The Interrelation of Environmental Factors and Heterologous Protein Production in *Pichia pastoris* -A Systems Biology Approach-

Dissertation

Zur Erlangung des Doktorgrades an der Universität für Bodenkultur Wien

Eingereicht von

Mag. Martin Dragosits

Betreuer:

Ao.Univ.Prof. DI Dr Diethard Mattanovich

Department für Biotechnologie

Universität für Bodenkultur Wien

Wien, Dezember 2009

Acknowledgements

I want to thank my supervisor Diethard Mattanovich for the great chance to work in his lab at the University of Natural Resources and Applied Life Sciences Vienna. Thanks for the possibility to get in touch with applied aspects of microbiology, for the guidance during my thesis but also for the freedom to realize my own ideas.

I want to thank my colleagues in the lab. They made the last years a wonderful experience by providing a competitive but also positive working atmosphere. Thanks for all the fruitful discussions, advices, relaxing coffee brakes, cookies, candies and parties. Special thanks to Gitti for her advice during the work in the lab and during the preparation of this thesis and M^2, Gerhard, Hansi and Stefan for their help and advice during bioreactor cultivations.

Thanks to all the people of the Genophys project I got to know in the last years for the motivating discussions, joint experiments and the nice conferences and meetings.

I want to thank my family and friends who accompanied me during the last years. Special thanks belong to my father. Thank you for encouraging me to make my own way and supporting me all along during my undergraduate as well as graduate studies. Special thanks also belong to Julia. Thank you for your patience and help in good as well as hard and uncertain times.

Table of Contents

Abstract

Microorganisms are well established for the production of technical and therapeutical proteins, with *Pichia pastoris* being on the forefront of fungal host organisms for recombinant protein expression. However, especially the production of complex proteins turned out to be difficult in microbial host cells.

Embedded in a trans-national research consortium this work aimed at the investigation of the impact of environmental factors on heterologous protein production in *Pichia pastoris*. Environmental factors such as temperature, osmolarity and oxygenation can have a tremendous effect on recombinant protein production. Furthermore, these parameters can be easily manipulated to improve process performance. Thus, the analysis of their effect on cellular physiology and their interrelation with product formation is of great interest. To gain insights into the physiological changes in *P. pastoris* during cultivation at different environmental conditions, DNA microarray analysis, proteome analysis and flux calculations were applied. These Systems Biology techniques represent valuable tools to assess biological regulation at different levels.

Temperature and oxygenation turned out to be of special interest in the context of recombinant protein production in *P. pastoris*. A decrease of both factors lead to increased specific productivity of the antibody fragment Fab 3H6, whereas osmolarity of the culture medium had no effect on recombinant Fab 3H6 production. It is known that manipulation of processes linked to protein folding and secretion can have a positive effect on protein production in microbial hosts. The application of proteome and transcriptome analysis resulted in the identification of additional potential bottlenecks that might hamper protein production in *P. pastoris*. The obtained data can serve as a valuable basis for future rational strain engineering.

Additionally, this Systems Biology approach highlighted the complex interrelation of the unfolded protein response (UPR) and the environmental factors temperature and osmolarity. The obtained data indicate that an increase of any of the two parameters during steady state cultivation lead to the induction of the unfolded protein response or at least to increased folding stress, thus pointing out the tremendous importance of the UPR for the capability of *P. pastoris* to cope with environmental changes.

Zusammenfassung

Mikroorganismen werden heutzutage erfolgreich zur Produktion von technisch und therapeutisch relevanten Proteinen eingesetzt. *Pichia pastoris* zählt dabei zu den bedeutendsten Organismen aus der Gruppe der Hefen. Trotz vieler Fortschritte auf dem Gebiet der Proteinproduktion in Mikroorganismen, gestaltet sich die Herstellung von komplexen Proteinen in mikrobiellen Zellen häufig als schwierig.

Integriert in ein internationales Forschungskonsortium war das Ziel dieser Arbeit den Einfluss von Umweltfaktoren auf die Produktion von rekombinanten Proteinen in *P. pastoris* zu untersuchen. Umweltfaktoren wie Temperatur, Osmolarität und Sauerstoffversorgung haben großen Einfluss auf das Wachstum und die Physiologie von Mikroorganismen und können in biotechnologischen Produktionsprozessen leicht manipuliert werden. Daher stellen sie eine gute Möglichkeit dar, um Produktionsprozesse zu optimieren. Durch die Anwendung von DNA Microarrays, Proteom Analyse und metabolischen Flux-Kalkulationen konnten wertvolle Einblicke in die physiologischen Änderungen in *P. pastoris* als Antwort auf unterschiedliche Umweltbedingungen gewonnen werden. Durch die Kombination dieser systembiologischen Methoden können zelluläre Regulationsmechanismen auf verschiedenen Ebenen untersucht werden.

Wie sich herausstelle, sind die Umweltfaktoren Temperatur und Sauerstoffversorgung von besonderem Interesse für die Produktion eines Antikörperfragments in *P. pastoris*. Sowohl die Absenkung von Temperatur als auch der Sauerstoffversorgung resultierten in einer erhöhten Produktion des Antikörperfragments Fab 3H6. Eine Änderung der Osmolarität hatte jedoch keinen Einfluss auf die Produktivität. Es ist bereits bekannt, dass der zelluläre Proteinfaltungs- und Sekretionsapparat einen Engpass bei der Produktion von rekombinanten Proteinen darstellen kann. Durch Anwendung der genannten Methoden konnten weitere potentielle Engpässe identifiziert werden. Die gewonnenen Daten könnten sich als wertvolle Basis für zukünftige Stammoptimierung von *P. pastoris* erweisen.

Weiters konnten wertvolle Einblicke in die komplexe Beziehung von Umwelteinflüssen und dem Stress verbunden mit Proteinfaltung gewonnen werden. Die Möglichkeit effizient auf Proteinfaltungsstress zu reagieren ist für *P. pastoris* auch von großer Bedeutung um auf umweltbedingten Stress, wie zum Beispiel erhöhte Temperatur oder erhöhte Osmolarität zu reagieren.

Aim of the Study

The aim of the study was to investigate the effect of different environmental growth conditions on antibody fragment secretion in the yeast *Pichia pastoris*. Furthermore, Systems Biology methods such as microarray analysis and 2D-DIGE were established for *P. pastoris* to gain insights into the physiological changes upon a shift of growth conditions on the whole cell level.

Keywords

Pichia pastoris
Heterologous Protein Production
Antibody fragment
Environmental Factors
Temperature
Osmolarity
Oxygenation
Systems Biology
2D-DIGE
Microarray Analysis
Flux Analysis

1. Introduction

Pichia pastoris is a well-established host organism for recombinant protein production. Both, intracellular and secretory production can be applied to produce proteins for a wide variety of applications such as structural analysis, enzyme characterization, technical and therapeutical applications (Lin Cereghino *et al.*, 2001). Although other production systems such as mammalian cell culture and plant systems offer some particular benefits, especially in the case of therapeutical proteins, microbial host systems still play an essential role in protein production for human administration (Ferrer-Miralles *et al.*, 2009). *P. pastoris* was the target organism for outstanding research to produce humanized glycosylations in yeast (Jacobs *et al.*, 2009, Hamilton & Gerngross, 2007, Hamilton *et al.*, 2006). These particular advances point out the importance and great potential of *P. pastoris* as host organism for recombinant protein production.

Despite the extensive use of *P. pastoris* as protein production platform, highlighted by approximately 2800 articles, which appear when '*Pichia pastoris*' is entered as keyword in the NCBI PubMed database, molecular biology knowledge of this yeast is rare. As *P. pastoris* is a methylotrophic yeast, the methanol assimilation pathway of *P. pastoris* is known. Furthermore, as growth on methanol as carbon source depends on the proliferation of peroxisomes, *P. pastoris* is also a model organism for studying peroxisome biogenesis and pexophagy (Dunn *et al.*, 2005).

Regarding recombinant protein production, especially secretion of heterologous proteins turned out to be a difficult task, resulting in unpredictable product yields for different recombinant proteins. However, it is well established that environmental factors can have a tremendous effect on protein production in *P. pastoris* and generally also in other microbial and higher eukaryotic expression systems. Alterations in temperature, oxygenation, osmolarity and pH can have a positive effect on production of recombinant proteins (Lin *et al.*, 2007, Li *et al.*, 2001, Wu *et al.*, 2004, Shi *et al.*, 2003). The physiological response to these factors results in a vast amount of adaptations of the host cell and consequently influences mechanisms such as the protein folding and the secretion machinery. Thus, heterologous protein production and the response to environmental factors are highly interrelated and need to be closely investigated.

With nearly no physiological information of *P. pastoris* at hand, many if not most approaches to improve heterologous protein production involve bioprocess engineering rather than rational strain engineering.

Nowadays modern molecular biology offers a wide variety of techniques to study cellular reactions on a systems wide level (Graf *et al.*, 2009). Therefore, it becomes feasible to investigate not only classical model organisms such as *Escherichia coli* and *Saccharomyces cerevisiae* but also other industrially important microbial species, in our case *P. pastoris*, on multiple levels to deepen our understanding of the physiology of this host as well.

1.1 The Genophys Project

This thesis was embedded in a transnational research consortium, the Genophys project. Researchers from different European countries, including Austria, Germany, Italy, Spain and Finland and with expertise in recombinant protein production in several microbial species contributed to this project. *S. cerevisiae*, *P. pastoris*, *Trichoderma reesei*, *E. coli and Pseudoalteromonas haloplanktis* were used as host organisms for the production of a dimeric model protein. The common aim was to gain insight into the physiological response of microorganisms to environmental factors and its interrelation with recombinant protein production. Although several publications on the effect of environmental factors on recombinant protein production are available, the multitude of different approaches that were used in these studies, makes it very difficult to identify common features and differences among different host organisms. To gain such information about potential common responses and species-specific differences, similar experimental setups were used in all partner laboratories.

Generally, recombinant protein production often results in quite different protein yields depending on the properties of the protein of interest. Nevertheless, especially the production of heterodimeric and / or disulfide-bonded proteins is difficult, resulting in rather low specific productivities when microbial host systems are used. Therefore, antibody fragments, such as Fab fragments, represent ideal model proteins for studies on the production of complex proteins. Fab fragments are heterodimeric proteins composed of a full length antibody light chain and a shortened antibody heavy chain, both chains linked by a disulfide bond.

Massive efforts were accomplished to produce antibody fragments and full length antibodies in prokaryotic and eukaryotic microbial host cells (Aldor *et al.*, 2005, Mazor *et al.*, 2009, Wacker *et al.*, 2002, Jacobs et al., 2009, Burgess-Brown *et al.*, 2008). In the current study the

same antibody fragment, namely the Fab 3H6 (Gach *et al.*, 2007, Kunert *et al.*, 2002) was expressed in all host organisms mentioned before. In the case of *P. pastoris* the Fab 3H6 was expressed under the control of the constitutive glyceraldahyde-3-phosphate (GAP) promoter and the alcohol oxidase (AOX1) terminator. The *S. cerevisiae* alpha factor signal sequence was used to secrete the Fab into the culture medium. Different temperatures, osmolarities and oxygenation conditions were applied during steady state cultivation to analyze the effect on Fab 3H6 productivity and the resulting physiological changes on a systems wide level using transcriptome and proteome analysis.

1.2 Recombinant Protein Production and Protein Folding Stress in Bacteria and Fungi

Recombinant protein production depends on the exploitation of the cellular transcription, translation and protein folding and secretion machinery of the host cell. Consequently, it is obvious that cellular protein folding, secretion and quality control mechanisms represent a severe bottleneck in heterologous protein production. However, in this section just a brief overview of this topic will be given as the main features of recombinant protein production in both prokaryotic and eukaryotic hosts are summarized in a recent review of the Genophys consortium (Gasser *et al.*, 2008).

In prokaryotic organisms such as the most prominent member *E. coli*, which lack intracellular membrane compartmentalization, intracellular protein folding is performed in the cytoplasm. Although recombinant proteins can be targeted to the periplasm of the bacterial cell, intracellular production is very common in prokaryotic hosts. After their synthesis, proteins have to reach their native conformational state. If this cannot be achieved, proteins are either targeted to degradation or form aggregates of insoluble protein, so-called inclusion bodies (IBs). Thus, intracellular production of heterologous protein production often results in the formation of these IBs (Baneyx & Mujacic, 2004). As soluble proteins are thought to be superior to insoluble protein in terms of quality, a very common approach is to reduce the amount of IB formation. This can be achieved by altered growth conditions such as decreased growth temperature (Vasina & Baneyx, 1997, Vasina & Baneyx, 1996). Nevertheless, prokaryotic cells possess mechanisms to support proteins on their way to their native conformation. In *E. coli* among this protein family of molecular chaperones the most important members are *DnaK*, *DnaJ*, *GrpE*, *GroEL* and *GroES*. Normally, these proteins are induced upon exposure to heat stress aiming at the reduction of protein folding stress (Hartl,

1996). In recombinant protein production co-overexpression of such heat shock proteins was also successfully applied to reduce the insoluble protein fraction and to increase the amount of soluble protein (Gupta *et al.*, 2009).

In the last years, the dogma of directly relating protein solubility and quality is getting pitted (Ferrer-Miralles et al., 2009). It was always of major importance to increase the soluble protein fraction as it was thought that it represented the native and thus biologically active conformational state. However, it has been shown in the last few years that biological activity of recombinant proteins produced in bacteria is not limited to the soluble protein fraction but that properly folded and active proteins also occur in bacterial IBs (Doglia *et al.*, 2008). Additionally, cultivation below the optimal growth temperature affects protein quality in IBs as well (Doglia *et al.*, 2008, Vera *et al.*, 2007). Thus, IBs may be directly applied to certain industrial applications (Martínez-Alonso *et al.*, 2009), as demonstrated by the crosslinked IB (CLIB) technology (Nahálka *et al.*, 2008).

Many recombinant proteins of interest harbor disulfide bonds. However, under the reducing conditions in the cytoplasm of *E. coli*, disulfide bonds cannot be produced. Thus, these circumstances display a major obstacle that hampers protein production in bacterial cells. To overcome these problems two commonly applied possibilities exist. As the periplasm and the culture supernatant represent more oxidized environments than the cytoplasm, recombinant proteins can be targeted into these compartments. Furthermore, genetic engineering was applied to generate a strain known as Origami. This strain carries mutations in the thioredoxin and glutathione pathways and expresses an intracellular form of the disulfide-bond isomerase Dbsc. The Origami strain has been successfully used to produce functional antibody fragments in the cytoplasm of *E. coli* (Jurado *et al.*, 2002).

Besides *E. coli* as the most common bacterial host organism for recombinant protein production other prokaryotic expression systems have been developed. Other Gram-negative species such as *P. haloplanktis* (Cusano *et al.*, 2006) and *Caulobacter crescentus* (Bingle *et al.*, 2000), but more important Gram-positive species like *Bacilli* such as *Bacillus subtilis* (Schumann, 2007) are used as well. The advantages and drawbacks of the most frequently used bacterial expression systems are outlined in more detail in a recent review by Kay Terpe (Terpe, 2006).

Although the same principles of protein folding apply to eukaryotic systems such as yeasts and filamentous fungi, the underlying regulatory mechanisms are more complex than in prokaryotic organisms due to intracellular compartmentalization. Eukaryotic cells evolved

specialized compartments to perform distinct tasks. Folding of proteins that are determined for secretion is performed in the endoplasmic reticulum (ER) and subsequent maturation and sorting takes place in the Golgi apparatus. Generally, targeting of recombinant proteins to the extracellular growth medium can be achieved more easily in yeasts and fungi than in Gram-negative bacteria such as *E. coli*.

Protein folding in the ER involves several hsp70 molecular chaperones such as BiP/Grp78 and calnexin. Furthermore, eukaryotic cells can perform oxidative protein folding (disulfide bond formation) in the ER by the action of protein disulfide isomerases. One major control point in eukaryotic protein folding is the exit from the ER (Shuster, 1991). Prolonged retention of proteins in the ER due to misfolding leads to the redirection of these proteins to the ER associated protein degradation pathway (ERAD) and subsequent degradation by the cytosolic 26S proteasome (Vembar & Brodsky, 2008).

If a protein is targeted to the secretory pathway, the next step involves translocation to the Golgi apparatus. One benefit of eukaryotic systems such as yeast is that they can produce post-translational modifications (PTMs) including glycosylation. Such PTMs are often essential for the biological activity or serum stability of recombinant proteins and do not naturally occur in prokaryotic host systems such as *E. coli*. Concerning such glycosylated secreted proteins, the core glycan is already added to the protein in the ER. However, to prevent misglycosylations the UDP-glucose:glycoprotein glycosyltransferase (UGT) system exists (Kleizen & Braakman, 2004). After the protein has successfully been moved to the Golgi apparatus PTMs are completed by several additions and trimming steps and then released into the supernantant (Munro, 2001, van Vliet *et al.*, 2003).

Protein folding and transport in yeasts is a very complex mechanism involving a large number of enzymes. Recombinant protein production eventually leads to protein misfolding and overload of the ER. To avoid negative effects of protein secretion on cell viability several rescue mechanisms evolved to protect the host cell from these detrimental effects (Gasser *et al.*, 2008). A key player in this rescue system is the unfolded protein response (UPR). When improperly folded proteins accumulate in the ER, a signaling cascade involving the chaperone BiP, Ire1 and Hac1 is triggered (Travers *et al.*, 2000, Mori *et al.*, 1996). The transcription factor Hac1 translocates into the nucleus and induces the transcription of chaperones and protein folding helpers such as calnexin, BiP and Pdi1. In *P. pastoris* ribosome and membrane proliferation is also induced by Hac1 (Graf *et al.*, 2008). Several research articles highlight that the co-production of proteins, which are involved in the UPR, can have a beneficial effect

on recombinant protein production in various fungal host organisms, such as *S. cerevisiae*, *P. pastoris*, *Hansenula polymorpha* and *T. reesei* (reviewed in Gasser et al., 2008).
Considering present day industrial processes, *S. cerevisiae* is still the most prominent fungal expression host, especially concerning therapeutic proteins (Gerngross, 2004) but alternatives are advancing. Fungal and yeast species such as *Aspergillus niger*, *T. reesei*, *P. pastoris*, *Hansenula polymorpha* and *Kluyveromyces lactis* are continuously improved and applied for heterologous protein production (Gerngross, 2004, Porro *et al.*, 2005). Especially in the last few years, non-conventional yeasts have entered the spotlight and are very promising hosts for the production therapeutic proteins. In contrast to bacterial cells, such as *E. coli*, no endotoxins (e.g. LPS) are produced and extensive glyco-engineering enables human like glycosylation of heterologous proteins (Hamilton & Gerngross, 2007).

Substantial improvements of heterologous protein production were achieved by influencing the protein folding machinery in prokaryotic and eukaryotic cells. However, different or even contradictory results were achieved for different model proteins as well as different host organisms, resulting rather in a trial and error approach than in targeted and rational strain engineering. One major obstacle towards directed strain engineering in alternative host systems seems to be the lack of detailed knowledge of host physiology, molecular biology and biochemistry. Affordable DNA sequencing methods and high throughput transcript profiling such as microarray analysis will be useful to close this gap for alternative hosts and uncharacterized strains. Thus, Systems Biotechnology will also become available for non-conventional but highly prospective host systems.

1.3 Systems Biotechnology

Since Systems Biotechnology derives from Systems Biology, first a short overview about the history of Systems Biology will be given. Per definition, Systems Biology aims at the global understanding and modeling of the entire network of reactions in a living cell or even multicellular organisms (Westerhoff & Palsson, 2004). Building a model of the cellular regulatory network depends on transcriptome, proteome as well as metabolome profiling. In contrast to a global Systems Biology approach, Systems Biotechnology rather aims at understanding the processes related to product formation and their improvement, therefore accepting gaps in the description of other cellular processes (Graf *et al*., 2009).

One of the major techniques that made a systems approach feasible, is the fast parallel profiling on the mRNA levels under different conditions. DNA microarrays or DNA chips were developed and became a major tool for comparative parallel analysis in transcript profiling. A DNA microarray is a device with thousands of nucleic acid sequences immobilized on a solid support, usually a glass slide. By hybridization of e.g. fluorescently labeled complementary nucleic acids, subsequent signal capturing and computational/mathematical analysis, it is possible to draw quantitative conclusions on mRNA levels when different biological samples are compared. The first approaches towards microarray analysis are somehow hard to find because of parallel developments but actually started in the late 80's and early 90's of the last century, roughly 20 years ago. In 1991 Stephen Fodor, which thereafter became the president of the company Affymetrix, a major distributor of microarrays, and co-workers published an article about a photolithographical technique for spatially addressed and parallel chemical synthesis (Fodor *et al.*, 1991). However, the first articles about microarrays attempted to introduce a new technique rather than to present biologically meaningful data. It took until the mid nineties until the first publications about biological results based on parallel transcript analysis were published (Schena *et al.*, 1995, Schena *et al.*, 1996). Nowadays a wide variety of DNA microarray platforms is available and eventually used for a wide variety of organisms. With the rise of microarrays to a powerful tool in biomedical research other problems related to data analysis, data interpretation and reproducibility emerged. For a more complete overview about the developments and pitfalls of microarray analysis an excellent review was written by Rogers and Cambrosio (Rogers & Cambrosio, 2007). Although DNA microarray analysis is now regularly performed in laboratories all over the world, it became obvious that analysis on the transcript levels can just help to understand a small part of the complex puzzle. The simple scheme "DNA → mRNA → protein" is only partially valid as at each point plenty of regulatory mechanisms interfere. Thus, further techniques on other levels are necessary to complement our knowledge in a systems type manner.

Quantitative proteomics is another major achievement with important contributions to Systems Biology. While transcript profiling is still important, the analysis of a cell's or tissue's proteome is of particular interest as proteins represent the effectors of cellular function and changes on the transcript level cannot be directly extrapolated to the protein level.

A major technique for proteome analysis during the last decades was two-dimensional gel electrophoresis (2D-GE). This technique is based on the combination of isoelectric focusing and conventional SDS-PAGE to separate proteins for the parallel analysis of hundreds or more than thousand of proteins. For comparative analysis of different samples classical 2D-GE relied on a large amount of replica gels and conventional silver staining. A major step forward in 2D-GE technology was the application of fluorescent labeling, similar to microarray analysis. This resulted in a technology named 2D fluorescence difference in gel electrophoresis (2D-DIGE) (Unlü *et al.*, 1997, Tonge *et al.*, 2001). The labeling of samples with fluorescent dyes allowed the application of up to three samples on a single gel and the application of internal standards. Furthermore, fluorescent signals show a much higher dynamic range than classical silver staining. Taken together 2D-DIGE is superior in data acquisition and statistical analysis. Nevertheless, although 2D-DIGE represents a huge improvement in proteome analysis, there are of course several issues. On a typical 2D Gel only a small part of the complete proteome is resolved. Proteins with a generally low intracellular concentration are underrepresented, resulting in poor quantitative results for these proteins. Nevertheless, 2D-DIGE can be easily established and currently represents a major pillar of proteome research. 2D-DIGE also relies on subsequent protein identification via N-terminal sequencing or more commonly used mass spectrometry based methods. Alternatively, several gel-free approaches using only mass spectrometry (MS) as a core technology were developed.

MS per se is not a quantitative method. Thus, several techniques were developed to enable MS based quantitative proteomics. One of the first methods used feeding of different stable nitrogen isotopes, ^{14}N and ^{15}N, to microorganisms such as bacteria and yeast (Oda *et al.*, 1999). An improvement of this method was then achieved by introduction of stable-isotope labeling by amino acid feeding in cell culture (SILAC) by Ong and co-workers (Ong *et al.*, 2002). In contrast to these *in vivo* labeling methods also *in vitro* labeling methods such as ‘isotope-coded affinity tags’ (ICAT) (Gygi *et al.*, 1999) and ‘isobaric tags for relative and absolute quantification’ (iTRAQ) (Ross *et al.*, 2004) were developed. A review discussing the advantages and pitfalls of these methods and also newer approaches such as label free MS based quantification has been published recently (Bachi & Bonaldi, 2008).

Quantification of proteins still represents only one part of proteomic research. Proteins are heavily modified more or less directly after translation but also during their action in the cellular metabolism. These modifications include glycosylations, which have been mentioned before but also other types of modifications such as phosphorylation, lipidation, methylation,

acetylation and ubiquitination. Additionally, the function and interaction partners of proteins should be known. Therefore, the analysis of their modifications and their interactions is also of particular importance to understand a complex biological system. To answer these questions methods such as two-hybrid screenings, affinity purification and protein arrays (Talapatra *et al.*, 2002, Uttamchandani *et al.*, 2006) as well as mass spectrometry are applied. These approaches are described in more detail in a recent review (Abu-Farha *et al.*, 2009).

Of course, the knowledge of the cellular transcriptome and the proteome does not tell the complete story of cellular metabolism. Metabolic Flux Analysis (MFA) was developed in the early nineties. It aims at 'the detailed quantification of all metabolic fluxes in the central metabolism of a (micro)organism' (Wiechert, 2001). Thus, metabolic fluxes are of great interest for the identification of potential bottlenecks, which are consequently targets for genetic engineering.

Metabolites are small molecular weight compounds such as sugar-compounds, amino acids and lipids and represent the actual targets of enzyme activity. Even microbial metabolomes consist of hundreds of major metabolites with different physico-chemical properties. It is therefore obvious that the parallel quantification of all these metabolites is very difficult, thus resulting in only subsets of analyzed metabolites in most studies (Kell, 2004). Several methods were developed and are currently applied to analyze cellular metabolomes. Both, NMR and MS based methods are used for metabolome analysis but due to its superior speed and sensitivity it has been noticed that only MS based methods might have the potential for high-throughput analysis (Sauer, 2004). Usually for MS based techniques chromatographic methods such as gas chromatography (GC) or liquid chromatography (LC) are coupled to MS instruments. Nevertheless, to avoid chromatographic separations, which may depend on derivatization of metabolites, also pure MS methods such as Fourier-transform ion cyclotron resonance (FT-ICR) were developed (Brown *et al.*,2005, Kosaka *et al.*, 2000).

For high throughput analysis not only adequate metabolomic methods and equipment have to be supplied but also adequate and efficient cell cultivation systems are necessary. Continuous cultivations, thus steady state conditions, are of extraordinary importance for transcriptome, proteome and metabolome analysis (Knijnenburg *et al.*, 2009, Regenberg *et al.*, 2006), but suffer the drawback that high-throughput attempts are not feasible yet. For large scale parallel analysis (quasi) steady states, using maximum exponential growth in batch culture might also be used (Sauer, 2004).

MFA relies on the determination of intracellular and extracellular fluxes to generate a mathematic flux model. By the combination of the measurement of real fluxes in *in-vivo*

experiments and the assumed fluxes of the simulated experiment, the individual parameters of the flux model can be fitted to produce an *in silico* model of a cellular fluxome (Wiechert, 2001).

Experiments using the systems approach with its diversity of techniques to create a 'complete' network model depend on genomic and biochemical data of the particular organism. It is therefore obvious that these experiments are limited to organisms where this information is available. To obtain this kind of data, DNA sequencing of the organism's genome is necessary. A huge effort was undertaken to obtain the full or partial genome sequence of important organisms. For many years genome sequencing itself depended on the Sanger sequencing method and consequently on enormous expenses to obtain genomic data. In the recent years several new sequencing techniques were developed, which allow fast and parallel sequencing and are time saving and relatively cheap compared to conventional Sanger sequencing. The principles of these so-called next generation sequencing (NGS) methods such as 454 sequencing (Roche), the Illumina Solexa system and the ABI SOLiD system have been reviewed recently (Ansorge, 2009). These sequencing techniques and newer approaches, so-called Third Generation Sequencing (TGS) methods, which aim at reducing the time and costs even more, will make genome data of less popular but very relevant organisms publicly available. Furthermore, due to their cost effectiveness NGS and TGS methods can also be applied for RNA sequencing, expression profiling and ChIP-Seq (chromatin immunoprecipitation-sequencing) experiments (Ansorge, 2009, Graf *et al*., 2009).

The described techniques are already available for several years, and Systems Biotechnology approaches for rational strain engineering have been applied in 'White Biotechnology'. Nevertheless, due to the complexity of the involved mechanisms it has been rarely applied for recombinant protein production (Graf *et al.*, 2009).

It should be mentioned that the presented approaches to analyze cells on a systems level generate a 'bioinformatic burden' that arises from the complexity and the enormous amount of data produced by these approaches. Nonetheless, the manifold of data yet available for common model organisms is a great advantage for the development of metabolic models of less characterized organisms such as *P. pastoris*. The knowledge acquired in previous studies represents some kind of rear cover for the fast and efficient generation of metabolic networks and biological process descriptions of less commonly used organisms in academic research and industry.

2 Results

In this section, only a minimum of references is cited as relevant literature is cited in the appended journal publications and submitted manuscripts. It should be noted that many people contributed to the work and the results presented in this thesis. For the section 2.3 'Analysis of the effect of *HAC1* overexpression by *P. pastoris* specific microarrays' the main part of the work was carried out by Alexandra Graf and Brigitte Gasser and for part 2.6 'The effect of oxygenation on *P. pastoris*' the lion's share was achieved by Kristin Baumann and Michael Maurer. Nevertheless, it has been included in this thesis as this work was essential for the development of the systems level approach and a thorough investigation of the effect of environmental factors on recombinant protein production in *P. pastoris*.

2.1 Establishment of Systems Biology tools for *P. pastoris*

To carry out a systems level analysis of the effect of environmental growth factors on *P. pastoris* physiology appropriate techniques had to be established. Whereas microarrays were already available for many species, *P. pastoris* specific microarrays were unavailable due to the lack of a publicly accessible genome sequence. Although heterologous microarray hybridization has been successfully applied for *P. pastoris* (Sauer *et al*., 2004), it was necessary to develop *P. pastoris* specific microarrays to utilize to full potential of microarray analysis (Graf *et al.*, 2008). Additionally, 2D-DIGE technology was established for *P. pastoris* (Dragosits *et al.*, 2009). Protein identification by LC-ESI-MS/MS resulted in a reference intracellular protein map, which so far comprises approximately 90 identified protein spots. Furthermore, by applying the same principals of 2D-GE and mass spectrometry, a map of the *P. pastoris* secretome could also be established. Last but not least, metabolic flux calculations, as a third pillar of systems level analysis, were performed in cooperation with Prof. Pau Ferrer's research group in Barcelona.

2.2 Special features of the protein expression host *P. pastoris*

One of the reasons why *P. pastoris* became attractive as a protein expression host system is its ability to grow on methanol as sole and formerly cheap carbon source. Furthermore, in contrast to *S. cerevisiae*, *P. pastoris* is a Crabtree negative yeast. Upon aerobic growth at high

glucose concentration negligible or no ethanol is produced. By avoiding this overflow metabolism in *P. pastoris*, higher cell densities can be achieved in batch and fed batch cultivations. This advantage of the *P. pastoris* expression systems is known for a long time, but with a genome sequence at hand (De Schutter *et al.*, 2009, Mattanovich *et al.*, 2009) a closer investigation of the reasons for these advantages is possible. The genome sequence of *P. pastoris* e.g. highlights a potential reason for the differences of growth kinetics in *S. cerevisiae* and *P. pastoris*. In contrast to *S. cerevisiae* fewer glucose transporters are present in *P. pastoris.* In *P. pastoris* only 2 instead of 20 hexose transcporters were found based on sequence similarity (Mattanovich *et al.*, 2009). Furthermore, two potential high affinity transporters with similarity to high affinity transporters of *Kluyveromyces lactis* were identified. As glucose uptake largely determines growth kinetics these results are in good agreement with the actual aerobic growth characteristics of *P. pastoris* on glucose.

Another advantage of *P. pastoris* is the generally low secretion of endogenous proteins into the culture medium. Thus, the extracellular proteome (secretome) was analyzed *in silico*, using signal sequence prediction tools, and *in vivo*. Two dimensional gel electrophoresis of the culture supernatant of glucose limited *P. pastoris* grown in chemostat culture revealed that only very few endogenous proteins were released into the culture medium. The predicted secretome comprises 88 proteins, whereas only 28 proteins could be verified under the culture conditions chosen (Mattanovich *et al.*, 2009). Among these proteins present in the culture supernatant, also most likely cell wall associated proteins such as Gas1 were identified. No proteases were identified by LC-ESI MS/MS analysis and the actual proteolytic activity of the culture supernatant was very low, most likely arising from the 2% of dead and lysed cells in chemostat cultures. These results demonstrate that although a methanol based feed to induce recombinant protein production is the most common method in *P. pastoris*, a glucose based feed is of great advantage as a minimum of proteins is secreted into the culture supernatant and cell viability is generally higher, thus resulting in a much lower contamination with intracellular proteases.

2.3 Analysis of the effect of *HAC1* overexpression by *P. pastoris* specific microarrays

A publicly available genome sequence was not present at the beginning of the project. However, a partially annotated sequence was commercially accessible through Integrated Genomics (www.integratedgenomics.com). Additional gene finding and annotation allowed the identification of approximately 4000 genes with annotated function and about 11000

potential open reading frames (ORFs). Based on this information *P. pastoris* specific microarrays using the Agilent platform were developed (Graf *et al.*, 2008).

These microarrays were applied to analyze the effect of DTT treatment and *HAC1* overexpression in *P. pastoris.* Both, DTT treatment and *HAC1* overexpression induced the core UPR, including chaperones and folding enzymes. The analysis of the effect of *HAC1* overexpression on cellular physiology is therefore of special interest considering recombinant protein production.

Apart from the core UPR, the obtained data indicated crucial differences in the transcriptional response upon exposure to DTT or *HAC1* overexpression. In contrast to DTT treatment, *HAC1* overexpression resulted in a massive upregulation of genes involved in ribosomal biogenesis, RNA metabolism and translation. Furthermore, only about 50% of the genes known to be involved in DTT response in *S. cerevisiae* behaved similar in *P. pastoris*.

These results highlight that, although DTT is regularly used to induce UPR, responses to such as chemical stimuli are also triggered, thus being different from the response induced by recombinant protein. Additionally, the substantial differences in the response to DTT in *S. cerevisiae* and *P. pastoris* highlight the necessity of improving our knowledge of other industrially relevant microorganisms. Due to these differences, systems level models and conclusions on the biology of microorganisms cannot be simply based only on classical model organisms such as *S. cerevisiae*.

