related to the protein processing in the endoplasmic reticulum (ER) were significantly up-regulated when comparing 0.2 to 0.1 h\(^{-1}\). Clearly ER functions were up-regulated at higher dilution rates, because more proteins were targeted into the ER. In order to identify different roles of ER functions at the different dilution rates, genes associated with protein processing in the ER are listed in Table 2. We found that genes related to the unfolded protein response (UPR) were up-regulated in both production strains at high dilution rates, and more importantly, the *HAC1* gene, which encodes the transcription factor that initiates UPR was also up-regulated.
The information in Table 2 indicates that when expressing recombinant proteins (IP and α-amylase in our case), the UPR level was activated for increasing specific growth rates, and when cells approached high specific growth rates (close to their maximum specific growth rate) they activate what could be called a super-UPR, where not only associated genes are up-regulated due to activation of Hac1, but *HAC1* is also itself transcriptionally up-regulated (Bernales et al. 2006).

*Proteasome and stress response have close relation with low growth rate and protein yield on substrates*

Reporter Transcription Factors (TFs) (Oliveira et al. 2008; Patil and Nielsen 2005) showed that, Figure S2, genes regulated by the general stress response, Msn2 and Msn4 (Moye-Rowley 2002; Causton et al. 2001), were expressed higher at the lowest dilution rate (D=0.05 h\(^{-1}\)), where the highest yields of both recombinant proteins on substrates were obtained. Similar result was found from Reporter KEGG pathway analysis, Figure S1, i.e. genes related to proteasome were highly expressed at the lowest dilution rate (D=0.05 h\(^{-1}\)). Since many stress regulated genes also are associate with protein processing, and protein degradation and turnover (Hatahet and Ruddock 2009; Haynes et al. 2004), the slow growth conditions may ensure that cells allocate sufficient resources for protein production and at the same time the moderate stress response ensures efficient post-translational processing. Similar experiments have been performed in *E. coli*, showing that inducing stress associated proteins could benefit protein folding (Thomas and Baneyx 2000; Hoffmann and Rinas 2000; Gill et al. 2001).

**Discussion**

Here we found that the productivity and yield (either on biomass or on substrate) of the recombinant proteins change at different specific growth rates, even though the transcription remains similar. In a previous study, we found that amylase and IP showed different
production kinetics when cultivated in diauxic batch fermentation with glucose as the carbon source: amylase was produced at much higher rate in the ethanol phase, whereas the production rate of IP dropped substantially after the diauxic shift (Liu et al. 2012). Furthermore, we found that the amylase yield on biomass increased more than 2 fold in fed-batch cultivations (0.08 h⁻¹ feed rate) compared to batch cultivation, whereas the IP yield on biomass did not show a clear difference (Hou et al. 2012a). We also found that amylase productivity was higher at anaerobic conditions compared with aerobic conditions (Liu et al. submitted for publication). All this taken together suggests that IP is produced in a growth associated manner, whereas amylase tends to gain higher production levels at lower cell growth conditions.

Based on the transcriptome data, we suggest that growth effects on recombinant protein production mainly rely on ER functions, stress responses and proteasome activities, as summarized in Figure 4a. The specific growth rate of 0.1 h⁻¹ appeared to be a shifting point between growth effects and protein production effects. At low specific growth rates (0.05-0.1 h⁻¹), the growth effects play the main role based on the nutrient-dependent stress and proteasome responses, whereas at higher specific growth rates (0.1-0.2 h⁻¹), the protein production starts to play the main role, as indicated by i) genes regulated by the general stress transcription factors (Msn2 and Msn4) showed similar expression levels indicating that when the specific growth rate increased above 0.1 h⁻¹ nutrient starvation response was deactivated; ii) genes related to oxidative stress (Yap1) were up-regulated when comparing growth rates of 0.2 to 0.1 h⁻¹; iii) more importantly, the super-UPR was activated at high specific growth rates, which might positively cause the up-regulation of genes associated with proteasome and protein processing in the ER.

In order to further identify the effects of protein production, we also performed reporter feature analysis where we compared the two strains grown at the three different dilution rates
(Figures S3 and S4). As summarized in Figure 4b, since α-amylase is a bigger protein compared to IP (478 vs 50 amino acids), even though it is produced in a moderate level, the amino acid metabolism and energy metabolism were up-regulated in the amylase producing strains. The effect of RPP in both strains also showed a shift around the dilution rate of 0.1 h⁻¹.

At low dilution rates (0.05 and 0.1 h⁻¹), where IP was produced at higher levels than α-amylase, the ER functions and oxidative stress related genes were expressed at higher levels compared to the α-amylase producing strains. On the other hand, at high dilution rates (0.2 h⁻¹) the α-amylase producing strain showed more induction of cell stress genes even though at this dilution rate there was lower protein production than in the IP producing strain. This is illustrated by the expression levels of genes related to Sod1, a cytosolic superoxide dismutase responsive to presence of reactive oxygen species (ROS) and cell aging. At the dilution rate 0.05 h⁻¹, where IP is produced at higher levels, the expression of these genes were higher in the IP producing strain, and at the dilution rate 0.1 h⁻¹ the expression level was about the same in the two protein producing strains, whereas at the dilution rate of 0.2 h⁻¹ the expression level of the Sod1 genes were higher in the α-amylase producing strain. Thus, even though IP was produced at a much higher level than amylase at the highest dilution rate, the α-amylase may become misfolded resulting in increased production of ROS and causing oxidative stress leading to up-regulation of the Sod1 genes. Taken together, amylase needs increased folding capacity and at lower specific growth rates (including the ethanol growth phase in batch fermentations) the secretory pathway may have additional capacity to handle the recombinant protein production, and in particular ensure protein turnover of misfolded proteins, whereas when the cells are growing fast there is not sufficient capacity to handle the additional burden imposed by α-amylase production, and this results in increased cellular stress.
In conclusion, our experiments report the effect of cell growth on recombinant protein production and demonstrate that the production of IP is positively correlated with growth, whereas the production of α-amylase is negatively correlated with growth. When expressing recombinant proteins, the UPR was activated with increasing specific growth rates, and the super-UPR might also be activated when approaching the cells maximum specific growth rate. The growth rate impact was protein specific and fermentation optimization should take into account the properties of produced proteins.

Acknowledgments

This work is financially supported by the European Research Council ERC project INSYSBIO (Grant no. 247013), the Novo Nordisk Foundation, the Chalmers Foundation, and the Knut and Alice Wallenberg Foundation.

Reference


Table 1. Physiological characterizations of engineered strains.

<table>
<thead>
<tr>
<th>D (h⁻¹)</th>
<th>Strains</th>
<th>$r_s$</th>
<th>$^{a}Y_{SX}$</th>
<th>$^{b}Y_{SG}$</th>
<th>$^{c}Y_{SA}$</th>
<th>$^{d}Y_{SE}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>AAC</td>
<td>0.096±0.004 0.52±0.02</td>
<td>n.d.</td>
<td>n.d.</td>
<td>n.d.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.111±0.001 0.45±0.01</td>
<td>0.012±0.001</td>
<td>0.005±0.001</td>
<td>n.d.</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>AAC</td>
<td>0.259±0.008 0.39±0.01</td>
<td>0.002±0.001</td>
<td>0.153±0.004</td>
<td>0.009±0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.245±0.012 0.41±0.02</td>
<td>0.023±0.003</td>
<td>0.126±0.018</td>
<td>0.013±0.002</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>AAC</td>
<td>0.649±0.054 0.31±0.02</td>
<td>0.115±0.011</td>
<td>0.077±0.001</td>
<td>0.173±0.005</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.645±0.018 0.31±0.01</td>
<td>0.051±0.007</td>
<td>0.070±0.004</td>
<td>0.175±0.004</td>
<td></td>
</tr>
</tbody>
</table>

AAC strands for amylase producing strain, and AIC stands for insulin producing strain. The specific glucose uptake rate is given as (g/g DW/h) ($r_s$). Yields (g/g glucose) calculated here only consider the steady state and the total consumed substrate. $^a$Biomass, $^b$Glycerol, $^c$Acetate, $^d$Ethanol. The data represent biological triplicates.
Table 2. Genes associated with protein processing in the ER that were transcriptionally changed at different dilution rates. Comparison between strain AAC and strain AIC.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Unfolded protein response</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common</td>
<td></td>
<td><strong>SSSI, DGK1, OST1, SIL1</strong></td>
</tr>
<tr>
<td>D0.1/D0.05</td>
<td><strong>WSC4, MCD4, LHS1, ERD2, PMT2</strong></td>
<td><strong>HMX1, MNT3, ALG1, ALG5, GPI8, GET1, OST4, OST5, SBH2, VRG4, PMT6, KRE27, PER1, TRS33, SRP21, SNL1, GSF2, ALG12, UBC6, FES1, SSE2, SSAR, RRT12, SSA3</strong></td>
</tr>
<tr>
<td>D0.2/D0.1</td>
<td><strong>HAC1, PDI1, KAR2, RSE1, UIL1, SCJ1, ERV25, ERV29, UBC7, PCM1, SEC24, SEC27, COS8, SFB2, PMT3, PMT5, DCR2, YIP3, DOG2, ALG6, ALG7, KTR1, SPF1, WSC4</strong></td>
<td><strong>CDC48, SEC23, FPR2, KEG1, HLJ1, GET3, CWH41, STT3, ERD1, EMP24, SAR1, KRE11, SRP101, SEC21, SVP26, MSC7, SHE3, SSM4, MID1, CSG2, OST4, BST1, USO1, SEC39, EPS1, ZRG17, HSP26</strong></td>
</tr>
</tbody>
</table>

**Bold:** genes up-regulated genes in each comparison, standard: genes down regulated in each comparison.
Figure Legends

Fig. 1. The heterologous protein production in chemostat cultivations. a) The productivity of amylase and insulin precursor at different growth rates. (Square), insulin producing strain AIC. (Circle), amylase producing strain AAC. b) Final biomass production of AAC and AIC at different growth rates. (Black bar), 0.05 h\(^{-1}\) dilution rate. (Grey bar), 0.1 h\(^{-1}\) dilution rate. (blank bar), 0.2 h\(^{-1}\) dilution rate.