2.4 The effect of growth temperature on *P. pastoris*

In chemostat cultures growth temperature had a significant impact on the specific productivity (q_P) of the antibody fragment Fab 3H6. We could show that a two-fold increase and a three-fold increase of q_P were achieved by decreasing the growth temperature from 30 to 25°C and from 30 to 20°C, respectively. As already outlined in the introduction, the positive influence of growth temperature below the so-called optimal growth temperature has been previously reported for *P. pastoris* and other organisms. However, concerning *P. pastoris*, the methanol inducible AOX1 systems is widely used and better established than the GAP system used in the current study. As methanol is a toxic compound and results in high cell lethality during a standard fed batch procedure, higher cell viability and a decrease of protease release were considered to be responsible for improved production upon temperature reduction during methanol based processes. Nevertheless, in the presented experimental setup, using glucose as carbon source, cell viability was >98% and the contamination of the culture broth with

intracellular proteases was insignificant. Thus, other parameters related to changes in the host cell physiology had to be responsible for the increased q_P during growth at low temperature. Therefore, the mentioned systems level analyses were applied to shed light onto the physiological reasons for the increased production of Fab 3H6.

Although 2D-DIGE might be technically limited by detecting mainly highly abundant proteins and leaving the majority of the cellular proteome unobservable, it proved to be a valuable tool to monitor the changes, which accompanied growth at decreased temperature. At the proteome level at least five cellular processes, which were affected by temperature, could be identified. Whereas proteins involved in ribosome biogenesis and assembly as well as amino acid metabolism showed higher levels at 20°C, proteins involved in energy metabolism, protein folding and oxidative stress response showed decreased levels at 20°C. Especially the impact on energy metabolism was quite unexpected. Proteins involved in the tricarboxylic acid (TCA) cycle, such as Aco1p, Cit1p, Fum1p and Mdh1p showed lower levels at 20°C. Consequently, with biomass data, metabolite concentrations in the culture broth and off gas data at hand, metabolic flux calculations were performed. The obtained data strengthened the results from 2D-DIGE as the metabolic model of the central carbon metabolism predicted a slight decrease of the glucose uptake rate but also a decrease in the flux through the TCA cycle at 20°C.

Many molecular chaperones and proteins involved in the unfolded protein response, which play a crucial role in protein folding and turnover showed lower levels at 20°C. Especially decreased levels of the UPR sensor protein Kar2p/BiP and the stress induced chaperones Ssa4p and Hsp82p lead to the conclusion that protein folding stress was greatly reduced at low temperature (Dragosits *et al.*, 2009). Consequently the decrease of proteins involved in oxidative stress response is reasonable as protein folding in the ER is a redox sensitive process, which may lead to the generation of reactive oxygen species (ROS). Additionally, increased levels of proteins related to ribosome biogenesis (Rpp0p and Gsp1p) were also observed, and a similar response to low growth temperature has been reported for *S. cerevisiae*.

Microarray experiments support the data obtained by 2D-DIGE (unpublished data). At the transcript level the upregulation of genes involved in translation, ribosome biogenesis and amino acid metabolism at 20°C was even more evident than on the protein level. Interestingly, genes involved in TCA cycle such as *ACO1*, *FUM1* and *MDH1* did not show significant changes upon a decrease of growth temperature, although the protein levels of

these genes evidently decreased at 20°C. These data indicated substantial post-transcriptional control mechanisms taking place during steady state cultivation of *P. pastoris*.

Additionally the measurement of intracellular solutes showed that trehalose levels decreased during growth at 20°C compared to 25 and 30°C (unpublished data). These data also pointed to a decrease of intracellular (protein folding) stress, because trehalose has been described to serve as a stress protectant to environmental stresses such as temperature in *S. cerevisiae* (Wiemken, 1990).

Altogether, the described events may lead to a decreased energy demand and a decreased flux through the TCA cycle, but also to an increased biomass yield and increased production of Fab 3H6 at 20°C. Although no increased mRNA levels for the Fab 3H6 heavy and light chain with decreased temperature were observed, a higher rate of mRNA translation due to higher translational activity may also contribute to the higher q_P at 20°C.

2.5 The effect of osmolarity on *P. pastoris*

For mammalian cell lines but also for *P. pastoris*, a positive effect of increased osmolarity has already been reported. Nevertheless, especially for mammalian cell culture increased osmolarity does not necessarily lead to increased product titers as increased osmotic pressure constitutes a severe stress condition leading to a decrease of cell viability. We were therefore interested whether increased osmolarity can also have a positive effect on the production of Fab 3H6 under the control of a glycolytic promoter in *P. pastoris*. By applying chemostat cultivation we could show that an increase of osmolarity did not have a significant effect on the q_P of Fab 3H6 in *P. pastoris*. As the response to environmental stresses is still a fundamental question, not yet addressed in *P. pastoris*, we performed 2D-DIGE and microarray analysis to track the differences between cultivation at different osmolarities. Furthermore, as the accumulation of intracellular solutes to compensate high osmotic pressure is a common response of microorganisms, with glycerol being of particular importance (Mager & Siderius, 2002), we were interested whether such compatible solutes were present in *P. pastoris*.

By measuring intracellular polyol concentrations with HPLC we could show that *P. pastoris* produces several intracellular polyols, but in contrast to *S. cerevisiae*, *P. pastoris* rather accumulated arabitol instead of glycerol upon exposure to elevated osmotic pressure. Furthermore, intracellular arabitol levels were higher than glycerol levels even at low osmolarity. This result once again demonstrated that substantial differences between these

two yeast species exist. A detailed investigation of the proteome and transcriptome data revealed further information, which indicated a different response to high osmolarity in *P. pastoris*.

It turned out that in the control strain of *P. pastoris*, which did not express the recombinant Fab fragment, there were more significant changes on the proteome level but also and even more prominent on the transcript level than in the Fab 3H6 producing strain. Whereas transcript data indicated that cell wall integrity signaling, cell wall composition and ribosome biogenesis were affected by high osmolarity, proteome data acquired by 2D-DIGE highlighted the induction of the unfolded protein response at elevated osmolarity in the control strain. Kar2p/BiP, the protein disulfide isomerase Pdi1p and some cytosolic and mitochondrial molecular chaperones, such as Scc1p, Sse1p, Ssz1p and Hsp60p, showed higher levels at increased osmolarity in the control but not in the recombinant protein expressing strain. It has already been shown that UPR induction can occur during osmotic stress in halotolerant yeast species such as *Rhodoturola mucilaginosa* (Lahav *et al.*, 2004) but it has not be reported to happen to this extend in *S. cerevisiae*. Thus we tested the growth of *P. pastoris* as well as *S. cerevisiae* on media containing different NaCl and KCl concentrations. The obtained data fit to the hypothesis that *P. pastoris* shares more similarities with halotolerant yeast species than *S. cerevisiae* as *P. pastoris* was able to grow on higher NaCl and KCl concentrations than *S. cerevisiae*.

It is well established that UPR induction occurs during recombinant protein production. This is due to an overload of the cellular protein folding capacity and therefore poses a major obstacle during recombinant protein production as described in the introduction. In this context and with the knowledge that UPR induction also plays an essential role during osmotic stress response, it seems plausible that a *P. pastoris* strain, which is permanently exposed to protein folding stress, needs less adaptations to cope with elevated osmolarity than a non-producing control strain. From previous microarray experiments it is known that UPR induction also results in upregulation of genes involved in ribosome biogenesis and translation in *P. pastoris* (Graf et al., 2008). Upon high osmolarity an upregulation of ribosome biogenesis was observed in the control strain but not in the Fab expressing strain. It can be concluded from the current data and available literature that a massive cross-talk or interrelation of the UPR with other signaling pathways such as the HOG pathway and the cell wall integrity pathway exits. Due to this crosstalk of signaling pathways and the particular importance of the UPR for osmotolerance, the recombinant protein expressing strain of *P. pastoris* was already pre-conditioned to growth at elevated osmolarity. Thus, the

physiological response on both the proteome and the transcriptional level was lower in the Fab 3H6 expressing strain.

Regarding the optimal osmolarity for *P. pastoris* growth, no clear conclusions can be drawn from the current data. During fed batch cultivations, osmolarities as high as 850mOsmol kg^{-1}, which corresponds to the medium osmolarity setpoint in the current study, are very common. However, as these conditions already lead to arabitol accumulation and transcriptional as well as proteome related responses, the optimal osmolarity may be near the physiological osmolarity. The physiological osmolarity relates to 300mOsmol kg^{-1}, a setpoint, which would be between the low and medium osmolarity setpoints of this study.

2.6 The effect of oxygenation on *P. pastoris*

It has been shown that, similar to growth temperature, decreased oxygen concentrations improved the specific production of antibody fragment Fab 3H6 in *P. pastoris* (Baumann *et al.*, 2008). Additionally, due to the switch from respirative to respirofermentative or fermentative metabolism increasing amounts of ethanol were produced. Furthermore, oxygen-limited and hypoxic growth conditions resulted in a strong decrease of yeast biomass during chemostat cultivation. Although the physiological reasons for the increase of q_P are still unknown but will also be investigated on a systems levels similar to temperature and osmolarity experiments, the results of the oxygenation experiment already lead to a practical application. Based on the data of *P. pastoris* grown in carbon limited chemostats at different oxygenation, an oxygenation/ethanol regulated fed batch strategy was developed (Baumann et al., 2008). This novel strategy resulted in increased volumetric productivity, shorter process times and lower biomass accumulation than a standard linear fed batch commonly applied for *P. pastoris*.

2.7 Comparison of antibody Fab 3H6 production in various microorganisms

The Genophys project, whereof this thesis is a part of, aimed at the comparison of Fab production and the effect of environmental factors in different host organisms. It turned out that *P. pastoris* was the best of the three fungal host organisms in terms of antibody fragment production. Furthermore, *P. pastoris* showed the second highest specific production of all host organisms analyzed in this project. Concerning specific production in the analyzed hosts, eurkaryotic expression systems are regarded to be better equipped for the production of

complex proteins such as heterodimeric Fab fragments. Nevertheless, as some differences in the expression strategies, such as choice of promoter and leader sequences, growth rate and biomass yield existed, the comparison of specific production in different host cells has to be done carefully and may not be representative. Nevertheless, very high titers of Fab 3H6 were achieved at very low biomass concentrations in *P. haloplanktis*. Thus, this study highlights *P. haloplanktis,* a prokaryotic expression system, as a very prospective host organism for recombinant protein production.

As a main goal of the project was not only to compare specific production in several hosts but to analyze the influence of environmental growth parameters, 4 out of 5 organisms, namely *S. cerevisiae, P. pastoris, T. reesei* and *E. coli,* were cultivated at different temperatures. Beside *P. pastoris,* a decrease of growth temperature in carbon limited chemostat cultures had also a positive effect on specific production of Fab 3H6 in *T. reesei* and *E. coli*. In *S. cerevisiae* no change of specific production was reported. Furthermore, *T. reesei* had, similar to *P. pastoris*, a higher biomass yield at lower growth temperature, whereas *S. cerevisiae* had lower biomass yield at lower temperature. The response to different cultivation temperatures might be to some extend species and also strain specific. In *S. cerevisiae* also reduced oxygenation did not have a positive effect on recombinant protein production as it was reported for *P. pastoris*. However, *P. pastoris, T. reesei* and *E. coli* showed increased specific production at lower temperature. Furthermore, *P. haloplanktis* is a psychrophilic bacterium and was cultivation at only one but low temperature, namely at 15°C. Thus, there is evidence that similar, well conserved, physiological responses, which for example lead to increased protein production at decreased growth temperature, might exist even between non-related species.

3 Conclusions

Hitherto unknown relations between recombinant protein production and environmental growth parameters could be established. Both, decreased temperature and decreased oxygenation had a positive effect on specific production of Fab 3H6 in *P. pastoris*. By applying Systems Biology techniques such as 2D-DIGE and microarray analysis some potential reasons for the increased productivity at decreased temperature were highlighted. Considering these data, it becomes obvious that, besides the protein folding and secretion machinery, many other cellular processes may represent targets for rational strain engineering and protein production in *P. pastoris*. As already reported in previous studies fine-tuning of

the translation machinery might be useful to increase specific as well as volumetric production of recombinant proteins in *P. pastoris* cultivation processes. The obtained data further highlight that, similar to bacterial systems, most notably *E. coli*, heterologous protein production obviously results in a metabolic burden for the host cell. So far, this metabolic burden on the cellular energy metabolism has been more or less neglected for yeast expression systems, because it is e.g. only slightly observable in the biomass yield. Nevertheless, the central carbon metabolism might represent an interesting target to manipulate and improve recombinant protein secretion in yeasts such as *P. pastoris* as well.

As a wide variety of genome scale methods is currently available, many studies applying either proteome, transcriptome or metabolome studies have been published in the recent years. However, all methods in the field of system biological research suffer major drawbacks and can only deliver information of a single level of the microbial or cellular control mechanisms. Taking this thesis as an example, it is of great importance to combine as many methods as possible to gain the maximum amount of information. Although substantial knowledge could be gained by analyzing either proteome or transcript data, a combination of both turned out to be much more powerful.

This Systems Biology approach of generating large amounts of data on multiple scales, together with the establishment of well-annotated and publicly available genomes, might contribute to a fast development of systems and metabolic models for alternative expression systems, such as the yeast *P. pastoris* or the psychrophilic bacterium *P. haloplantkis*. Such metabolic models could aid to design *P. pastoris* on a pure rational basis without the need of trial and error attempts for strain engineering. In the last years, slow but substantial progress has been made by the identification and co-overexpression of single genes to improve protein production. With precise metabolic models at hand, it might become feasible to concentrate rather on the manipulation of entire cellular regulatory circuits than on the overexpression and/or deletion of single genes.

Different proteins might cause different physiological responses of the host systems. Thus, various model proteins as well as various expression promoters, terminators and secretion signals should be included in future studies. The combined analysis of such data could lead to the discovery of core parameters that generally determine heterologous protein production and secretion in microbial hosts. Furthermore, this will help to deepen our knowledge of the physiology and the effect of recombinant protein production in *P. pastoris*.

4 Abbreviations

2D-GE	two-dimensional gel electrophoresis
2D-DIGE	two-dimensional difference in gel electrophoresis
ChIP-Seq	chromatin immunoprecipitation-sequencing
ER	endoplasmatic reticulum
ERAD	endoplasmatic reticulum associated degradation
FT-ICR	Fourier-transform ion cyclotron resonance
GC	gas chromatography
HOG	high osmolarity glycerol
HPLC	high performance liquid chromatography
IB(s)	inclusion body(ies)
ICAT	isotope-coded affinity tags'
ITRAQ	isobaric tags for relative and absolute quantification
LC	liquid chromatography
LC-ESI MS	liquid chromatography – electrospray ionization mass spectrometry
LPS	lipopolysaccharide
MFA	metabolic flux analysis
MS	mass spectrometry
NGS	next generation sequencing
ORF(s)	open reading frame(s)
PTM(s)	post-translational modification(s)
q_P	specific productivity
ROS	reactive oxygen species
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SILAC	stable-isotope labeling by amino acid feeding in cell culture
TCA	tricarboxylic acid
TGS	third generation sequencing
UPR	unfolded protein response

5 References

Abu-Farha, M., F. Elisma, H. Zhou, R. Tian, M. Asmer & D. Figeys, (2009) Proteomics: from technology developments to biological applications. *Anal Chem* **81**: 4585-4599.

Aldor, I., D. Krawitz, W. Forrest, C. Chen, J. Nishihara, J. Joly & K. Champion, (2005) Proteomic profiling of recombinant Escherichia coli in high-cell-density fermentations for improved production of an antibody fragment biopharmaceutical. *Appl Environ Microbiol* **71**: 1717-1728.

Ansorge, W., (2009) Next-generation DNA sequencing techniques. *N Biotechnol* **25**: 195-203.

Bachi, A. & T. Bonaldi, (2008) Quantitative proteomics as a new piece of the systems biology puzzle. *J Proteomics* **71**: 357-367.

Baneyx, F. & M. Mujacic, (2004) Recombinant protein folding and misfolding in Escherichia coli. *Nat Biotechnol* **22**: 1399-1408.

Baumann, K., M. Maurer, M. Dragosits, O. Cos, P. Ferrer & D. Mattanovich, (2008) Hypoxic fed-batch cultivation of Pichia pastoris increases specific and volumetric productivity of recombinant proteins. *Biotechnol Bioeng* **100**: 177-183.

Bingle, W., J. Nomellini & J. Smit, (2000) Secretion of the Caulobacter crescentus S-layer protein: further localization of the C-terminal secretion signal and its use for secretion of recombinant proteins. *J Bacteriol* **182**: 3298-3301.

Brown, S., G. Kruppa & J. Dasseux, (2005) Metabolomics applications of FT-ICR mass spectrometry. *Mass Spectrom Rev* **24**: 223-231.

Burgess-Brown, N., S. Sharma, F. Sobott, C. Loenarz, U. Oppermann & O. Gileadi, (2008) Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study. *Protein Expr Purif* **59**: 94-102.

Cusano, A., E. Parrilli, G. Marino & M. Tutino, (2006) A novel genetic system for recombinant protein secretion in the Antarctic Pseudoalteromonas haloplanktis TAC125. *Microb Cell Fact* **5**: 40.

De Schutter, K., Y. Lin, P. Tiels, A. Van Hecke, S. Glinka, J. Weber-Lehmann, P. Rouzé, Y. Van de Peer & N. Callewaert, (2009) Genome sequence of the recombinant protein production host Pichia pastoris. *Nat Biotechnol* **27**: 561-566.

Doglia, S., D. Ami, A. Natalello, P. Gatti-Lafranconi & M. Lotti, (2008) Fourier transform infrared spectroscopy analysis of the conformational quality of recombinant proteins within inclusion bodies. *Biotechnol J* **3**: 193-201.

Dragosits, M., J. Stadlmann, J. Albiol, K. Baumann, M. Maurer, B. Gasser, M. Sauer, F. Altmann, P. Ferrer & D. Mattanovich, (2009) The Effect of Temperature on the Proteome of Recombinant Pichia pastoris. *J Proteome Res* **8**: 1380-1392.

Dunn, W. J., J. Cregg, J. Kiel, I. van der Klei, M. Oku, Y. Sakai, A. Sibirny, O. Stasyk & M. Veenhuis, (2005) Pexophagy: the selective autophagy of peroxisomes. *Autophagy* **1**: 75-83.

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

Fodor, S., J. Read, M. Pirrung, L. Stryer, A. Lu & D. Solas, (1991) Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**: 767-773.

Gach, J., M. Maurer, R. Hahn, B. Gasser, D. Mattanovich, H. Katinger & R. Kunert, (2007) High level expression of a promising anti-idiotypic antibody fragment vaccine against HIV-1 in Pichia pastoris. *J Biotechnol* **128**: 735-746.

Gasser, B., M. Saloheimo, U. Rinas, M. Dragosits, E. Rodríguez-Carmona, K. Baumann, M. Giuliani, E. Parrilli, P. Branduardi, C. Lang, D. Porro, P. Ferrer, M. Tutino, D. Mattanovich & A. Villaverde, (2008) Protein folding and conformational stress in

microbial cells producing recombinant proteins: a host comparative overview. *Microb Cell Fact* **7**: 11.

Gerngross, T., (2004) Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. *Nat Biotechnol* **22**: 1409-1414.

Graf, A., M. Dragosits, B. Gasser & D. Mattanovich, (2009) Yeast systems biotechnology for the production of heterologous proteins. *FEMS Yeast Res* **9**: 335-348.

Graf, A., B. Gasser, M. Dragosits, M. Sauer, G. Leparc, T. Tuechler, D. Kreil & D. Mattanovich, (2008) Novel insights into the unfolded protein response using Pichia pastoris specific DNA microarrays. *BMC Genomics* **9**: 390.

Gupta, P., A. Ghosalkar, S. Mishra & T. Chaudhuri, (2009) Enhancement of over expression and chaperone assisted yield of folded recombinant aconitase in Escherichia coli in bioreactor cultures. *J Biosci Bioeng* **107**: 102-107.

Gygi, S., B. Rist, S. Gerber, F. Turecek, M. Gelb & R. Aebersold, (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**: 994-999.

Hamilton, S., R. Davidson, N. Sethuraman, J. Nett, Y. Jiang, S. Rios, P. Bobrowicz, T. Stadheim, H. Li, B. Choi, D. Hopkins, H. Wischnewski, J. Roser, T. Mitchell, R. Strawbridge, J. Hoopes, S. Wildt & T. Gerngross, (2006) Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science* **313**: 1441-1443.

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

Hartl, F., (1996) Molecular chaperones in cellular protein folding. *Nature* **381**: 571-579.

Jacobs, P., S. Geysens, W. Vervecken, R. Contreras & N. Callewaert, (2009) Engineering complex-type N-glycosylation in Pichia pastoris using GlycoSwitch technology. *Nat Protoc* **4**: 58-70.

Jurado, P., D. Ritz, J. Beckwith, V. de Lorenzo & L. Fernández, (2002) Production of functional single-chain Fv antibodies in the cytoplasm of Escherichia coli. *J Mol Biol* **320**: 1-10.

Kell, D., (2004) Metabolomics and systems biology: making sense of the soup. *Curr Opin Microbiol* **7**: 296-307.

Kleizen, B. & I. Braakman, (2004) Protein folding and quality control in the endoplasmic reticulum. *Curr Opin Cell Biol* **16**: 343-349.

Knijnenburg, T., J. Daran, M. van den Broek, P. Daran-Lapujade, J. de Winde, J. Pronk, M. Reinders & L. Wessels, (2009) Combinatorial effects of environmental parameters on transcriptional regulation in Saccharomyces cerevisiae: a quantitative analysis of a compendium of chemostat-based transcriptome data. *BMC Genomics* **10**: 53.

Kosaka, T., T. Takazawa & T. Nakamura, (2000) Identification and C-terminal characterization of proteins from two-dimensional polyacrylamide gels by a combination of isotopic labeling and nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* **72**: 1179-1185.

Kunert, R., R. Weik, B. Ferko, G. Stiegler & H. Katinger, (2002) Anti-idiotypic antibody Ab2/3H6 mimics the epitope of the neutralizing anti-HIV-1 monoclonal antibody 2F5. *AIDS* **16**: 667-668.

Lahav, R., A. Nejidat & A. Abeliovich, (2004) Alterations in protein synthesis and levels of heat shock 70 proteins in response to salt stress of the halotolerant yeast Rhodotorula mucilaginosa. *Antonie Van Leeuwenhoek* **85**: 259-269.

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

Lin Cereghino, G., A. Sunga, J. Lin Cereghino & J. Cregg, (2001) Expression of foreign genes in the yeast Pichia pastoris. *Genet Eng (N Y)* **23**: 157-169.

Lin, H., T. Kim, F. Xiong & X. Yang, (2007) Enhancing the production of Fc fusion protein in fed-batch fermentation of Pichia pastoris by design of experiments. *Biotechnol Prog* **23**: 621-625.

Mager, W. & M. Siderius, (2002) Novel insights into the osmotic stress response of yeast. *FEMS Yeast Res* **2**: 251-257.

Martínez-Alonso, M., N. González-Montalbán, E. García-Fruitós & A. Villaverde, (2009) Learning about protein solubility from bacterial inclusion bodies. *Microb Cell Fact* **8**: 4.

Mattanovich, D., A. Graf, J. Stadlmann, M. Dragosits, A. Redl, M. Maurer, M. Kleinheinz, M. Sauer, F. Altmann & B. Gasser, (2009) Genome, secretome and glucose transport highlight unique features of the protein production host Pichia pastoris. *Microb Cell Fact* **8**: 29.

Mazor, Y., T. Van Blarcom, B. Iverson & G. Georgiou, (2009) Isolation of full-length IgG antibodies from combinatorial libraries expressed in Escherichia coli. *Methods Mol Biol* **525**: 217-239, xiv.

Mori, K., T. Kawahara, H. Yoshida, H. Yanagi & T. Yura, (1996) Signalling from endoplasmic reticulum to nucleus: transcription factor with a basic-leucine zipper motif is required for the unfolded protein-response pathway. *Genes Cells* **1**: 803-817.

Munro, S., (2001) What can yeast tell us about N-linked glycosylation in the Golgi apparatus? *FEBS Lett* **498**: 223-227.

Nahálka, J., A. Vikartovská & E. Hrabárová, (2008) A crosslinked inclusion body process for sialic acid synthesis. *J Biotechnol* **134**: 146-153.

Oda, Y., K. Huang, F. Cross, D. Cowburn & B. Chait, (1999) Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A* **96**: 6591-6596.

Ong, S., B. Blagoev, I. Kratchmarova, D. Kristensen, H. Steen, A. Pandey & M. Mann, (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**: 376-386.

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

Regenberg, B., T. Grotkjaer, O. Winther, A. Fausbøll, M. Akesson, C. Bro, L. Hansen, S. Brunak & J. Nielsen, (2006) Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in Saccharomyces cerevisiae. *Genome Biol* **7**: R107.

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

Robinson, A., V. Hines & K. Wittrup, (1994) Protein disulfide isomerase overexpression increases secretion of foreign proteins in Saccharomyces cerevisiae. *Biotechnology (N Y)* **12**: 381-384.

Rogers, S. & A. Cambrosio, (2007) Making a new technology work: the standardization and regulation of microarrays. *Yale J Biol Med* **80**: 165-178.

Ross, P., Y. Huang, J. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson & D. Pappin, (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**: 1154-1169.

Sauer, M., P. Branduardi, B. Gasser, M. Valli, M. Maurer, D. Porro & D. Mattanovich, (2004) Differential gene expression in recombinant Pichia pastoris analysed by heterologous DNA microarray hybridisation. *Microb Cell Fact* **3**: 17.

Sauer, U., (2004) High-throughput phenomics: experimental methods for mapping fluxomes. *Curr Opin Biotechnol* **15**: 58-63.

Schena, M., D. Shalon, R. Davis & P. Brown, (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467-470.

Schena, M., D. Shalon, R. Heller, A. Chai, P. Brown & R. Davis, (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci U S A* **93**: 10614-10619.

Schumann, W., (2007) Production of recombinant proteins in Bacillus subtilis. *Adv Appl Microbiol* **62**: 137-189.

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

Shuster, J., (1991) Gene expression in yeast: protein secretion. *Curr Opin Biotechnol* **2**: 685-690.

Talapatra, A., R. Rouse & G. Hardiman, (2002) Protein microarrays: challenges and promises. *Pharmacogenomics* **3**: 527-536.

Terpe, K., (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **72**: 211-222.

Tonge, R., J. Shaw, B. Middleton, R. Rowlinson, S. Rayner, J. Young, F. Pognan, E. Hawkins, I. Currie & M. Davison, (2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* **1**: 377-396.

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

Unlü, M., M. Morgan & J. Minden, (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**: 2071-2077.

Uttamchandani, M., J. Wang & S. Yao, (2006) Protein and small molecule microarrays: powerful tools for high-throughput proteomics. *Mol Biosyst* **2**: 58-68.

van Vliet, C., E. Thomas, A. Merino-Trigo, R. Teasdale & P. Gleeson, (2003) Intracellular sorting and transport of proteins. *Prog Biophys Mol Biol* **83**: 1-45.

Vasina, J. & F. Baneyx, (1996) Recombinant protein expression at low temperatures under the transcriptional control of the major Escherichia coli cold shock promoter cspA. *Appl Environ Microbiol* **62**: 1444-1447.

Vasina, J. & F. Baneyx, (1997) Expression of aggregation-prone recombinant proteins at low temperatures: a comparative study of the Escherichia coli cspA and tac promoter systems. *Protein Expr Purif* **9**: 211-218.

Vembar, S. & J. Brodsky, (2008) One step at a time: endoplasmic reticulum-associated degradation. *Nat Rev Mol Cell Biol* **9**: 944-957.

Vera, A., N. González-Montalbán, A. Arís & A. Villaverde, (2007) The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnol Bioeng* **96**: 1101-1106.

Wacker, M., D. Linton, P. Hitchen, M. Nita-Lazar, S. Haslam, S. North, M. Panico, H. Morris, A. Dell, B. Wren & M. Aebi, (2002) N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science* **298**: 1790-1793.

Westerhoff, H. & B. Palsson, (2004) The evolution of molecular biology into systems biology. *Nat Biotechnol* **22**: 1249-1252.

Wiechert, W., (2001) 13C metabolic flux analysis. *Metab Eng* **3**: 195-206.

Wiemken, A., (1990) Trehalose in yeast, stress protectant rather than reserve carbohydrate. *Antonie Van Leeuwenhoek* **58**: 209-217.

Wu, M., G. Dimopoulos, A. Mantalaris & J. Varley, (2004) The effect of hyperosmotic pressure on antibody production and gene expression in the GS-NS0 cell line. *Biotechnol Appl Biochem* **40**: 41-46.

6 Publications

6.1 Reviews

Gasser B, Saloheimo M, Rinas U, Dragosits M, Rodríguez-Carmona E, Baumann K, Giuliani M, Parrilli E, Branduardi P, Lang C, Porro D, Ferrer P, Tutino ML, Mattanovich D, Villaverde A. Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microb Cell Fact. 2008 Apr 4;7:11

Graf A, Dragosits M, Gasser B, Mattanovich D. Yeast systems biotechnology for the production of heterologous proteins. FEMS Yeast Res. 2009 May;9(3):335-48

Microbial Cell Factories

BioMed Central

Review

Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview

Brigitte Gasser[1], Markku Saloheimo[2], Ursula Rinas[3], Martin Dragosits[1], Escarlata Rodríguez-Carmona[4], Kristin Baumann[5], Maria Giuliani[6], Ermenegilda Parrilli[6], Paola Branduardi[7], Christine Lang[8], Danilo Porro[7], Pau Ferrer[5], Maria Luisa Tutino[6], Diethard Mattanovich*[1] and Antonio Villaverde[4]

Address: [1]University of Natural Resources and Applied Life Sciences Vienna, Department of Biotechnology, Vienna, Austria, [2]VTT Technical Research Centre, Espoo, Finland, [3]Helmholtz Center for Infection Research, Braunschweig, Germany, [4]Autonomous University of Barcelona, Institute for Biotechnology and Biomedicine, Department of Genetics and Microbiology, and CIBER-BBN Network in Bioengineering, Biomaterials and Nanomedicine, Barcelona, Spain, [5]Autonomous University of Barcelona, Department of Chemical Engineering, Barcelona, Spain, [6]University of Naples Federico II, School of Biotechnological Sciences, Naples, Italy, [7]University of Milano-Bicocca, Department of Biotechnology and Bioscience, Milan, Italy and [8]Technical University Berlin, Faculty III, Institute for Microbiology and Genetics, Berlin, Germany

Email: Brigitte Gasser - brigitte.gasser@boku.ac.at; Markku Saloheimo - markku.saloheimo@vtt.fi; Ursula Rinas - ursula.rinas@helmholtz-hzi.de; Martin Dragosits - martin.dragosits@boku.ac.at; Escarlata Rodríguez-Carmona - escarlata.rodriguez@uab.cat; Kristin Baumann - kristin.baumann@uab.es; Maria Giuliani - maria.giuliani@unina.it; Ermenegilda Parrilli - erparril@unina.it; Paola Branduardi - paola.branduardi@unimib.it; Christine Lang - christine.lang@tu-berlin.de; Danilo Porro - danilo.porro@unimib.it; Pau Ferrer - pau.ferrer@uab.cat; Maria Luisa Tutino - tutino@unina.it; Diethard Mattanovich* - diethard.mattanovich@boku.ac.at; Antonio Villaverde - avillaverde@servet.uab.es

* Corresponding author

Published: 4 April 2008

Microbial Cell Factories 2008, **7**:11 doi:10.1186/1475-2859-7-11

Received: 18 December 2007
Accepted: 4 April 2008

This article is available from: http://www.microbialcellfactories.com/content/7/1/11

Abstract

Different species of microorganisms including yeasts, filamentous fungi and bacteria have been used in the past 25 years for the controlled production of foreign proteins of scientific, pharmacological or industrial interest. A major obstacle for protein production processes and a limit to overall success has been the abundance of misfolded polypeptides, which fail to reach their native conformation. The presence of misfolded or folding-reluctant protein species causes considerable stress in host cells. The characterization of such adverse conditions and the elicited cell responses have permitted to better understand the physiology and molecular biology of conformational stress. Therefore, microbial cell factories for recombinant protein production are depicted here as a source of knowledge that has considerably helped to picture the extremely rich landscape of in vivo protein folding, and the main cellular players of this complex process are described for the most important cell factories used for biotechnological purposes.

Review

One of the main bottlenecks in recombinant protein production is the inability of the foreign polypeptides to reach their native conformation in heterologous host cells, which usually results into their prevalence in the insoluble cell fraction. The unusually high and non-phys-

iological rates of recombinant protein production and the occurrence of significant amounts of misfolded protein species drive the cells to a global conformational stress condition. This situation is characterized by a series of individual physiological responses provoked in order to minimize any toxicity of misfolded protein species and to restore cellular folding homeostasis. The generalized use of microbial cell factories for biological synthesis of proteins and the growing interest in the physiological aspects of conformational stress have converted recombinant cells into schools of protein folding, from which scientists are learning about the cell-protein relationships during the complex process of in vivo protein folding.

The purpose of this review is to summarize the major concepts of the cell biology of protein folding. For that, eukaryotic cells, illustrated by yeasts and filamentous fungi are dissected regarding the mechanics and composition of their folding machinery, misfolding stress responses and strategies to cope with conformational stress. The complexity of the folding, trafficking and secretion machineries of these cell factories is presented versus the relatively simple folding scheme in bacterial cells such as *Escherichia coli* that are also common hosts for recombinant protein production. Despite the existing obvious differences, evolutionary conserved physiological traits regarding folding stress can be identified when comparing eukaryotic and prokaryotic hosts. Furthermore, practical implications of all these findings to improve protein production processes are discussed in their biotechnological context.

Protein folding and conformational stress in eukaryotic cells

Yeasts and filamentous fungi are among the most frequently used eukaryotic cell systems for recombinant protein production, in part due to the performance of post-translational modifications that bacteria cannot perform, that are, in most cases, required for proper protein activity. In eukaryotic cells, endoplasmatic reticulum (ER) resident proteins are responsible for correct protein folding. The list of such folding-assistant proteins includes calnexin, chaperones of the hsp70 and hsp90 families (e.g. BiP/Grp78, Grp94), the protein disulfide isomerases (Pdi) which catalyze the formation of disulfide bonds and the peptidyl-prolyl-isomerases. Some of the post-translational modifications such as N-glycosylation are initiated in the ER lumen. Both natural and recombinant proteins are only exported to the Golgi by vesicular transport when their correct conformation has been assured by a glucose-dependent surveillance mechanism of the ER. Unless there is a differing signal, proteins intended for secretion are directed from the Golgi to the outside of the plasma membrane by specific transport vesicles [1,2]. A schematic overview of the protein folding processes is presented in Figure 1, while the responses to secretion stress are summarized in Figure 2.