Fig. 2. Relative expression levels of insulin precursor and α-amylase genes at different dilution rates. (Square), insulin producing strain AIC. (Circle), amylase producing strain AAC. Transcript levels were determined by qPCR, from independent biological duplicates.

Fig. 3. Analysis of the genome wide transcription data. a) Comparison of different dilution rates for both protein producing strains FDR<0.001. b) Comparison of amylase and insulin strains at three different dilution rates, FDR<0.01. AAC strands for amylase producing strain, and AIC stands for insulin producing strain.

Fig. 4. Common and distinct pathways that were significantly changed in growth-dependent and protein production-dependent manners at different dilution rates. a) When comparing each strain at different dilution rates. b) When comparing the two strains grown at the three different dilution rates. Reporter KEGG pathways and Reporter TFs were selected from figure 3, 4, S3 and S4. (Red), up-regulated. (Green), down-regulated. (blue), both up- and down-regulated.
**Fig. 1**

- **a**
  - Graph showing productivity (µmol/g DCW/h) vs. dilution rate (/h).
  - Productivity increases significantly with increasing dilution rate.

- **b**
  - Bar graph showing final biomass (g/L) for AAC and AIC.
  - AAC has a higher biomass than AIC at most dilution rates.

**Fig. 2**

- Graph showing gene expression (Fold Change) vs. dilution rate (/h).
  - Gene expression increases with increasing dilution rate for both AAC and AIC.
  - The increase is more pronounced for AAC compared to AIC.
Fig. 3

(a) Venn diagram with numbers indicating the overlap between AAC and AIC.

(b) Venn diagram with numbers indicating the overlap between AAC and AIC.

Fig. 4

(a) Table showing the functions, KEGG pathways, and transcription factors (TFs) with AAC/AIC ratios.

(b) Table showing the functions, KEGG pathways, and TFs with AAC/AIC ratios.

Legend for Fig. 3:
- AAC_0.1/AAC_0.05
- AIC_0.1/AIC_0.05
- AAC_0.2/AAC_0.1
- AIC_0.2/AIC_0.1

Legend for Fig. 4:
- AAC/AI 0.05
- AAC/AI 0.1
- AAC/AI 0.2
- AAC/AI 0.1
PAPER VI

Diagram:

- UPR Effect
- Host Engineering
- Random Mutation
- RPP IP and amylase
- Expression Design
- Leader
- Promoter
- Marker
- Aeration
- Cell growth
- Fermentation Analysis
MINIREVIEW

Metabolic engineering of recombinant protein secretion by Saccharomyces cerevisiae

Jin Hou1,2, Keith E.J. Tyo1,3, Zihe Liu1, Dina Petranovic1 & Jens Nielsen1

1Department of Chemical and Biological Engineering, Chalmers University of Technology, Göteborg, Sweden; 2State Key Laboratory of Microbial Technology, Shandong University, Jinan, China; and 3Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, USA

Correspondence: Jens Nielsen, Department of Chemical and Biological Engineering, Chalmers University of Technology, Kemivägen 10, SE-412 96 Göteborg, Sweden. Tel.: +46 31 772 3804; fax: +46 31 772 3801; e-mail: nielsenj@chalmers.se

Received 16 January 2012; revised 19 April 2012; accepted 22 April 2012.

DOI: 10.1111/j.1567-1364.2012.00810.x

Editor: Hyun Ah Kang

Keywords
protein secretion; systems biology; yeast; genetic engineering.

Abstract

The yeast Saccharomyces cerevisiae is a widely used cell factory for the production of fuels and chemicals, and it is also provides a platform for the production of many heterologous proteins of medical or industrial interest. Therefore, many studies have focused on metabolic engineering S. cerevisiae to improve the recombinant protein production, and with the development of systems biology, it is interesting to see how this approach can be applied both to gain further insight into protein production and secretion and to further engineer the cell for improved production of valuable proteins. In this review, the protein post-translational modification such as folding, trafficking, and secretion, steps that are traditionally studied in isolation will here be described in the context of the whole system of protein secretion. Furthermore, examples of engineering secretion pathways, high-throughput screening and systems biology applications of studying protein production and secretion are also given to show how the protein production can be improved by different approaches. The objective of the review is to describe individual biological processes in the context of the larger, complex protein synthesis network.

Introduction

The introduction of genetic engineering in the 1970s resulted in the establishment of an efficient biotech industry with one of the foci being the production of recombinant proteins for therapeutic use. Today more than 50 pharmaceutical proteins are being produced using recombinant technologies, and many of these are blockbuster pharmaceuticals (Walsh, 2010). These recombinant proteins can be produced using a range of different cell factories, including bacteria, yeast, filamentous fungi, insect cells, and mammalian cells. Many different yeast and fungal systems have been compared for protein production for pharmaceutical, food, and other industries (Madzak et al., 2004; Porro et al., 2005). At the industrial level, there is a consolidation in the choice of cell factory, so most of the production is achieved in Escherichia coli, Pichia pastoris, Saccharomyces cerevisiae, and Chinese hamster ovary cells (CHO cells). This consolidation that provides a limited number of general production platforms, allows faster optimization and scale up of protein production by the given cell factory. Furthermore, engineering of these microorganisms is driven by the desire to improve productivity and the ability to produce new products with optimal pharmacokinetic properties, for example, strains of P. pastoris that can produce proteins with human glycosylation patterns (Gerngross, 2004; Hamilton et al., 2006; Li et al., 2006). This development of more efficient and improved cell factories is driven by metabolic engineering, which involves directed genetic engineering of cell factories with the objective to change and improve their properties (Kim et al., 2012).

The yeast S. cerevisiae is a widely used cell factory for the production of fuels and chemicals, such as bioethanol – by far the largest volume fermentation product. It is also used for the production of several recombinant proteins, for example, human insulin, hepatitis vaccines, and human papillomavirus (HPV) vaccines. Saccharomyces cerevisiae also serves as an important model eukaryote, and many fundamental studies have therefore been performed on this organism. It was also the first eukaryotic organism to have its genome sequenced, and a number of
high-throughput studies have been pioneered using this organism as a model (Nielsen & Jewett, 2008). Owing to its model organism status and use in industry for recombinant protein production, there have been many studies on both (1) the basic cell and molecular biology of protein secretion and (2) strategies for engineering these processes for improved protein production.

There are many examples of engineering of *S. cerevisiae* for improved protein production, including optimizing of fermentation process, selecting the expression vectors systems, choosing the signal sequence for extracellular targeting and engineering host strains for better folding and post-translational modification (Idiris et al., 2010). Largely improvement of the heterologous protein has been achieved from milligrams to grams per liter based on these engineering in the past decade. However, as illustrated in these reviews, many of these attempts have given rather specific conclusions: rational targets for over-expression or deletion have been chosen, but often it was found that the strategy worked successfully only for one (or a few) protein(s), and the same engineered strain could not be used as a general cell factory platform for the production of a range of different recombinant proteins. This can be explained by the complexity of protein processing and secretion pathways. Folding, glycosylation, disulfide bond formation, and vesicle trafficking must all be accomplished while maintaining quality control feedback loops and avoiding situations that will perturb cellular homeostasis. Each process must be tuned to a specific state based on the secreted protein’s physical properties, for example, number of disulfide bonds, protein size, protein hydrophobicity, etc. Through detailed understanding of the individual processes and integrated analysis of the interplay between these processes, it should be possible to derive general models for protein secretion that can be used for engineering the secretion pathway and thereby result in improved cell factories for recombinant protein production. Therefore, genetic engineering combining with systems biology approach has become more and more useful for effective recombinant protein production (Graf et al., 2009).

Systems biology approaches are increasingly valuable for metabolic engineering of cell factories for metabolite production (Nielsen & Jewett, 2008). This is particularly due to the robustness of genome-scale metabolic models (GEMs). Whether these concepts can be expanded into use for improving recombinant protein production is still to be seen, but considering the complexity and many interaction partners involved in protein synthesis, protein folding, protein processing and secretion, it is very likely that systems biology approaches may substantially impact this field, both in terms of gaining system-level understanding and in terms of identifying engineering targets using these system-level models.

Our review focuses on systematically organizing and interconnecting secretory processes, that is, mapping of key components in post-translational modification process. This scaffold moves us toward a systems level of the large and complicated process of protein production. Different examples of recombinant protein production by yeast, including an overview of the different tools available for efficient protein production, will highlight the parameters that can be altered in these systems and potential outcomes. There are very few systems biology studies on protein secretion, but we will give examples on the use of omics analysis for studying specific processes, and we will also provide overall flowcharts for protein secretion processing that may be used as scaffolds for building more detailed models of protein secretion.

### A scaffold for protein post-translational modifications

For secreted proteins and proteins targeted to the plasma membrane and organelles of the endosome membrane system, there are many steps after translation before the protein is matured and trafficked to the correct location. A common pathway, called the secretory pathway, is used to complete the protein maturation process. Correct folding, post-translational modifications, and trafficking are required for membrane-bound, ER, Golgi, vacuole, cell outer membrane, cell wall, or secreted proteins. The secretory pathway primarily relies on local interactions (e.g. receptor-secreted protein interactions or chemical alterations to the secreted protein) to make decisions on the fate of the secretory protein, rather than transcriptional responses (e.g. transcriptional activation of a secretory machinery). Exhaustive catalogs of secretory/vps/endocytotic factors have been obtained by forward genetics, suppressors screens, screens of null mutant collections, and synthetic genetic analysis (Bard & Malhotra, 2006; Weerapana & Imperiali, 2006) (and at yeastrgenome.org), so in the following, we will focus on the ‘decision making’ components of the secretory pathway that interact directly with proteins traversing this pathway. By this, we map how the inherent biochemistry and the state of a secretory protein (amino acid sequence, folding, oxidation, glycosylation, etc.) determine the response by the secretory pathway. Many recognition complexes that are responsible for directing the vesicle to the correct organelle operate independently from the cargo that is in the vesicle and operate at a higher level of organization than is covered here. These aspects are covered in reviews by Pfeffer (Pfeffer & Aivazian, 2004). In many of the reviews discussed earlier, aspects of the yeast secretory pathway...
are melded with higher eukaryotic secretory pathways, but here we will focus on delineating yeast specific processes. Figure 1 shows an overview of the secretory pathway and the major processes involved. Supporting Information, Figure S1 through Fig. S5 and Tables 1 and 2 break down specific processes and catalog the secretory proteins associated with this pathway.

**Targeting to the endoplasmic reticulum**

After ribosomal synthesis begins, a protein bound for the secretory pathway must be selectively targeted to the ER, the first organelle the protein has to pass through in this pathway. The presignal sequence, an N-terminal 15–50 amino acid sequence, determines this step. Varying hydrophobicity of the central region of the presignal can lead to one of three fates (Fig. S1; Martoglio & Dobberstein, 1998). The first, default route uses a hydrophilic presignal (or the lack of a presignal) to ensure cytosolic translation of the protein. A second route uses highly hydrophobic signals to initiate cotranslational translocation at the ER/cytosol interface. In this process, the presignal is bound by the signal recognition particle (SRP) during translation (Table 1; Ng et al., 1996; Mason et al., 2000). SRP will pause translation and direct the ribosome to the ER membrane-bound SRP receptor (SR) (Table 1). Once the ribosome/SRP complex has docked at the SR, cotranslational translocation proceeds, that is, the polypeptide is synthesized as it passes through the Sec61 complex into the ER lumen (Rapiejko & Gilmore, 1997). The energy to drive the polypeptide into the ER is generated by GTP hydrolysis during translation (Osborne et al., 2005). Membrane-bound proteins will be inserted into the ER membrane during cotranslational translocation. After the N-terminal presignal has been inserted into the Sec61 complex, the various hydrophobic regions of the polypeptide chains can leave the Sec61 pore and enter the lipid phase of the ER membrane (Van den Berg et al., 2004). A third pathway exists for presignals that are weakly hydrophobic (Fig. S1). These presignals are not bound by the SRP, and translation is carried out in the cytosol, and the unfolded polypeptide chain is stabilized...
In this scenario, the presignal will interact directly with the Sec61 and Sec62/63 complexes (Table 1), independent of SRP (Plath et al., 1998). The newly synthesized, but unfolded, protein is pulled through the Sec61 complex by being bound to Kar2p, a yeast ER chaperone homolog of Bip/GRP78 (Matlack et al., 1999). As Kar2p binds more and more of the polypeptide chain in the ER, the protein is pulled from the cytosol to the ER. This third, SRP-independent pathway appears to be sufficient to traffic ER-bound proteins necessary for growth and survival. SRP null mutants are viable, but grow slowly, indicating that the second route (involving SRP particle) is important but not strictly required for viability (Brown et al., 1994; Rapoport, 2007).

**Endoplasmic reticulum processing**

By either of the routes described previously, the polypeptide will begin to enter the ER. During translocation, many structural and chemical modifications will be occurring to manage folding and quality control (Fig. S2). For soluble proteins, the presignal is cleaved by the signal peptidase complex (SPC) immediately (Table 1; YaDeau et al., 1991). For membrane proteins with multiple transmembrane regions, the presignal remains until all membrane spanning regions have been synthesized. Finally, folding chaperones will begin to cover exposed hydrophobic patches (Simons et al., 1995).

By cytosolic chaperones. In this scenario, the presignal will interact directly with the Sec61 and Sec62/63 complexes (Table 1), independent of SRP (Plath et al., 1998). The newly synthesized, but unfolded, protein is pulled through the Sec61 complex by being bound to Kar2p, a yeast ER chaperone homolog of Bip/GRP78 (Matlack et al., 1999). As Kar2p binds more and more of the polypeptide chain in the ER, the protein is pulled from the cytosol to the ER. This third, SRP-independent pathway appears to be sufficient to traffic ER-bound proteins necessary for growth and survival. SRP null mutants are viable, but grow slowly, indicating that the second route (involving SRP particle) is important but not strictly required for viability (Brown et al., 1994; Rapoport, 2007).

**Table 1. Proteins involved in cytosolic and ER decisions**

<table>
<thead>
<tr>
<th>Protein complex or grouping</th>
<th>Proteins involved</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal recognition particle (SRP)</td>
<td>Srp14p, Srp21p, Srp54p, Srp65p, Srp68p, scR1 RNA</td>
<td>Recognize presignal, direct to SR</td>
</tr>
<tr>
<td>SR receptor (SR)</td>
<td>Srb10p, Srb102p</td>
<td>ER receptor for SR</td>
</tr>
<tr>
<td>Sec61 complex</td>
<td>Sec61p,Shb1p,Sec62p</td>
<td>Cotranslational translocation pore</td>
</tr>
<tr>
<td>Sec62/63 complex</td>
<td>Sec62p,Sec63p,Sec71p,Sec72p</td>
<td>Post-translational translocation pore</td>
</tr>
<tr>
<td>Signal peptidase complex (SPC)</td>
<td>Sec11p,Spc1p,Spc2p,Spc3p</td>
<td>Presignal cleavage</td>
</tr>
<tr>
<td>Oligosaccharyl transferase (OST)</td>
<td>Wbp1p,Swp1p, Ost2p, Ost1p, Ost5p, Stt3p, Ost3p, Ost6p, Ost4p</td>
<td>N-linked glycosylation</td>
</tr>
<tr>
<td>Protein O-mannosyl transferases (PMT)</td>
<td>Pmt1p,Pmt2p, Pmt3p, Pmt4p, Pmt5p, Pmt6p, Pmt7p</td>
<td>O-linked glycosylation</td>
</tr>
<tr>
<td>ER chaperones</td>
<td>Kar2p, Sip1p, Sip1p</td>
<td>Protein folding</td>
</tr>
<tr>
<td>ER Redox enzymes</td>
<td>Ero1p, Pdi1p, Eug1p, Mpd1p, Mpd2p, Eps1p</td>
<td>Oxidation/reduction of disulfide bonds</td>
</tr>
<tr>
<td>N-linked glycan trimming</td>
<td>Cwh41p, Ret2p, Mns1p</td>
<td>Misfolded protein sensing</td>
</tr>
<tr>
<td>Hrd1p complex</td>
<td>Hrd1p, Hrd3p, Usa1p, Der1p</td>
<td>Misfolded protein sensing/trafficking</td>
</tr>
<tr>
<td>COPII cargo receptors</td>
<td>Sec24p, Sfb2p, Sfb3p, Shr3p, Chs7p, Vma22p, Uso1p, Ypt1p</td>
<td>Traffic proteins from ER to Golgi</td>
</tr>
</tbody>
</table>

**Table 2. Proteins involved in Golgi and post-Golgi decisions**

<table>
<thead>
<tr>
<th>Protein complex or grouping</th>
<th>Proteins involved</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-linked mannan polymerases</td>
<td>Mnn9p, Van1p</td>
<td>Mannose extension (2–10)</td>
</tr>
<tr>
<td>O-linked mannosylases</td>
<td>Anp1p, Mnn9p, Mnn10p, Mnn11p, Hoc1p</td>
<td>Mannose extension (11–40)</td>
</tr>
<tr>
<td>COPI complex</td>
<td>Mnn1p, Mnn2p, Mnn5p</td>
<td>Mannose extension (+40)</td>
</tr>
<tr>
<td>KO complex</td>
<td>Ktr1p, Ktr3p, Ktr1p/Kre2p, Mnn1p</td>
<td>Mannose extension (5) for secretory proteins</td>
</tr>
<tr>
<td>AP-1 complex</td>
<td>Cop1p(a), Sec27p(b), Sec21p(y), Ret2p(a)</td>
<td>Receptors for retrotransport from cis-Golgi to ER</td>
</tr>
<tr>
<td>AP-3 complex</td>
<td>Apl1p, Apl2p, Apl4p, Apm1p</td>
<td>CPY pathway to vacuole</td>
</tr>
<tr>
<td>GGA complex</td>
<td>Gga1p, Gga2p</td>
<td>CPY &amp; ALP pathway to vacuole</td>
</tr>
<tr>
<td>ESCRT-0 complex</td>
<td>Vps27p, Hse1p</td>
<td>Ubiquitin-based sorting to vacuole</td>
</tr>
<tr>
<td>ESCRT-1 complex</td>
<td>Gag2p, Vps28p, Vps36</td>
<td>Ubiquitin-based sorting to vacuole</td>
</tr>
<tr>
<td>ESCRT-2 complex</td>
<td>Stp22p, Snf2p, Vps28p, Mvb12p</td>
<td>Ubiquitin-based sorting to vacuole</td>
</tr>
<tr>
<td>ESCRT-3 complex</td>
<td>Vps20, Vps24, Did4p, Snf7p</td>
<td>Ubiquitin-based sorting to vacuole</td>
</tr>
</tbody>
</table>
tion occurs at the hydroxyl groups of serine and threonine and is catalyzed by protein O-mannosyltransferases (PMTs) (Table 1; Strahl-Bolsinger et al., 1999). PMTs transfer a single mannose to the serine/threonine in the ER, but more mannoses may be added later in the Golgi. O-linked glycosylation appears to occur before N-linked glycosylation, resulting in O-linked glycosylation on the serine/threonine of the N-linked recognition sequence (N-X-[S/T]). This implies that N-linked asparagine glycosylation and O-linked serine/threonine glycosylation may be in competition (Ecker et al., 2003).