The protein folding process and subsequent secretion is a rather complex process involving many interacting participants. Due to this interdependence, genetically increasing the rate of one step can lead to rate-limitation of another one, which can then become the bottleneck of the expression system. Moreover, in most cases the rate limiting step in the eukaryotic secretion pathway has been identified to be the exit of proteins from the ER [3]. Linked to this control point is a mechanism called ER-associated protein degradation (ERAD), which is responsible for the retention of misfolded or unmodified non-functional proteins in the ER and their subsequent removal. Protein degradation is executed by linking the misfolded protein to ubiquitin after it has been re-translocated into the cytosol through the same ER translocon pore where it had been imported. The ubiquitin-marked protein is then recognized and degraded by the 26S proteasome in the cytosol (recently reviewed by [4,5].

Two quality control systems in the ER ensure that only correctly folded, modified and assembled proteins travel further along the secretory pathway. The UDP-glucose:glycoprotein glucosyltransferase (UGT) is a central player of glycoprotein quality control in the ER (reviewed among others by [6]). After addition of the core glycan (GlcNac2-Man9-Glc3) to specific asparagine residues of the nascent polypeptide, the three terminal glucose residues have to be clipped off before the protein can exit the ER. Non-native polypeptides are tagged for reassociation with the ER-lectin calnexin by readdition of the terminal glucose onto the N-glycan mediated by UGT. This enzyme specifically recognizes and binds to molten globule-like folding intermediates, thereby acting as sensor of the protein folding status. Re-glucosylation of erroneous glycoproteins prevents their release from the calnexin cycle and subsequent secretion. Upon persistent misfolding, N-glycosylated polypeptides are slowly released from calnexin and enter a second level of retention-based ER quality control by aggregating with the BiP chaperone complex [7]. This correlates with the loss in the ability to emend misfolding. The BiP complex is involved in co-translational translocation of the nascent polypeptide into the ER lumen and preferentially binds to hydrophobic patches. Prolonged binding to either calnexin or the BiP complex targets the polypeptides to the ERAD, however, the exact mechanisms remain elusive (reviewed by [6]). The fact that accumulation of proteins in the ER is able to influence the synthesis of foldases and chaperones such as BiP and Pdi by transcriptional activation in the nucleus lead to the conclusion early on that there must be an intracellular signalling pathway from the ER to the nucleus,

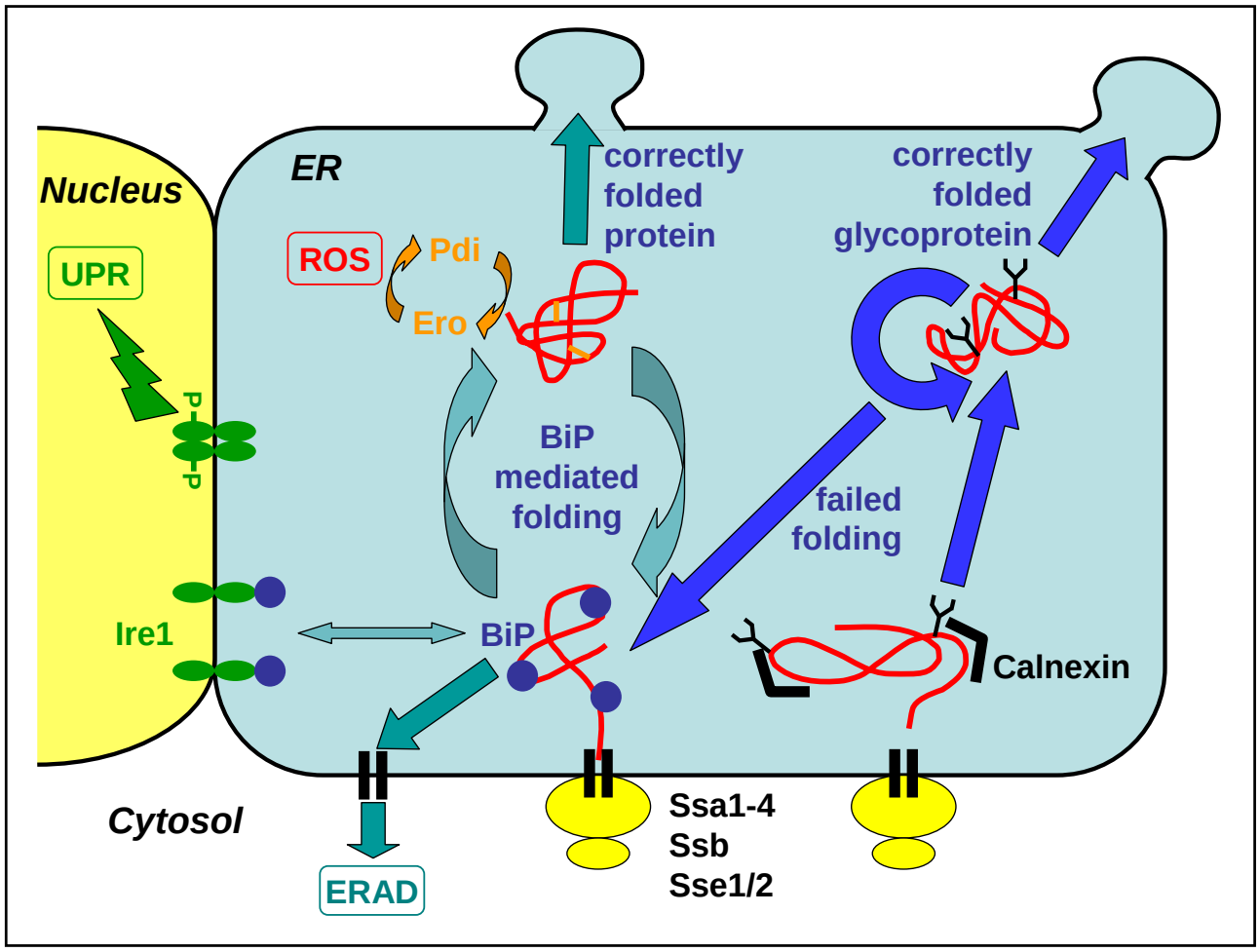


Figure 1
Schematic representation of protein folding, quality control, degradation and secretion in yeast (as an example for lower eukaryotic cells). Secretory proteins are transported into the ER through the Sec61 translocon complex of the ER membrane either co-translationally or post-translationally. In the latter case, cytosolic chaperones (Ssa1-4, Ssb, Sse1/2) support solubility and prevent aggregation of the polypeptide chains. After translocation to the ER, nascent polypeptides are bound by BiP and mediated to mature folding in an ATP-dependent cyclic process of release of and binding to BiP. The formation of correct disulfide bonds is mediated in a cycle of Pdi and Ero activity, which may lead to the formation of reactive oxygen species (ROS). Correctly folded protein is released to transport vesicles, while prolonged BiP binding, indicating misfolding, leads to retrograde translocation to the cytosol and proteasomal degradation (ERAD). Nascent glycoproteins are bound by calnexin and mediated to correct folding and processing of the N-glycans. Failed folding leads to binding by the BiP complex and targeting to ERAD, while correctly folded and processed glycoproteins are released to transport vesicles. Prolonged binding of BiP to partially misfolded proteins leads to the induction of the unfolded protein response (UPR), mediated by Ire1 (see also figure 2).

called the unfolded protein response (UPR) (for reviews see [8,9]).

After having passed ER quality control successfully, proteins intended for secretion have to be transported to the Golgi network. Specialized cargo vesicles that selectively incorporate these proteins bud from the ER and are targeted to the Golgi membrane by the activity of the coat protein complex II (COPII). In the Golgi network proteins undergo additional post-translational modifications and are subjected to sorting mechanisms that finally target them to their final destination. Possible trafficking routes include direction to the plasma membrane, to the endosomal compartments, to the vacuole, as well as retrograde transport to the ER (review by [10]). Secretory proteins are then delivered to the cell surface by specialized post-Golgi

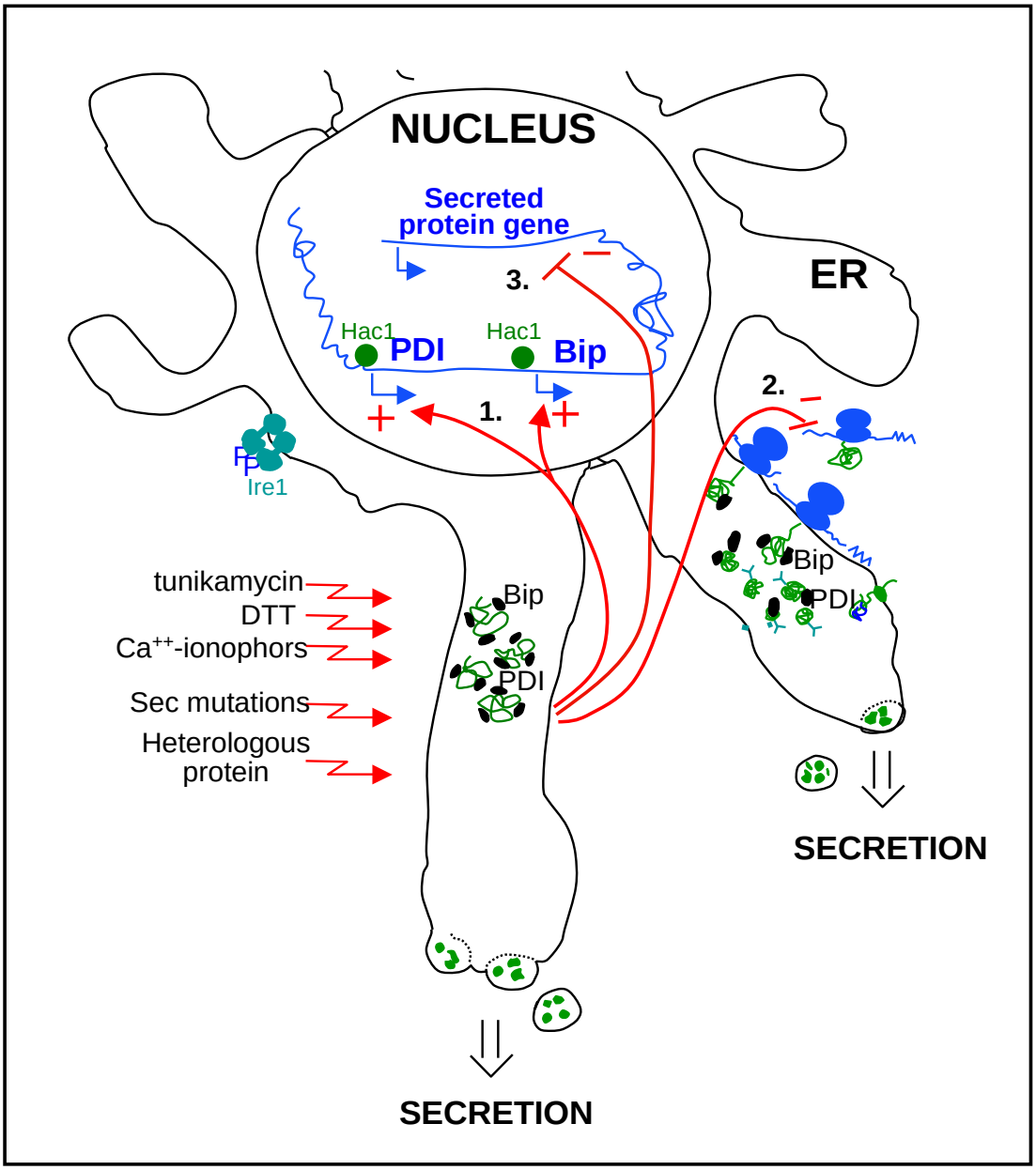


Figure 2
Schematic representation of secretion stress responses in eukaryotes Secretory proteins are translocated to the ER either during their translation or post-translationally. Folding of these proteins in the ER can be disturbed by environmental factors or it can be inhibited experimentally by agents inhibiting protein folding like dithiothreitol (DTT) and Ca-ionophores or agents inhibiting glycosylation like tunicamycin. It has been observed that foreign proteins often do not fold well and cause conformational stress. Several responses of the cell to impaired protein folding in the ER have been discovered: 1.) Unfolded protein response (UPR). Genes encoding folding helpers like the chaperone Bip and the foldase protein disulfide isomerase Pdi, and a large number of other genes involved in other functions of the secretory pathway are induced. The proteins Ire1 and Hac1 involved in this signal transduction pathway are shown in the figure. 2.) Translation attenuation. The translation initiation factor eIF2 alpha is phosphorylated, and subsequently translation initiation is inhibited. This reduces the influx of proteins into the ER. This response is only known from mammalian cells. 3.) Repression under secretion stress (RESS). The mRNA levels of genes encoding secreted proteins are down-regulated during ER stress. This response has been discovered in filamentous fungi, but there is evidence for its occurrence in plants.

secretory vesicles that dock to and fuse with the plasma membrane. The process called exocytosis includes targeting of the secretory vesicles to the appropriate membrane mediated by the Exocyst, a multiprotein complex, and by interaction of the v-SNAREs (vesicle, in yeast: Snc1/2 proteins) and t-SNAREs (target membrane; Sso1/2p and Sec9p) and release of the cargo proteins outside the cell after fusion of the secretory vesicle with the plasma membrane.

Impact of the environment on folding and folding stress

During the recent years, it has become evident that a variety of metabolic and environmental stresses may have a strong impact on recombinant protein production. Both types of stress factors occurring during industrial production processes in yeasts, along with potential metabolic and cell engineering approaches to overcome production constraints, were reviewed in Mattanovich et al. [11]. Among environmental factors that affect protein folding and secretion, especially temperature, low pH, high osmolarity and oxidative stress may play an important role.

While many studies have been performed on optimizing fermentation conditions for maximum specific productivity in yeasts, data correlating increased product yields to improved protein folding and secretion mechanisms are still missing. Similar reports regarding the impact of cultivation conditions on protein production in filamentous fungi remain scarce and usually limited to case studies [12-14]. Wang et al. [15] reviewed the impacts of bioprocess strategies on recombinant protein production in filamentous fungi, and concluded that the major effect of the environmental changes correlates to varying morphological forms, which exhibit different secretory capacities.

Temperature

Temperature has a profound impact on cell metabolism and abundance/regulation of folding-related genes/proteins (hsp70 family, ER-membrane proteins, etc.). Lowering the cultivation temperature from 30 to 20–25°C has been reported to increase product titers in yeasts in several cases [16-19]. While it may be speculated that a lower growth temperature is leading to lower specific growth rates, thus enabling folding of the recombinant proteins at a lower rate, it was shown recently in chemostat cultures that actually gene regulatory events take place. In continuous cultures of *Pichia pastoris* expressing a human antibody Fab fragment specific productivity of the heterologous protein was significantly increased during the chemostat process at lower temperature (1.4-fold on average). Several genes related to protein targeting to the ER and folding (*SSA4*, *SEC53*, *KAR2*, *ERO1*) and core metabolism genes were found among the genes down-regulated at 20°C, as were the product genes [20]. Transcription of genes involved in the regulation of vesicular transport, exocytosis, ER-associated protein degradation as well as markers for response to oxidative and hyperosmotic stress was enhanced in comparison to 25°C steady state. The reduction in transcriptional activity of the core metabolism is a likely explanation for the reduced mRNA levels of the product genes (LC and HC), which were under control of the glycolytic GAP (glyceraldehyde 3-phosphate dehydrogenase) promoter. The authors hypothesized that at lower temperature a reduced amount of folding stress is imposed on the cells, consequently leading to a higher rate of correctly folded product. Although lower temperature has been shown to improve protein secretion rates, this still depends on the nature of the heterologous protein. Production of a hyperthermophilic enzyme was improved by cultivation at higher temperature (40°C) in *Saccharomyces cerevisiae*, thereby reducing ER folding stress [21].

Additionally to regulatory events, many positive effects of temperature shifts on protein production might be linked to cell wall composition (porosity) and cell cycle. Indeed, increased levels of chitin and cell wall linking beta-glucans have been determined in yeast cells grown at 37°C compared to 22°C in batch cultures [22].

Generally, it turns out that cultivation at an optimized temperature is one of the crucial parameters for improved specific productivity, as it is likely to direct carbon fluxes towards heterologous protein production, and maintains the cells in the more secretion competent phases of the cell cycle.

Oxygenation

Redox processes play a major role in heterologous protein production, both related to the oxidation of the product to form disulfide bonds, and to oxidative stress of the host cell during cultivation. Cultivation of methylotrophic yeasts like *P. pastoris* on methanol leads to significant oxidative stress, which may be relieved by the addition of antioxidants like ascorbic acid [23]. Similarly, the expression of antioxidant enzymes like superoxide dismutase was reported to relieve oxidative stress [24].

Apart from environmental stressors, oxidative stress can be imposed on the host cells by intrinsic factors such as leakage in the respiratory pathway, beta-oxidation of lipids, or accumulation of misfolded protein in the ER. There is strong evidence that oxidative stress is connected to growth temperature. While in most cases lower growth temperature results in lower oxidative stress, Gasser et al. [20] showed that the genes coding for the key regulatory enzymes of both the cell redox homeostasis (thioredoxin reductase *TRR1*, thioredoxin peroxidase *TSA1*, glutathione oxidoreductase *GLR1*) and osmoregulation (mitogen-activated protein (MAP) kinase *HOG1*) were

induced at the lower temperature where higher secretion rates occur. Generally, the secretory pathway compartments maintain a higher oxidized status compared to the cytosol in order to enable disulfide-bond formation. Finally the electrons generated during the oxidative folding cycles are transferred to molecular oxygen and may lead to the formation of reactive oxygen species [25].

Interestingly, it was shown recently that very low oxygen supply enhances the secretion rate of heterologous proteins in *P. pastoris* significantly, which led to the development of a hypoxic fed batch strategy with over 2-fold increased productivity [26].

Osmolarity

So far no clear connection between medium osmolarity and protein folding has been established. Previous data indicate that the response is extremely transient [27]; and even less is known of the effect of osmolarity on heterologous protein production. Mager and Siderius [28] describe temporary cell growth arrest (either at G1 or G2/M) upon hyperosmotic stress conditions accompanied by the induction of the high osmolarity glycerol (HOG) kinase pathway in *S. cerevisiae*. Intracellular glycerol levels are increased in order to adjust osmo-balance through the modification of cell wall integrity. Unlike in animal cells where an osmotic shock leads to increased exocytosis [29], and hyperosmotic GS-NS0 mammalian cells that exhibit an increased specific production rate (albeit decreased growth rate) as compared with iso-osmotic cultures [30], osmo-regulated secretion behaviour in fungi remains unproven. In methanol grown *P. pastoris* cells, salt stress prior to induction was shown to have a positive effect on single chain antibody scFv titers [19], while Lin et al. [18] reported a negative effect of salt supplementation on the secretion of an Fc fusion protein.

pH

Osmolarity and pH seem to trigger highly interrelated responses. From an industrial point of view the main desired effect of low pH is to reduce the activity of host proteases which can lead to severe protein degradation (reviewed among others by [31]), but no uniform picture has been assigned to the correlation of pH and protease activity in the culture broth. Both in yeasts and filamentous fungi changing the pH of the culture medium can significantly improve protein yields, however, this effect is most probably not directly associated with improved protein folding mechanisms. On the other hand, lower extracellular pH requires higher energy to maintain intracellular pH values constant/physiological, thereby delaying cell growth and enforcing the cell wall barrier [22,32,33]. Subsequently this more rigid cell wall may diminish secretion efficiency of the pH stressed cells. Lin at al. [18] tested different pH values (ranging from 3.0 to 7.2) during fed batch production of a Fc fusion protein in *P. pastoris* and reported detection of the heterologous protein only at the highest pH of 7.2, however, the authors conclude that the pH optimum is strongly protein and strain dependent.

Folding stress and heterologous protein production

The ER-resident chaperone BiP (binding protein, in yeast encoded by *KAR2*) belongs to Hsp70 family of heat shock proteins and it is present in the lumen of the endoplasmatic reticulum of all eukaryotes. The yeast homologue is sometimes referred to as Grp78. Binding to BiP prevents the nascent part of secretory or transmembrane proteins from misfolding, until synthesis of the protein is finished. It has been suggested that BiP is not only involved in the translocation of the nascent polypeptides across the ER membrane into the ER lumen, but that it is a key element of an ER-resident quality control mechanism that prevents unfolded proteins from leaving the ER [34]. Other functions associated to BiP are the solubilisation of folding precursors, stabilization of unassembled protein subunits and redirecting misfolded polypeptide chains to the cytosol for ubiquitin-labeling and subsequent degradation by the proteasome (ERAD, ER-associated protein degradation, [35]). Besides a basal constitutive expression level, BiP transcription is induced by the presence of mutant and misfolded proteins in the ER lumen and by stress effects that result in the accumulation of unfolded proteins [36], presumably including the high level expression of heterologous proteins. A saturation of the secretory pathway seems possible, as extractable levels of free folding assistants BiP and Pdi1 decrease when heterologous proteins are overexpressed in *S. cerevisiae* [37]. Kauffman et al. [38] observed an induction of BiP during the expression of a scFv fragment in this yeast species, and Hohenblum et al. [39] have reported increased levels of BiP upon expression of recombinant human trypsinogen in *P. pastoris*. Likewise, biPA and pdiA transcript levels were increased due to heterologous protein overexpression, as well as upon high level secretion of homologous enzymes in filamentous fungi [40-42].

ER-associated protein degradation is a complex process in which misfolded proteins in the ER are redirected to the translocon for retranslocation to the cytosol, where they are subjected to proteasomal degradation. Additionally, excess subunits of multimeric proteins that are unable to assemble are degraded through the ERAD mechanism. According to Plemper et al. [43], the malfolded proteins are retro-translocated through the Sec61-complex translocon pore, through which they had entered the lumen of the ER before, accompanied by ubiquitination at the cytosolic side of the ER membrane. The labeling of substrates destined for degradation by the cytosolic 26S proteasome requires an Ub (ubiquitin) activating enzyme, an

Ub conjugating enzyme and an Ub ligase besides ubiquitin itself. In *P. pastoris* three essential components of the ERAD pathway have been shown to be up-regulated upon production of an antibody Fab fragment in correlation to higher protein secretion rates: *HRD1*, coding for an Ub protein ligase, that is able to recruit Ub conjugating enzymes (such as the gene product of *UBC1*) next to the Sec61 translocon pore complex [20].

Prolonged ER retention of misfolded proteins entails repetitive rounds of oxidative protein folding attempts by foldases such as Pdi and consequently results in the generation of reactive oxygen species (ROS). Alleviation of the ER stress is accomplished by the upregulation of the UPR and subsequent induction of the ERAD, however, prolonged UPR induction can also contribute to the stress situation by the accumulation of ROS. In this context, both oxidative stress and ERAD occur in addition to UPR induction when hydrophobic cutinase accumulates in the ER of *S. cerevisiae* [44], while hirudin production in *P. pastoris* lead to increased levels of ROS [23]. Recently it has been shown that overstraining or failure of the ERAD components leads to persistent ER stress conditions and subsequent cell death in both yeasts and higher eukaryotic cells [45,46].

The unfolded protein response pathway is activated by a unique mechanism not known in any other signal transduction pathway (for a recent review see [47]). The sensor protein Ire1p [48] resides in the ER membrane and possesses both kinase and endonuclease activities. When unfolded proteins accumulate in the ER, Ire1p undergoes autophosphorylation and oligomerisation, and catalyses the cleavage of the mRNA encoding the UPR transcription factor, called Hac1/hacA in yeasts and filamentous fungi [49,50] or Xbp1 in mammalian cells [51]. In this way Ire1p initiates an unconventional intron splicing event that has been shown in *S. cerevisiae* to be completed by tRNA ligase [52]. Splicing of yeast *HAC1* mRNA removes a translation block mediated by the intron [53] and enables formation of the activator protein. For mammalian Xbp1 it has been shown that the unspliced mRNA produces an unstable protein that represses the UPR target genes, whereas the spliced mRNA is translated to a potent, stable activator protein [51]. In the filamentous fungi *Trichoderma reesei*, *Aspergillus nidulans* [50] and *Aspergillus niger* [54], the hac1/hacA mRNA is truncated at the 5' flanking region during UPR induction, in addition to the unconventional intron splicing. This truncation removes upstream open reading frames from the mRNAs, most probably increasing translation initiation at the start codon of the *HAC1*/*HACA* open reading frame. Kincaid and Cooper [46] identified a novel function of Ire1p in the degradation of mRNAs encoding selected secretory proteins thus avoiding potential overload of the ER and the translocon complex *a priori*.

ER-associated stress responses such as UPR and ERAD were reported to be induced upon overexpression of several heterologous proteins, *e.g.*, human tissue plasminogen activator (t-PA) in *T. reesei* [55] and *A. niger* [56], and bovine chymosin in *A. nidulans* [57]. Similarily, overexpression of Fab fragments [20] and *Rhizopus oryzae* lipase [58] revealed UPR induction in *P. pastoris*.

In another layer of ER stress regulation, mammalian cells can attenuate translation initiation during unfolded protein accumulation into the ER, in order to reduce the influx of proteins to the ER. This regulation pathway is initiated by the ER membrane kinase PERK that has some similarity with Ire1 [59]. PERK phosphorylates the translation initiation factor eIF2alpha, resulting in drastic reduction in translation. This mechanism is not known in yeasts or filamentous fungi, and PERK orthologues can not be found in the genomes of the lower eukaryotes. Interestingly, the filamentous fungi *T. reesei* [60] and *A. niger* [61] have an alternative mechanism for controlling the protein influx to the ER. In conditions of ER stress the mRNAs encoding secreted proteins are rapidly down-regulated. This mechanism called RESS (repression under secretion stress) was shown to be dependent on the promoters of the genes encoding secreted proteins, and thus it probably functions at the level of transcription [60]. It has been observed that in *Arabidopsis thaliana* a large number of genes encoding secreted proteins are down-regulated when cells are exposed to ER stress [62], implying the possibility that RESS might also exist in plants.

Overcoming folding stress for improved protein production

Although promising expectations emerged that increased BiP levels would result in increased folding capacity in the ER, and thus improved secretion rates, the findings were rather inconsistent and unpredictable. Some studies emphasize that overproduction of BiP stimulates protein secretion in *S. cerevisiae* (5-fold increase in secretion of human erythropoietin [63], 26-fold increase in bovine prochymosin [64], 2.5-fold increase in the titer of antithrombotic hirudin due to 2.5 times higher biomass yields [65]). While the secretion level of plant thaumatin in *Aspergillus awamori* was increased up to 2.5-fold compared to a wild type strain due to bipA overexpression [66], the secretory behaviour of the same protein was not affected by overexpression of *KAR2* in *S. cerevisiae* [64]. According to Wittrup and coworkers, a reduction of BiP levels leads to decreased secretion of foreign proteins, however, no effect was observed upon a 5-fold overexpression of BiP on secretion levels of three different recombinant proteins in *S. cerevisiae* [67], and neither for cutinase in *A. awamori* [68]. Other reports even suggest a

negative impact of BiP overexpression, as extracellular levels of *A. niger* glucose oxidase (GOX) decreased 10-fold upon BiP overexpression in *Hansenula polymorpha* [69]. As prolonged binding to BiP seems to direct proteins rather to degradation than to the secretory pathway, it becomes more obvious why the overexpression of this chaperone alone does not result in higher levels of secreted foreign proteins, but can negatively influence expression levels, as reported by Kauffman *et al*. [38] and van der Heide *et al*. [69]. Interestingly, *Pyrococcus furiosus* beta-glucosidase secretion in *S. cerevisiae* is diminished with increased BiP levels, but benefited from higher protein disulfide isomerase (Pdi) levels, although the protein did not contain any disulfide bonds [70], pointing at the chaperone activity of Pdi, as discussed below.

Conesa et al. [71] examined the impact of overexpression of two ER quality control factors, BiP and calnexin, on the secretion of glycosylated *Phanerochaete chrysosporium* manganese peroxidase (MnP) in *A. niger*, as the expression levels of these genes were induced upon recombinant protein production. While BiP overproduction diminished manganese peroxidase secretion levels severely, overexpression of calnexin resulted in a four- to fivefold increase in the extracellular MnP levels. Higher levels of calnexin also showed beneficial effects in mammalian and baculo virus expression systems [72,73]. Recently, the co-overexpression of calnexin was shown to stimulate the secretion of three glycoproteins and one unglycosylated product (HSA) in *H. polymorpha* (2–3 fold on average; [74]). On the other hand, secretion of human serum albumin (HSA) remained unaffected by raising calnexin levels in *Schizosaccharomyces pombe* [75], while in *S. cerevisiae* deletion of the calnexin gene *CNE1* was reported to enhance secretion of both antitrypsin [76] and unstable lysozymes [77,78].

Protein disulfide isomerase (Pdi) is a multifunctional protein resident in the ER lumen that is responsible for the correct formation of disulfide bonds during oxidative folding and the isomerisation of uncorrectly folded disulfides. Apart from this foldase activity, Pdi also acts as a chaperone. An additional *PDI* gene copy in *S. cerevisiae* successfully improved secretion of human growth factor by 10-fold, of *S. pombe* acid phosphatase by 4-fold [63] and of human lysozyme by around 30–60% [79]. Human lysozyme as well as HSA production could also be enhanced by the same strategy in *Kluyveromyces lactis* (1.8 fold and 15 fold, respectively; [80,81]). Both *S. cerevisiae PDI1* and the *P. pastoris* own homolog were proven to be functional in *P. pastoris* by facilitating secretion of the human parathyroid hormone (hPTH, [82]), human anti HIV1 2F5 Fab [83], and *Necator americanus* secretory protein Na-ASP1 [84], the latter reporting a correlation between the secretory enhancement and the *PDI* copy number. Generally, no clear picture emerged from the co-overexpression of the two folding helpers, BiP and Pdi. Whereas synergistic action of BiP and Pdi was suggested regarding the improvement of the secretion of various single chain fragments (scFv) in *S. cerevisiae* [85], a 2-fold increase in secretion of the A33scFv in *P. pastoris* was only achieved by additional copies of *KAR2*, but not *PDI*, and not by the combination of both [86], in analogy to the antagonistic effect observed in CHO cells [87]. Coexpression of *KAR2*, *PDI1* or *SSO2* exhibited no effect on secretion of gamma-Interferon (IFNgamma) in *H. polymorpha* [88]. Moreover, coexpression of *cypB*, which encodes a foldase of the ER secretory pathway [89], did no increase production of tissue plasminogen activator (t-PA) in *A. niger*, although t-PA production elicited an UPR response detectable through elevated transcript levels of *bip*, *pdi* and *cypB* [90]. Thus, it seems that the effect of coexpression of chaperone and foldase genes strongly depends on the properties of the target protein and, moreover, it seems that fine-tuned overexpression of these genes are required to generate a functional secretory network to improve foreign protein overproduction. For example, in *A. niger*, overexpression of *bip* to a certain threshold was beneficial for plant sweet protein thaumatin production, but above this threshold level thaumatin production decreased [66]. Similarly, defined levels of Pdi were required for optimum thaumatin secretion in *A. niger* [91].

The flavoenzyme Ero1 is required for oxidation of protein dithiols in the ER. It is oxidized by molecular oxygen and acts as a specific oxidant of protein disulfide isomerase (Pdi). Disulfides generated de novo within Ero1 are transferred to Pdi and then to substrate proteins by dithiol-disulfide exchange reactions [92]. Duplication of either *KlPDI1* or *KlERO1* genes led to a similar increase in HSA yield in *K. lactis*, while duplication of both genes accelerated the secretion of HSA and improved cell growth rate and yield. Increasing the dosage of *KlERO1* did not affect the production of human interleukin 1beta, a protein that has no disulfide bridges [93].

Finally, another approach to stimulate the secretory pathway concertedly is to overexpress the unfolded protein response (UPR) activating transcription factor Hac1. Transcriptional analyses in *S. cerevisiae* revealed that up to 330 genes are regulated by Hac1, most of them belonging to the functional groups of secretion or the biogenesis of secretory organelles (e.g. ER-resident chaperones, foldases, components of the translocon). Interestingly, genes encoding proteins involved in protein degradation, vesicular trafficking, lipid biogenesis and vacuolar sorting are also induced by Hac1 [94]. In this context, Higashio and Kohno [95] describe the stimulation of ER-to-Golgi transport through the UPR by inducing COPII vesicle formation. The homologs of *S. cerevisiae HAC1* in *T. reesei*

(hac1) and *A. nidulans* (hacA) have been identified [50] and the effects of UPR induction by constitutive overexpression of these genes have been evaluated. The heterologous overexpression of *T. reesei* hac1 in *S. cerevisiae* yielded a 2.4-fold improvement in *Bacillus* α-amylase secretion, and a slight increase in the yeast endogenous invertase as well as in total protein secretion. *S. cerevisiae HAC1* overexpression was shown to enhance secretion of the endogenous invertase (2-fold), and recombinant α-amylase (70% increase), but did not effect secretion of *T. reesei* EGI, a protein supposed to accumulate in the ER. Disruption of *HAC1* in *S. cerevisiae* led to a reduced secretion of the two recombinant proteins (α-amylase -75%, EGI -50%), but not of the endogenous invertase [96]. Similar results could also be seen in *A. awamori*, where overproduction of *A. awamori* hacA ameliorated secretion of *Trametes versicolor* laccase and bovine preprochymosin 7-fold and 2.8 fold, respectively [97], and in *P. pastoris*, where heterologous expression of *S. cerevisiae HAC1* increased the secretion rate of a Fab antibody fragment [83].

Novel strategies: genome wide-screening

All these approaches are limited to the existing knowledge base. Novel processes might be identified and targeted to improve secretion (including non-UPR regulated genes) through different approaches. In this regard, high throughput flow cytometry and cell sorting are valuable tools to isolate overproducing clones [98]. One approach is to screen overexpression libraries for improved secretion of heterologous protein, which is anchored to the cell surface via agglutinin (Aga2p) and detected by immunofluorescent staining. Shusta et al. [99] showed that the levels of surface-displayed single chain T-cell receptors correlated strongly with the soluble expression of the respective proteins. A 3-fold higher secreting clone could be isolated out of a library potentially as large as 10^8 in a couple of weeks [100]. Screening of a yeast cDNA library in *S. cerevisiae* surface display strains identified cell wall proteins, translational components and the folding assistant Ero1 as beneficial for the secretion of various antibody fragments [101]. However, one potential drawback of this high throughput method is that the display efficiency of the protein of interest can be dominated by its fusion partner Aga2p, as BiP and *PDI* overexpression had no effect on surface display levels of the scFvs although they increased soluble expression levels [85].