After translocation, the nascent protein must undergo a series of folding and disulfide bond-forming steps. Quality control sensing determines if the correct structures have been formed before the folded protein is allowed to leave the ER for the Golgi (Fig. S2). Protein chaperones assist the polypeptides along the path to correct folding and help to remove them from the ER when a protein has terminally misfolded (Table 1). Kar2p (BiP), a Hsp70 family molecular chaperone, binds exposed hydrophobic stretches of amino acids (Blond-Elguindi et al., 1993). These hydrophobic regions are generally on the interior of a protein and are only exposed in incorrectly folded proteins. Kar2p repeatedly binds/releases these hydrophobic regions while hydrolyzing ATP (Gething, 1999). When Kar2p is bound by ATP, the Kar2p protein binds weakly to misfolded proteins, while ADP-bound Kar2p binds misfolded proteins tightly.

Disulfide bond formation must correctly pair distal cysteines of the polypeptide chain to form and stabilize the protein in its mature conformation (Fig. S2 and Table 1). Electrons are transferred from the newly formed disulfide bond to protein disulfide isomerase (PDI, Pdi1p in S. cerevisiae) which in turn passes the electrons to the FAD-bound Oxidoreductin 1 (Ero1p). Finally, the electrons are passed to the terminal electron acceptor O$_2$ (Tu & Weissman, 2002). This mechanism forms disulfide bridges at random, and the correct pairings must be found by a trial and error process, involving the repeated oxidation/reduction of cysteines by Pdi1p and its homologs (Tu & Weissman, 2004).

Exit from the ER can proceed by two pathways, (1) to the degradation pathway, ER-associated degradation (ERAD), for misfolded proteins (Fig. S3), and (2) to the Golgi, for properly folded proteins (Fig. S2). The exact biochemical mechanisms for these two pathways have not been completely determined in yeast. However, many parts of the decision making process have been identified. Detection of misfolded proteins and subsequent degradation is accomplished by several pathways (Fig. S3). Glycosylation structures of glycoproteins can traffic proteins to degradation. N-linked glycosylation trimming by glucosidase I (Cwh41p) and glucosidase II (Rot2p) are accomplished quickly and are observed for proteins that exit the ER (Fig. S3 and Table 1; Herscovics, 1999). ER mannosidase I (Mns1p) appears to be a gatekeeper for this degradation pathway. Mns1p removes a single mannose that is involved with targeting for the degradation pathway. Mns1p activity is lower than Cwh41p and Rot2p (Jakob et al., 1998), and this may result in a residence-time clock for proteins that are attempting to be folded. If a protein remains in the ER for too long, the mannose will be removed from the glycoprotein, and the protein will be retranslocated to the cytosol for degradation (Knop et al., 1996). Yos9p, Htm1p, and Mnl1p are believed to act as lectins for targeting de-mannosylated proteins to the ERAD (Fig. S3; Jakob et al., 2001). Kar2p and the Sec61 complex also are involved in the ERAD pathway, with Kar2p-binding acting as a residence-time clock similar to Mns1p, causing terminally misfolded proteins to be shuttled out of the ER (Brodsky et al., 1999). Membrane-bound misfolded proteins can be trafficked to degradation by three different pathways, depending on if the misfolding takes place in the ER lumen, intramembrane space, or on the cytosolic side (Fig. S3; Carvalho et al., 2006). When misfolding occurs on the ER luminal side, Der1p recruits the misfolded protein to the Hrd1p complex for ubiquitination (Table 1). When misfolding occurs inside the membrane, the Hrd1p complex ubiquinates in a Der1p-independent manner. Finally, cytosolic misfolding is managed by the Dna10p ubiquitin ligase. Ubiquinated proteins are trafficked to cytosolic proteosome activities. Calnexin/calreticulin systems have been elucidated in mammalian systems. However, the calnexin homolog in yeast (Cne1p) does not appear to have the same function but does have chaperone activity and is involved in the protein degradation pathway (Xu et al., 2004).

For a protein to exit to the Golgi, it must by-pass the degradation pathways mentioned previously and be recognized by receptors for export in COPII vesicles (Fig. S2). These COPII vesicles will traverse from the ER to Golgi where the membrane-bound or soluble proteins are further processed (for a recent review, see Dancourt & Barlowe, 2010). Sar1p acts as a trigger for the structural formation of the COPII vesicles, recruiting Sec13, Sec23p, Sec24p, and Sec31p) to complete the bud formation (Matsuoka et al., 1998). Importantly, several recognition signals are used to specifically bind export-ready proteins inside the forming vesicle. Soluble proteins are trafficked by: Sec24p binding to the di-acidic DXE cargo-sorting signal (Mossessova et al., 2003), and Emp24p, Erv14p, Erv25p, Erv26p, and Erv29p receptors binding to other unidentified motifs (Schimmoller et al., 1995; Belden & Barlowe, 1996). Membrane-bound proteins have cytosolic signals that are recognized by the Sec23-Sec24 complex (Table 1; Bonifacino & Glick, 2004). Sfb2p and Sfb3p, which are Sec24p homologs, are believed to bind other
cargo-sorting signals (Roberg et al., 1999; Kurihara et al., 2000; Peng et al., 2000). Shr2p, Chs7p, and Vma22p associate specifically with secretory proteins and may be involved in sorting their target proteins to the Golgi (Herrmann et al., 1999). Glycosylphosphatidylinositol (GPI)-anchored proteins are sorted to the Golgi by Uso1p and/or Ypt1p (Morsomme et al., 2003). Still other proteins appear to be captured nonspecifically and are transported to the Golgi by bulk flow (Malkus et al., 2002).

After the COPII vesicle buds off the ER, it traverses to the Golgi by diffusion (Preuss et al., 1992).

**Golgi processing**

In *S. cerevisiae*, the Golgi apparatus exists as individual cisternae scattered throughout the cell, which changes from cis cisternae to trans cisternae, in contrast to higher eukaryotes that have well-ordered stacked cisternae (Matsuura-Tokita et al., 2006). Regardless of the localization, many important modifications are made to the proteins in the Golgi, and these modifications affect the post-Golgi trafficking (Fig. S4). Glycoproteins are mannosylated (sometimes exceeding 50 mannoses) on the N-linked and O-linked sugar structures (Fig. S4; Hashimoto & Yoda, 1997; Jungmann & Munro, 1998). Mannoses are added to N-linked sugars in consecutive order by Och1p (one mannosyl), mannan polymerase I complex (M-Pol I) (10 mannosyl), mannan polymerase II complex (M-Pol II) (40 mannosyl), and finally, Mnn1p, Mnn2p, and Mnn5p which can add more mannosyl (Table 2; Hashimoto & Yoda, 1997; Jungmann & Munro, 1998). O-linked glycans have more stringent mannosylation, and only five mannoses are added, and only to proteins that will be on the exterior of the cell (Table 2; Strahl-Bolsinger et al., 1999). The O-mannosylations are believed to be a signal for trafficking to the exocytosis pathways.

Maturation of the protein in the Golgi also involves cleaving the polypeptide chain. Three Golgi-resident proteases can cleave the polypeptide based on different recognition sites (Fig. S4, lower part). Kex1p cleaves C-terminal arginine or lysine (Cooper & Bussey, 1989). Kex2p, the most well-studied protease, cleaves a (K/R)-R motif (Rockwell et al., 2002). Ste13p is a dipeptidyl aminopeptidase that cleaves repeated X-A motifs (Julius et al., 1983). These polypeptide cleavages allow the following: maturation of proteins, activation of catalytic activity, and changed conformation for binding the intended receptor.

**Post-Golgi sorting**

After the Golgi maturation processes are completed, the most important sorting processes will take place on the exit from the Golgi. Trafficking from the Golgi can go in many directions, depending on the final destination of the protein, retrograde to ER, transport to early endosome, late endosome, vacuole, plasma membrane, or extracellular space (Fig. S5). Retrograde transport from the cis-Golgi to the ER is important to return membrane area, ER SNAREs, cargo adapter proteins, and membrane components to the ER, otherwise these resources would be depleted from the ER. COPI vesicles are responsible for the retrograde transport from cis-Golgi to ER (Table 2). Soluble proteins in the Golgi that must be transferred back to the ER contain an HDEL sequence that is bound by the COPII protein Erd2p (Aoe et al., 1997). A range of COPII subunits can recognize cytoplasmic motifs of membrane proteins, such as α and β’ to KXXX, γ to FF or K[K/R][X of p24 protein, and δ to the δl motifs (WXX[W/Y/F]) (Eugster et al., 2004).

Another motif, RKR, on the cytoplasmic side of potassium transporters Trk1p/Trk2p causes retrograde transport to the ER, although the receptor is not known (Zerangue et al., 1999).

Three pathways exit from the trans-Golgi network (TGN), (1) the carboxypeptidase (CPY) pathway, (2) the Golgi-localized, γ-Ear-containing, ADP-ribosylation factor-binding proteins (GGAs)-associated pathway, and (3) alkaline phosphatase (ALP) pathway. The default route to the vacuole is via the CPY pathway, a two-step process using adaptor protein (AP) complexes 1 (Fig. S5 and Table 2). AP-1 complex vesicles can transfer proteins from the TGN to the early endosome (Valdivia et al., 2002; Abazed Fuller, 2008). In vitro studies, Kex2p is sorted via AP-1 complex to the early then late endosome (Abazed Fuller, 2008). From the early endosome, the default route moves proteins from the late endosome to the vacuole (Dell’Angelica et al., 1997). Data suggest that proteins not having a sorting signal are automatically sorted to the CPY pathway, such as recombinant secretory proteins (Cowles et al., 1997).