Furthermore, genome-wide analytical tools like DNA microarrays are regarded as data mining source for physiological effects, stress regulation and host engineering. Sauer et al. [102] have analysed the differential transcriptome of a *P. pastoris* strain overexpressing human trypsinogen versus a non-expressing strain. 13 out of the 524 significantly regulated genes were selected, and their *S. cerevisiae* homologs were overexpressed in a *P. pastoris* strain producing a human antibody Fab fragment [103]. Five previously characterized secretion helpers (*PDI1*, *ERO1*, *SSO2*, *KAR2*/BiP and *HAC1*), as well as 6 novel, hitherto unidentified, factors, more precisely Bfr2 and Bmh2 involved in protein transport, the chaperones Ssa4 and Sse1, the vacuolar ATPase subunit Cup5 and Kin2, a protein kinase connected to exocytosis, proved their benefits for practical application in lab scale production processes by increasing both specific production rates as well as volumetric productivity of an antibody fragment up to 2.5-fold in fed batch fermentations of *P. pastoris* [103].

Protein folding and conformational stress in prokaryotic cells

Since early recombinant DNA times, bacteria (especially *E. coli*) have been the most widely used microorganisms for recombinant protein production due to genetic simplicity, fast growth rate, high cell density production and availability of a spectrum of genetic systems, among others. For production processes being efficient, foreign genes are expressed from plasmids and under the control of inducible promoters, what results into non physiological and unusually high transcription rates. Strong production of recombinant proteins can lead to a stressful situation for the host cell, with the extent of the bacterial stress response being determined by the specific properties of the recombinant protein, and by the rates of transcription and translation [104]. This fact has a clear and profoundly negative impact on productivity and probably protein quality. In addition, recombinant proteins fail, very often, to reach their native conformation when produced in bacteria [105]. This is caused by a coincidence of diverse events impairing protein folding including bottlenecks in transcription and translation, undertitration of chaperones and proteases, improper codon usage and inability of disulfide-bond formation [106,107]. Misfolded protein species usually deposit as amorphous masses of insoluble material called inclusion bodies [108], recorded as by-products of bacterial protein production processes. Inclusion bodies are mainly formed by the deposition of unfolded or partially misfolded protein species that interact through hydrophobic patches unusually exposed to the solvent and with high amino acid sequence homology [109,110]. The specificity in protein aggregation makes inclusion bodies highly pure in composition and therefore enriched in the recombinant protein itself. However, also truncated versions of the recombinant product, other plasmid-encoded proteins, but also defined host cell proteins can get entrapped within bacterial inclusion bodies [111-116]. Moreover, the presence of folding assistant proteins in inclusion bodies [117-119] confirm that specific interactions lead to entrappment of cellular proteins in these aggregates. The high purity of inclusion bodies makes them a convenient source of easily extractable pro-

tein that must be refolded in vitro by denaturing procedures, a fact that has been largely exploited for biotechnology purposes [120]. The potential routes of a newly synthesized protein in the bacterial cytosol are illustrated in Figure 3.

Although inclusion bodies are mainly found in the cytoplasm, they occur also in the bacterial periplasm if proteins have been engineered to present a leader peptide for secretion [121]. In fact, a control quality system mostly separated from that acting in the cytoplasm assist folding of secreted proteins in the periplasmic space of gram negative bacteria. This is regulated through the combined activity of two partially overlapping systems, regulated by the alternate σ factor σ^E and by the Cpx envelope stress signalling system, that intricately combine the activity of chaperones and proteases [122,123]. However, the simultaneous activation of stress signals in both bacterial compartments upon the production of misfolding prone proteins strongly suggest a close physiological and genetic connection between cytoplasmic and extracytoplasmic systems [124]. The quality control and conformational stress in the periplasmic space has been extensively revised elsewhere [121,125].

Different to the unfolded protein response (UPR) described in eukaryotic cells, the physiological reaction to conformational stress in the bacterial cytoplasm has not

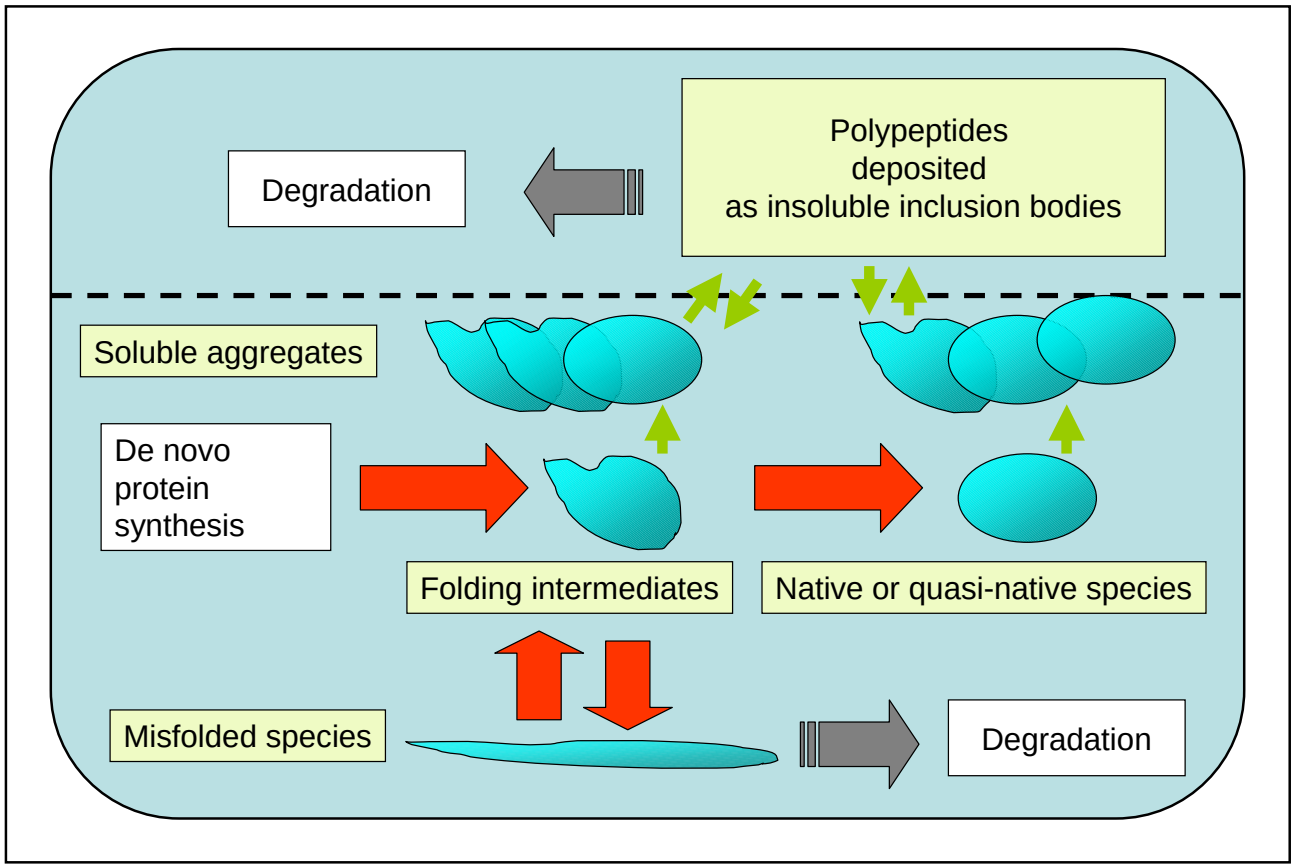


Figure 3
Schematic representation of protein folding and aggregation in recombinant *E. coli*. After de novo synthesis, a fraction of recombinant proteins (especially heterologous proteins with conformationally complex disulfide bridges) do not reach their native conformation and aggregate as insoluble deposits named inclusion bodies. Protein aggregates already exist in the soluble cell fraction, and can involve native or quasi-native protein species. The main cytoplasmic chaperones involved in the protein folding process (red arrows) include the trigger factor, DnaK, DnaJ, GrpE, GroEL and GroES. Both soluble aggregates and individual protein species can enter the virtual insoluble cell fraction (indicated by a dashed line) and deposit as inclusion bodies, in a fully reversible process (green arrows). Protein release from inclusion bodies is mainly controlled by DnaK, ClpB and IbpA,B. Proteases (Ion, ClpP and others) attack both soluble and insoluble species with folding defects. In particular, proteases degrade inclusion body proteins in situ, or through a more complex process intimately related to the protein release process, and therefore, strongly dependent on DnaK.

received any similar precise name. Transcriptome analysis of recombinant *E. coli* has resulted in a catalogue of genes up-regulated during protein production [126,127]. Among them several heat shock genes have been identified (including those encoding the proteases Lon, ClpP, HslV and HslU, and the chaperones IbpA, IbpB, DnaK, DnaJ, ClpB, HtpG, MopA and MopB among others) but also other ones not directly involved in protein quality (such as YagU, YojH, YbeD and others) and whose precise role remains to be identified. This fact indicates that the conformational stress imposed by protein production is more complex and physiologically distinguishable from that caused by thermal denaturation, namely the heat shock response [128,129], and includes several overlapping stress responses [104] Well characterized stress events have been observed during recombinant protein production such as SOS DNA repair [130] and stringent responses [131], although it is still be solved whether such reactions are directly associated to the prevalence of unfolded or misfolded protein species and the eventual connection with the σ^{32}-regulated heat shock response. The expression of some of these stress genes is being used as a convenient marker of conformational stress in recombinant cells [132].

The bacterial conformational stress itself has been poorly characterized from its physiological side. Instead, many efforts have been addressed to a rather practical issue such as minimizing aggregation, what in turn has resulted in a better comprehension of in vivo protein folding processes. Since solubility has been considered for a long time being synonymous with protein quality, increasing the relative yield of soluble protein has been targeted by physicochemical approaches. From already classical studies, it is well known that high temperatures impair protein folding and favour aggregation of the recombinant proteins as inclusion bodies [133,134]. Therefore, reducing the growth temperature has been a general strategy used to minimize inclusion body formation [135-137] that, like other strategies, has rendered moderately positive, but unpredictable and product-dependent results [107]. Fusion of folding-reluctant species to highly soluble homologous or thermostable proteins has in some cases, resulted in moderate enhancement of the passenger protein solubility [106,138,139].

Chaperones and protein degradation

Folding failures of recombinant proteins produced in *E. coli* is generally attributed to a limitation in the cell concentration of folding assistant elements, which cannot process the newly synthesized aggregation prone polypeptides. This assumption is physiologically supported by the overexpression of chaperone genes, in particular of chaperone genes from the heat-shock protein family, in response to recombinant protein overproduction [126,127,133]. Thus, coproduction of the main heat shock chaperones (specially GroEL and DnaK) together with the target protein has been largely explored as a way to minimize aggregation and to enhance the solubility of the recombinant protein product (reviewed in [140-143]). In many cases, solubility has been significantly enhanced by coexpression of individual chaperone genes, while in others an even negative effect on product stability and host viability has been observed. Selection of the suitable chaperone(s) is still a trial-and-error process. However, more recent results indicate that complete chaperones gene sets rather than individual chaperone genes with synergistic and/or cooperative activities (such as DnaK-DnaJ-GrpE and GroEL-GroES sets) will lead to a more predictable improvement of target protein solubility. [144-147].

Interestingly, when producing enzymes or fluorescent proteins in DnaK$^-$ cells, the biological activities and therefore the conformational quality of aggregated polypeptides is much more close to that of soluble versions, compared to wild type cells [148-150]. Furthermore, the overexpression of the *dnaK* gene along with a model GFP recombinant protein dramatically reduces the specific fluorescence of a GFP fusion in both soluble and insoluble versions [151]. This indicates that DnaK directly or indirectly impairs the folding state of the aggregated proteins. In this regard, the production of GFP variants in absence of DnaK results in highly fluorescent inclusion bodies [152]. In these cells, both the protein yield and quality were dramatically enhanced although the solubility is lower than in the wild type, as expected. This occurs by the inhibition of GFP proteolysis mediated by the proteases Lon and ClpP, which participate in the in vivo disintegration of inclusion bodies in absence of protein synthesis [153,154]. Probably, such proteases act coordinately in a disaggregation complex [155-157] in which DnaK, ClpB and IbpAB remove aggregated polypeptides for proteolytic digestion. Therefore, although solubility can be indeed enhanced by high levels of DnaK, GroEL and other chaperones it occurs at expenses of quality and yield, probably by generally stimulating proteolysis [116]. In fact, solubility and conformational quality are not only non coincident parameters [158] but highly divergent protein features [152].

Disulfide-bond formation in recombinant E. coli

Usually, the cytoplasmic space of *E. coli* is a reducing environment. Therefore, disulfide-bonds within proteins are not formed, a fact that represents an additional obstacle for proper folding of many recombinant proteins. There are two approaches to produce disulfide-bonded proteins in *E. coli* expression, namely *in vitro* refolding of inclusion body proteins under appropriate redox conditions [120] or manipulating *in vivo* conditions by either converting

the cytoplasm into an oxidizing environment or secreting the protein into the periplasmic space or even further into the culture medium (less reducing environments). Correct disulfide bond formation in the periplasm of *E. coli* is a catalyzed process, where the oxidation of cysteine pairs occurs through the transfer of disulfides from the highly oxidizing DsbA/DsbB proteins to the proof-reading proteins DsbC/DsbD which are able to rearrange non-native disulfides to their native configuration [159]. In particular, overexpression of DsbC has been shown to increase the yield of correctly disulfide-bonded proteins in the periplasm of *E. coli* [160-162]. The co-expression of eukaryotic protein disulfide isomerases in *E. coli* can also favour the formation of disulfide bonds in the periplasmic space [163,164].

Disulfide bond formation in the cytoplasm of *E. coli* can occur when the genes encoding thioredoxin reductase (*trxB*) and glutathione oxido-reductase (*gor*) are inactivated [165,166]. A double-mutant strain containing appropriate mutations, known as Origami, has been used, for example, to generate active variants of tissue-type plasminogen activator [165] and functional antibody fragments in the *E. coli* cytoplasm [167,168]. In some cases, recovery of functional cytoplasmic disulfide-bonded proteins can be further enhanced by coexpressing signal sequence deficient periplasmic chaperones and/or disulfide-bond isomerases such as DsbC [165,167,169,170]. Unfortunately, *trxB gor* mutants exhibit impaired growth characteristics [112,165], but, at least for antibody fragments it has been shown that expression yields of correctly disulfide-bonded proteins in the cytoplasm can be similar to those obtained by secretion into the periplasmic space [171].

Protein folding and secretion in non-conventional bacterial expression systems

Although *E. coli* is still the most commonly used prokaryotic organism for heterologous protein production, other bacterial hosts are becoming more and more attractive.

Gram-positive Bacilli strains are also frequently employed at industrial level. In contrast to *E. coli*, their outer envelope has no lipopolysaccharides, also called "endotoxins" since they exert a pyrogenic activity in humans or other mammals. Therefore, many pharmaceutically relevant products have been obtained in several strains [172]. In addition, the Bacilli strains are attractive hosts because they have a naturally high secretion capacity, as they export proteins directly into the extracellular medium. Amongst Bacilli species, the protein secretion pathway in *B. subtilis* have been deeply investigated at molecular level and a comprehensive literature survey is reported in [173]. Several bottlenecks in the expression and secretion of heterologous proteins have been highlighted [174]. The secretory pathway of proteins can be divided into three functional stages: the early stage, involving the synthesis of secretory pre-proteins, their interaction with chaperones and binding to the secretory translocase complex; the second stage, consisting in translocation across the cytoplasmic membrane; and the last stage, including removal of the N-terminal signal peptide, protein refolding and passage through the cell wall. A pivotal role in the whole secretion process is played by molecular chaperones [175]. *B. subtilis* has two types of molecular chaperones, intracellular and extra-cytoplasmic molecular chaperones. GroES, GroEL, DnaK, DnaJ and GrpE are intracellular molecular chaperones. Besides being involved in and largely responsible for protein folding and minimizing aggregation, these chaperones maintain pre-proteins (the products to be secreted) in translocation-competent conformations [176]. PrsA is the only known extracytoplasmic folding factor in *B. subtilis*. PrsA is a lipoprotein that consists of a 33 kDa lysine-rich protein part and the N-terminal cysteine with a thiol-linked diacylglycerol anchoring the protein to the outer leaflet of the cytoplasmic membrane [177]. Subsequent folding of a secreted mature protein into a stable and active conformation usually requires PrsA protein. In *prsA* mutants, the secretion and stability of some model proteins is decreased, if not abolished, while overproduction of PrsA enhances the secretion of exoproteins engineered to be expressed at a high level [178].

There is, however, a physiologic limit to the overloading of *B. subtilis* secretory pathway. The massive production of homologous and heterologous exoproteins has been reported to induce a phenomenon called "protein secretion stress response" [179]. The CssRS two-component regulatory system is able to detect the presence of partially folded or unfolded exo-protein intermediates and activates the transcription of several genes, among which a key role is played by *htrAB*. These genes encode two membrane localised serine proteases involved in the proteolysis of aberrant products [180].

Several gene expression systems using non-conventional prokaryotic organisms as host cells have been developed over the last decades. Each bacterial host was generally implemented to overcome defined problems/bottlenecks observed during the recombinant production of specific protein classes in conventional systems, such as *E. coli* and *B. subtilis*. The use of such non-conventional systems is still very limited and largely suffers from the lack of molecular details concerning host physiology and any other phenomenon related to massive recombinant protein production. Notwithstanding, some of them may represent useful model systems to further investigate on the optimization of recombinant protein folding and quality.

In this context, some interest has been raised by the implementation of an Antarctic Gram negative bacterium as production host. *Pseudoalteromonas haloplanktis* TAC125 was isolated from a sea water sample collected in the vicinity of the Dumont d'Urville Antarctic station, in Terre Adélie. It is characterised by fast growth rates, combined with the ability to reach very high cell densities even in uncontrolled laboratory growth conditions and to be easily transformed by intergeneric conjugation [181]. There features made *P. haloplanktis* TAC125 an attracting host for the development of an efficient gene-expression system for the recombinant protein production at low temperatures of thermal-labile and aggregation-prone proteins [182]. Furthermore, it was the first Antarctic Gram-negative bacterium which genome was fully determined and carefully annotated [183].

Since high temperatures have a general negative impact on protein folding due to the strong temperature dependence of hydrophobic interactions that mainly drive the aggregation reaction [184], and favour conformational stress, the production of recombinant proteins at low temperatures represents an exciting model to study the dynamics of protein folding and misfolding and to improve the quality of the products. The growth of *E. coli* below 37°C has been often explored to minimize aggregation but without consistent, protein-irrespective results. Also, the use of suboptimal growth temperatures might negatively affect the biology of the host cell and the performance of the process in undesirable and non predictable ways. Recombinant protein production in psychrophilic bacteria, i.e. at temperature as low as 4°C, may minimize undesired hydrophobic interactions during protein folding, desirably resulting in enhancing the yield of soluble and correctly folded products while operating close to the optimal growth range. Furthermore, with respect to mesophilic cells growing at suboptimal temperatures, psychrophiles contain a full set of folding factors already adapted to assist optimally, when required, protein folding at freezing temperatures.

The efficiency of the cold-adapted expression system was tested by producing several aggregation-prone products in *P. haloplanktis* TAC125, such as a yeast α-glucosidase [182], the mature human nerve growth factor [182], and a cold adapted lipase [185]. All the recombinant products resulted to be fully soluble and biologically competent.

Concluding remarks

In vivo protein folding is a very complex issue that involves many cellular proteins and physiological responses. During recombinant protein production, conformational stress conditions elicited by the synthesis of aggregation prone polypeptides profoundly alter the physiology of the host cell, triggering mechanisms addressed to manage potentially toxic misfolding protein species and to recover the cell folding homeostasis. The use of different microorganims as factories for recombinant protein production, including yeast, filamentous fungi and bacteria has resulted in dramatic gains of information about the biology of such stress responses, and has provided valuable information to better understand the mechanics of in vivo protein folding and aggregation.

However, so far it has not been possible to create the "perfect folding environment". Especially with respect to industrial protein production processes, the direct impact of altered process conditions on recombinant protein folding remains unclear. Ongoing research in the authors' labs is targeted to elucidate the physiological responses of different eukaryotic and prokaryotic microbial hosts on a genome wide level in order to interrelate environmental stresses to protein folding/aggregation mechanisms and eliminate bottlenecks.

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

All authors contributed equally to this manuscript, and read and approved the final version.

Acknowledgements

The authors gratefully acknowledge the financial support provided by MEC, Spain and Austria Science Fund through the EUROCORES Programmes EuroSCOPE of the European Science Foundation, which is also funded by the European Commission, Contract no. ERAS-CT-2003-980409 of the Sixth Framework Programme. References of received grants are:

to AV: BIO2005-23732-E and BIO2007-61194 MEC, and 2005SGR-00956 AGAUR, Spain

to PF: Complementary Actions Plan from the Spanish MEC, Project No. BIO2005-23733-E. AGAUR (Generalitat de Catalunya, Spain), Grant 2005SGR 00698 and Reference Network in Biotechnology (XerBa).

to DM: Austrian Science Fund, project I37.

References

1. Gething MJ, Sambrook J: **Protein folding in the cell.** *Nature* 1992, **355(6355):**33-45.
2. Stryer , Lubert : **Biochemie.** *Spektrum der Wissenschaft VerlagsGmbH* 1995.
3. Shuster JR: **Gene expression in yeast: protein secretion.** *Curr Opin Biotechnol* 1991, **2(5):**685-690.
4. Brodsky JL, Scott CM: **Tipping the delicate balance: defining how proteasome maturation affects the degradation of a substrate for autophagy and endoplasmic reticulum associated degradation (ERAD).** *Autophagy* 2007, **3(6):**623-625.
5. Meusser B, Hirsch C, Jarosch E, Sommer T: **ERAD: the long road to destruction.** *Nat Cell Biol* 2005, **7(8):**766-772.
6. Kleizen B, Braakman I: **Protein folding and quality control in the endoplasmic reticulum.** *Curr Opin Cell Biol* 2004, **16(4):**343-349.
7. Molinari M, Galli C, Vanoni O, Arnold SM, Kaufman RJ: **Persistent glycoprotein misfolding activates the glucosidase II/UGT1-**

driven calnexin cycle to delay aggregation and loss of folding competence. *Mol Cell* 2005, **20(4):**503-512.
8. Ma Y, Hendershot LM: **The unfolding tale of the unfolded protein response.** *Cell* 2001, **107(7):**827-830.
9. Patil C, Walter P: **Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals.** *Curr Opin Cell Biol* 2001, **13(3):**349-355.
10. Nakano A: **Yeast Golgi apparatus – dynamics and sorting.** *Cell Mol Life Sci* 2004, **61(2):**186-191.
11. Mattanovich D, Gasser B, Hohenblum H, Sauer M: **Stress in recombinant protein producing yeasts.** *J Biotechnol* 2004, **113(1–3):**121-135.
12. Garcia-Kirchner O, Segura-Granados M, Rodriguez-Pascual P: **Effect of media composition and growth conditions on production of beta-glucosidase by Aspergillus niger C-6.** *Appl Biochem Biotechnol* 2005, **121–124:**347-359.
13. Liu CQ, Chen QH, Cheng QJ, Wang JL, He GQ: **Effect of cultivating conditions on alpha-galactosidase production by a novel Aspergillus foetidus ZU-G1 strain in solid-state fermentation.** *J Zhejiang Univ Sci B* 2007, **8(5):**371-376.
14. Wang L, Ridgway D, Gu T, Moo-Young M: **Effects of process parameters on heterologous protein production in Aspergilus niger fermentation.** *Journal of Chemical Technology and Biotechnology* 2003, **78(12):**1259-1266.
15. Wang L, Ridgway D, Gu T, Moo-Young M: **Bioprocessing strategies to improve heterologous protein production in filamentous fungal fermentations.** *Biotechnol Adv* 2005, **23(2):**115-129.
16. Hackel BJ, Huang D, Bubolz JC, Wang XX, Shusta EV: **Production of soluble and active transferrin receptor-targeting single-chain antibody using Saccharomyces cerevisiae.** *Pharm Res* 2006, **23(4):**790-797.
17. Li Z, Xiong F, Lin Q, d'Anjou M, Daugulis AJ, Yang DS, Hew CL: **Low-temperature increases the yield of biologically active herring antifreeze protein in Pichia pastoris.** *Protein Expr Purif* 2001, **21(3):**438-445.
18. Lin H, Kim T, Xiong F, Yang X: **Enhancing the production of Fc fusion protein in fed-batch fermentation of Pichia pastoris by design of experiments.** *Biotechnol Prog* 2007, **23(3):**621-625.
19. Shi X, Karkut T, Chamankhah M, Alting-Mees M, Hemmingsen SM, Hegedus D: **Optimal conditions for the expression of a single-chain antibody (scFv) gene in Pichia pastoris.** *Protein Expr Purif* 2003, **28(2):**321-330.
20. Gasser B, Maurer M, Rautio J, Sauer M, Bhattacharyya A, Saloheimo M, Penttilä M, Mattanovich D: **Monitoring of transcriptional regulation in Pichia pastoris under protein production conditions.** *BMC Genomics* 2007, **8:**179.
21. Smith JD, Richardson NE, Robinson AS: **Elevated expression temperature in a mesophilic host results in increased secretion of a hyperthermophilic enzyme and decreased cell stress.** *Biochim Biophys Acta* 2005, **1752(1):**18-25.
22. Aguilar-Uscanga B, Francois JM: **A study of the yeast cell wall composition and structure in response to growth conditions and mode of cultivation.** *Lett Appl Microbiol* 2003, **37(3):**268-274.
23. Xiao A, Zhou X, Zhou L, Zhang Y: **Improvement of cell viability and hirudin production by ascorbic acid in Pichia pastoris fermentation.** *Appl Microbiol Biotechnol* 2006, **72(4):**837-844.
24. Li JR, Yu P: **Expression of Cu, Zn-superoxide dismutase gene from Saccharomyces cerevisiae in Pichia pastoris and its resistance to oxidative stress.** *Appl Biochem Biotechnol* 2007, **136(1):**127-139.
25. Tu BP, Weissman JS: **Oxidative protein folding in eukaryotes: mechanisms and consequences.** *J Cell Biol* 2004, **164(3):**341-346.
26. Baumann K, Maurer M, Dragosits M, Cos O, Ferrer P, Mattanovich D: **Hypoxic fed batch cultivation of Pichia pastoris increases specific and volumetric productivity of recombinant proteins.** *Biotechnol Bioeng* 2008, **100(1):**177-83.
27. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO: **Genomic expression programs in the response of yeast cells to environmental changes.** *Mol Biol Cell* 2000, **11(12):**4241-4257.
28. Mager WH, Siderius M: **Novel insights into the osmotic stress response of yeast.** *FEMS Yeast Res* 2002, **2(3):**251-257.
29. Ollivier H, Pichavant-Rafini K, Puill-Stephan E, Calves P, Nonnotte L, Nonnotte G: **Effects of hyposmotic stress on exocytosis in isolated turbot, Scophthalmus maximus, hepatocytes.** *J Comp Physiol [B]* 2006, **176(7):**643-652.
30. Wu MH, Dimopoulos G, Mantalaris A, Varley J: **The effect of hyperosmotic pressure on antibody production and gene expression in the GS-NS0 cell line.** *Biotechnol Appl Biochem* 2004, **40(Pt 1):**41-46.
31. Jahic M, Veide A, Charoenrat T, Teeri T, Enfors SO: **Process technology for production and recovery of heterologous proteins with Pichia pastoris.** *Biotechnol Prog* 2006, **22(6):**1465-1473.
32. Fernandes AR, Mira NP, Vargas RC, Canelhas I, Sa-Correia I: **Saccharomyces cerevisiae adaptation to weak acids involves the transcription factor Haa1p and Haa1p-regulated genes.** *Biochem Biophys Res Commun* 2005, **337(1):**95-103.
33. Kapteyn JC, ter Riet B, Vink E, Blad S, De Nobel H, Van Den Ende H, Klis FM: **Low external pH induces HOG1-dependent changes in the organization of the Saccharomyces cerevisiae cell wall.** *Mol Microbiol* 2001, **39(2):**469-479.
34. Dorner A, Kaufman R: **The levels of endoplasmic reticulum proteins and ATP affect folding and secretion of selective proteins.** *Biologicals* 1994, **22(2):**103-112.
35. Nishikawa S, Fewell S, Kato Y, Brodsky J, Endo T: **Molecular chaperones in the yeast endoplasmic reticulum maintain the solubility of proteins for retrotranslocation and degradation.** *J Cell Biol* 2001, **153(5):**1061-1070.
36. Kozutsumi Y, Segal M, Normington K, Gething M, Sambrook J: **The presence of malfolded proteins in the endoplasmic reticulum signals the induction of glucose-regulated proteins.** *Nature* 1988, **332(6163):**462-464.
37. Robinson A, Wittrup K: **Constitutive overexpression of secreted heterologous proteins decreases extractable BiP and protein disulfide isomerase levels in Saccharomyces cerevisiae.** *Biotechnol Prog* 1995, **11(2):**171-177.
38. Kauffman K, Pridgen E, Doyle Fr, Dhurjati P, Robinson A: **Decreased protein expression and intermittent recoveries in BiP levels result from cellular stress during heterologous protein expression in Saccharomyces cerevisiae.** *Biotechnol Prog* 2002, **18(5):**942-950.
39. Hohenblum H, Gasser B, Maurer M, Borth N, Mattanovich D: **Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant Pichia pastoris.** *Biotechnol Bioeng* 2004, **85(4):**367-375.
40. Nyyssönen E, Keränen S: **Multiple roles of the cellulase CBHI in enhancing production of fusion antibodies by the filamentous fungus Trichoderma reesei.** *Curr Genet* 1995, **28(1):**71-79.
41. Pakula T, Salonen K, Uusitalo J, Penttilä M: **The effect of specific growth rate on protein synthesis and secretion in the filamentous fungus Trichoderma reesei.** *Microbiology* 2005, **151(Pt 1):**135-143.
42. Punt P, van Gemeren I, Drint-Kuijvenhoven J, Hessing J, van Muijlwijk-Harteveld G, Beijersbergen A, Verrips C, van den Hondel C: **Analysis of the role of the gene bipA, encoding the major endoplasmic reticulum chaperone protein in the secretion of homologous and heterologous proteins in black Aspergilli.** *Appl Microbiol Biotechnol* 1998, **50(4):**447-454.
43. Plemper RK, Bordallo J, Deak PM, Taxis C, Hitt R, Wolf DH: **Genetic interactions of Hrd3p and Der3p/Hrd1p with Sec61p suggest a retro-translocation complex mediating protein transport for ER degradation.** *J Cell Sci* 1999, **112(Pt 22):**4123-4134.
44. Sagt C, Müller W, van der Heide L, Boonstra J, Verkleij A, Verrips C: **Impaired cutinase secretion in Saccharomyces cerevisiae induces irregular endoplasmic reticulum (ER) membrane proliferation, oxidative stress, and ER-associated degradation.** *Appl Environ Microbiol* 2002, **68(5):**2155-2160.
45. Haynes CM, Titus EA, Cooper AA: **Degradation of misfolded proteins prevents ER-derived oxidative stress and cell death.** *Mol Cell* 2004, **15(5):**767-776.
46. Kincaid MM, Cooper AA: **ERADicate ER stress or die trying.** *Antioxid Redox Signal* 2007, **9(12):**2373-2387.
47. Ron D, Walter P: **Signal integration in the endoplasmic reticulum unfolded protein response.** *Nat Rev Mol Cell Biol* 2007, **8(7):**519-529.
48. Cox JS, Shamu CE, Walter P: **Transcriptional induction of genes encoding endoplasmic reticulum resident proteins requires a transmembrane protein kinase.** *Cell* 1993, **73(6):**1197-1206.
49. Cox JS, Walter P: **A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response.** *Cell* 1996, **87(3):**391-404.