The GGA-associated pathway traffics vesicles directly from the TGN to the late endosome. A QRPL motif followed by ubiquitination appears to be the common signals for targeting through this pathway. Gga1p and Gga2p are the sorting proteins and this pathway traffic Vps10p and other vacuole resident proteins to the late endosome (Valls et al., 1990). Rsp5p is a broad-range ubiquitin ligase responsible for ubiquinating these proteins (Dunn & Hicke, 2001; Wang et al., 2001). The ubiquitin-binding domain of Gga1p and Gga2p (Table 2) appears important in the trafficking process (Costaguta et al., 2006). At the late endosome, the GGA pathway converges with the CPY pathway in default transport to the vacuole.

Finally, an additional route exists to traffic proteins from the TGN directly to the vacuole, namely the ALP pathway.
The ALP pathway transports membrane proteins using AP-3, independent of the endosome (Fig. 55; Piper et al., 1997). This pathway relies on a 13–16 amino acid (arginine- and lysine-rich) cytoplasmic signal and was identified by ALP sorting aberrant mutants (Cowles et al., 1997).

Endosomal sorting complex required for transport (ESCRT) complexes, four complexes in all, can also bind ubiquinated proteins and form luminal vesicles that are trafficked to the vacuole (Table 2). The ESCRT-0 complex (Vps27p and Hse1p) has ubiquitin interacting motifs that recruit the other ESCRT complexes (Bilodeau et al., 2002). These complexes recruit a deubiquinating enzyme (Doa4p), necessary for maintaining ubiquitin homeostasis in the cytosol, and structural proteins that create the luminal vesicles for vacuolar degradation that are characteristic of the multivesicle bodies (MVB) (Dupre & Hagenauer-Tsapis, 2001; Luhtala & Odorizzi, 2004).

Exocytosis

For proteins that will follow the exocytotic pathway from the trans-Golgi, two pathways exist (Fig. 55). From density-based separation experiments, two types of vesicles are known to merge with the cell membrane and are named light density secretory vesicles (LDSV) and heavy density secretory vesicles (HDSV) (Harsay & Bretscher, 1995). LDSV are known to carry constitutively expressed cell membrane proteins, such as Bgl2p, Pma1p, and Gas1p. LDSV are believed to emerge from the trans-Golgi and transit directly to the cell membrane (Gurunathan et al., 2002). This process takes around 30 min. LDSV may be the final step in lipid raft-based sorting that begins in the ER (Bagnat et al., 2000). Specific cell membrane proteins partition to high sterol-rich domains of the ER membrane. These rafts are directed through the secretory pathway and are finally merged with the cell membrane. Conversely, HDSV package soluble, secreted proteins, such as invertase (Suc2p) and acid phosphates (Pho11p, Pho12p, and Pho5p) that are transcriptionally regulated and induced under certain conditions. HDSV move from the endosome to the cell membrane and are thus subject to many of the mutations that block movement to and through the early/late endosome (Gurunathan et al., 2002). These mutants, which block the HDSV pathway, were shown to use the LDSV pathway for the secretion of proteins normally bound for HDSV pathway (Harsay & Schekman, 2002).

Unfolded protein response–transcriptional control of the secretory pathway

While much of the secretory pathway is managed on the basis of protein–protein interactions (such as ubiquinonization of misfolded proteins) and chemical modifications to the trafficked protein (such as glycosylation and disulfide bond formation), these processes occur in unstressed conditions during normal cell growth. However, when protein folding stress begins to overwhelm the processing machinery of the ER, large scale transcriptional alterations become necessary to bring the secretory pathway back into homeostasis. This transcriptional response, the Unfolded Protein Response (UPR) is a large scale orchestrated response that increases the capacity of the secretory pathway, clearance of misfolded proteins, and oxidative conditions in the ER (Bard & Malhotra, 2006).

The UPR broadly consists of an upstream sensing mechanism and a downstream activation mechanism to coordinate this broad stress response. The upstream mechanism has been studied in great detail and is primarily controlled by two key proteins, the ER transmembrane protein, Ire1p, and the transcriptional activator, Hac1p. Ire1p contains an ER luminal domain that binds Kar2p/BiP and a cytosolic domain that has kinase and endonuclease activity. Misfolded proteins in the ER are detected when large amounts of Kar2p are recruited away from Ire1p. Under normal conditions, a portion of Kar2p is associated with immature protein, allowing them to fold completely, while the majority of Kar2p is associated with Ire1p. This association with Ire1p causes steric effects that prevent dimerization of Ire1p. However, under stress conditions, most Kar2p molecules are associated with unfolded protein, while simultaneously unfolded proteins are bound to Ire1p. This exchange of Kar2p for unfolded protein causes Ire1p to dimerize. Upon dimerization, the cytoplasmic portion of Ire1p phosphorylates itself, which in turn, activates an endonuclease domain on the cytoplasmic portion of Ire1p. This endonuclease activity is specific to an mRNA sequence in HAC1u, the transcribed RNA from HAC1. Unactivated HAC1u mRNA is constitutively expressed in the cell. However, because of the presence of a 3′ RNA hairpin, HAC1u cannot be translated. Activated Ire1p cleaves HAC1u (becoming HAC1′, for induced) to remove the hairpin, which is followed by R1g1p ligation (tRNA ligase), allowing translation to proceed. Hac1p can then be expressed as a functional transcriptional activator. Recent study revealed that ER-lumenal domain of yeast Ire1 can bind to unfolded proteins directly, drive Ire1 dimerization and activate the UPR (Gardner & Walter, 2011).

A mathematical model has been developed to describe the upstream/activation portion of the UPR. Raden et al. (2005) use a series of ordinary differential equations to describe the Ire1p activation, as it relates to its Kar2p binding state. The model predicted steric effects, by only Kar2p, are not adequate to explain the dynamics of UPR...
activation. A key facet of this work was that the model considered the relative concentrations of Ire1p and Kar2p in the ER, combined with expected kinetics. The model predicted that with Kar2p over-expression, the cell should tolerate higher amounts of unfolded protein before inducing the UPR. This prediction was tested experimentally, and it was found that the amplitude of UPR activation was decreased, but the UPR induction threshold occurred at the same unfolded protein levels. A revised model, which included an unknown secondary effector (presumably unfolded protein binding to Ire1p), was able to capture the experimental observations. This model should be useful in understanding the conditions that lead to upstream UPR activation and the level of activation that can be expected. The downstream portion of the UPR is characterized by a large, multifaceted response to bring the secretory pathway back to homeostasis (Tyo et al., 2012). Hac1p is a transcriptional activator that is known to interact with three binding sequence (in coordination with Gcn4p) to regulate many different activities within the cell in an attempt to correct the misfolded protein problem in the ER (Mori et al., 1996; Travers et al., 2000; Patil et al., 2004). In all, the expression of approximately 380 genes is altered in the UPR response, although only half have Hac1p binding sequences in the promoter (Travers et al., 2000; Kimata et al., 2006). The upstream/detection part of the UPR pathway has been elucidated, however, the downstream/implementation part of the response has been limited to identifying promoter sequences that are specific to UPR and DNA microarray analysis that has identified genes altered by the UPR (Travers et al., 2000; Kimata et al., 2006).

Many cellular responses are activated simultaneously. Broadly, the UPR (1) increases capacity of the secretory pathway and (2) clears unwanted/unnecessary proteins. In the ER, folding rate is increased by upregulating chaperones, such as Fkb2p, Lhs1p, and Kar2p, and disulfide bond formation by Ero1p, Pdi1p, and others. To accommodate increased disulfide bond formation activity and the subsequent reactive oxygen species that can damage the cell (Haynes et al., 2004), oxidative stress response genes are also activated. Glycosylation processing elements of the ER and Golgi are also upregulated to increase processing capacity of the secretory pathway, as these glycosylations are required for proper folding of many proteins. Trafficking components used in COPI, COPII, and post-Golgi vesicles are upregulated. Finally, metabolic pathways for lipid and inositol are upregulated, to increase the amount of membrane. Membrane, while often not considered to be an active component of the secretory pathway, provides essential surface area that is essential for almost all secretory pathway processes.

Aside from increasing secretion capacity, the UPR also clears unfolded protein and reduces the demand for the secretory pathway. To remove misfolded proteins, elements of the ERAD and ubiquitin/proteosome system are upregulated (Travers et al., 2000). Interestingly, cotranslational translocation and post-translational translocation are increased at the ER-cytosol interface, but this is most likely to facilitate the transport of misfolded proteins back to the cytosol for proteolysis, not transport into the ER. Misfolded proteins may also be cleared from the ER in a “feed forward” manner by moving them through the Golgi to the vacuole, as COPII vesicle components are upregulated. Evidence indicates that misfolded proteins can be degraded independent of ERAD, as mutants that abolish ERAD are constitutively activated for UPR and misfolded proteins can be targeted to the vacuole (Hong et al., 1996; Travers et al., 2000). Kimata et al. (2006) also found a number of exocytosis-targeted proteins were downregulated in the UPR. For example, acid phosphatases (Pho3p and Pho5p), various transporters (Ato3p, Fet3p, Fre1p, and Tpo1p), and α-factor, which consume secretory pathway capacity, are reduced to help secretory pathway stress.