50. Saloheimo M, Valkonen M, Penttila M: **Activation mechanisms of the HAC1-mediated unfolded protein response in filamentous fungi.** *Mol Microbiol* 2003, **47(4):**1149-1161.
51. Yoshida H, Matsui T, Yamamoto A, Okada T, Mori K: **XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor.** *Cell* 2001, **107(7):**881-891.
52. Sidrauski C, Cox JS, Walter P: **tRNA ligase is required for regulated mRNA splicing in the unfolded protein response.** *Cell* 1996, **87(3):**405-413.
53. Ruegsegger U, Leber JH, Walter P: **Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response.** *Cell* 2001, **107(1):**103-114.
54. Mulder HJ, Saloheimo M, Penttila M, Madrid SM: **The transcription factor HACA mediates the unfolded protein response in Aspergillus niger, and up-regulates its own transcription.** *Mol Genet Genomics* 2004, **271(2):**130-140.
55. Arvas M, Pakula T, Lanthaler K, Saloheimo M, Valkonen M, Suortti T, Robson G, Penttila M: **Common features and interesting differences in transcriptional responses to secretion stress in the fungi Trichoderma reesei and Saccharomyces cerevisiae.** *BMC Genomics* 2006, **7:**32.
56. Guillemette T, van Peij N, Goosen T, Lanthaler K, Robson G, van den Hondel C, Stam H, Archer D: **Genomic analysis of the secretion stress response in the enzyme-producing cell factory Aspergillus niger.** *BMC Genomics* 2007, **8:**158.
57. Sims AH, Gent ME, Lanthaler K, Dunn-Coleman NS, Oliver SG, Robson GD: **Transcriptome analysis of recombinant protein secretion by Aspergillus nidulans and the unfolded-protein response in vivo.** *Appl Environ Microbiol* 2005, **71(5):**2737-2747.
58. Resina D, Bollók M, Khatri N, Valero F, Neubauer P, Ferrer P: **Transcriptional response of P. pastoris in fed-batch cultivations to Rhizopus oryzae lipase production reveals UPR induction.** *Microb Cell Fact* 2007, **6:**21.
59. Harding HP, Zhang Y, Ron D: **Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase.** *Nature* 1999, **397(6716):**271-274.
60. Pakula TM, Laxell M, Huuskonen A, Uusitalo J, Saloheimo M, Penttila M: **The effects of drugs inhibiting protein secretion in the filamentous fungus Trichoderma reesei. Evidence for down-regulation of genes that encode secreted proteins in the stressed cells.** *J Biol Chem* 2003, **278(45):**45011-45020.
61. Al-Sheikh H, Watson AJ, Lacey GA, Punt PJ, MacKenzie DA, Jeenes DJ, Pakula T, Penttila M, Alcocer MJ, Archer DB: **Endoplasmic reticulum stress leads to the selective transcriptional down-regulation of the glucoamylase gene in Aspergillus niger.** *Mol Microbiol* 2004, **53(6):**1731-1742.
62. Martinez IM, Chrispeels MJ: **Genomic analysis of the unfolded protein response in Arabidopsis shows its connection to important cellular processes.** *Plant Cell* 2003, **15(2):**561-576.
63. Robinson A, Hines V, Wittrup K: **Protein disulfide isomerase overexpression increases secretion of foreign proteins in Saccharomyces cerevisiae.** *Biotechnology (NY)* 1994, **12(4):**381-384.
64. Harmsen M, Bruyne M, Raué H, Maat J: **Overexpression of binding protein and disruption of the PMR1 gene synergistically stimulate secretion of bovine prochymosin but not plant thaumatin in yeast.** *Appl Microbiol Biotechnol* 1996, **46(4):**365-370.
65. Kim M, Han K, Kang H, Rhee S, Seo J: **Coexpression of BiP increased antithrombotic hirudin production in recombinant Saccharomyces cerevisiae.** *J Biotechnol* 2003, **101(1):**81-87.
66. Lombraña M, Moralejo F, Pinto R, Martín J: **Modulation of Aspergillus awamori thaumatin secretion by modification of bipA gene expression.** *Appl Environ Microbiol* 2004, **70(9):**5145-5152.
67. Robinson A, Bockhaus J, Voegler A, Wittrup K: **Reduction of BiP levels decreases heterologous protein secretion in Saccharomyces cerevisiae.** *J Biol Chem* 1996, **271(17):**10017-10022.
68. van Gemeren I, Beijersbergen A, van den Hondel C, Verrips C: **Expression and secretion of defined cutinase variants by Aspergillus awamori.** *Appl Environ Microbiol* 1998, **64(8):**2794-2799.
69. van der Heide M, Hollenberg C, van der Klei I, Veenhuis M: **Overproduction of BiP negatively affects the secretion of Aspergillus niger glucose oxidase by the yeast Hansenula polymorpha.** *Appl Microbiol Biotechnol* 2002, **58(4):**487-494.
70. Smith J, Tang B, Robinson A: **Protein disulfide isomerase, but not binding protein, overexpression enhances secretion of a non-disulfide-bonded protein in yeast.** *Biotechnol Bioeng* 2004, **85(3):**340-350.
71. Conesa A, Jeenes D, Archer DB, van den Hondel CA, Punt PJ: **Calnexin overexpression increases manganese peroxidase production in Aspergillus niger.** *Appl Environ Microbiol* 2002, **68(2):**846-851.
72. Chung JY, Lim SW, Hong YJ, Hwang SO, Lee GM: **Effect of doxycycline-regulated calnexin and calreticulin expression on specific thrombopoietin productivity of recombinant Chinese hamster ovary cells.** *Biotechnol Bioeng* 2004, **85(5):**539-546.
73. Higgins MK, Demir M, Tate CG: **Calnexin co-expression and the use of weaker promoters increase the expression of correctly assembled Shaker potassium channel in insect cells.** *Biochim Biophys Acta* 2003, **1610(1):**124-132.
74. Klabunde J, Kleebank S, Piontek M, Hollenberg CP, Hellwig S, Degelmann A: **Increase of calnexin gene dosage boosts the secretion of heterologous proteins by Hansenula polymorpha.** *FEMS Yeast Res* 2007, **7(7):**1168-1180.
75. Marechal A, Tanguay PL, Callejo M, Guerin R, Boileau G, Rokeach LA: **Cell viability and secretion of active proteins in Schizosaccharomyces pombe do not require the chaperone function of calnexin.** *Biochem J* 2004, **380(Pt 2):**441-448.
76. Parlati F, Dominguez M, Bergeron JJ, Thomas DY: **Saccharomyces cerevisiae CNE1 encodes an endoplasmic reticulum (ER) membrane protein with sequence similarity to calnexin and calreticulin and functions as a constituent of the ER quality control apparatus.** *J Biol Chem* 1995, **270(1):**244-253.
77. Arima H, Kinoshita T, Ibrahim HR, Azakami H, Kato A: **Enhanced secretion of hydrophobic peptide fused lysozyme by the introduction of N-glycosylation signal and the disruption of calnexin gene in Saccharomyces cerevisiae.** *FEBS Lett* 1998, **440(1–2):**89-92.
78. Song Y, Sata J, Saito A, Usui M, Azakami H, Kato A: **Effects of calnexin deletion in Saccharomyces cerevisiae on the secretion of glycosylated lysozymes.** *J Biochem (Tokyo)* 2001, **130(6):**757-764.
79. Hayano T, Hirose M, Kikuchi M: **Protein disulfide isomerase mutant lacking its isomerase activity accelerates protein folding in the cell.** *FEBS Lett* 1995, **377(3):**505-511.
80. Bao W, Fukuhara H: **Secretion of human proteins from yeast: stimulation by duplication of polyubiquitin and protein disulfide isomerase genes in Kluyveromyces lactis.** *Gene* 2001, **272(1–2):**103-110.
81. Iwata T, Tanaka R, Suetsugu M, Ishibashi M, Tokunaga H, Kikuchi M, Tokunaga M: **Efficient secretion of human lysozyme from the yeast, Kluyveromyces lactis.** *Biotechnol Lett* 2004, **26(23):**1803-1808.
82. Vad R, Nafstad E, Dahl L, Gabrielsen O: **Engineering of a Pichia pastoris expression system for secretion of high amounts of intact human parathyroid hormone.** *J Biotechnol* 2005, **116(3):**251-260.
83. Gasser B, Maurer M, Gach J, Kunert R, Mattanovich D: **Engineering of Pichia pastoris for improved production of antibody fragments.** *Biotechnol Bioeng* 2006, **94(2):**353-361.
84. Inan M, Aryasomayajula D, Sinha J, Meagher M: **Enhancement of protein secretion in Pichia pastoris by overexpression of protein disulfide isomerase.** *Biotechnol Bioeng* 2006, **93(4):**771-778.
85. Shusta E, Raines R, Plückthun A, Wittrup K: **Increasing the secretory capacity of Saccharomyces cerevisiae for production of single-chain antibody fragments.** *Nat Biotechnol* 1998, **16(8):**773-777.
86. Damasceno L, Anderson K, Ritter G, Cregg J, Old L, Batt C: **Co-overexpression of chaperones for enhanced secretion of a single-chain antibody fragment in Pichia pastoris.** *Appl Microbiol Biotechnol* 2006.
87. Borth N, Mattanovich D, Kunert R, Katinger H: **Effect of increased expression of protein disulfide isomerase and heavy chain binding protein on antibody secretion in a recombinant CHO cell line.** *Biotechnol Prog* 2005, **21(1):**106-111.
88. Klabunde J, Hollenberg C, Gellissen G: **Impact of overexpressed secretory-pathway genes on the secretion of IFNgamma in**

Hansenula polymorpha applying an rDNA-cointegration approach for assessment. *FEMS Yeast Res* 2005.
89. Derkx P, Madrid S: **The foldase CYPB is a component of the secretory pathway of Aspergillus niger and contains the endoplasmic reticulum retention signal HEEL.** *Mol Genet Genomics* 2001, **266(4):**537-545.
90. Wiebe M, Karandikar A, Robson G, Trinci A, Candia J, Trappe S, Wallis G, Rinas U, Derkx P, Madrid S, *et al.*: **Production of tissue plasminogen activator (t-PA) in Aspergillus niger.** *Biotechnol Bioeng* 2001, **76(2):**164-174.
91. Moralejo F, Watson A, Jeenes D, Archer D, Martín J: **A defined level of protein disulfide isomerase expression is required for optimal secretion of thaumatin by Aspegillus awamori.** *Mol Genet Genomics* 2001, **266(2):**246-253.
92. Gross E, Kastner D, Kaiser C, Fass D: **Structure of Ero1p, source of disulfide bonds for oxidative protein folding in the cell.** *Cell* 2004, **117(5):**601-610.
93. Lodi T, Neglia B, Donnini C: **Secretion of human serum albumin by Kluyveromyces lactis overexpressing KlPDI1 and KlERO1.** *Appl Environ Microbiol* 2005, **71(8):**4359-4363.
94. Travers K, Patil C, Wodicka L, Lockhart D, Weissman J, Walter P: **Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation.** *Cell* 2000, **101(3):**249-258.
95. Higashio H, Kohno K: **A genetic link between the unfolded protein response and vesicle formation from the endoplasmic reticulum.** *Biochem Biophys Res Commun* 2002, **296(3):**568-574.
96. Valkonen M, Penttilä M, Saloheimo M: **Effects of inactivation and constitutive expression of the unfolded- protein response pathway on protein production in the yeast Saccharomyces cerevisiae.** *Appl Environ Microbiol* 2003, **69(4):**2065-2072.
97. Valkonen M, Ward M, Wang H, Penttilä M, Saloheimo M: **Improvement of foreign-protein production in Aspergillus niger var. awamori by constitutive induction of the unfolded-protein response.** *Appl Environ Microbiol* 2003, **69(12):**6979-6986.
98. Mattanovich D, Borth N: **Applications of cell sorting in biotechnology.** *Microb Cell Fact* 2006, **5:**12.
99. Shusta EV, Kieke MC, Parke E, Kranz DM, Wittrup KD: **Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency.** *J Mol Biol* 1999, **292(5):**949-956.
100. Rakestraw JA, Baskaran AR, Wittrup KD: **A flow cytometric assay for screening improved heterologous protein secretion in yeast.** *Biotechnol Prog* 2006, **22(4):**1200-1208.
101. Wentz AE, Shusta EV: **A novel high-throughput screen reveals yeast genes that increase secretion of heterologous proteins.** *Appl Environ Microbiol* 2007, **73(4):**1189-1198.
102. Sauer M, Branduardi P, Gasser B, Valli M, Maurer M, Porro D, Mattanovich D: **Differential gene expression in recombinant Pichia pastoris analysed by heterologous DNA microarray hybridisation.** *Microb Cell Fact* 2004, **3(1):**17.
103. Gasser B, Sauer M, Maurer M, Stadlmayr G, Mattanovich D: **Transcriptomics-based identification of novel factors enhancing heterologous protein secretion in yeasts.** *Appl Environ Microbiol* 2007, **73(20):**6499-6507.
104. Hoffmann F, Rinas U: **Stress induced by recombinant protein production in Escherichia coli.** *Adv Biochem Eng Biotechnol* 2004, **89:**73-92.
105. Baneyx F, Mujacic M: **Recombinant protein folding and misfolding in Escherichia coli.** *Nat Biotechnol* 2004, **22(11):**1399-1408.
106. Sorensen HP, Mortensen KK: **Advanced genetic strategies for recombinant protein expression in Escherichia coli.** *J Biotechnol* 2005, **115(2):**113-128.
107. Sorensen HP, Mortensen KK: **Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli.** *Microb Cell Fact* 2005, **4(1):**1.
108. Villaverde A, Carrio MM: **Protein aggregation in recombinant bacteria: biological role of inclusion bodies.** *Biotechnol Lett* 2003, **25(17):**1385-1395.
109. Carrio M, Gonzalez-Montalban N, Vera A, Villaverde A, Ventura S: **Amyloid-like properties of bacterial inclusion bodies.** *J Mol Biol* 2005, **347(5):**1025-1037.
110. Speed MA, Wang DI, King J: **Specific aggregation of partially folded polypeptide chains: the molecular basis of inclusion body composition.** *Nat Biotechnol* 1996, **14(10):**1283-1287.
111. Hart R, Rinas U, Bailey J: **Protein composition of Vitreoscilla hemoglobin inclusion bodies produced in Escherichia coli.** *J Biol Chem* 1990, **265(21):**12728-12733.
112. Neubauer A, Soini J, Bollok M, Zenker M, Sandqvist J, Myllyharju J, Neubauer P: **Fermentation process for tetrameric human collagen prolyl 4-hydroxylase in Escherichia coli: Improvement by gene optimisation of the PDI/beta subunit and repeated addition of the inducer anhydrotetracycline.** *J Biotechnol* 2007, **128:**308-321.
113. Rinas U, Bailey JE: **Protein compositional analysis of inclusion bodies produced in recombinant Escherichia coli.** *Appl Microbiol Biotechnol* 1992, **37(5):**609-614.
114. Rinas U, Boone TC, Bailey JE: **Characterization of inclusion bodies in recombinant Escherichia coli producing high levels of porcine somatotropin.** *J Biotechnol* 1993, **28(2–3):**313-320.
115. Rinas U, Bailey JE: **Overexpression of bacterial hemoglobin causes incorporation of pre-beta-lactamase into cytoplasmic inclusion bodies.** *Appl Environ Microbiol* 1993, **59(2):**561-566.
116. Rinas U, Hoffmann F, Betiku E, Estape D, Marten S: **Inclusion body anatomy and functioning of chaperone-mediated in vivo inclusion body disassembly during high-level recombinant protein production in Escherichia coli.** *J Biotechnol* 2007, **127(2):**244-257.
117. Allen SP, Polazzi JO, Gierse JK, Easton AM: **Two novel heat shock genes encoding proteins produced in response to heterologous protein expression in Escherichia coli.** *J Bacteriol* 1992, **174(21):**6938-6947.
118. Carrio MM, Villaverde A: **Construction and deconstruction of bacterial inclusion bodies.** *J Biotechnol* 2002, **96(1):**3-12.
119. Carrio MM, Villaverde A: **Localization of chaperones DnaK and GroEL in bacterial inclusion bodies.** *J Bacteriol* 2005, **187(10):**3599-3601.
120. Vallejo LF, Rinas U: **Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins.** *MicrobCell Fact* 2004, **3(1):**11.
121. Miot M, Betton JM: **Protein quality control in the bacterial periplasm.** *MicrobCell Fact* 2004, **3(1):**4.
122. Connolly L, De Las Penas A, Alba B, Gross C: **The response to extracytoplasmic stress in Escherichia coli is controlled by partially overlapping pathways.** *Genes Dev* 1997, **11(15):**2012-2021.
123. Hunke S, Betton JM: **Temperature effect on inclusion body formation and stress response in the periplasm of Escherichia coli.** *Mol Microbiol* 2003, **50(5):**1579-1589.
124. Zahrl D, Wagner M, Bischof K, Koraimann G: **Expression and assembly of a functional type IV secretion system elicit extracytoplasmic and cytoplasmic stress responses in Escherichia coli.** *J Bacteriol* 2006, **188(18):**6611-6621.
125. Duguay A, Silhavy T: **Quality control in the bacterial periplasm.** *Biochim Biophys Acta* 2004, **1694(1–3):**121-134.
126. Jurgen B, Lin HY, Riemschneider S, Scharf C, Neubauer P, Schmid R, Hecker M, Schweder T: **Monitoring of genes that respond to overproduction of an insoluble recombinant protein in Escherichia coli glucose-limited fed-batch fermentations.** *Biotechnol Bioeng* 2000, **70(2):**217-224.
127. Lesley SA, Graziano J, Cho CY, Knuth MW, Klock HE: **Gene expression response to misfolded protein as a screen for soluble recombinant protein.** *Protein Eng* 2002, **15(2):**153-160.
128. Arsene F, Tomoyasu T, Bukau B: **The heat shock response of Escherichia coli.** *Int J Food Microbiol* 2000, **55(1–3):**3-9.
129. Bukau B: **Regulation of the Escherichia coli heat-shock response.** *Mol Microbiol* 1993, **9(4):**671-680.
130. Aris A, Corchero JL, Benito A, Carbonell X, Viaplana E, Villaverde A: **The expression of recombinant genes from bacteriophage lambda strong promoters triggers the SOS response in Escherichia coli.** *Biotechnol Bioeng* 1998, **60(5):**551-559.
131. Harcum SW, Bentley WE: **Heat-shock and stringent responses have overlapping protease activity in Escherichia coli. Implications for heterologous protein yield.** *Appl Biochem Biotechnol* 1999, **80(1):**23-37.
132. Schultz T, Martinez L, de MA: **The evaluation of the factors that cause aggregation during recombinant expression in E. coli is simplified by the employment of an aggregation-sensitive reporter.** *MicrobCell Fact* 2006, **5:**28.
133. Hoffmann F, Rinas U: **Kinetics of heat-shock response and inclusion body formation during temperature-induced produc-**

tion of basic fibroblast growth factor in high-cell-density cultures of recombinant Escherichia coli. *Biotechnol Prog* 2000, **16(6):**1000-1007.
134. Strandberg L, Enfors SO: **Factors influencing inclusion body formation in the production of a fused protein in Escherichia coli.** *Appl Environ Microbiol* 1991, **57(6):**1669-1674.
135. Vasina JA, Baneyx F: **Recombinant protein expression at low temperatures under the transcriptional control of the major Escherichia coli cold shock promoter cspA.** *Appl Environ Microbiol* 1996, **62(4):**1444-1447.
136. Vasina JA, Baneyx F: **Expression of aggregation-prone recombinant proteins at low temperatures: a comparative study of the Escherichia coli cspA and tac promoter systems.** *Protein Expr Purif* 1997, **9(2):**211-218.
137. Vasina JA, Peterson MS, Baneyx F: **Scale-up and optimization of the low-temperature inducible cspA promoter system.** *Biotechnol Prog* 1998, **14(5):**714-721.
138. De Marco V, Stier G, Blandin S, de Marco A: **The solubility and stability of recombinant proteins are increased by their fusion to NusA.** *Biochem Biophys Res Commun* 2004, **322(3):**766-771.
139. Huang H, Liu J, de MA: **Induced fit of passenger proteins fused to Archaea maltose binding proteins.** *Biochem Biophys Res Commun* 2006, **344(1):**25-29.
140. Baneyx F, Palumbo JL: **Improving heterologous protein folding via molecular chaperone and foldase co-expression.** *Methods Mol Biol* 2003, **205:**171-197.
141. Hoffmann F, Rinas U: **Roles of heat-shock chaperones in the production of recombinant proteins in Escherichia coli.** *Adv Biochem Eng Biotechnol* 2004, **89:**143-161.
142. Schlieker C, Bukau B, Mogk A: **Prevention and reversion of protein aggregation by molecular chaperones in the E. coli cytosol: implications for their applicability in biotechnology.** *J Biotechnol* 2002, **96(1):**13-21.
143. Thomas JG, Ayling A, Baneyx F: **Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from E. coli. To fold or to refold.** *Appl Biochem Biotechnol* 1997, **66(3):**197-238.
144. de Marco A, de Marco V: **Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning.** *J Biotechnol* 2004, **109(1–2):**45-52.
145. de Marco A, Deuerling E, Mogk A, Tomoyasu T, Bukau B: **Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli.** *BMC Biotechnol* 2007, **7(1):**32.
146. de Marco A: **Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in Escherichia coli.** *Nat Protoc* 2007, **2(10):**2632-2639.
147. Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T: **Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in Escherichia coli.** *Appl Environ Microbiol* 1998, **64(5):**1694-1699.
148. Garcia-Fruitos E, Gonzalez-Montalban N, Morell M, Vera A, Ferraz RM, Aris A, Ventura S, Villaverde A: **Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins.** *MicrobCell Fact* 2005, **4:**27.
149. Garcia-Fruitos E, Carrio MM, Aris A, Villaverde A: **Folding of a misfolding-prone beta-galactosidase in absence of DnaK.** *Biotechnol Bioeng* 2005, **90(7):**869-875.
150. Gonzalez-Montalban N, Garcia-Fruitos E, Ventura S, Aris A, Villaverde A: **The chaperone DnaK controls the fractioning of functional protein between soluble and insoluble cell fractions in inclusion body-forming cells.** *MicrobCell Fact* 2006, **5:**26.
151. Martinez-Alonso M, Vera A, Villaverde A: **Role of the chaperone DnaK in protein solubility and conformational quality in inclusion body-forming Escherichia coli cells.** *FEMS Microbiol Lett* 2007, **273(2):**187-195.
152. Garcia-Fruitos E, Martinez-Alonso M, Gonzalez-Montalban N, Valli M, Mattanovich D, Villaverde A: **Divergent Genetic Control of Protein Solubility and Conformational Quality in Escherichia coli.** *J Mol Biol* 2007, **374:**195-205.
153. Carrio MM, Villaverde A: **Role of molecular chaperones in inclusion body formation.** *FEBS Lett* 2003, **537(1–3):**215-221.
154. Vera A, Aris A, Carrio M, Gonzalez-Montalban N, Villaverde A: **Lon and ClpP proteases participate in the physiological disintegration of bacterial inclusion bodies.** *J Biotechnol* 2005, **119(2):**163-171.
155. Mogk A, Schlieker C, Friedrich KL, Schonfeld HJ, Vierling E, Bukau B: **Refolding of substrates bound to small Hsps relies on a disaggregation reaction mediated most efficiently by ClpB/DnaK.** *J Biol Chem* 2003, **278(33):**31033-31042.
156. Mogk A, Deuerling E, Vorderwulbecke S, Vierling E, Bukau B: **Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation.** *Mol Microbiol* 2003, **50(2):**585-595.
157. Mogk A, Bukau B: **Molecular chaperones: structure of a protein disaggregase.** *Curr Biol* 2004, **14(2):**R78-R80.
158. Gonzalez-Montalban N, Garcia-Fruitos E, Villaverde A: **Recombinant protein solubility-does more mean better?** *Nat Biotechnol* 2007, **25(7):**718-720.
159. Nakamoto H, Bardwell JC: **Catalysis of disulfide bond formation and isomerization in the Escherichia coli periplasm.** *Biochim Biophys Acta* 2004, **1694:**111-119.
160. Hu X, O'Dwyer R, Wall JG: **Cloning, expression and characterisation of a single-chain Fv antibody fragment against domoic acid in Escherichia coli.** *J Biotechnol* 2005, **120:**38-45.
161. Maskos K, Huber-Wunderlich M, Glockshuber R: **DsbA and DsbC-catalyzed oxidative folding of proteins with complex disulfide bridge patterns in vitro and in vivo.** *J Mol Biol* 2003, **325:**495-513.
162. Qiu J, Swartz JR, Georgiou G: **Expression of active human tissue-type plasminogen activator in Escherichia coli.** *Appl Environ Microbiol* 1998, **64:**4891-4896.
163. Ostermeier M, De Sutter K, Georgiou G: **Eukaryotic protein disulfide isomerase complements Escherichia coli dsbA mutants and increases the yield of a heterologous secreted protein with disulfide bonds.** *J Biol Chem* 1996, **271:**10616-10622.
164. Zhan X, Schwaller M, Gilbert HF, Georgiou G: **Facilitating the formation of disulfide bonds in the Escherichia coli periplasm via coexpression of yeast protein disulfide isomerase.** *Biotechnol Prog* 1999, **15:**1033-1038.
165. Bessette PH, Aslund F, Beckwith J, Georgiou G: **Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm.** *Proc Natl Acad Sci USA* 1999, **96:**13703-13708.
166. Stewart EJ, Aslund F, Beckwith J: **Disulfide bond formation in the Escherichia coli cytoplasm: an in vivo role reversal for the thioredoxins.** *EMBO J* 1998, **17:**5543-5550.
167. Jurado P, Ritz D, Beckwith J, de LV, Fernandez LA: **Production of functional single-chain Fv antibodies in the cytoplasm of Escherichia coli.** *J Mol Biol* 2002, **320:**1-10.
168. Santala V, Lamminmaki U: **Production of a biotinylated single-chain antibody fragment in the cytoplasm of Escherichia coli.** *J Immunol Methods* 2004, **284:**165-175.
169. Heo MA, Kim SH, Kim SY, Kim YJ, Chung J, Oh MK, Lee SG: **Functional expression of single-chain variable fragment antibody against c-Met in the cytoplasm of Escherichia coli.** *Protein Expr Purif* 2006, **47:**203-209.
170. Levy R, Weiss R, Chen G, Iverson BL, Georgiou G: **Production of correctly folded Fab antibody fragment in the cytoplasm of Escherichia coli trxB gor mutants via the coexpression of molecular chaperones.** *Protein Expr Purif* 2001, **23(2):**338-347.
171. Fernandez LA: **Prokaryotic expression of antibodies and affibodies.** *Curr Opin Biotechnol* 2004, **15:**364-373.
172. Terpe K: **Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems.** *Appl Microbiol Biotechnol* 2006, **72(2):**211-222.
173. van Wely K, Swaving J, Freudl R, Driessen A: **Translocation of proteins across the cell envelope of Gram-positive bacteria.** *FEMS Microbiol Rev* 2001, **25(4):**437-454.
174. Li W, Zhou X, Lu P: **Bottlenecks in the expression and secretion of heterologous proteins in Bacillus subtilis.** *Res Microbiol* 2004, **155(8):**605-610.
175. Bolhuis A, Tjalsma H, Smith HE, de Jong A, Meima R, Venema G, Bron S, van Dijl JM: **Evaluation of Bottlenecks in the Late Stages of Protein Secretion in Bacillus subtilis.** *Appl Environ Microbiol* 1999, **65(7):**2934-2941.
176. Yuan G, Wong S: **Isolation and characterization of Bacillus subtilis groE regulatory mutants: evidence for orf39 in the dnaK operon as a repressor gene in regulating the expression of both groE and dnaK.** *J Bacteriol* 1995, **177(22):**6462-6468.

177. Leskelä S, Wahlström E, Kontinen V, Sarvas M: **Lipid modification of prelipoproteins is dispensable for growth but essential for efficient protein secretion in Bacillus subtilis: characterization of the Lgt gene.** *Mol Microbiol* 1999, **31(4):**1075-1085.
178. Kontinen V, Sarvas M: **The PrsA lipoprotein is essential for protein secretion in Bacillus subtilis and sets a limit for high-level secretion.** *Mol Microbiol* 1993, **8(4):**727-737.
179. Hyyryläinen H, Bolhuis A, Darmon E, Muukkonen L, Koski P, Vitikainen M, Sarvas M, Prágai Z, Bron S, van Dijl J, *et al.*: **A novel two-component regulatory system in Bacillus subtilis for the survival of severe secretion stress.** *Mol Microbiol* 2001, **41(5):**1159-1172.
180. Westers H, Westers L, Darmon E, van Dijl J, Quax W, Zanen G: **The CssRS two-component regulatory system controls a general secretion stress response in Bacillus subtilis.** *FEBS J* 2006, **273(16):**3816-3827.
181. Duilio A, Tutino M, Marino G: **Recombinant protein production in Antarctic Gram-negative bacteria.** *Methods Mol Biol* 2004, **267:**225-237.
182. Parrilli E, Duilio A, Tutino ML: **Heterologous protein expression in psychrophilic hosts.** In *Psychrophiles: from Biodiversity to Biotechnology* Edited by: RMea. Berlin Heidelberg: Springer-Verlag; 2008:365-379.
183. Médigue C, Krin E, Pascal G, Barbe V, Bernsel A, Bertin P, Cheung F, Cruveiller S, D'Amico S, Duilio A, *et al.*: **Coping with cold: the genome of the versatile marine Antarctica bacterium Pseudoalteromonas haloplanktis TAC125.** *Genome Res* 2005, **15(10):**1325-1335.
184. Kiefhaber T, Rudolph R, Kohler H, Buchner J: **Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation.** *Biotechnology (NY)* 1991, **9(9):**825-829.
185. de Pascale D, Cusano AM, Autore F, dP Parrilli E, Marino G, Tutino ML: **The cold-active Lip1 lipase from the Antarctic bacterium Pseudoalteromonas haloplanktis TAC125 is a member of a new bacterial lipolytic enzyme family.** *Extremophiles* 2008.

MINIREVIEW

Yeast systems biotechnology for the production of heterologous proteins

Alexandra Graf[1,2], Martin Dragosits[1], Brigitte Gasser[1] & Diethard Mattanovich[1,2]

[1]Department of Biotechnology, Institute of Applied Microbiology, University of Natural Resources and Applied Life Sciences, Vienna, Austria; and [2]School of Bioengineering, University of Applied Sciences FH-Campus, Vienna, Austria

Correspondence: Diethard Mattanovich, Department of Biotechnology, Institute of Applied Microbiology, University of Natural Resources and Applied Life Sciences, Muthgasse 18, A-1190 Vienna, Austria.
Tel.: +43 1 360 06 6569;
fax: +43 1 369 7615;
e-mail: diethard.mattanovich@boku.ac.at

Received 9 December 2008; revised 23 February 2009; accepted 26 February 2009.
First published online 1 April 2009.

DOI:10.1111/j.1567-1364.2009.00507.x

Editor: Teun Boekhout

Keywords
systems biotechnology; heterologous protein; yeast; *Pichia pastoris*; systems biology; metabolic engineering.

Abstract

Systems biotechnology has been established as a highly potent tool for bioprocess development in recent years. The applicability to complex metabolic processes such as protein synthesis and secretion, however, is still in its infancy. While yeasts are frequently applied for heterologous protein production, more progress in this field has been achieved for bacterial and mammalian cell culture systems than for yeasts. A critical comparison between different protein production systems, as provided in this review, can aid in assessing the potentials and pitfalls of applying systems biotechnology concepts to heterologous protein producing yeasts. Apart from modelling, the methodological basis of systems biology strongly relies on postgenomic methods. However, this methodology is rapidly moving so that more global data with much higher sensitivity will be achieved in near future. The development of next generation sequencing technology enables an unexpected revival of genomic approaches, providing new potential for evolutionary engineering and inverse metabolic engineering.

Introduction

The unexpectedly fast progress in genome sequencing over the last decade has provided an invaluable source of information on the physiology of microorganisms, including a comprehensive overview on cellular endowment with metabolic enzymes. Simultaneously, metabolic modelling has been developed and applied to the mathematical description of the central metabolic processes of bacteria (Edwards & Palsson, 2000) and yeast (Förster *et al.*, 2003). Together with extensive work on genomic data to address ideally all metabolic processes of a cell, these metabolic models led to the concept of systems biology (Westerhoff & Palsson, 2004). Several systems biology models of *Saccharomyces cerevisiae* have previously been described and recently unified to one comprehensive model (Herrgård *et al.*, 2008, and references therein). To acquire data for dynamic modelling, postgenomic analyses at the transcriptomics, proteomics and metabolomics level are implemented.

These models offer the opportunity to predict cellular processes and are therefore regarded as highly valuable resources for strain optimization. Since 1990, the concepts of metabolic engineering have been developed and applied as the knowledge-based improvement of cell factories using genetic engineering (Bailey, 1991; Nielsen, 2001). Extending the concepts of metabolic engineering to a broad system basis has led to the conception of systems biotechnology (Lee *et al.*, 2005; Nielsen & Jewett, 2008). While systems biology aims ideally at the global understanding and modelling of the entire cellular network of reactions, systems biotechnology will rather accept gaps in the description of cellular processes, as long as the processes related to product formation can be mapped. The systems biotechnology approach can be seen as an iterative, cyclic process, integrating high throughput data generation with metabolic modelling and production strain optimization (Fig. 1).

The concepts of metabolic engineering were initially mostly applied to the production of metabolites. In 2000, heterologous proteins were introduced as a new class of products to be addressed by metabolic engineering (Ostergaard *et al.*, 2000). It is obvious that the complexity of the protein production and secretion process (Fig. 2) renders it

much more challenging to be addressed by tools of rational and quantitative analysis, as summarized in Table 1.

In cases where the genetic traits controlling complex cellular responses are not known, researchers have applied random mutagenesis and selection schemes to engineer metabolic pathways by modifying enzymes, transporters or regulatory proteins. This method has been termed evolutionary engineering (Sonderegger & Sauer, 2003), but its application to protein production is not straight forward as protein overexpression is usually not advantageous for the cell. Single cell sorting of large cell populations may be applied to overcome this limitation (Mattanovich & Borth, 2006). However, as the changes are not directed, it is often difficult to determine the genetic modification that is responsible for the improvement (Nevoigt, 2008). Understanding the biological system as a whole greatly facilitates the rational understanding of such mutants. Genome-wide analysis methods also made another biotechnological approach, namely inverse metabolic engineering (Bailey *et al.*, 1996), much more feasible, where phenotypic differences serve as the basis for elucidating genetic modifications needed to optimize production strains (Bro & Nielsen, 2004).

Yeasts are attractive hosts for production of heterologous proteins (Porro *et al.*, 2005). However, a number of bottlenecks and stress factors limit the full potential of this class of organisms (Mattanovich *et al.*, 2004), and systems biotechnology will offer new opportunities for modelling, analysis and optimization of protein production systems. In the

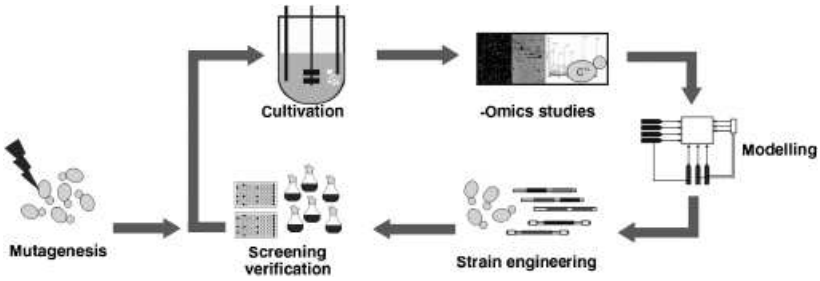


Fig. 1. The systems biotechnology circle. A primary production strain is cultivated under relevant conditions. Omics methods feed models, which aid the design of strain engineering and screening of new, improved strains. Random mutagenesis serves to increase variability, which can also be achieved by evolutionary engineering.