The downstream portion of the UPR is ripe for systems biology modeling. As discussed, the UPR initiates and coordinates many processes in the cell to bring the secretory pathway back to homeostasis. While the transcription factor Hac1p is known to signal the UPR, the specific transcription factors that initiate the many sub-tasks of the UPR have not been identified. As well, the biological information flow should be useful to engineer the secretory pathway for greater recombinant protein productivity. Recently, integrative systems biology analysis was used to identify Hac1p, Fhl1p, and Skn7p as significant transcription factors in the UPR response (Tyo et al., 2012). Fhl1p shows us the role in the coordinated down-regulation of ribosomal protein and ribosomal rRNA, thereby decreasing the total translational capacity of the cell. Skn7p is responsible for managing oxidative and osmotic stress responses in the cell. In a UPR stress response, Skn7p is used to upregulate oxidative stress response, thereby mitigating ROS, while downregulating osmotic stress response. Downregulating the osmotic stress response results in fewer cell wall proteins being processed in the secretory pathway, freeing up additional secretion capacity. Further study should lead to scaffold models to map all major branches of the UPR.

Biotechnology: parameters to increase secretion

Through detailed knowledge of the secretion pathway, it has become possible to improve the secretion yield and
efficiency through a combination of different molecular techniques (Idiris et al., 2010): (1) engineering signal sequences, (2) optimizing the ER folding environment, (3) affecting vesicle transport, and (4) reducing protease activities. High-throughput screening approach is also frequently used to improve the secretory capability, and in the future, it will be interesting to exploit systems biology tools for the evaluation of improved mutants with the objective to find novel metabolic engineering targets.

Nowadays, the secretion level of recombinant protein secretion in *S. cerevisiae* is still in the order of mg L\(^{-1}\), although some industrial companies have managed to elevate the titers of certain proteins to the g L\(^{-1}\) range. A summary of recombinant protein secretion systems in *S. cerevisiae* is presented in Table S1 and a more detailed review of different strategies is given in the following.

### Engineering the signal sequence

#### The leader sequence

The leader sequence determines, in part, the trafficking of a secreted protein. The presequence determines whether cotranslational translocation or post-translational translocation occurs for entrance to the ER and the pro-sequence determines the sorting mechanisms in the trans-Golgi network. Native *S. cerevisiae* leader sequences, foreign leader sequences, and leader sequence devised from theory (synthetic leader) have been used to target heterologous proteins for secretion.

Native leaders often possess certain advantages, which is proved by many cases including human serum albumin (HSA) (Sleep et al., 1990), human interferon (IFN) (Piggott et al., 1987), and *Aspergillus niger* glucose oxidase (GOD) (De Baetselier et al., 1992). However, recombinant proteins produced by *S. cerevisiae* are often hyperglycosylated and retained in the periplasmic space (Spear & Ng, 2003; Schmidt, 2004). It is therefore sometimes preferred to choose highly glycosylated leaders, such as the *S. cerevisiae* α-factor leader, which has proven to be very efficient in some cases, for example, for the secretion of human epidermal growth factor (hEGF) (Chigira et al., 2008), human platelet derived growth factor (PDGF) (Robinson et al., 1994), and *Schizosaccharomyces pombe* acid phosphatase (Baldari et al., 1987). However, it is not possible to predict which leader is best suited for efficient secretion of a given protein. It is therefore often required to experimentally evaluate different leaders. This is illustrated by a study of Li et al. (2002), who evaluated various leader sequences including INU1, SUC2, PHO5, and MEL1, to secrete either green fluorescent protein (GFP) or GFP-hexokinase fusions. In all cases, the majority of the protein accumulated in the vacuole or endosome (Li et al., 2002). However, using a viral leader from the K28 preprotoxin, secretion was efficient (Eiden-Plach et al., 2004). Another example of this is a study which showed that the yeast invertase signal SUC2 was correctly cleaved from all secreted IFN molecules (Parekh & Wittrup, 1997) unlike when using the native IFN leader that only resulted in 64% cleavage (Hitzeman et al., 1983). However, when using the same SUC2 leader to secrete human α-1-antitrypsin (α-AT), approximately 80% of the protein accumulated in the secretory pathway (Moir & Dumais, 1987).

Synthetic leaders are often used to solve secretion problems, such as (1) inefficient processing of pre- or pro-leaders, (2) hyperglycosylation protein accumulation, and (3) incorrect trafficking in the secretory pathway. Examples of synthetic pre- and pro-leaders include the expression of insulin precursor (IP) (Kjeldsen, 2000), human adenosine A2a receptor (A2aR) (Butz et al., 2003), bovine pancreatic trypsin inhibitor (BPTI) (Parekh & Wittrup, 1997), and single-chain antibody (scFv) (Shusta et al., 1998). Recently, we performed a comparison of a synthetic leader with the α-factor leader and found the synthetic leader to be slightly more efficient for the secretion of insulin precursor and α-amylase (Liu et al., 2012).

There have also been several studies on the importance of both the pre- and pro-regions for different secretion strategies. For most proteins, for example, human insulin-like growth factor 1 (hIGF-1) (Romanos et al., 1992) and α-globin (Rothblatt et al., 1987), both the pre- and pro-leader should be applied to achieve an optimal secretion. However, there are some exceptions. Ernst et al. found that the pro-region of the α factor leader has only a minor effect on secreting aminoglycoside phosphotransferase (APH) and granulocyte colony-stimulating factor (GCSF), whereas for interleukin-1β, the preregion decreased Kex2p processing efficiency compared with the case when only the pro-region was applied (Ernst, 1988). One possible explanation is that the pro-region may help to stabilize the mRNA or facilitate transcription process (Gabrielsen et al., 1990), however, more studies are still needed to further look into the roles of the different parts of the leader sequence.

#### Spacers for leader sequences

To achieve a correct final product, the specific proteases need to efficiently cut the pre- and pro-proteins at the correct places. This affects sorting as well as product quality. Recombinant protein secretion directed by pre-pro-leader sequences typically relies on Kex2p endoprotease activity, which is often limiting. Inefficient Kex2p processing results in the secretion of hyperglycosylated unprocessed pro-proteins (Fabre et al., 1991; Kjeldsen et al., 2001).
et al., 1996). There are many ways to solve this problem. In some cases, spacer residues were included to provide a hydrophilic environment that improves cleavage by Kex2p (Guisez et al., 1991). Another approach, modifying the protein coding sequence, such as to include an alanine N-terminal to the human Interleukin-6 (hIL-6), can also improve cleavage (Guisez et al., 1991). Kjeldsen et al. (1996) tried either to apply a spacer peptide between the leader and the insulin precursor or to apply a “mini C-peptide” (Kjeldsen et al., 1999), and both approaches were found to increase the efficiency of Kex2p endoprotease processing. However, a spacer at the N-terminus of the secreted protein is not always helpful, and in one study, it was found that this approach resulted in 5% intracellular retention of hEGF and 50% for IFN (Singh et al., 1984). Another approach is to over-express the protease genes. Barr et al. (1987) over-expressed the KEX2 gene, and this resulted in improved secretion of correctly processed transforming growth factor-α (TGFα) into the culture medium. Over-expression of S. cerevisiae aspartyl protease (YAP3) (Egel-Mitani et al., 1990) or dipeptidyl aminopeptidase (STE13) (Julius et al., 1983) was also found to improve pro-sequence cleavage. In general, the spacer should have an absence of nonspecific interaction sequences (Fuchs et al., 1997), optimal proteolytic accessibility (Leong & Chen, 2007), and protection of the interface from hydrophobic fragments (Reiter et al., 1994).

**Engineering protein folding and glycosylation**

Glycosylation takes place in the ER and Golgi and can be engineered based on the amino acid sequence of the protein or the glycosylation enzymes (Tables 1 and 2). Glycosylation mitigates aggregation (Parthasarathy et al., 2006) and hydrolysis (Robinson et al., 2004), and also increases interaction affinity and selectivity (Rudd et al., 1999), but it is still not fully clarified how glycosylation affects secretion level.

Glycosylation seems to have no significant effect on the secretion of α-amylase (Nieto et al., 1999) and IL-1α (Livi et al., 1990). While on the other hand, missing one essential glycosylation site of CD47 reduced its surface expression level by more than 90% (Parthasarathy et al., 2006). Glycosylation has been shown to facilitate protein folding of EGF (Demain & Vaishnav, 2009) and immunoglobulin (Rudd et al., 1999) and keep the activity of interleukin-1β (Livi et al., 1991). Furthermore, introducing extra N-glycosylation sites can improve secretion, as illustrated by the secretion of cutinase, where a fivefold or 1.8-fold increase in secretion was obtained after introducing a N-glycosylation site in the N-terminal and C-terminal regions, respectively (Sagt et al., 2000).

When no glycosylation sites can be added or engineered in the coding region of the protein, an alternative solution is to apply a leader sequence which contains N-glycosylation sites (Chen et al., 1994). N-glycosylation has been shown to be very important for α-glycosylases, especially for the pro-region, when directing insulin secretion (Caplan et al., 1991; Kjeldsen et al., 1998). A synthetic leader LA19 with two N-glycosylation sites has also been developed (Fabre et al., 1991) and demonstrated optimal glycosylation for insulin secretion (Kjeldsen et al., 1998). In addition to engineering glycosylation to improve secretory efficiency, important improvements have been made in engineering humanized glycosylation in yeasts. Wildt and Gerngross review this topic in detail (Wildt & Gerngross, 2005).

The number of disulfide bonds is another factor that affects protein secretion (Hober & Ljung, 1999). For example, the expression level of insulin-like growth factor-1 (IGF1) decreased by about one-third when removing either Cys23p or Cys96p, which are likely to be involved in disulfide bond formation (Steube et al., 1991). The expression level and affinity of CD47 decreased by 30% when the core disulfide bond is missing (Parthasarathy et al., 2006).