Table 1. Challenges for systems biotechnology research for heterologous protein production

Key elements for systems biotechnology	Challenges faced in protein expression and secretion
Metabolites and enzymes	Molecular players only partly defined
Metabolic pathways	Pathways less clear than those for metabolic processes
Stoichiometry of metabolic reactions	Stoichiometry difficult to define
Metabolic fluxes	Fluxes and concentrations of participating 'metabolites' difficult or not yet measurable

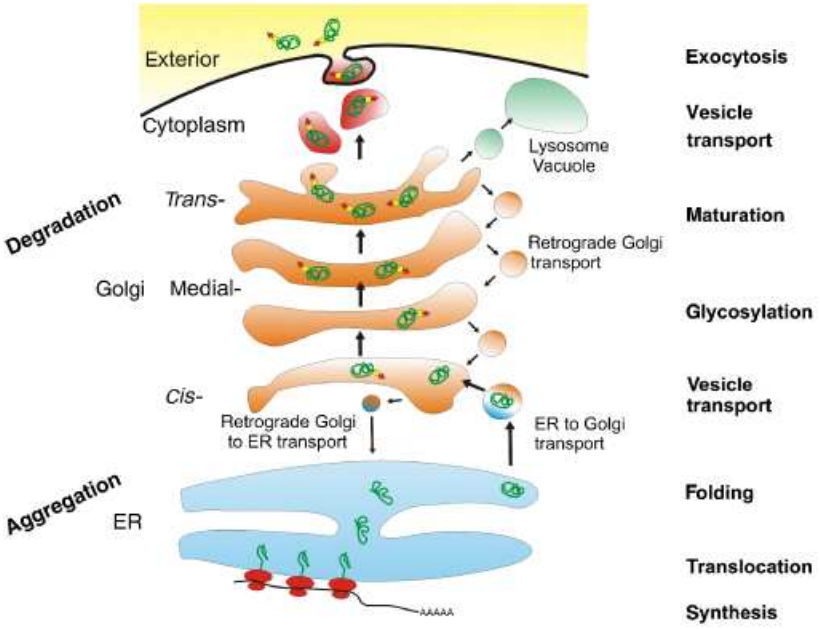


Fig. 2. Protein synthesis and secretion in eukaryotic cells. Pathways and key 'metabolic' processes of secreted proteins are indicated, including unproductive steps such as degradation or aggregation. These pathways should serve as a basis for the development of a model of the protein production and secretion process.

following we will describe the applications of systems biotechnology to yeasts with a strong emphasis on heterologous protein production, highlighting work on other classes of host organisms also, and provide an outlook as to where the development and integration of new methodology can lead this field.

Impact of systems biology on yeast biotechnological processes

Systems biology is not a purely academic research area as the quantitative description of microbial production cell lines is also already of interest for biotechnological industry. As pointed out before, yeasts producing heterologous proteins have rarely been investigated on a systems level so far, while the importance of systems biotechnology for industrial applications others than protein production has been well documented in the recent years (Pizarro *et al*., 2007; Takors *et al*., 2007; Mukhopadhyay *et al*., 2008). These applications span mainly the production of primary metabolites such as amino acids, alcohols or organic acids, often employing bacteria as production hosts. Global analysis of the host cell metabolism can help to improve the production of metabolites, which may consequently require further cell engineering to resist high concentrations of possibly toxic chemical compounds. Alper *et al*. (2006) recently showed how engineering of the global transcription machinery can help to improve ethanol resistance in yeast cells.

Additionally, systems biotechnology already impacts on the production of yeast-based alcoholic beverages as a system-wide understanding of the molecular basis of the production process can lead to improved sensory qualities for consumer demands. Proteomic and transcriptomic methods have been applied to investigate wine and beer fermentations (Kobi *et al*., 2004; Beltran *et al*., 2006; Zuzuarregui *et al*., 2006). It becomes clear from the system-wide analysis of wine fermentations that microbial cells encounter many different kinds of stresses during the fermentation process. However, it is obvious that a deeper understanding of the cellular reactions to environmental stresses is also crucial for other biotechnologically relevant batch and fed-batch processes such as amino acid, biofuel and, of course, protein production. The yeast stress response has been a topic of detailed investigations in the recent years. Both, transcriptomic and proteomic approaches have been used to investigate the effect of temperature (Gasch *et al*., 2000; Tai *et al*., 2007), high osmolarity (Blomberg, 1995; Chen *et al*., 2003; Gori *et al*., 2007; Kim *et al*., 2007), hydrostatic pressure (Fernandes *et al*., 2004) and nutrient limitations (Kolkman *et al*., 2006) in several yeast and fungal species in recent years. It has been outlined earlier that cellular reactions to environmental stresses are mainly a transient response, on which most studies have focused. However, during industrial processes, rather constant suboptimal growth conditions that are far from the natural environment of the cells are imposed (Mattanovich *et al*., 2004). For example, several reports indicate a positive effect of reduced growth temperature on the production rate of heterologous proteins in the yeast *Pichia pastoris* (Li *et al*., 2001; Jahic *et al*., 2003; Shi *et al*., 2003). However, these data are not fully conclusive as the authors suggest that lower activity of proteases in the culture supernatant or decreased cell death rate is responsible for increases in productivity. On the other hand, Hohenblum *et al*. (2003) showed that, at least for *P. pastoris*, significant cell death occurs at low pH while temperature does not influence viability. None of these studies applied system-wide analysis of the host organism, so that the underlying biology remained unexplored. Newer studies may shed light on such contradictory data. Recently, Tai *et al*. (2007) performed steady-state cultivations of *S. cerevisiae* at different temperatures and analysed them with microarrays. Although not using heterologous protein-producing strains, these experiments allow conclusions on the long-term adaptation of production cells to suboptimal conditions. The authors observed an upregulation of ribosome biogenesis genes and a downregulation of environmental stress response genes at a low temperature, which differed largely from previous results on rapid changes of temperature. Similarly, Gasser *et al*. (2007a) found stress response genes downregulated at lower temperature in steady-state cultures of *P. pastoris* expressing an antibody Fab fragment, while the specific productivity of this protein was increased.

However, transcriptome, proteome and metabolome data from small-scale laboratory experiments might differ significantly from cellular regulatory patterns as they occur during large-scale industrial processes. Furthermore there might be crucial genotypic differences between laboratory and industrial strains. As the majority of the mentioned studies were performed in laboratory strains, the direct applicability of these results on industrial strains is questionable. Production strain optimization can be achieved if the cellular metabolism and the regulatory networks of an industrial cell line are considered and investigated (Takors *et al*., 2007). Up to now, such results are rarely obtained in academia, as the genomic information of important industrial strains is still missing. Omics approaches as well as fast modern sequencing techniques bear the potential to change the methodical approaches here, as will be discussed below.

The external stresses mentioned above, and intrinsic stress mediated by protein overproduction, play a major role for the physiological constraints of a protein production system (Mattanovich *et al*., 2004). As these constraints share similar patterns among different classes of host organisms, we will also highlight research with non-yeast hosts for protein production in the following chapter.

Application of systems biotechnology to heterologous protein production

Engineering of recombinant yeasts based on genome-wide analysis

Applications of genome-wide technologies in yeasts are scarce in the field of recombinant protein production. Some of the rare examples analysing cellular responses due to protein overproduction are reported for the nonconventional yeasts *P. pastoris* (Sauer *et al.*, 2004; Gasser *et al.*, 2007a; Dragosits *et al.*, 2009) and *Kluyveromyces lactis* (van Ooyen *et al.*, 2006). The analysis of the cellular proteome during a fermentation of a chymosin expressing *K. lactis* strain indicated stress during protein production (upregulation of Hsp26p and Sod2p; van Ooyen *et al.*, 2006), whereas the *P. pastoris* work outlines how environmental factors such as temperature and pH affect protein expression and secretion on a transcriptomic (Sauer *et al.*, 2004; Gasser *et al.*, 2007a) and proteomic level (Dragosits *et al.*, 2009). Alternatively, metabolic flux analyses of protein-producing yeasts were performed. These focused, on the one hand, on the synthesis of high levels of intracellular human superoxide dismutase in *S. cerevisiae* (Gonzalez *et al.*, 2003), and, on the other, on core metabolic processes of *P. pastoris* during growth on glycerol and methanol (Solà *et al.*, 2004, 2007). However, apart from one exception, no strain improvement strategies resulted out of all these studies so far (Table 2).

The comparison of the differential transcriptome of a *P. pastoris* strain overexpressing human trypsinogen vs. a nonexpressing strain did reveal a network of genes being influenced due to the exploitation of the cellular expression machinery. This knowledge was further exploited to elucidate novel secretion helper factors that allowed the removal of bottlenecks in protein expression. Thirteen out of the 524 significantly regulated genes were selected and overexpressed in a *P. pastoris* strain producing a human antibody Fab fragment. Five previously characterized secretion helpers (Pdi1, Ero1, Sso2, Kar2/BiP and Hac1), as well as six novel, hitherto unidentified, factors, more precisely Bfr2 and Bmh2 involved in protein transport, the chaperones Ssa4 and Sse1, the vacuolar ATPase subunit Cup5 and Kin2, a protein kinase connected to exocytosis, increasing both specific production rates as well as volumetric productivity of an antibody fragment up to 2.5-fold in fed batch fermentations (Gasser *et al.*, 2007b). Very recently, a similar approach was leading to improved membrane protein production in *S. cerevisiae*, based on engineered expression of *BMS1*, involved in ribosomal subunit assembly (Bonander *et al.*, 2009). A convincing example of evolutionary engineering was based on random mutagenesis and screening for overproduction of human serum albumin in *S. cerevisiae*, followed by the identification of four genes related to Kar2 ATPase activity, which were upregulated in the selected mutant strain. Overexpression of these genes in other *S. cerevisiae* strains led to increased production of three different heterologous proteins (Payne *et al.*, 2008).

Protein folding and secretion appear to be major limitations for yeast expression systems, while the main concerns for other systems are growth, viability and metabolic burden (see Mammalian cells and Bacteria). For yeast production hosts these problems are not as crucial for the production of recombinant proteins, which can be regarded as one reason why systems biotechnology-based strain engineering has hardly been applied so far in this area.

Another important aspect that explains the lack of omics-based cell engineering in yeasts is the degree of availability of omics tools for yeast and other fungal species. Out of the 82 presently sequenced ascomycetes genomes, only 15% are biotechnologically used organisms, whereas the majority (54%) were pathogens, and the remaining were sequenced for comparative genomic studies (*Saccharomyces sensu stricto* group). The lack of published genome sequences is reflected in a lack of commercially available microarrays for most yeasts species. Some exceptions are arrays available for *S. cerevisiae* and *Schizosaccharomyces pombe* (Affymetrix and Agilent), or *Candida albicans* (Washington University, St. Louis). Proteomic studies are also hampered as they rely on annotated genome sequences for efficient performance. Alternatively, research groups performed transcriptional profiling by either heterologous hybridization to commercial *S. cerevisiae* arrays (e.g. for *P. pastoris*, Sauer *et al.*, 2004; and for *K. lactis*, Rosende *et al.*, 2008), or by designing custom microarrays. As an example, *P. pastoris* microarrays have been developed by our group, and are available for research applications (Graf *et al.*, 2008). While the first approach can only capture genes that are in common with *S. cerevisiae* [therefore excluding species-specific genes such as the assimilation pathways for methanol (*P. pastoris* and *Hansenula polymorpha*), hydrocarbons (*Yarrowia lipolytica*), or xylose (*Pichia stipitis*)], the latter often made the custom-made arrays unavailable for other groups, thereby limiting research activities in the field. Consistently, proteomics were mainly performed for pathogenic species (reviewed by Josic & Kovac, 2008).

Another drawback in the fungal kingdom is the high genetic diversity between the individual species, even within the phylum *Ascomycota*. The average sequence identity of orthologous proteins among the hemiascomycota is 50–60%, which is less than the *c.* 70% identity between man and fish, not to speak of 94% identity between man and mouse (Dujon, 2006). Accordingly, the DNA sequence identity among rodents is much higher, making heterologous omics between the hamster-derived Chinese hamster ovarian cells (CHO) or baby hamster kidney (BHK) cells and mouse or rat more feasible than among yeasts.

Table 2. Overview of systems biotechnological approaches for improved recombinant protein production in different hosts

Host	Analyses	Engineering	Results	References
E. coli	Transcriptome, proteome	Ribosomal genes, amino acid biosynthesis genes, IbpAB	Up to fourfold higher productivity	Reviewed in Park *et al*. (2005)
E. coli	Proteome	Controlled coexpression of PspA		Aldor *et al*. (2005)
E. coli	Proteome, transcriptome	Use of rare codons	Eight times more recombinant protein	Lee & Lee (2005)
E. coli	Proteome	New promoter (aldA)	30-fold higher product levels	Han *et al*. (2008)
E. coli	Secreted proteome	OsmY as fusion partner	High-level secretion	Qian *et al*. (2008)
B. megaterium	Metabolic fluxes	Pyruvate as carbon source	17-fold more secreted product, less protease activity	Fürch *et al*. (2007)
S. cerevisiae	Random mutagenesis	Co-chaperone genes related to Kar2 activity	1.5-fold increase of expression	Payne *et al*. (2008)
P. pastoris	Transcriptome	overexpression of secretion factors	2.5-fold increase of secretion	Gasser *et al*. (2007b)
A. niger	'Genomic methods' genome sequence and transcriptome	Disruption of protease genes	1.4-fold increased secretion	Wang *et al*. (2008)
A. niger	Transcriptome and proteome 'integrative genomics'	Knock-out of ERAD factor doaA and overexpression of oligosaccharyltransferase sstC	Improved intracellular production	Jacobs *et al*. (2009)
A. oryzae	Transcriptome	Knock-down of protease genes	1.2-fold increased secretion	Kimura *et al*. (2008)
CHO	Transcriptome	Overexpression of antiapoptotic and knock-down of proapoptotic genes	Higher viability leading to 2.5-fold higher titers	Wong *et al*. (2006)
CHO	Transcriptome	Stress markers for early clone screening	Time for clone establishment? Better and earlier clone selection	Trummer *et al*. (2008)
NS0	Transcriptome, proteome	Genes related to cholesterol dependence	Cholesterol-independent cell lines	Seth *et al*. (2006)

Filamentous fungi

A recent publication summarized 'The first 50 microarray studies in filamentous fungi', starting with an incomplete microarray for *Trichoderma reesei* in 2002 (Breakspear & Momany, 2007) and stated that for filamentous fungi, so far no engineering based on omics existed. However, in 2008 the first reports of proteome and transcriptome studies resulting in clear engineering strategies were published (Table 2): 132 protease genes were monitored during the production of human lysozyme in *Aspergillus oryzae* on a microarray and compared with degradation conditions, and upregulated protease cluster were identified. Out of these disruption targets three genes were already known to improve heterologous protein production, but the knock-down of one novel protease improved secreted yields of human lysozyme by 22% (Kimura *et al*., 2008). Wang *et al*. (2008) identified four protease genes of *Aspergillus niger* with genomic methods, which, upon disruption, led to increased protein secretion up to 40%. Another study determined the effect of enzyme overproduction in three different strains of *A. niger*, and extracted two engineering targets out of the upregulated genes and proteins involved in protein folding and the endoplasmic reticulum (ER)-associated protein degradation (ERAD) pathway. Combined engineering by knock-out of the ERAD factor doaA and overexpression of the oligosaccharyltransferase sttC led to improved production of a heterologous protein (Jacobs *et al*., 2009).

However, it should be noted that a number of genome-wide studies have revealed important aspects concerning protein production and its limitation in fungal expression systems. While not being applied directly, they have contributed significantly to the understanding of the protein production process, and led to strain engineering later on. The transcriptomic changes upon protein overexpression have been described for *Aspergillus nidulans* (Sims *et al*., 2005), *T. reesei* (Arvas *et al*., 2006) and *A. niger* (Guillemette *et al*., 2007), highlighting the impact of unfolded protein response (UPR) on protein folding, glycosylation, vesicle transport and ERAD, and identifying significant differences of UPR regulation between *S. cerevisiae* and filamentous fungi.

If we look beyond fungi, examples for systems biotechnological approaches in the field of recombinant protein production become more prevalent. Since the late 1990s, transcriptomics and proteomics were applied to bacterial and mammalian cultures used for heterologous protein production in order to elucidate cell physiology. Although numerous studies exist that describe the physiological

behaviour of cells to certain stresses – data also available for *S. cerevisiae* – activity beyond pure description is concentrated to a limited number of research groups. Recent reviews by these groups highlight that it is crucial to use state-of-the-art omics tools for physiological understanding and gaining insights into the host, as only detailed understanding of host cell physiology makes subsequent metabolic or cell engineering possible.

Mammalian cells

In mammalian cell culture most proteomic and transcriptomic analyses were performed to address problems or phenomena that have been observed previously on a 'macroscopic' level (e.g. metabolic shift, fed batch cultivation, apoptosis and stress conditions brought up by elevated osmolarity, sodium butyrate and low temperature). Since Korke *et al.* (2002) predicted the implementation of genomics and proteomics in cell culture engineering – for example for the selection of production cell lines (identification of gene regulation leading to adaptation to serum-free growth, or to adaptation to suspension growth) – and for bioprocess engineering both methodology and answers have evolved. Several excellent reviews summarizing these recent advances have been published during the last years (Griffin *et al.*, 2007; Gupta & Lee, 2007; Kuystermans *et al.*, 2007; Jaluria *et al.*, 2008). A list of papers dealing with genome-wide analysis of different mammalian cell lines (mainly murine myeloma cells NS0 and CHO) can be found in Kuystermans *et al.* (2007). In agreement with Gupta & Lee (2007), who pinpoint that a large number of omics approaches only generate lists of genes without direct application, we confirm that only four out of the 21 cited studies have resulted in an actual strategy for cell line engineering (Table 2).

Dinnis & James (2005) asked if one should learn 'lessons from nature' for engineering of antibody secreting mammalian cells. At least two lessons have been learned: induction of the UPR in order to reflect B-cell development (van Anken *et al.*, 2003) leads to increased secretion of several recombinant proteins (reviewed by Dinnis & James, 2005; Khan & Schroder, 2008). Another lesson that has been learned was the overexpression of anti-apoptotic genes, and knock-down of pro-apoptotic genes identified by microarray analysis of CHO cells, leading to prolonged cell viability and consequently up to 2.5-fold higher titres of interferon γ (Wong *et al.*, 2006).

Engineering mammalian cell culture based on genome-scale technologies was mainly applied to improve cell metabolism and growth (Griffin *et al.*, 2007), upstream cell culture conditions (e.g. temperature, hyperosmotic pressure and impact of small chemical compounds), downstream product quality (mainly assessed by proteomics), and cell culture media requirements (Gupta & Lee, 2007). As an example, the transcriptional analysis and proteomics did not stop at the identification of the responsible genes for the cholesterol dependence of NS0 cells. Subsequent engineering of the identified genes allowed cholesterol-independent cell growth (Seth *et al.*, 2006). Cross-species microarrays of high producer clones of EPO-Fc producing CHO lead to the identification of three ER stress marker genes correlated to insufficient resistance to shear stress in the early stage of clone selection before the respective phenotype could be observed (Trummer *et al.*, 2008).

Proteomic and transcriptional profiling of high- and low-productivity cell lines, or cells cultivated under conditions that lead to high specific productivity (q_P) (e.g. treatment with sodium butyrate and low-temperature cultivation) were carried out to discover the target genes leading to the super-secreting cells. Common features correlated to high productivity were the upregulation of secretory pathway proteins (especially chaperones and foldases) and cytoskeletal proteins in high-producing cell lines, as well as higher abundance of proteins belonging to the functional groups redox balance and vesicular transport. Additionally, decreased growth rate-related genes/proteins and decreased levels of stress genes were reported to occur in concordance with higher q_P (Kuystermans *et al.*, 2007; Seth *et al.*, 2007). While there were speculations that high-producing cell lines are likely to have a higher vesicle traffic and membrane recycling activity (Yee *et al.*, 2008), other attempts to identify correlations between single genes and improved secretory capacity failed. When trying to integrate all available genome-scale information of high-producing mammalian cell lines to find the genetic events leading to the super-secreting cells, it had to be concluded that there is no direct relation between a distinct set of genes and a trait, that there are no 'hyperproductivity master genes' (Seth *et al.*, 2007). On the contrary, multiple contributing pathways, even alternative pathways may lead to improved q_P. Therefore the authors highlight the importance of data analysis approaches going beyond the identification of differentially expressed genes such as pattern discovery, pathway and network analysis in order to grasp the complexity of the gene–trait relationship (Seth *et al.*, 2007).

Bacteria

A general overview about whole systems level metabolic engineering in bacteria based on omics, including potential applications, was given by Park *et al.* (2005) and Gupta & Lee (2007). One common feature in bacteria, for example *Escherichia coli*, is that protein overproduction usually leads to a (severe) decrease in specific growth rate due to a shortage of energy and precursors.

Very early proteomic work in *E. coli* and *Bacillus subtilis* overproducing a heterologous protein accumulating in the cytoplasm as inclusion bodies showed that both species react to the recombinant protein with increased levels of heat

shock proteins and chaperones, whereas no clear picture regarding the regulation of ribosomal proteins emerged, as higher abundance in *B. subtilis*, and decreased levels in *E. coli* were observed (Jürgen *et al*., 2000, 2001). These studies can be seen as initiating a comprehensive understanding of the cellular responses to protein overproduction in bacteria, and although no direct engineering benefits were achieved, they gave rise to improvement of production strains (reviewed by Chou, 2007).

Since then various studies investigated the response of bacteria to several different proteins, but hardly any new hypotheses or applications evolved out of these studies. Some rare exceptions to this include the overexpression of ribosomal genes downregulated during insulin-like growth factor 1 expression in high cell density cultivation of *E. coli* leading to enhanced productivity and the engineering of small heat shock proteins IbpAB identified in inclusion bodies during overexpression of recombinant proteins in *E. coli* (all summarized in Park *et al*., 2005). By proteome profiling Aldor *et al*. (2005) identified the phage shock protein PspA to be coregulated with heterologous protein expression, and improved the yield by controlled coexpression of PspA. However, most studies are conducted on a case-by-case basis, and are not leading to improved production platforms. Alternatively, genome-wide analysis of cellular reactions to protein production may allow for the identification of marker genes that signal cellular stress as a response to protein overexpression. Their monitoring during protein production processes should allow to react on the bioprocess level before the stressful conditions, and consequently reduced cell growth and reduced viability will occur (Dürrschmid *et al*., 2008; Nemecek *et al*., 2008).

Interestingly, genome-wide analyses were also performed to analyse the behaviour of mutant strains with superior production characteristics. As an example, the proteome of a *Bacillus megaterium* chemical mutant exhibiting higher production levels of recombinant intracellular dextranesucrase and better cultivation behaviour, showed higher abundance of proteins related to protein synthesis and protein translocation (Wang *et al*., 2006). The observation of reduced levels of tRNA synthetases both on the proteomic and the mRNA level of an *E. coli* mutant secreting four times more α-haemolysin (HylA) in comparison with its parental strain led to an alternative metabolic engineering strategy, namely to use rare codons to slow down translation, which improved HylA secretion eight times in the parental strain (Lee & Lee, 2005).

In a different application of omics technologies, proteome analysis of *E. coli* in response to oleic acid was used to select oleic acid-inducible promoters for recombinant protein production. The use of the aldehyde dehydrogenase *aldA* promoter increased green fluorescent protein fluorescence intensity 30-fold compared with the IPTG-inducible tac promoter while applying the cheaper inducer, oleic acid (Han *et al*., 2008). On the other hand, a screen of the extracellular proteome of *E. coli* identified naturally secreted proteins as fusion partner for recombinant proteins in order to stimulate secretion. Out of 12 tested low-molecular-weight fusion partners, OsmY proved to be the best secretion partner resulting in high-level excretion of three model proteins into the culture supernatant of *E. coli* (Qian *et al*., 2008).

Alternatively, metabolic flux calculations can be carried out with the aim to identify bottlenecks in the fermentation that may need to be eliminated by genetic engineering. A recent study investigated the influence of two different carbon sources (glucose and pyruvate) on metabolic fluxes and productivity in *B. megaterium*, and concluded that pyruvate improves recombinant protein production 17-fold as less protease secretion and enhanced energy and reduction equivalent metabolism occurred. Additionally, the authors state that the overproduction of the recombinant protein increases the flux through the TCA and glycolysis, and reduces the flux through gluconeogenesis and the pentose phosphate pathway (Fürch *et al*., 2007).

New (post) genomic approaches to systems biotechnology

Systems biology as well as application-oriented systems biotechnology depend essentially on omics methods. A critical overview on current developments in this area and the potentials and pitfalls of current and upcoming methods is provided in the following, and summarized in Table 3.

Genomes

As already discussed in the previous section systems biotechnology has focused on certain organisms simply because their genome was sequenced and at least partly annotated. This is understandable because many omics methods can only be utilized to their full extent if the genome sequence is accessible and information about the positions of functional elements in the DNA is available. Furthermore, though the genome of *S. cerevisiae* has been very well studied, it is not a typical example for many yeast species that are used for protein production (Blank *et al*., 2005). Therefore, it is vital for systems biotechnology to create reference genomes with high-quality annotation of yeast species used in protein production. Sequencing technologies have made tremendous progress in the last few years, rendering the technology significantly cheaper, faster and more flexible than the traditional Sanger method, making it feasible for small scale studies with limited resources. With a reference genome at hand resequencing becomes an integral part of the workflow in systems biotechnology as shown in Fig. 3, thus expanding the systems biotechnology cycle. First studies applying the technology to selected mutants to understand the genetic

Table 3. Critical summary of omics methods for systems biotechnology

Field	Methods	Advantages and disadvantages
Genomics	NGS	Fast and cheap method for whole-genome (re)sequencing without cloning bias
		Advantageous for mutation and subsequent strain analysis for inverse metabolic engineering purposes
Transcriptomics	Expression microarrays	Whole genome transcriptomics, cheap solution for in-house pipeline
		Susceptible to noise and bias
	Ref-Seq	Better correlation to qPCR results, large dynamic range limited only by sequencing depth
		Little background noise
		Can be used to detect splicing variants and 5′- and 3′-UTR boundaries
		Quantification is feasible even for mRNAs expressed at low levels
		Loss of strand-specific information
Proteomics	DIGE (gel-based systems)	Large number of different proteins over a large mass range can be detected
		Information about physicochemical properties
		Expensive and biased towards high-abundant proteins
		Membrane-bound and hydrophobic as well as small proteins cause problems
	MS	Mass and structure information of proteins
		Amino acid composition
		Detection of post-translational modifications
	Quantitative MS	Labelling (*in vivo* – SILAC, *in vitro* – ICAT, iTRAQ): increases the dynamic range of the analysis, but more expensive and detected proteins depend on the labelling method; *in vivo* labelling is not suitable for industrial processes
		Label-free quantification: quantification of a large number of proteins to characterize cells in different states; but less accurate and problematic to identify low-abundant proteins
	Protein Microarrays	Can only detect selected proteins due to lack of highly specific capture reagents and a lack in sensitivity
		Difficulties in retaining protein functionality
Interactomics	ChIP-chip	Regulatory DNA–protein binding interactions
		Chromatin packaging
	ChIP-Seq	Better resolution, less input material needed than ChIP-chip
		Usable for organisms without available genomic sequence
		Quantification is possible
Metabolomics	Metabolic modelling	Understanding regulatory pathways
		Identifying key players
		Simulation of system- wide reactions (either through logical networks or flux analysis) before biotechnological engineering is possible
		Creation of metabolic/signaling networks is complex and time consuming

background of their improved phenotype are already available for ethanol-producing *P. stipitis* (Smith *et al.*, 2008).

Next generation sequencing (NGS)

Though NGS has a high potential of revolutionizing genetics, it comes with a set of pitfalls. All NGS methods create much shorter sequence reads (35–400 bp) than the Sanger method (*c.* 750–900 bp). This is especially a problem for *de novo* sequencing because even short repeats will make an assembly impossible, resulting in a high number of contigs. A comprehensive summary of the NGS technologies that are currently available was published by Mardis (2008). Third-generation sequencing (TGS) or also called next-next generation sequencing methods aim to further reduce the cost and run time of sequencing while improving the ease of handling the method. Additionally, most of these new technologies promise to have much longer reads than the NGS methods and thereby eliminating the problems related to short read length. Variants of currently pursued TGS technologies could be commercially available within the next 5 years (Gupta, 2009).

After the first excitement of the 1990s about the possibilities of sequencing and the completion of the human genome project in 2003, it was believed that the postgenomic era had begun. Now, a few years later, the picture looks somewhat different. It seems that with the emergence of NGS, the chapter of genomics has to be reopened again. At the moment, there are 873 completely sequenced genomes, of which about 83% are bacterial species having a rather small genome, and 4135 ongoing genome projects with a much lower proportion (50%) of bacterial species (http://www.genomesonline.org/gold.cgi). The rapidly increasing amount of available genomic data poses a challenging problem for data storage, management and interpretation, shifting the bottleneck towards bioinformatics, annotation and analysis tools.

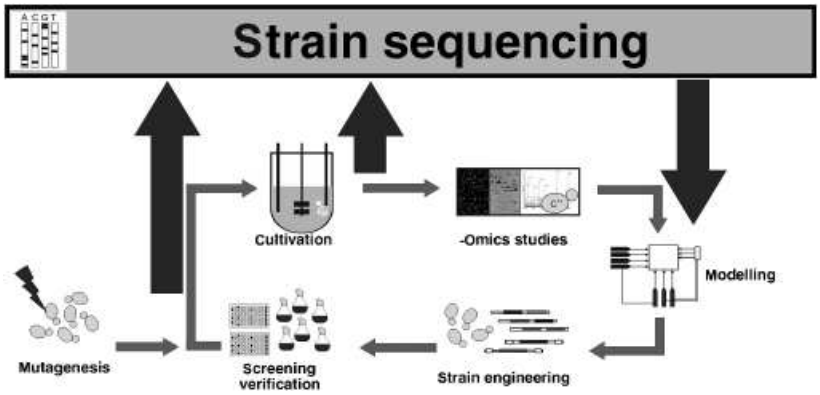


Fig. 3. A second genomics based systems biotechnology circle. NGS methods enable the resequencing of selected mutants to superimpose genotype on phenotype. Thus, evolutionary engineering and inverse metabolic engineering will gain enormous new potential.

Transcriptomics

Apart from *de novo* sequencing, NGS and TGS methods can be applied to many other questions that are relevant for the optimization of protein production, for example copy number variation, transcription factor binding, noncoding RNAs as well as expression profiles. Up to now DNA microarrays were the technologies of choice in the field of transcriptomics. For organisms for which microarrays are not available but that have an accessible genome sequence, bioinformatics tools (gene prediction and oligo design) make it possible to generate comprehensive expression data at a relatively low cost, as recently shown for *P. pastoris* (Graf *et al.*, 2008). Microarray experiments generally suffer from the existence of nonbiological variation, high background noise and only limited comparability due to a multitude of possible data-processing methods (Kawasaki, 2006). Sources for nonbiological variability and possible computational solutions have been extensively discussed in the literature (e.g. Draghici *et al.*, 2006; Johnson *et al.*, 2008). NGS solves many of the technical problems that microarrays suffer from but is still more expensive and many research groups and companies have an established microarray-processing pipeline in place. Also, some companies that offer NGS-sequencing machines offer microarrays as well, which indicates that a complementary use of both technologies is the most probable future.

Proteomics

Changes in the behaviour of a production system are often affected by post-translational modifications of proteins and therefore not visible on a transcriptome level. Besides there is no quantitative correlation between transcript expression levels and the amount of protein in the cell (Hartmann *et al.*, 2008). Analysing the genome and transcriptome alone is therefore not sufficient to understand or predict regulatory mechanism of cells or organisms. Unfortunately, though proteomics provides valuable insights for systems biotechnology, studies of changes during protein production are largely still missing (Josic & Kovac, 2008). The classical technique of proteome research, two-dimensional (2D) gel electrophoresis, suffers from a bias towards specific protein classes and towards highly abundant proteins (Bro & Nielsen, 2004) on the experimental level. On the data-processing level analysis of 2D gel images comprises many of the problems of microarray analysis, especially the difficulty to compare results due to differences between platforms and data-processing methods (Elrick *et al.*, 2006). In recent years, proteomics moved from a local (analysing a limited selection of proteins) to a global (analysing the whole proteome) technology. Quantitative mass spectroscopy has become the method of choice, and coupled with more sensitive labelling methods it facilitates high throughput proteome analysis (Elrick *et al.*, 2006). Stable isotope labelling by amino acids in cell culture (SILAC) is an *in vivo* labelling technique that requires the cells to be cultivated in media-containing labelled amino acids, thereby rendering it not feasible for large-scale production analyses. *In vitro* labelling techniques are more promising, although they still have shortcomings. In isotope-coded affinity tags (ICAT) cysteine residues are tagged. Because these residues are rare, it simplifies the peptide mixtures but proteins that do not contain cystein cannot be measured, and, furthermore, the small number of peptides for each protein compromises measurement reliability. Another *in vitro* labelling method is isobaric tags for relative and absolute quantification (iTRAQ). Here the N-terminus and sidechain amines are tagged with at least four different masses. Because these amines are more frequent, the cysteine-based restriction of ICAT is removed (Bachi & Bonaldi, 2008). Because all approaches involving isotopes are cost-intensive, label-free quantification methods using spectrum counts, integrated ion intensities or spectral feature analysis are sometimes preferable. The drawbacks of label-free methods include increased computational complexity due to lower accuracy and reproducibility of the data and the inability to quantify low-abundance proteins (Nesvizhskii *et al.*, 2007). With the move towards high throughput, it is essential to develop and

Table 4. Standards for the publication of omics data

Short name	Designation	References or web addresses
MIAME	Minimum information about microarray experiments	http://www.mged.org/Workgroups/MIAME/miame.html
MIAPE	The minimum information about a proteomics experiment	http://www.psidev.info/index.php?q=node/91
MIRIAM	Minimum information requested in the annotation of biochemical models	http://www.ebi.ac.uk/compneur-srv/miriam/, proposed by Le Novère *et al*. (2005)
MIAMET	Minimum information on a metabolomics experiment	Proposed by Bino *et al*. (2004)

use the proper bioinformatics tools to efficiently process and statistically validate the generated data.

Metabolomics and metabolic modelling

Metabolomics and fluxomics go a step further and use information gained from the other omics to build a model of certain processes of the cell or ideally of the whole cell. Whereas metabolomics focuses on the metabolites involved, fluxomics predicts flux distributions within the cell based on measured rates of metabolites and their mass balance (Kim *et al*., 2008). Several types of metabolic networks exist. Stoichiometric models and dynamic or kinetic models are the more traditional approach for which comprehensive knowledge about the players and their relationship is necessary. These models consist of a relatively small number of reactions or elements and their quality depends to a large extent on the quality of experimental data (Steuer, 2007). The advantage is of course that, if a high-quality model is achieved, reasonable predictions about phenotypic behaviour can be made. Early metabolic networks were limited to a few pathways of the core metabolism, but the availability of genome sequences and extensive omics data made it possible to fully describe the core metabolism and move to new areas such as signalling (Arga *et al*., 2007) or lipid metabolism (Nookaew *et al*., 2008). Protein expression and secretion networks are still lacking. Some of the reasons why this topic was avoided by scientists for so long are listed in Table 1. Despite these pitfalls, it can be anticipated that model development of the protein production pathways will significantly contribute to a better quantitative understanding of the contributing reactions and their relevance.

Topological networks on the other hand consist of many nodes that represent genes or proteins and edges representing the connection between those nodes. Such models can be much larger than stoichiometric or kinetic models and can make use of high throughput data but are only static descriptions and contain no information about the type, time or place of interaction between two nodes. At least the type of interaction can be modelled using control logic, described by Schlitt & Brazma (2005). With a sequenced and annotated genome at hand, a topological or control logic network can be computed for all known cellular functions. The challenge now is to combine the different approaches into a predictive dynamic model of the whole cell. The first such approach was taken for *A. niger* (Andersen *et al*., 2008), while Herrgård *et al*. (2008) took the first step towards such a model for *S. cerevisiae* by combining two existing metabolic models (at the control logic state) into a consensus yeast metabolic network and implementing it in systems biology markup language (SBML), which is a widely used data format for metabolic networks.

With the pace at which omics methods develop and the amount of data they produce, it is important to keep in mind that each of the omics fields only shows us an isolated part of the picture. To improve our understanding of the function of organisms, it is essential to be able to give meaning to the data representing concentrations of genes, proteins and metabolites. Therefore, it will become even more important in the future to merge information from different omics sources into a coherent picture. More effort has to be put in developing methods that can integrate and validate these data as well as help managing the fast-increasing flood of information. This in mind, several standards have been developed to avoid a confusing mess of data sources, treatment and analysis variants (Table 4). While these standards define a minimal frame on experimental and computational quality and data deposition, they cannot cope with systematic differences between different omics platforms. These systematic differences call for cautiousness in data comparison between omics platforms and data-processing methods. It should be emphasized that the aim of standardization is not to define data quality, but rather the information on data assessment, and to guarantee that raw data are made available to the scientific community.

Conclusions

Current status of systems biotechnology for protein production

Systems biotechnology has proven its value for strain design and optimization. System-wide approaches to complex cellular processes such as protein production are still in their infancy. Interestingly, more progress in this field has been made for bacteria and mammalian cells than for yeasts. This may be due to a lack of genomic information and postgenomic tools for industrially relevant yeast species. The

revival of genomics through NGS methods is about to close this gap, and a critical review of the state of research with bacterial and cell culture host systems provides guidance as to where to direct research with yeasts in this field.

Lessons to learn from non-yeast expression systems

At present, most limitations in protein production in yeasts are attributed to bottlenecks during folding and secretion. Thus, engineering of yeast protein factories mainly means knowledge-based engineering of chaperones and ER resident folding catalysts. While these approaches were verified by transcriptomic profiling of yeasts and other fungal production hosts recently, they may not convey the full picture. As can be seen in non-yeast expression systems, the application of integrated genomic approaches allows looking beyond the borders of the secretory compartments during the production of recombinant proteins. While engineering strategies leading to higher viability and higher stress resistance in mammalian cells may not be directly applied to yeast systems, they reveal the potential that apparently unconnected cellular processes can be manipulated in order to increase protein yields/productivity. Bacterial studies revealed a shortage of ribosomes, energy and precursors during recombinant protein production, a possibility yet underestimated for fungal hosts. The availability of improved metabolic models – as they exist for bacteria – for *A. niger* and *S. cerevisiae*, and their applicability for related fungal species, may allow a considerable progress regarding the behaviour of the core metabolism and energy supply during protein overexpression. Alternatively, production processes can be monitored by methods of metabolic flux analysis and controlled to improve protein yields.

The importance of systems level screening during clone selection in mammalian cells may also be converted to fungal production hosts, as soon as the respective genetic/molecular traits leading to high secretory capacity are identified. Once again, the importance of thorough and comparative data analysis should be stressed in this respect, as the pathways identified upon protein overproduction in yeasts and filamentous fungi are very similar to those regulated in mammalian cells.

With the advent of cheap and fast sequencing technologies, and consequently higher coverage of biotechnologically relevant fungal species, resequencing of improved mutant strains and subsequent inverse metabolic engineering also become feasible. Additionally, this information may contribute to the big search for the 'holy grail' of protein production – the 'hyperproductivity master genes' – or at least lead to a better understanding of the cellular pathways influencing productivity.

Finally, the impact of systems biotechnology on the improvement of bioprocess performance, for example by the identification of novel stress markers such as those shown in *E. coli*, better strain performance (mammalian cells) or prevention of proteolytic degradation by disruption of cellular proteases as reported for filamentous fungi and bacterial systems should be highlighted. All these approaches can be readily transferred to fungal production processes by applying the respective systems biotechnological tools.

In contrast to mammalian and bacterial protein production, application of systems level approaches for the targeted engineering of yeasts and other fungal production hosts is still at an early stage. However, expectations are high that the recent advances emerging in the fungal field will just be the beginning of the 'systems biotechnological' age for improved protein production strains.

Outlook

A major focus of future work should be the quantitative understanding of molecular principles behind protein synthesis, modification and secretion, derived from basic production strains as well as mutants and rationally engineered strains.

NGS methods provide the tool to rationalize inverse metabolic engineering approaches so that they can be implemented in future into rational system-wide modelling and optimization strategies.

Acknowledgements

Research on heterologous protein production in yeasts in our laboratory is supported by the Austrian Science Fund (project no. I37-B03), the European Science Foundation (programme EuroSCOPE), the Austrian Research Promotion Agency (programme FHplus), Polymun Scientific GmbH and Boehringer Ingelheim Austria GmbH.

References

Aldor I, Krawitz D, Forrest W, Chen C, Nishihara J, Joly J & Champion K (2005) Proteomic profiling of recombinant *Escherichia coli* in high-cell-density fermentations for improved production of an antibody fragment biopharmaceutical. *Appl Environ Microb* **71**: 1717–1728.

Alper H, Moxley J, Nevoigt E, Fink G & Stephanopoulos G (2006) Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314**: 1565–1568.

Andersen M, Nielsen M & Nielsen J (2008) Metabolic model integration of the bibliome, genome, metabolome and reactome of *Aspergillus niger*. *Mol Syst Biol* **4**: 178.

Arga K, Onsan Z, Kirdar B, Ulgen K & Nielsen J (2007) Understanding signaling in yeast: insights from network analysis. *Biotechnol Bioeng* **97**: 1246–1258.

Arvas M, Pakula T, Lanthaler K *et al.* (2006) Common features and interesting differences in transcriptional responses to secretion stress in the fungi *Trichoderma reesei* and *Saccharomyces cerevisiae*. *BMC Genomics* **7**: 32.

Bachi A & Bonaldi T (2008) Quantitative proteomics as a new piece of the systems biology puzzle. *J Proteomics* **71**: 357–367.

Bailey J (1991) Toward a science of metabolic engineering. *Science* **252**: 1668–1675.

Bailey J, Sburlati A, Hatzimanikatis V, Lee K, Renner W & Tsai P (1996) Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. *Biotechnol Bioeng* **52**: 109–121.

Beltran G, Novo M, Leberre V *et al.* (2006) Integration of transcriptomic and metabolic analyses for understanding the global responses of low-temperature winemaking fermentations. *FEMS Yeast Res* **6**: 1167–1183.

Bino R, Hall R, Fiehn O *et al.* (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci* **9**: 418–425.

Blank L, Lehmbeck F & Sauer U (2005) Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts. *FEMS Yeast Res* **5**: 545–558.

Blomberg A (1995) Global changes in protein synthesis during adaptation of the yeast *Saccharomyces cerevisiae* to 0.7 M NaCl. *J Bacteriol* **177**: 3563–3572.

Bonander N, Darby RAJ, Grgic L *et al.* (2009) Altering the ribosomal subunit ratio in yeast maximizes recombinant protein yield. *Microb Cell Fact* **8**: 10.

Breakspear A & Momany M (2007) The first fifty microarray studies in filamentous fungi. *Microbiology* **153**: 7–15.

Bro C & Nielsen J (2004) Impact of 'ome' analyses on inverse metabolic engineering. *Metab Eng* **6**: 204–211.

Chen D, Toone W, Mata J *et al.* (2003) Global transcriptional responses of fission yeast to environmental stress. *Mol Biol Cell* **14**: 214–229.

Chou C (2007) Engineering cell physiology to enhance recombinant protein production in *Escherichia coli*. *Appl Microbiol Biot* **76**: 521–532.

Dinnis D & James D (2005) Engineering mammalian cell factories for improved recombinant monoclonal antibody production: lessons from nature? *Biotechnol Bioeng* **91**: 180–189.

Draghici S, Khatri P, Eklund A & Szallasi Z (2006) Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet* **22**: 101–109.

Dragosits M, Stadlmann J, Albiol J *et al.* (2009) The effect of temperature on the proteome of recombinant *Pichia pastoris*. *J Proteome Res* **8**: 1380–1392.

Dujon B (2006) Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends Genet* **22**: 375–387.

Dürrschmid K, Reischer H, Schmidt-Heck W, Hrebicek T, Guthke R, Rizzi A & Bayer K (2008) Monitoring of transcriptome and proteome profiles to investigate the cellular response of E. coli towards recombinant protein expression under defined chemostat conditions. *J Biotechnol* **135**: 34–44.

Edwards J & Palsson B (2000) The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *P Natl Acad Sci USA* **97**: 5528–5533.

Elrick M, Walgren J, Mitchell M & Thompson D (2006) Proteomics: recent applications and new technologies. *Basic Clin Pharmacol* **98**: 432–441.

Fernandes P, Domitrovic T, Kao C & Kurtenbach E (2004) Genomic expression pattern in *Saccharomyces cerevisiae* cells in response to high hydrostatic pressure. *FEBS Lett* **556**: 153–160.

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

Fürch T, Wittmann C, Wang W, Franco-Lara E, Jahn D & Deckwer W (2007) Effect of different carbon sources on central metabolic fluxes and the recombinant production of a hydrolase from *Thermobifida fusca* in *Bacillus megaterium*. *J Biotechnol* **132**: 385–394.

Gasch A, Spellman P, Kao C *et al.* (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* **11**: 4241–4257.

Gasser B, Maurer M, Rautio J *et al.* (2007a) Monitoring of transcriptional regulation in *Pichia pastoris* under protein production conditions. *BMC Genomics* **8**: 179.

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

Gonzalez R, Andrews B, Molitor J & Asenjo J (2003) Metabolic analysis of the synthesis of high levels of intracellular human SOD in *Saccharomyces cerevisiae* rhSOD 2060 411 SGA122. *Biotechnol Bioeng* **82**: 152–169.

Gori K, Hébraud M, Chambon C, Mortensen H, Arneborg N & Jespersen L (2007) Proteomic changes in *Debaryomyces hansenii* upon exposure to NaCl stress. *FEMS Yeast Res* **7**: 293–303.

Graf A, Gasser B, Dragosits M *et al.* (2008) Novel insights into the unfolded protein response using *Pichia pastoris* specific DNA microarrays. *BMC Genomics* **9**: 390.

Griffin T, Seth G, Xie H, Bandhakavi S & Hu W (2007) Advancing mammalian cell culture engineering using genome-scale technologies. *Trends Biotechnol* **25**: 401–408.

Guillemette T, van Peij N, Goosen T *et al.* (2007) Genomic analysis of the secretion stress response in the enzyme-producing cell factory *Aspergillus niger*. *BMC Genomics* **8**: 158.

Gupta P (2009) Single-molecule DNA sequencing technologies for future genomics research. *Trends Biotechnol* **26**: 602–611.

Gupta P & Lee K (2007) Genomics and proteomics in process development: opportunities and challenges. *Trends Biotechnol* **25**: 324–330.

Han M, Lee J, Lee S & Yoo J (2008) Proteome-level responses of *Escherichia coli* to long-chain fatty acids and use of fatty acid

inducible promoter in protein production. *J Biomed Biotechnol* **2008**: 735101.

Hartmann M, Roeraade J, Stoll D, Templin M & Joos T (2008) Protein microarrays for diagnostic assays. *Anal Bioanal Chem* **393**: 1407–1416.

Herrgård M, Swainston N, Dobson P *et al.* (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol* **26**: 1155–1160.

Hohenblum H, Borth N & Mattanovich D (2003) Assessing viability and cell-associated product of recombinant protein producing *Pichia pastoris* with flow cytometry. *J Biotechnol* **102**: 281–290.

Jacobs DI, Olsthoorn MM, Maillet I *et al.* (2009) Effective lead selection for improved protein production in *Aspergillus niger* based on integrated genomics. *Fungal Genet Biol* **46**: 141–152.

Jahic M, Wallberg F, Bollok M, Garcia P & Enfors S (2003) Temperature limited fed-batch technique for control of proteolysis in *Pichia pastoris* bioreactor cultures. *Microb Cell Fact* **2**: 6.

Jaluria P, Chu C, Betenbaugh M & Shiloach J (2008) Cells by design: a mini-review of targeting cell engineering using DNA microarrays. *Mol Biotechnol* **39**: 105–111.

Johnson D, Li W, Gordon D *et al.* (2008) Systematic evaluation of variability in ChIP-chip experiments using predefined DNA targets. *Genome Res* **18**: 393–403.

Josic D & Kovac S (2008) Application of proteomics in biotechnology – microbial proteomics. *Biotechnol J* **3**: 496–509.

Jürgen B, Lin H, Riemschneider S *et al.* (2000) Monitoring of genes that respond to overproduction of an insoluble recombinant protein in *Escherichia coli* glucose-limited fed-batch fermentations. *Biotechnol Bioeng* **70**: 217–224.

Jürgen B, Hanschke R, Sarvas M, Hecker M & Schweder T (2001) Proteome and transcriptome based analysis of *Bacillus subtilis* cells overproducing an insoluble heterologous protein. *Appl Microbiol Biot* **55**: 326–332.

Kawasaki E (2006) The end of the microarray Tower of Babel: will universal standards lead the way? *J Biomol Tech* **17**: 200–206.

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

Kim T, Sohn S, Kim H & Lee S (2008) Strategies for systems-level metabolic engineering. *Biotechnol J* **3**: 612–623.

Kim Y, Nandakumar M & Marten M (2007) Proteome map of *Aspergillus nidulans* during osmoadaptation. *Fungal Genet Biol* **44**: 886–895.

Kimura S, Maruyama J, Takeuchi M & Kitamoto K (2008) Monitoring global gene expression of proteases and improvement of human lysozyme production in the nptB gene disruptant of *Aspergillus oryzae*. *Biosci Biotech Bioch* **72**: 499–505.

Kobi D, Zugmeyer S, Potier S & Jaquet-Gutfreund L (2004) Two-dimensional protein map of an "ale"-brewing yeast strain: proteome dynamics during fermentation. *FEMS Yeast Res* **5**: 213–230.

Kolkman A, Daran-Lapujade P, Fullaondo A, Olsthoorn M, Pronk J, Slijper M & Heck A (2006) Proteome analysis of yeast response to various nutrient limitations. *Mol Syst Biol* **2**: 2006. 0026.

Korke R, Rink A, Seow T, Chung M, Beattie C & Hu W (2002) Genomic and proteomic perspectives in cell culture engineering. *J Biotechnol* **94**: 73–92.

Kuystermans D, Krampe B, Swiderek H & Al-Rubeai M (2007) Using cell engineering and omic tools for the improvement of cell culture processes. *Cytotechnology* **53**: 3–22.

Lee P & Lee K (2005) Engineering HlyA hypersecretion in *Escherichia coli* based on proteomic and microarray analyses. *Biotechnol Bioeng* **89**: 195–205.

Lee S, Lee D & Kim T (2005) Systems biotechnology for strain improvement. *Trends Biotechnol* **23**: 349–358.

Le Novère N, Finney A, Hucka M *et al.* (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotechnol* **23**: 1509–1515.

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

Mardis E (2008) The impact of next-generation sequencing technology on genetics. *Trends Genet* **24**: 133–141.

Mattanovich D & Borth N (2006) Applications of cell sorting in biotechnology. *Microb Cell Fact* **5**: 12.

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

Mukhopadhyay A, Redding A, Rutherford B & Keasling J (2008) Importance of systems biology in engineering microbes for biofuel production. *Curr Opin Biotech* **19**: 228–234.

Nemecek S, Marisch K, Juric R & Bayer K (2008) Design of transcriptional fusions of stress sensitive promoters and GFP to monitor the overburden of *Escherichia coli* hosts during recombinant protein production. *Bioproc Biosyst Eng* **31**: 47–53.

Nesvizhskii A, Vitek O & Aebersold R (2007) Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Meth* **4**: 787–797.

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

Nielsen J (2001) Metabolic engineering. *Appl Microbiol Biot* **55**: 263–283.

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

Nookaew I, Jewett M, Meechai A *et al.* (2008) The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Syst Biol* **2**: 71.

Ostergaard S, Olsson L & Nielsen J (2000) Metabolic engineering of *Saccharomyces cerevisiae*. *Microbiol Mol Biol R* **64**: 34–50.

Park S, Lee S, Cho J, Kim T, Lee J, Park J & Han M (2005) Global physiological understanding and metabolic engineering of microorganisms based on omics studies. *Appl Microbiol Biot* **68**: 567–579.

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

Pizarro F, Vargas F & Agosin E (2007) A systems biology perspective of wine fermentations. *Yeast* **24**: 977–991.

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

Qian Z, Xia X, Choi J & Lee S (2008) Proteome-based identification of fusion partner for high-level extracellular production of recombinant proteins in *Escherichia coli*. *Biotechnol Bioeng* **101**: 587–601.

Rosende SS, Becerra M, Salgado MT, Lamas-Maceiras M, González M & Picos MAF (2008) Growth phase-dependent expression of *Kluyveromyces lactis* genes and involvement of 3′-UTR elements. *Process Biochem* **43**: 1153–1157.

Sauer M, Branduardi P, Gasser B, Valli M, Maurer M, Porro D & Mattanovich D (2004) Differential gene expression in recombinant *Pichia pastoris* analysed by heterologous DNA microarray hybridisation. *Microb Cell Fact* **3**: 17.

Schlitt T & Brazma A (2005) Modelling gene networks at different organisational levels. *FEBS Lett* **579**: 1859–1866.

Seth G, Ozturk M & Hu W (2006) Reverting cholesterol auxotrophy of NS0 cells by altering epigenetic gene silencing. *Biotechnol Bioeng* **93**: 820–827.

Seth G, Charaniya S, Wlaschin K & Hu W (2007) In pursuit of a super producer-alternative paths to high producing recombinant mammalian cells. *Curr Opin Biotechnol* **18**: 557–564.

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

Sims AH, Gent ME, Lanthaler K, Dunn-Coleman NS, Oliver SG & Robson GD (2005) Transcriptome analysis of recombinant protein secretion by *Aspergillus nidulans* and the unfolded-protein response *in vivo*. *Appl Environ Microb* **71**: 2737–2747.

Smith D, Quinlan A, Peckham H *et al.* (2008) Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Res* **18**: 1638–1642.

Solà A, Maaheimo H, Ylönen K, Ferrer P & Szyperski T (2004) Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*. *Eur J Biochem* **271**: 2462–2470.

Solà A, Jouhten P, Maaheimo H, Sánchez-Ferrando F, Szyperski T & Ferrer P (2007) Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates. *Microbiology* **153**: 281–290.

Sonderegger M & Sauer U (2003) Evolutionary engineering of *Saccharomyces cerevisiae* for anaerobic growth on xylose. *Appl Environ Microb* **69**: 1990–1998.

Steuer R (2007) Computational approaches to the topology, stability and dynamics of metabolic networks. *Phytochemistry* **68**: 2139–2151.

Tai S, Daran-Lapujade P, Walsh M, Pronk J & Daran J (2007) Acclimation of *Saccharomyces cerevisiae* to low temperature: a chemostat-based transcriptome analysis. *Mol Biol Cell* **18**: 5100–5112.

Takors R, Bathe B, Rieping M, Hans S, Kelle R & Huthmacher K (2007) Systems biology for industrial strains and fermentation processes – example: amino acids. *J Biotechnol* **129**: 181–190.

Trummer E, Ernst W, Hesse F *et al.* (2008) Transcriptional profiling of phenotypically different Epo-Fc expressing CHO clones by cross-species microarray analysis. *Biotechnol J* **3**: 924–937.

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

van Ooyen A, Dekker P, Huang M, Olsthoorn M, Jacobs D, Colussi P & Taron C (2006) Heterologous protein production in the yeast *Kluyveromyces lactis*. *FEMS Yeast Res* **6**: 381–392.

Wang W, Hollmann R & Deckwer W (2006) Comparative proteomic analysis of high cell density cultivations with two recombinant *Bacillus megaterium* strains for the production of a heterologous dextransucrase. *Proteome Sci* **4**: 19.

Wang Y, Xue W, Sims AH *et al.* (2008) Isolation of four pepsin-like protease genes from *Aspergillus niger* and analysis of the effect of disruptions on heterologous laccase expression. *Fungal Genet Biol* **45**: 17–27.

Westerhoff H & Palsson B (2004) The evolution of molecular biology into systems biology. *Nat Biotechnol* **22**: 1249–1252.

Wong D, Wong K, Nissom P, Heng C & Yap M (2006) Targeting early apoptotic genes in batch and fed-batch CHO cell cultures. *Biotechnol Bioeng* **95**: 350–361.

Yee J, de Leon Gatti M, Philp R, Yap M & Hu W (2008) Genomic and proteomic exploration of CHO and hybridoma cells under sodium butyrate treatment. *Biotechnol Bioeng* **99**: 1186–1204.

Zuzuarregui A, Monteoliva L, Gil C & del Olmo M (2006) Transcriptomic and proteomic approach for understanding the molecular basis of adaptation of *Saccharomyces cerevisiae* to wine fermentation. *Appl Environ Microb* **72**: 836–847.

6.2 Research publications

Dragosits M, Stadlmann J, Albiol J, Baumann K, Maurer M, Gasser B, Sauer M, Altmann F, Ferrer P, Mattanovich D. The effect of temperature on the proteome of recombinant *Pichia pastoris.*

Dragosits M, Stadlmann J, Graf A, Maurer M, Gasser B, Sauer M, Kreil DP, Altmann F, Mattanovich D. The unfolded protein response is involved in osmotolerance of *Pichia pastoris.* **submitted manuscript**

Dragosits M, Frascotti G, Bernard-Granger L, Vázquez F, Giuliani M, Baumann K, Rodríguez-Carmona E, Tokkanen J, Parrilli E, Wiebe MG, Kunert R, Maurer M, Gasser B, Sauer M, Branduardi P, Pakula T, Saloheimo M, Penttilä M, Ferrer P, Tutino ML, Villaverde A, Porro D, Mattanovich D. Comparative analysis of antibody fragment production in diverse microbial host cells**. in preparation**

Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, Kleinheinz M, Sauer M, Altmann F, Gasser B. Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris.*

Graf A, Gasser B, Dragosits M, Sauer M, Leparc GG, Tüchler T, Kreil DP, Mattanovich D. Novel insights into the unfolded protein response using *Pichia pastoris* specific DNA microarrays.

Baumann K, Maurer M, Dragosits M, Cos O, Ferrer P, Mattanovich D. Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins.

research articles Journal of proteome research

The Effect of Temperature on the Proteome of Recombinant *Pichia pastoris*

Martin Dragosits,[†] Johannes Stadlmann,[‡] Joan Albiol,[§] Kristin Baumann,[§] Michael Maurer,[†,||] Brigitte Gasser,[†] Michael Sauer,[†,||] Friedrich Altmann,[‡] Pau Ferrer,[§] and Diethard Mattanovich*[,†,||]

Department of Biotechnology, BOKU-University of Natural Resources and Applied Life Sciences, Vienna, Austria, Department of Chemistry, BOKU-University of Natural Resources and Applied Life Sciences, Vienna, Austria, Departament d'Engiyeria Química, Escola Tècnica Superior d'Enginyeria, Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès), Spain, and School of Bioengineering, University of Applied Sciences FH-Campus Wien, Vienna, Austria

Received September 10, 2008

The impact of environmental factors on the productivity of yeast cells is poorly investigated so far. Therefore, it is a major concern to improve the understanding of cellular physiology of microbial protein production hosts, including the methylotrophic yeast *Pichia pastoris.* Two-Dimensional Fluorescence Difference Gel electrophoresis and protein identification via mass spectrometry were applied to analyze the impact of cultivation temperature on the physiology of a heterologous protein secreting *P. pastoris* strain. Furthermore, specific productivity was monitored and fluxes through the central carbon metabolism were calculated. Chemostat culture conditions were applied to assess the adaption to different growth temperatures (20, 25, 30 °C) at steady-state conditions. Many important cellular processes, including the central carbon metabolism, stress response and protein folding are affected by changing the growth temperature. A 3-fold increased specific productivity at lower cultivation temperature for an antibody Fab fragment was accompanied by a reduced flux through the TCA-cycle, reduced levels of proteins involved in oxidative stress response and lower cellular levels of molecular chaperones. These data indicate that folding stress is generally decreased at lower cultivation temperatures, enabling more efficient heterologous protein secretion in *P. pastoris* host cells.

Keywords: *Pichia pastoris* • temperature • secretion • proteome • carbon flux

Introduction

Pichia pastoris is a well-established yeast host for the production of heterologous proteins, due the availability of strong inducible and constitutive promoter systems and efficient high cell density fermentation protocols.[1] There is evidence of a major bottleneck for heterologous protein secretion in protein folding and the cellular secretion machinery.[2−4] In a recent study, it was shown by Gasser and co-workers that overexpression of genes involved in protein folding and secretion can lead to improved expression of recombinant proteins.[4] Despite the frequent use of *P. pastoris* for protein production, physiological data allowing a deeper understanding of this bottleneck are rare. Previous studies report that decreased growth temperature can lead to improved protein production in batch, fed-batch and chemostat culture systems.[5,6] Concerning the positive effect of reduced growth temperature on specific productivity of heterologous proteins, the common opinion is that at least during cultivation on methanol a reduction of cell lysis and proteolytic activity is responsible for higher product yields at lower cultivation temperature, although no detailed studies have been performed so far. For other environmental stresses, such as the effect of osmolarity on protein production performance, there are few and ambiguous data available for various host organisms.[5−7] Taken together, the impact of environmental factors on the expression of heterologous proteins, especially in *P. pastoris*, is poorly characterized.[8]

Hitherto the immediate response of microorganisms to sudden changes of environmental conditions and not after adaption in steady-state conditions has been closely investigated. For example, the environmental stress response (ESR) to sudden changes in growth conditions of *Saccharomyces cerevisiae* and other yeasts is already well-studied,[9,10] but recent investigations indicate that this stress is alleviated after a short period of adaption in chemostat cultures. Indeed, ESR reactive genes can show opposed expression patterns when comparing the expression levels directly after the environmental shock and

* To whom correspondence should be addressed. University of Natural Resources and Applied Life Sciences, Department of Biotechology, Muthgasse 18, Vienna, Austria. Tel: (+43)1/36006-6569. Fax: (+43) 1 3697615. E-mail: diethard.mattanovich@boku.ac.at.

[†] Department of Biotechnology, BOKU-University of Natural Resources and Applied Life Sciences.

[‡] Department of Chemistry, BOKU-University of Natural Resources and Applied Life Sciences.

[§] Universitat Autònoma de Barcelona.

[||] University of Applied Sciences FH-Campus Wien.

10.1021/pr8007623 CCC: $40.75

Published on Web 02/13/2009

the expression levels after adaption to the new growth condition.[11] It is therefore essential to investigate the effect of environmental factors that are imposed mainly by fermentation conditions, such as temperature, osmolarity and oxygenation, on the physiology of recombinant host cells at steady-state growth.

In the present study, the effect of growth temperature on the proteome of recombinant *P. pastoris* expressing the Fab fragment of antibody 3H6[12] and of a nonexpressing control strain is described. As for other high-throughput approaches, growth rate-associated effects, which usually occur during shake flask cultivation, are critical and interfere with data analysis.[13,14] Therefore chemostat cultivation was the method of choice and cells were grown carbon-limited at a constant dilution rate to avoid any growth rate-related effects. Two-Dimensional Fluorescence Difference Gel Electrophoresis (2D-DIGE) was applied to monitor the changes of the intracellular proteome during growth at temperature setpoints that correspond to 100%, 80% and 60% of the maximum specific growth rate of *P. pastoris* in batch culture. Carbon fluxes through the central carbon metabolism were calculated based on a stoichiometric model and specific substrate and O_2 consumption rates, as well as specific biomass, metabolites and CO_2 production rates to verify our findings on the *P. pastoris* proteome.

Materials and Methods

Materials. All chemicals for yeast cultivations were molecular biology grade and purchased from Roth, Germany. CyDyes, 2D-Cleanup Kit, 2D-Quantkit and 24 cm IPG Drystrips 3−11 NL were purchased from GE Healthcare. All chemical reagents for two-dimensional gel electrophoresis were high purity grade and purchased from Sigma, unless stated otherwise.

Yeast Strains. Two strains were used in this study. The strain expressing the Fab fragment of the human anti-HIV antibody 3H6 was previously described by Baumann et al.[15] In short, both chains of the Fab fragment were expressed using the constitutive GAP-promoter and secreted via the *S. cerevisiae* α-mating factor leader secretion signal. For the nonexpressing control strain, *P. pastoris* X-33 was transformed with an empty pGAPzαA vector as described by Gasser et al.[2]

Chemostat Cultivation. For chemostat cultivations, a 3.5 L benchtop bioreactor (MBR, Switzerland) was used at a working volume of 1.5 L. A 1000 mL shake flask containing 150 mL of YPG medium (2% (w/v) peptone, 1% (w/v) yeast extract, 1% (w/v) glycerol) was inoculated with 1 mL of cryostock of the respective *P. pastoris* clones. The culture was grown for approximately 24 h at 28 °C and shaking at 170 rpm, and then used to inoculate the reactor to an optical density (OD_{600}) of 1.0. Batch medium contained per liter: 39.9 g of glycerol, 1.8 g of citric acid, 12.6 g of $(NH_4)_2HPO_4$, 0.022 g of $CaCl_2 \cdot 2H_2O$, 0.9 g of KCl, 0.5 g of $MgSO_4 \cdot 7H_2O$, 2 mL of Biotin (0.2 g L^{-1}), 4.6 mL of trace salts stock solution. The pH was set to 5.0 with 25% (w/w) HCl. Chemostat medium contained per liter: 50 g of glucose$\cdot 1H_2O$, 0.9 g of citric acid, 4.35 g of $(NH_4)2HPO_4$, 0.01 g of $CaCl_2 \cdot 2H_2O$, 1.7 g of KCl, 0.65 g of $MgSO_4 \cdot 7H_2O$, 1 mL of Biotin (0.2 g L^{-1}), and 1.6 mL of trace salts stock solution. The pH was set to 5.0 with 25% (w/w) HCl. Trace salts stock solution contained per liter: 6.0 g of $CuSO_4 \cdot 5H_2O$, 0.08 g of NaI, 3.0 g of $MnSO_4 \cdot H_2O$, 0.2 g of $Na_2MoO_4 \cdot 2H_2O$, 0.02 g of H_3BO_3, 0.5 g of $CoCl_2$, 20.0 g of $ZnCl_2$, 5.0 g of $FeSO_4 \cdot 7H_2O$, and 5.0 mL of H_2SO_4 (95−98% w/w).

After a batch phase of approximately 24 h, the continuous culture was started at a dilution rate of $D = 0.1$ h^{-1} (flow rate of 150 g h^{-1}). pH was controlled at 5.0 with 25% ammonium hydroxide (w/w). Gas flow rate was kept constant at 1.5 vvm (volume gas per volume medium and minute) and dissolved oxygen was kept at 20% by controlling the stirrer speed. Temperature for cultivations was maintained at 30 ± 0.6 °C, 25 ± 0.4 °C and 20 ± 1 °C, respectively. Off-gas was measured using a BCP O_2 and a BCP CO_2 gas sensor (BlueSens, Germany).

As there were no differences in biomass concentration and off-gas data between 3 and 5 residence times, it was assumed that steady-state conditions were reached after 5 residence times. Therefore, all samples were taken after cultivation for 5 residence times after a temperature shift. Three biological replicas were performed for each growth temperature, whereat three individual continuous cultures were inoculated for each strain. Although it is known that prolonged cultivation of yeasts in chemostat cultures can lead to evolutionary adaption,[16,17] we assumed that no adaption occurred during cultivation, because the maximum number of generations during chemostat cultivation did not exceed 21. Furthermore, the temperature regime was changed for each replicate to avoid any bias toward adaptive evolution (20/25/30 °C, 30/25/20 °C and 25/20/30 °C).

Samples for Biomass Determination, Supernatant, and 2D Gel Electrophoresis. For biomass determination, chemostat samples (2 × 10 mL) were collected by centrifugation. Cell pellets were washed in 10 mL of sterile H_2O and transferred to a weighed beaker. Biomass was dried at 105 °C until constant weight. For two- dimensional gel electrophoresis, aliquots of 2 mL of chemostat culture were centrifuged at 4 °C, supernatants were removed, and cell pellets were transferred to −80 °C immediately. Culture supernatants were stored separately at −20 °C until analysis.

Determination of Total RNA Content. Total RNA determination was performed as previously described by Benthin et al.[18] In short, cell pellets were washed with 0.7 M ice-cold $HClO_4$ and resuspended in 3 mL of KOH. Cells were incubated at 37 °C for 1 h. Afterward, samples were put on ice and 1 mL of 3 M $HClO_4$ was added. Samples were centrifuged to collect cell pellets and supernatants were transferred to a fresh Eppendorf tube. Additionally, cell pellets were washed twice with 0.5 M $HClO_4$ and pooled with the initial supernatants. RNA concentration was determined on a spectrophotometer by reading absorbance at 260 nm. Total RNA content (in %) is related to yeast dry mass.

Determination of Total Protein Content. Cell pellets from 2 mL of chemostat culture were washed twice with phosphate buffered saline pH 7.0 (PBS) and resuspended in 1 mL of PBS. Protein extraction occurred as previously described,[19] using NaOH and incubation at 95 °C to break up cells, and addition of 0.8 M HCl. Cellular debris were collected by centrifugation at 4 °C and supernatants were kept for determination of protein content using a BCA Protein Assay Kit (Pierce) applying bovine serum albumine as standard. Total protein content (in %) is related to yeast dry mass.

Determination of Proteolytic Activity in Culture Supernatants. For the detection of proteolytic activity in culture supernatants, the QuantiCleave Protease assay kit (Pierce) was used. The kit is based on succinylated casein as substrate and trinitrobenzene sulfonic acid (TNBSA), which forms a colored adduct with peptide fragments. Trypsin was used as standard protease.