Protein folding in the ER is often considered the flux controlling step in the secretion pathway (Lim et al., 2002), and over-expression of chaperones, especially Kar2p and PDI, therefore often allows for improved secretion. Kar2p acts as a folding chaperone by binding to exposed hydrophobic sequences (Ma et al., 1990) and also as an ER detergent functioning in the ERAD process (Robinson et al., 1996). On the other hand, PDI catalyzes disulfide bonds formation and isomerization (Labossiere et al., 1995). The soluble levels of PDI decrease upon over-expressing recombinant proteins, implying it functions not only as a catalyst, but also as a chaperone, binding to the heterologous proteins (Robinson & Wittrup, 1995). Over-expression of either Kar2p or PDI improves secretion levels in many cases (Table 3). Over-expression of PDI also improves secretion for proteins that do not contain disulfide bonds, for example, *Pyrococcus furiosus* β-glucosidase (Smith & Robinson, 2002), suggesting that PDI may act in a chaperone-like capacity or cooperate with the folding or degradation mechanisms on nondisulfide bonded protein (Powers & Robinson, 2007).

Sometimes, Kar2p and PDI work together to ensure proper folding, and Mayer et al. (2000) suggested that Kar2p may maintain the protein in an unfolded state by binding to the protein, and this makes the cysteine residues accessible for PDI activity. This Kar2p/PDI cooperativity increased secretion of scFv (Xu et al., 2005) and β-glucosidase (Smith et al., 2004). However, in
other cases, over-expression yields only a minor increase or even a decrease in the secretion, as illustrated for plant thaumatin (Harmsen et al., 1996), IFN-α2a and A2aR (Butz et al., 2003). These differences can be explained by each protein’s unique characteristics, such as the presence of glycosylation sites and the number of disulfide bonds.

Besides Kar2p, the cochaperones that are involved in regulating the ATPase activities of Kar2p, like DnaJ-like chaperone Jem1p, Scj1p, and nucleotide exchange factor Sll1p and Lhs1p, are also reported to increase the protein production. By single or multiple over-expression of these chaperones, the secretion levels of recombinant human albumin (rHA) granulocyte–macrophage colony-stimulating factor (GM-CSF), and recombinant human transferrin were improved significantly (Payne et al., 2008). Another approach to engineering the protein folding and secretion is to activate UPR by manipulation of the HAC1 gene. Over-expression of S. cerevisiae HAC1 resulted in a 70% increase in Bacillus amyloliquefaciens α-amylase secretion, but did not increase the secretion of ER-accumulated Trichoderma reesei endoglucanase EGI (Valkonen et al., 2003). Over-expressing T. reesei HAC1 in yeast resulted in a 2.4-fold increase in α-amylase secretion (Higashio & Kohno, 2002). It indicates the effect of UPR activation by HAC1 over-expression is protein specific and dependent on protein properties and regulation impact.

### Engineering protein trafficking and minimizing protein degradation

High-level expression of recombinant proteins often results in misfolding and accumulation of protein at certain steps in the secretion pathway. However, different proteins accumulate in different compartments, hepatitis B surface antigen (HBsAg) (Biemans et al., 1991), α-1-antitrypsin (Moir & Dumais, 1987), and erythropoietin (Elliott et al., 1989) accumulate in the ER compartment, but soybean proglycinin is retained in the Golgi (Utsumi et al., 1991). Secretion of heterologous proteins may also interfere with native protein secretion, for example, the secretion of host acid phosphatase gets disturbed by the secretion of tissue-type plasminogen activator (tPA) (Hinnen et al., 1989), probably due to induction of cell stress and lack of capacity in the secretion pathway. Secretion of heterologous genes may also cause increased ER stress that may link to other cellular processes and hereby result in reduced overall productivity.

Other proteins also assist with secretion. For example, Over-expression of the PDI oxidant Ero1p and a cell wall protein Ccw12p, has been reported to optimize the secretion of scTCR by 5.1- and 7.9-fold, respectively (Wentz & Shusta, 2007). Over-expression of the UBI4 gene, increase the secretion level of elafin by 10-fold (Chen et al., 1994).

Over-expression of SSO1 and SSO2, which are crucial for vesicle fusion to plasma membrane, increased α-amylase

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Amino acid</th>
<th>Disulfide bond</th>
<th>N-glycosylation site</th>
<th>BiP+ (by fold)</th>
<th>PDI+ (by fold)</th>
<th>BiP+ PDI+ (by fold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDGF-B</td>
<td>109</td>
<td>5</td>
<td>1</td>
<td>–</td>
<td>10 (Robinson et al., 1994)</td>
<td>–</td>
</tr>
<tr>
<td>Hirudin</td>
<td>65</td>
<td>3</td>
<td>0</td>
<td>2.5 (Kim et al., 2003)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>BPTI</td>
<td>58</td>
<td>3</td>
<td>0</td>
<td>1 (Robinson et al., 1996)</td>
<td>1 (Kowalski et al., 1998)</td>
<td>–</td>
</tr>
<tr>
<td>scFv</td>
<td>244</td>
<td>2</td>
<td>1</td>
<td>2.4 (Shusta et al., 1998)</td>
<td>2.3 (Shusta et al., 1998)</td>
<td>10.4 (Hackel et al., 2006)</td>
</tr>
<tr>
<td>scTCR</td>
<td>240</td>
<td>1</td>
<td>3</td>
<td>2 (Shusta et al., 2000)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>A2aR</td>
<td>412</td>
<td>0</td>
<td>2</td>
<td>1 (Butz et al., 2003)</td>
<td>75% (Butz et al., 2003)</td>
<td>1 (Butz et al., 2003)</td>
</tr>
<tr>
<td>rhG-CSF</td>
<td>174</td>
<td>2</td>
<td>0</td>
<td>1 (Robinson et al., 1996)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>PHO</td>
<td>435</td>
<td>8</td>
<td>9</td>
<td>1 (Robinson et al., 1996)</td>
<td>4 (Robinson et al., 1994)</td>
<td>–</td>
</tr>
<tr>
<td>P. furiosus β-glucosidase</td>
<td>421</td>
<td>1 Cys</td>
<td>0</td>
<td>1 (Smith &amp; Robinson, 2002)</td>
<td>1 (Smith &amp; Robinson, 2002)</td>
<td>1.6 (Smith et al., 2004)</td>
</tr>
<tr>
<td>Bovine prochymosin</td>
<td>345</td>
<td>4 Cys</td>
<td>2</td>
<td>20 (Harmsen et al., 1996)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Plant thaumatin</td>
<td>235</td>
<td>8</td>
<td>0</td>
<td>1 (Harmsen et al., 1996)</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Several data points come from Protein Knowledgebase (UniProtKB).
secretion by 2-fold (Larsson et al., 2001; Toikkanen et al., 2004). Co-over-expression of COG6, COY1, and IMH1, all genes related to Golgi vesicle transport, enhance Fab production by 1.2-fold (Gasser et al., 2007). Mutation of the cell wall protein Gas1p strongly improved the secretion of IGF1 (Brinkmann et al., 1993), and a mutation of PMR1, a Golgi-resident calcium ATPase gene (Rudolph et al., 1989), increased the secretion of prochymosin (Harmsen et al., 1996) and propapain (Ramjee et al., 1996). Recently, we showed that it is also possible to improve protein secretion by over-expression of SNARE regulating proteins Sec1/Munc18 (SM) proteins that modulate vesicle transport (Hou et al., 2012).

Proteins, targeted to the vacuole by a group of vacuolar sorting proteins (VPS) (Graham, 1991) and degraded, can hence not be exported. Interestingly, the intracellular sorting is dependent on the catalytic activity of Kex2p (Zhang et al., 2001). Deleting VPS4, VPS8, VPS13, VPS35, VPS36, or PEP4, all encoding vacuolar proteinases, resulted in higher yields of an insulin-containing fusion protein (ICFP) (Zhang et al., 2001). Single deletion of the extracellular protease Ski5p successfully improved the secretion level of killer toxin (Bussey et al., 1983), and disruption of YAP3 alone or together with KEX2 reduced the degradation of HSA and HSA-human growth hormone fusion protein. As well, a single deletion of KEX2 had a minor effect (Geisow et al., 1991).

Besides vacuolar sorting, some proteins may undergo proteasome-based protein degradation. This has been seen for cutinase production in yeast (Sagt et al., 2002). Delta’s strains have mutant genomic UBC4 gene, which encodes the ubiquitin-conjugating enzyme, resulting in extremely high plasmid copy number and over-expression of different proteins (Sleep et al., 2001).

**High-throughput screening for secretory pathway mutants**

**Random mutagenesis and screening**

Random mutagenesis and screening is another powerful tool to optimize protein expression level, stability, function and antigen-binding affinity (Wittrup, 2001; Vasserot et al., 2003). This can be mutagenesis of either (1) the recombinant protein to be secreted, or (2) the host strain to alter synthesis and secretory properties.

Concerning mutagenesis of the recombinant protein, Zhang et al. (2003) studied single- and double-point mutations within the insulin B-chain and suggested that failure to properly form disulfide bonds should contribute to altered intracellular trafficking. Kowalski et al. (1998) created all possible single and pairwise mutants of a BPTI cysteine and concluded that 5–55 disulfide bond is essential for protein folding and secretion.

When pursuing mutagenesis of the host strain, Smith et al. (1985) found four possible targets by screening mutagenized bovine growth hormone (rBGH) secretion strains and reported that mutations in two genes in particular, SSC1 and SCC2, yield the highest increase in around 15-fold compared with reference strains. Arffman et al. (1990) successfully isolated a strain that could secrete 70-fold more endoglucanase I (EGI) compared with a reference strain through multiple rounds of mutagenesis and selections.

**Screening through yeast surface display system**

Yeast surface display is a useful technology for the screening of improved protein expression, and it has been used for selecting high-secretion mutants of tumor necrosis factor receptor (TNFR) (Schweickhardt et al., 2003) and scFv (Starwalt et al., 2003). In yeast surface display, the target protein is bound to the mating agglutinin Aga2p by a pair of disulfide bonds. Then, the fusion is displayed on the surface of the cell by binding to the cell wall protein Aga1p (Huang & Shusta, 2005). Surface display data correlates well with secretion data (Shusta et al., 1999), and the technology can therefore be used for the screening of efficient secretion clones. Wentz & Shusta (2007) performed a genome-wide screening through flow cytometric scan by combining yeast cDNA libraries with yeast surface display and found five gene products that promoted display level of a single-chain T-cell receptor (sTCR), including cell wall proteins (Ccw12p, Cwp2p, and Sed1p), ribosomal protein (Rpp0p), and an ER oxidase (Ero1p).

**Omics analysis application for recombinant protein secretion**

Genome-wide systems analysis is becoming a very powerful tool to understand the cellular responses to protein production and assess the potential strategies for improving secretion. Bonander et al. (2009) analyzed the transcriptome data of eukaryotic glycerol facilitator (Fsp1) producing strains and showed that tuning BMS1 transcript levels resulted in a change of ribosomal subunit ratio and could be used to optimize yields of functional membrane and soluble protein targets. Gonzalez et al. (2003) used metabolic flux analysis to compare a human superoxide dismutase (SOD) production strain to a wild-type strain and showed that the flux of precursors to amino acids and nucleotides was higher, and the activities of the pentose phosphate (PP) pathway and TCA cycle were lower in the recombinant strain. They demonstrated that using the growth associated expression system, ideal
conditions for SOD synthesis were either active growth condition during respirofermentative metabolism or transition phase from a growing to a nongrowing state. The data indicated an increase in SOD flux could be achieved using a nongrowth-associated expression system that can eliminate part of the metabolic burden. Recently, our study analyzed secretory pathway dysfunction resulting from heterologous production of human insulin precursor or α-amylase in HAC1 dependent and independent manner by transcriptome and flux analysis. This study revealed that the oxidative radical production because of a futile cycle of disulfide formation and breaking and provided implication on engineering recombinant protein secretion, like engineering the post-Golgi sorting, and balancing the protein folding rates and oxidation rates (Tyo et al., 2012).

Besides S. cerevisiae, the systems biology approach was also used to analyze the secretion capability of P. pastoris. The comparison of the transcriptome of a P. pastoris strain producing human trysinogen with a nonproducing strain revealed a set of secretion helper genes. Thirteen of 524 upregulated genes were selected and the respective S. cerevisiae homologs were cloned and over-expressed in a P. pastoris strain expressing human antibody Fab fragment. Besides five previously characterized secretion helpers (PDI, Ero1p, Sso2p, Kar2/BiP, and Hac1p), another six proteins, more precisely Bfr2p and Bmh2p involved in protein transport, the chaperones Ssa4p and Sse1p, the vacuolar ATPase subunit Cup5p and Kin2p, a protein kinase connected to exocytosis proved their benefits in protein production (Gasser et al., 2007). Through modeling and measuring intracellular fluxes of secreted recombinant protein in P. pastoris with a 34S procedure, Pfeffer et al. demonstrated that 58% protein produced intracellularly were degraded within the cell, 35% were secreted to exterior and 7% were inherited to the daughter cells. This study provides insights of bottle-necks of recombinant protein production and is useful to determine the suitable strategy for secretion improvement (Pfeffer et al., 2012). Although there are not many examples on omics-based cell engineering, as the requirement for advanced cell factory platforms for protein production become greater, these systems biology tools will be highly useful to provide genomic-wide understanding of protein production processes and lead to further rational engineering in yeast, and the studies mentioned earlier provide excellent illustrations of the power of systems biology for studying the complex protein secretory pathway.

**Conclusions and perspectives**

From the discussions above, it is clear that there are many examples where engineering different parts of the protein secretion pathway has resulted in improvement of heterologous protein production by S. cerevisiae. The availability of efficient expression systems, fermentation techniques, combined with the advances in systems and synthetic biology has secured yeast as an important platform for many protein productions.

To obtain higher yields and higher quality proteins, secretion pathway engineering will be further applied to increase the protein secretion capability. Additional studies on quality control mechanism in ER are required to understand the cellular response to protein folding burden.

Still current engineering strategies are often only successful for a single protein, and they do not result in the establishment of a generally improved cell factory platform for heterologous protein production. Thus, with the objective to establish such a platform, there is clearly a need for improved knowledge about how the flux through the secretory pathway is controlled by the individual steps in the pathway.

Considering the complexity of protein production and secretion with the involvement of a very large number of components, such knowledge can only be obtained through integrated analysis of the complete system/pathway. Such integrated analysis should preferentially be performed using different engineered strains producing different types of proteins to understand the full spectrum of states the yeast protein production system can express. This kind of study could be carried out through expressing several different types of proteins, at best involving small nonglycosylated proteins like human insulin and more complex proteins such as highly glycosylated proteins with a large number of disulfide bonds like erythropoietin, in many different engineered strains, for example, strains that have over-expression of different foldases and isomerases. Through detailed analysis of these strains, for example, using different omics techniques and quantitative analysis of the secretion kinetics, using, for example, pulse-chase experiments, grown at different environmental conditions, it will be possible to establish a large dataset that would allow for advanced correlation analysis. Such correlation analysis could, for example, lead to identification of whether there is a correlation between expression and production for small and simple proteins or whether there is consistently an UPR for more complex proteins, independent of expression strength. Such correlations may lead to a number of hypotheses that can then form the basis for more detailed experiments, for example, on the role of individual proteins (or group of proteins) on protein synthesis and secretion. Results from these experiments can further be evaluated in the context of specific models for protein synthesis and secretion, and the end result of this kind of study may be a rather detailed mathematical model for...
these pathways, in analogy with models build for metabolism (Soh et al., 2012). Besides allowing for quantitative analysis of the role of the different steps in the pathways, such models can be used to guide engineering design of new cell factories (Tyo et al., 2010). Another path often used for in metabolic engineering for improved metabolite production is a combination of adaptive evolution (Çakar et al., 2012) combined with detailed phenotypic analysis to identify novel metabolic engineering targets, an approach generally referred to as inverse metabolic engineering (Oud et al., 2012).

Even though there are already some examples of mathematical models for specific subprocesses, for example, transcription and translation, there are currently no detailed mathematical models for the overall protein production process in yeast. An obvious first step would be to use existing mathematical models for glycosylation in CHO cells (Shelikoff et al., 1996; Umana & Bailey, 1997; Krambeck & Betenbaugh, 2005) and expand them to predict glycosylation in yeast. By this, we would better understand how both native and heterologous proteins are glycosylated and could use this knowledge to enhance our understanding of late secretory pathway sorting. Furthermore, there are relatively few studies where omics technologies have been used to their full potential to study the global effect on cellular function to, for example, the UPR. Compared with metabolism, where very detailed mathematical models have been set up and are used for designing pathway engineering strategies, there is much development needed before similar strategies can be used for designing novel engineering strategies for improving protein production. Thus, we conclude that even though there are currently very few examples of how systems biology has contributed to both our basic understanding and engineering of protein synthesis and secretion, systems biology has much to offer in this research field.

Acknowledgements

We thank NIH F32 Kirschstein NRSA fellowship, The Knut and Alice Wallenberg Foundation, EU Framework VII project SYSINBIO (Grant no. 212766), European Research Council project INSYSBIO (Grant no. 247013), and the Chalmers Foundation for funding.

Authors’ contribution

J.H., K.E.J.T. and Z.L. contributed equally to this work.

References


Rudolph HK, Antebi A, Fink GR et al. (1989) The yeast secretory pathway is perturbed by mutations in PMR 1, a member of a Ca2+ ATPase family. Cell 58: 133–145.


therapeutics by directed evolution. Drug Discov Today 8: 118–126.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Cytosolic processing depends on the protein pre-signal and its interaction with the signal recognition particle.

Fig. S2. ER processing includes folding, glycosylation and disulfide bond formation.

Fig. S3. Many degradation pathways can be utilized to remove misfolded proteins from the ER.

Fig. S4. Golgi processing includes additional glycosylations, pro-signal cleavage, and targeting proteins for vacuole or exocytosis.

Fig. S5. Post Golgi processing takes place in the trans-Golgi network.

Table S1. Expression levels for different heterologous proteins in S. cerevisiae.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.
Proteins, including enzymes and building blocks of life, play crucial roles in cell signaling, immune systems and the cell cycle. Many human proteins have important values or great potentials as biopharmaceutical. Since the first recombinant pharmaceutical, recombinant human insulin produced from *Escherichia coli*, was approved for clinical use, recombinant DNA technology and protein engineering have established an efficient tailor-made industry for protein production. Now there are over 300 biopharmaceuticals proteins and antibodies on the market, with more than $100 billion of sales. In addition, around 240 monoclonal antibody products and 120 recombinant proteins are in clinical trials.

The yeast *Saccharomyces cerevisiae* is a widely used cell factory for the production of fuels, chemicals, and it also provides a platform for the production of many heterologous proteins of medical or industrial interest. In this thesis, random and rational approaches, such as vector design, host engineering, fermentation analysis, UV mutation, coupled with high-throughput systems biology techniques (including whole genomic sequencing, microarray analysis and flux analysis) and integrated analysis (Reporter feature technique), were employed to engineer cellular properties more effectively and purposefully to construct cell factories for protein production. Our research provided a deep understanding of the processing of protein secretory pathway, proposed targets for future engineering, as well as shed lights for basic cellular metabolisms.

Zihe Flora Liu
Systems and Synthetic Biology (Sys²Bio)
Department of Chemical and Biological Engineering
Chalmers University of Technology
Sweden