Quantification of 3H6 Fab by ELISA. Quantification of the antibody was performed as previously described by Baumann and co-workers.[15]

Protein Solubilization, Cleanup and Labeling for Two-Dimensional Gel Electrophoresis. Cell pellets were washed with ice-cold 1× PBS, resuspended in cell lysis buffer (7 M urea, 2 M thiourea, 30 mM Tris, 4% (w/v) Chaps, pH 8.5) and combined with 500 μL of glass beads (0.5 mm, Sartorius). For cell disruption, samples were treated in a Thermo Savant Fastprep FP120 twice (speed 6.00 for 30 s; cooling interval of 30 s between treatments). Afterward, cellular debris were pelleted by centrifugation at 13 000 rpm at 4 °C for 5 min. Supernatants, containing total cellular protein, were transferred to a new tube and purified using the 2D-CleanUp Kit (GE Healthcare) and quantified via 2D-QuantKit (GE Healthcare) according to manufacturer's recommendations. For fluorescence labeling, 50 μg of protein extract from different cultivation temperatures or an internal standard (a combination of equal protein amounts of all samples for a 2D-DIGE run) was used and labeled with Cy2, Cy3 or Cy5, respectively, according to manufacturer's manual, whereas Cy2 was used for labeling of the internal standard protein pool (CyDye Minimal labeling kit, GE Healthcare).

Isoelectric Focusing. Prior to isoelectric focusing, 24 cm IPG DryStrips 3−11 NL (GE Healthcare) were equilibrated for 8−10 h at room temperature (7 M urea, 2 M thiourea, 2% (w/v) CHAPS, 4 mM DTT and 0.5% (v/v) IPG Buffer (pH 3−11). For isoelectric focusing, a total protein amount of 150 μg was loaded per IPG Drystrip (Cup loading). Focusing was performed on a IPGPhor isoelectric focusing unit (GE Healthcare) at the following conditions: 20 °C, 50 μA per strip: 6 h at 500 V, 3 h at 500−1000 V, 8 h at 1000−8000 V and 8000 V for a total of 58 kVh. After focusing, IPGDryStrips were either frozen at −80 °C or directly moved to second-dimension separation.

SDS-PAGE. IPG DryStrips were washed in re-equilibration solution I (150 mM Tris-HCl, pH 8.8, 6 M Urea, 30% (v/v) Glycerol, 2% (w/v) SDS and 0.5% (w/v) DTT) for 15 min. Afterward, drystrips were washed in re-equilibration solution II (150 mM Tris-HCl, pH 8.8, 6 M urea, 30% (v/v) glycerol, 2% (w/v) SDS and 4.5% (w/v) iodacetamide) for 15 min. Then, they were loaded onto a standard 12% polyacrylamide gel for SDS-PAGE at 12 W for 30 min and 100 W for 3−4 h.

Scanning, Image Analysis, and Data Handling. Gels were scanned on a Typhoon 9400 Imager (Amersham Biosciences) and images were analyzed using DeCyder Software v.5 (GE Healthcare). To identify differentially regulated proteins, a fold change of at least ≥1.5 and an 1-ANOVA p-value of ≤0.05 were considered statistically significant. DeCyder XML toolbox (GE Healthcare) was used for data export to Microsoft Excel. The R software package (R 2.6.0, http://www.r-project.org/) was used for Principal Component Analysis (PCA), and EPClust (Expression Profile Data CLUSTering and Analysis, http://www.bioinf.ebc.ee/EP/EP/EPCLUST/) was used for hierarchical clustering. For data clustering, the mean spot volume over all temperature setpoints for each protein spots was used as a reference.

Identification of Protein Spots by LC-ESI-QTOF Tandem MS. Protein spots, which passed the significance thresholds, were manually picked from CBB stained gels. The respective protein spots were processed as described previously.[20,21] Briefly, protein spots were excised, destained, carbamidomethylated, and digested with trypsin and the resulting tryptic peptides were extracted from the gel pieces and dried in a Speed-vac concentrator. Subsequently, the tryptic peptides were dissolved in 6 μL of 0.1% formic acid and subjected to capillary reversed-phase chromatography (BioBasic C18, 5 μm, 100 × 0.18 mm, Thermo). A linear gradient from 5% to 80% acetonitrile (containing 0.1% formic acid) was formed with an UltiMate 3000 Capillary LC-system (Dionex) with a flow rate of 1 μL/min. ESI MS/MS analysis was performed using a quadrupole time-of-flight (Q-TOF) Ultima Global (Waters Micromass) mass spectrometer. MS/MS acquisitions were performed over a range of 50−2000 m/z. Collision gas was argon at a collision voltage of 10−75V. *De novo* sequencing and BLAST searches against the *P. pastoris* sequence information available through Integrated Genomics[22] were performed, employing the Protein Lynx Global Server 2.1 Software (Waters Micromass), applying the following constraints: up to one missed cleavage per peptide, a fragment ion tolerance of 0.1 Da, carbamidomethylation of cysteine as fixed modification, oxidation of methionine as well as deamidation of asparagine as variable modifications. All matching spectra were reviewed manually. The measured parent and fragment ion masses were typically within 0.05 Da of the calculated values.

Identification of Pdi1 and Kar2. For the identification of Pdi1 and Kar2, a nonfluorescent 2D gel was prepared according to the described methods and proteins were semidry blotted onto a nitrocellulose membrane at 15 V for 30 min. Pdi1 and Kar2 were detected using an mouse anti-HDEL(2EF) antibody (Sigma) and goat anti-mouse AP conjugated antibody (Sigma). After detection of the signals, corresponding spots were picked from CBB stained 2D gels and the identity was verified by mass spectrometry as described in the previous section.

Quantification of Secreted Metabolites. Extracellular metabolites in culture supernatants (glucose, citrate, pyruvate, glycerol and ethanol) were measured by HPLC using a Aminex HPX-87H ion exchange column (Bio-Rad) and a LC Module I Plus (Waters). For data analysis, Millenium32 software (Waters) was used.

Metabolic Flux Calculations. For determination of metabolic fluxes, a simplified stoichiometric model of the central carbon metabolism was built upon a stoichiometric model of *S. cerevisiae*[23,24] and adapted to the present case. For the adaptation and simplification, the following main points were considered: aerobic glucose metabolization was performed mainly via glycolysis, pentose phosphate cycle and tricarboxylic acid cycle (TCA). All equations describing the draining of intermediate metabolites for biomass formation were added up to only one equation, thus, assuming *de facto* a nonsignificant change in biomass macromolecular composition during the performance of the tests. Compartmentation between mitochondria and cytoplasm was taken into account for intermediates of the TCA cycle including specification of cytoplasmic and mitochondrial pools for pyruvate, oxalacetate, α-ketoglutarate, NADPH and acetyl-CoA. According to previous studies,[25,26] supply of metabolic intermediates to the mitochondrial TCA cycle in aerobic glucose cultures was assumed to be mainly via oxaloacetate and pyruvate, while activity of the malic enzyme was assumed to be negligible according to previous studies[25,26] and preliminary microarray experiments of our laboratory (unpublished data). $NADH_2$ balance was lumped into one global pool. ATP balance was not considered. Potential differences in lipid composition, due to cultivation temperature, were not considered. To calculate the intracellular fluxes, mass balances around intracellular metabolites were considered, assuming a pseudo-steady state for the intracellular concentra-

Table 1. Physiological Properties of *P. pastoris* Grown at 20, 25 and 30 °C in Aerobic Chemostat Cultures[a]

strain	growth temperature (°C)	YDM (g L^{-1})	biomass yield (g_{dw} g_{gluc}^{-1})	whole cell protein (%)	RNA content (%)	residual glucose (mg L^{-1})	proteolytic activity ($\mu g_{trypsin}$ $mL_{supernatant}^{-1}$)	specific productivity ($mg_{3H6\ Fab}$ g_{YDM}^{-1} h^{-1})
X33 3H6 Fab	30	26.1 ± 0.4	0.52 ± 0.01	20 ± 0.7	4.4 ± 0.08	nd	0.17 ± 0.003	0.02 ± 0.001
X33 3H6 Fab	25	26.7 ± 0.3	0.53 ± 0.01	21 ± 1.2	4.6 ± 0.03	nd	0.19 ± 0.08	0.04 ± 0.002
X33 3H6 Fab	20	27.4 ± 0.4	0.55 ± 0.01	22 ± 0.0	4.8 ± 0.05	nd	0.17 ± 0.003	0.06 ± 0.002
X33 Control	30	26.5 ± 0.3	0.53 ± 0.01	21 ± 0.6	4.0 ± 0.31	nd	–	–
X33 Control	25	27.7 ± 0.2	0.56 ± 0.01	22 ± 0.7	4.4 ± 0.08	nd	–	–
X33 Control	20	27.3 ± 0.3	0.55 ± 0.01	19 ± 1.9	4.6 ± 0.09	nd	–	–

[a] 3H6 Fab producing strain (X33 3H6 Fab) and a control strain (X33 Control) grown in glucose-limited aerobic chemostat cultures at $D = 0.1$ h^{-1}. Values represent the mean ± standard error from three biological replicas. YDM, yeast dry mass; gluc, glucose; Nd, not detected; -, not determined.

tions of metabolites. The resulting metabolic model consisted in 39 metabolic reaction equations with 41 metabolites (33 internal and 8 external metabolites including biomass; for details see Supplemental Data 1 in Supporting Information). The derived stoichiometric matrix together with the measured input−output fluxes formed an overdetermined system of linear equations. Determination of the metabolic fluxes was performed solving the described linear system of equations as described previously in the literature.[27,28] Mathematical calculations were done using Matlab v7.4 (2007a, http://www.mathworks.com).

Northern Blot Analysis. Northern blot analysis was performed as previously described by Sauer and co-workers.[29] In short, total RNA was isolated using the Trireagent/chloroform extraction method. Total RNA was precipitated with isopropanol and washed twice with 70% ethanol. Afterward, RNA was resuspended in RNase free water and concentration was determined on a ND-1000 Nanodrop photometer (Thermo Scientific). Ten micrograms of total RNA was used for Northern blot analysis. RNA was separated on a formaldehyde containing agarose gel and transferred to a Nylon membrane (Nytran Supercharge, Schleicher & Schuell) by capillary transfer. Probes (3H6 heavy chain, 3H6 light chain, SSA4 and FUM1) were PCR amplified from plasmid DNA and genomic *P. pastoris* X-33 DNA, respectively, and DIG-labeled using a DIG labeling mix (Roche). Prehybridization and hybridization occurred at 42 °C in high SDS hybridization buffer. After appropriate washing steps, blots were stained using anti-Digoxigenin-alkaline phosphatase Fab fragments (Roche) and CDP Star reagent (New England Biolabs) according to manufacturer's manual. Chemiluminescent signals were visualized on a Lumi Imager F1 (Boehringer Mannheim) and data were analyzed using the LumiAnalyst software 3.0 (Boehringer Mannheim).

Student's *t* Test. In case that a Student's *t* test was performed, the cutoff *p*-value for significance was set to $p \leq 0.05$.

Results and Discussion

Physiological Properties and Specific Productivity at Different Growth Temperatures. A recombinant *P. pastoris* X-33 strain, expressing the 3H6 Fab fragment, and a control strain were grown in glucose limited aerobic chemostat cultures at 20, 25 and 30 °C in triplicate. These temperature set-points correspond to 60, 80 and 100% of the maximum specific growth rate (μ_{max}) as determined in shake flask cultures in synthetic batch medium (data not shown). As already reported previously,[6,30−33] there is a positive effect of reduced growth temperature on the secretion of heterologous proteins into the culture supernatant. A reduction of temperature from 30 to 25 °C resulted in a 2-fold increase of the specific productivity (q_P) of the 3H6 Fab fragment and a reduction from 30 to 20 °C yielded a 3-fold increase of q_P (Table 1).

Both, for the expressing strain and the control strain, a slight but significant increase in biomass yield was observed when growth temperature was reduced from 30 to 20 °C. The protein content did not significantly differ between cultivation at different temperature setpoints, neither for the expressing nor for the control strain. Also the total RNA content seems largely unaffected by growth temperature, except between 20 and 30 °C in the 3H6 Fab strain (Student's *t* test *p*-value = 0.011). Since the differences for the heterologous protein secreting strain were very small (an increase of 10% in total RNA amount at 20 °C compared to 30 °C cultivation), it is not clear whether they are of biological relevance. Furthermore, there was no detectable residual glucose present in culture supernatants throughout the different growth conditions, resulting in an almost constant glucose uptake rate of 1.54 ± 0.04 mmol g^{-1} h^{-1} (glucose per biomass per hour). No differences in the proteolytic activity could be seen in the culture supernatants of the 3H6 Fab secreting strain at the three cultivation temperatures. Because this proteolytic activity was temperature-independent and constant, it was not determined for the control strain.

2D-Gel Electrophoresis and Protein Identification. Two-dimensional gel electrophoresis yielded very reproducible spot patterns for the 3H6 Fab expressing strain and the control strain 2D-DIGE experiments. Approximately 1300 spots were present on all spot maps (Figure 1). Samples from all three temperature setpoints were compared with each other using DeCyder software v.5 (GE Healthcare). Spots exhibiting a fold change of ≥1.5 and 1-ANOVA *p*-value ≤0.05 in at least one comparison were chosen for identification via mass spectrometry. About 150 protein spots passed these criteria. The majority of differentially abundant spots could be identified when comparing 20 and 30 °C setpoints, as it was expected. The comparison of 25 and 30 °C yielded fewer proteins that were significantly different: 25 protein spots differed in the expressing strain and 76 protein spots in the control strain. The difference between expressing and control strain 2D-DIGE experiment was most probably due to differing gel image quality and the fact that many differentially abundant spots represented spots of low intensity. Most of these protein spots show also significant differences when comparing 20 and 30 °C temperature setpoints (see Supplemental Data 2 in Supporting Information for details).

Subsequently, out of these approximately 150 differentially abundant protein spots, 49 spots were identified by *de novo* sequencing and BLAST search (Table 2). It should be noted that more than these 49 spots were analyzed by mass spectrometry but failed to be identified based on several reasons. First, many spots showed low protein amount and therefore could not be identified due to the sensitivity limits of 2D-DIGE technology and the applied mass spectrometry equipment.

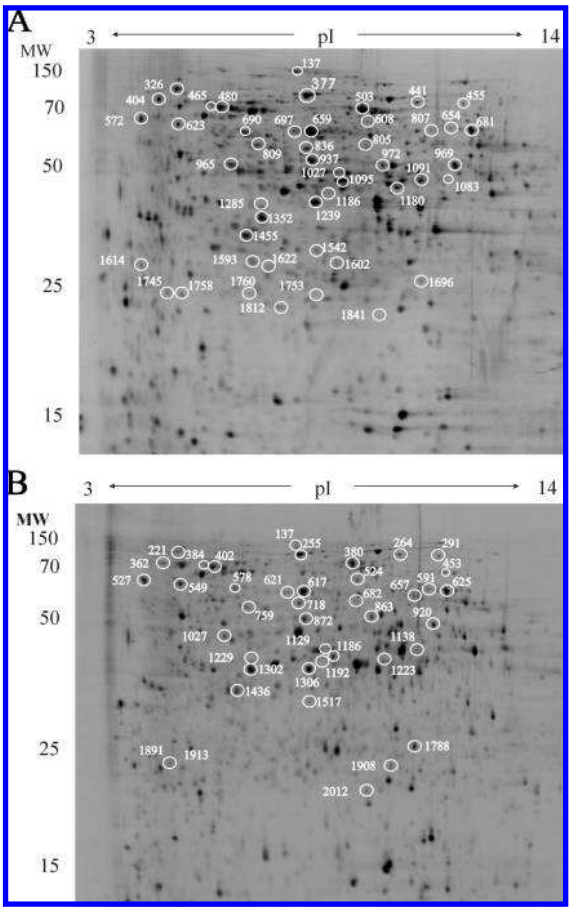


Figure 1. Two representative gels from the production strain (A) and the control strain (B) experiments. Spots that are highlighted in white circles with corresponding master numbers were identified via mass spectrometry. p*I*, isoelectric point; MW, molecular weight [kDa].

Second, some proteins did not pass the thresholds as described in the Material and Methods and no assured identification was possible. Furthermore, for some protein spots, it turned out that spot overlap occurred and peptides matching two or more different proteins were present in a single spot, which did not allow unambiguous identification.

Out of the 49 identified protein spots, 44 represented the predicted full-length size according to the position on the two-dimensional gels, whereas 5 protein spots represented fragments of full-length proteins (Ssa4 and Ssb1, Table 2). Additionally, in some cases, several spots with different isoelectric points (p*I*) but essentially identical molecular weights, matched the same particular protein (Ald4 and Cta1, Table 2). This is a common phenomenon in 2D-gel electrophoresis and has already been demonstrated in previous studies.[34,35] Protein fragments and p*I* shifts may be due to post-translational events such as protein phosphorylation and degradation. Table 3 shows the alterations in spot volume ratios and corresponding *p*-values for the comparison of the 20 and 30 °C cultivations. The comparison of 25 and 30 °C is not shown in the table as only four of the identified proteins are of statistical relevance and no conclusive interpretation was possible (for the production strain, Ssb1, Met6, Ssp120 and Sar1; for the control strain, Ssb1, Ssa4, Hsp60, Cta1; a complete list of spot volume ratios is available in Supplemental Data 2 in Supporting Information).

Overall Changes in Cellular Proteome at Different Temperatures. The greatest impact on cellular protein levels, as already mentioned, was observed when comparing 20 and 30 °C cultivations. Principal Component Analysis (PCA) also highlights that the major differences on the proteomic scale appeared at the comparison of 30/20 and 25/20 °C cultivations, whereas the 25 and 30 °C setpoints are close to each other. Additionally, it can be seen that, at least for the 49 identified protein spots, cultivation temperature constituted the major influence on cellular protein levels and not expression of the heterologous protein (Figure 2).

Hierarchical Cluster Analysis using EPClust resulted in an distinct separation of the 49 proteins (Figure 3). It shows, as well as PCA, that most of the identified protein spots behaved very similarly in the heterologous protein expressing and the control strain, resulting in a very homogeneous picture with two large clusters: (i) proteins with increased levels at lower cultivation temperature (20 °C) and (ii) proteins with decreased levels at lower temperature. Within the first cluster, at least two subclusters could be seen. The first subcluster comprises mainly protein fragments and proteins, which are involved in amino acid metabolism. The second subcluster mainly includes proteins involved in RNA and ribosome biogenesis, genome maintenance and a proteasomal component. For the second large cluster, the protein abundance ratios are even more similar in the production and the control strain, resulting in no significant subclusters.

Further discussion will focus on the results obtained from the comparison of 20 and 30 °C cultivations. Although the reduction of the growth temperature from 30 to 25 °C at steady-state conditions resulted in a 2-fold increase of q_P, the response of intracellular protein levels is much lower than between 20 and 30 °C. The fact that we cannot draw reasonable conclusions from the 25−30 °C comparison might highlight the limits of 2D-DIGE and subsequent mass spectrometry based protein identification. The unidentified less abundant proteins may bear additional and important information.

The Influence of Temperature on Particular Cellular Pathways. According to the present data (Table 3), there are at least five important cellular processes that were affected by cultivation temperature: Energy metabolism, oxidative stress response, protein folding, amino acid metabolism and RNA/ribosomal biogenesis. Whereas components involved in amino acid metabolism and RNA/ribosomal biogenesis showed higher levels at 20 °C, proteins involved in energy metabolism, stress response and protein folding showed decreased levels (Cluster 1 and 2 in Figure 2, respectively).

Although glucose uptake rates were quite constant, there was a great impact on proteins involved in glucose metabolism. At least five proteins, which are directly involved in oxidative decarboxylation or the tricarboxylic acid (TCA) cycle (Lpd1, Aco1, Cit1, Fum1 and Mdh1, respectively) showed lower levels at 20 °C as compared to 30 °C.

In contrast to proteins involved in TCA-cycle, the amount of glycerol kinase 1 (Gut1) increased at low temperature in both strains. Gut1 is responsible for the reversible formation of glycerol-3-phosphate from glycerol, which can be further metabolized in glycolysis and gluconeogenesis. Nevertheless, no glycerol was detected in culture supernatants (Supplemental Data 1 in Supporting Information). Since Gut1 expression is regulated by Ino2, Ino4, Adr1 and Opi1[36] in *S. cerevisiae*, gene

Table 2. Identification of Protein Spots by *de Novo* Sequencing and BLAST

3H6 Fab master no.	control master no.	homologous *S. cerevisiae* protein	standard name	M_r (kDa)/p*I*	peptides matched	sequence coverage (%)
377	255	Aconitate hydratase	Aco1	84.5/5.93	23	36.3
503	380	Acetyl CoA synthetase	Acs1	73.8/6.02	31	49.1
659	617	Aldehyde dehydrogenase #1	Ald4	42.1/8.45	19	52.7
697	621	Aldehyde dehydrogenase #2	Ald4	42.2/8.45	12	33.0
836	718	Aldehyde dehydrogenase #3	Ald6	54.3/5.71	21	53.7
441	264	Alcohol Oxidase	Aox1	73.8/6.41	12	24.1
969	920	Citrate synthase	Cit1	51.9/8.32	7	21.8
972	863	Fumarase	Fum1	52.6/6.79	12	32.4
809	759	Glycerol Kinase	Gut1	68.15/5.33	7	14.7
608	524	Isocitrate lyase	Icl1	61.5/6.15	10	24.6
1352	1302	Malate dehydrogenase	Mdh1	32.67/4.96	15	43.2
805	682	Dihydrolipoamide dehydrogenase	Lpd1	52.3/6.69	3	8
1455	1436	Cytochrome C peroxidase	Ccp1	41.9/7.01	9	23.1
654	591	Catalase −1	Cta1	57.8/6.98	6	11
681	625	Catalase −2	Cta1	57.8/6.98	10	21.8
–	1908	HSP31	Hsp31	51.9/6.05	6	12.7
623	549	HSP60/GrOEL	Hsp60	60.2/5.08	3	8.1
326	221	HSP82	Hsp82	80.9/4.87	14	25.6
404	362	ER ATPase BIP	Kar2	74.2/4.79	2	3.5
572	527	Protein disulfide isomerase	Pdi1	57.8/4.63	1	3.7
480	402	HSP SSA4	Ssa4	70.3/5.12	14	28.5
1027	1129	HSP SSA4 50 kDa fragment	Ssa4	70.3/5.12	8	15.1
1758	1913	HSP SSA4 25 kDa fragment	Ssa4	70.3/5.12	2	4.6
465	384	HSP SSB1	Ssb1	66.5/5.12	16	34.1
965	1027	HSP SSB1 50 kDa fragment	Ssb1	66.5/5.12	4	11.7
1745	1891	HSP SSB1 25 kDa fragment	Ssb1	66.5/5.12	8	16.3
1091	1138	Ala-Glyoxylate Aminotransferase	Agx1	30.9/6.73	10	48.8
1186	1192	Glutamate dehydrogenase #1	Gdh1	49.3/5.67	7	22.5
937	872	Glutamate dehydrogenase #2	Gdh1	49.3/5.67	8	25.8
1285	1229	Acetohydroxyacid reductoisomerase	Ilv5	44.3/8.18	9	29.1
807	657	Homocysteine methyltransferase #1	Met6	85.7/6.18	5	9.5
1542	1517	Homocysteine methyltransferase #2	Met6	85.7/6.18	5	9.1
1180	1223	S-adenosylmethionine syntetase	Sam2	42.3/6.43	11	34.1
1696	1788	GTP binding protein ran	Gsp1	24.1/6.98	7	25.8
1753	−	LSU Ribosomal Protein L10P	Rpp0	33.7/4.63	4	14.7
1622	−	LSU Ribosomal Protein L18P	Rpl8	34.1/8.59	6	18.9
1095	1186	Unknown, dehydrogenase	−	38.6/6.04	13	57.4
1239	1306	YPR127W	−	31.8/5.85	15	52.2
1841	2012	Adenosine Kinase	Ado1	17.2/6.10	4	36.9
–	453	Carnitine O-acetyltransferase	Cat2	70.3/7.19	5	10.5
137	207	EF3	Yef3	116.5/5.72	22	22.7
455	291	Long-chain-fatty-acid--CoA ligase	Faa2	25.4/6.73	3	17.2
1083	−	high mobility group-T protein	Hmo1	31.2/7.77	3	15.8
690	578	Myo-inositol-1-phosphate synthase	Ino1	58.4/5.26	3	7.2
1812	−	GTP Binding protein SAR1	Sar1	20.7/6.04	5	40.8
1602	−	20S Proteasome subunit	Scl1	27.6/6.05	7	29
1760	−	SNZ2	Snz2	31.9/5.40	3	16.2
1614	−	SSP120	Ssp120	25.6/4.79	2	12
1593	−	vacuolar ATPase subunit E	Vma4	33.1/8.79	12	33.3

products regulating the expression of phospholipid biosynthetic genes, this regulation might indicate changes in phospholipid metabolism. This assumption is supported by the fact that proteins involved in fatty acid metabolism, Faa2 and Cat2, are also significantly regulated. Faa2, a long chain fatty acyl-CoA synthase, and Cat2, which is involved in the transfer of acetyl-groups, showed lower protein levels at 20 °C.

Also other components of basic energy metabolism such as Acs1 (responsible for the formation of acetyl-CoA from acetate and the nuclear source of acetyl-CoA), cytosolic Ald4 and mitochondrial Ald6, which are responsible for interconversion of acetaldehyde and acetate, showed lower levels at 20 °C. Icl1, which catalyzes a key step in the glyoxylate cycle, the conversion of isocitrate to succinate and glyoxylate, was down-regulated on the protein level at 20 °C. The glyoxylate cycle is a bypass of the TCA cycle and serves for the utilization of C2 compounds. Interestingly, a basic level of alcohol oxidase 1 (Aox1), which is responsible for the utilization of methanol in *P. pastoris*, was also detected in this proteome study, although glucose was used as sole carbon source. This basic level of Aox1 was also affected by growth temperature and decreased during cultivation at 20 °C.

The expression of Aox1 and Icl1 is usually repressed by glucose.[37,38] Generally, chemostat cultivation results in extremely low levels of residual glucose and derepression of the *AOX1* and *ICL1* loci and the appearance of low levels of the corresponding proteins is therefore plausible. It has been shown in a chemostat-based *S. cerevisiae* study that a decrease

Table 3. Changes in Protein Levels (Ratios) and *p*-Values for the Identified Proteins of the 2D-DIGE Experiment by Comparing 20 and 30 °C Cultivations[a]

			20/30 °C 3H6 Fab		20/30 °C Control	
	homologous protein	standard name	average ratio	*p*-value	average ratio	*p*-value
Energy Metabolism	Aconitate hydratase	Aco1	−2.45	4.90×10^{-4}	−1.50	4.50×10^{-3}
	Acetyl CoA synthetase	Acs1	−2.02	9.40×10^{-4}	−1.70	4.50×10^{-6}
	Aldehyde dehydrogenase #1	Ald4	−1.92	6.90×10^{-4}	−1.51	3.00×10^{-5}
	Aldehyde dehydrogenase #2	Ald4	−1.95	5.70×10^{-5}	−1.46	3.40×10^{-2}
	Aldehyde dehydrogenase #3	Ald6	−1.96	4.50×10^{-4}	−1.48	3.30×10^{-4}
	Alcohol Oxidase	Aox1	−2.66	1.70×10^{-3}	−2.73	$5.90E \times 10^{-4}$
	Citrate synthase	Cit1	−1.66	2.10×10^{-2}	−3.05	1.10×10^{-4}
	Fumarase	Fum1	−2.21	1.60×10^{-4}	−1.42	$2.50E \times 10^{-2}$
	Glycerol Kinase	Gut1	2.04	5.20×10^{-3}	1.57	1.10×10^{-2}
	Isocitrate lyase	Icl1	−2.93	6.70×10^{-5}	−1.39	4.10×10^{-3}
	Malate dehydrogenase	Mdh1	−1.62	5.50×10^{-4}	−1.98	1.10×10^{-3}
	Dihydrolipoamide dehydrogenase	Lpd1	−2.3	8.80×10^{-3}	−2.05	5.00×10^{-5}
Oxidative Stress	Cytochrome C peroxidase	Ccp1	−1.81	1.70×10^{-4}	−2.41	8.20×10^{-6}
	Catalase −1	Cta1	−2.01	3.90×10^{-3}	−2.27	$3.00E \times 10^{-5}$
	Catalase −2	Cta1	−3.33	2.10×10^{-3}	−2.37	7.10×10^{-4}
Protein Folding	HSP31	Hsp31	−1.01	9.70×10^{-1}	−1.83	8.10×10^{-5}
	HSP60/GrOEL	Hsp60	−1.75	9.50×10^{-3}	−1.33	1.40×10^{-3}
	HSP82	Hsp82	−1.81	3.90×10^{-2}	−1.53	6.00×10^{-2}
	ER ATPase BIP	Kar2	−1.7	1.30×10^{-2}	−1.24	3.40×10^{-2}
	Protein disulfide isomerase	Pdi1	−1.17	1.80×10^{-1}	1.16	4.90×10^{-2}
	HSP SSA4	Ssa4	−4.22	1.20×10^{-3}	−4.63	2.70×10^{-6}
	HSP SSA4 50 kDa fragment	Ssa4	−1.93	2.80×10^{-3}	1.24	7.50×10^{-3}
	HSP SSA4 25 kDa fragment	Ssa4	3.09	6.60×10^{-4}	2.46	2.60×10^{-4}
	HSP SSB1	Ssb1	−3.26	1.30×10^{-3}	−3.32	2.40×10^{-4}
	HSP SSB1 50 kDa fragment	Ssb1	1.12	5.00×10^{-1}	2.66	4.90×10^{-3}
	HSP SSB1 25 kDa fragment	Ssb1	2.53	2.00×10^{-2}	2.56	9.90×10^{-5}
Amino acid Metabolism	Ala-Glyoxylate Aminotransferase	Agx1	1.89	4.50×10^{-3}	−1.93	$7.10E \times 10^{-4}$
	Glutamate dehydrogenase #1	Gdh1	2.06	8.90×10^{-4}	2.44	2.20×10^{-4}
	Glutamate dehydrogenase #2	Gdh1	1.44	1.70×10^{-2}	2.39	1.40×10^{-3}
	Acetohydroxyacid reductoisomerase	Ilv5	−1.87	7.90×10^{-3}	−1.26	2.60×10^{-1}
	Homocysteine methyltransferase #1	Met6	1.38	8.70×10^{-2}	2.18	2.50×10^{-6}
	Homocysteine methyltransferase #2	Met6	2.67	1.80×10^{-4}	2.2	1.90×10^{-7}
	S-adenosylmethionine syntetase	Sam2	1.83	4.50×10^{-4}	−1.05	4.40×10^{-1}
Ribosome/RNA	GTP binding protein ran	Gsp1	2.66	1.60 × 10−2	1.08	8.10×10^{-1}
Biogenesis	LSU Ribosomal Protein L10P	Rpp0	2.88	0.0016	−	−
	LSU Ribosomal Protein L18P	Rpl8	2.13	0.0053	−	−
Miscellaneous	Unknown, dehydrogenase	−	1.17	2.80×10^{-1}	−2.16	1.20×10^{-6}
	YPR127W	−	−1.93	4.00×10^{-3}	−3.3	1.50×10^{-6}
	Adenosine Kinase	Ado1	2.14	4.20×10^{-3}	*1.01*	*9.50×10^{-1}*
	Carnitine O-acetyltransferase	Cat2	−	−	−2.34	5.10×10^{-5}
	EF3	Yef3	−2.08	2.10×10^{-2}	−1.57	1.20×10^{-2}
	Long-chain-fatty-acid--CoA ligase	Faa2	−2.24	5.20×10^{-3}	−2.2	1.10×10^{-3}
	high mobility group-T protein	Hmo1	2.53	4.60×10^{-2}	−	−
	Myo-inositol-1-phosphate synthase	Ino1	*1.04*	*1.80×10^{-1}*	1.62	0.00019
	GTP Binding protein SAR1	Sar1	−2.39	2.80×10^{-3}	−	−
	20S Proteasome subunit	Scl1	2.24	1.20×10^{-3}	−	−
	SNZ2	Snz2	2.19	2.00×10^{-2}	−	−
	SSP120	Ssp120	2.38	9.20×10^{-4}	−	−
	vacuolar ATPase subunit E	Vma4	1.82	5.40×10^{-3}	−	−

[a] Expressing strain (3H6 Fab) and control strain (Control); -, indicates that a distinct spot could not be identified on the corresponding gels, whereas nonsignificant fold changes and *p*-values are highlighted in italic characters.

of growth temperature from 30 to 12 °C results in increased intracellular levels of metabolites such as glucose-6-phosphate.[39] It can be expected that such an increase of key-metabolites appears in *P. pastoris* chemostat cultivation during growth at low temperature and results in decreased levels of Aox1 and Icl1.

The impact of growth temperature on energy metabolism is obvious for both, the Fab secreting and the control strain, but abundance ratios and 1-ANOVA *p*-values are in many cases more significant in the Fab secreting strain.

Additionally, also proteins involved in oxidative stress response showed decreased levels during growth at 20 °C. Ccp1, which acts in mitochondria and degrades reactive oxygen species, and catalase A (Cta1), which is responsible for detoxification of hydrogen peroxide during beta-oxidation of fatty acids in peroxisomes in *S. cerevisiae*, showed lower protein levels at lower cultivation temperature. For Cta1, two spots that only differ in their p*I* and probably represent two isoforms of the protein were present on two-dimensional gels and both of them show decreased levels at low temperature growth.

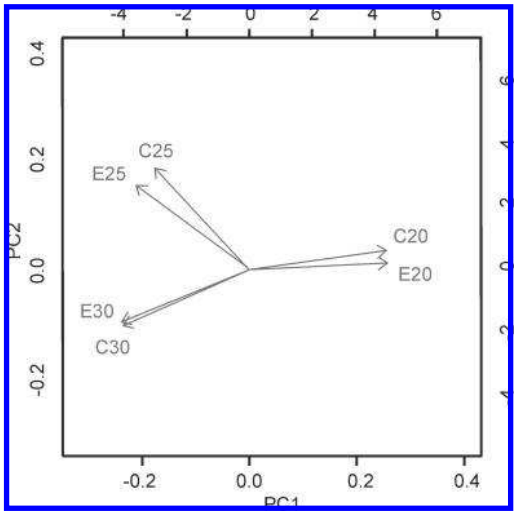


Figure 2. Results of principal component analysis. Biplot of principal component 1 (PC1) and principal component 2 (PC2). Production (E) and control strain (C) cultivated at 30 °C (E30, C30), 25 °C (E25, C25) and 20 °C (E20, C20).

P. pastoris Cta1 is also responsible for the detoxification of hydrogen peroxide, which is generated during growth on methanol. The down-regulation of proteins involved in oxidative stress response (Cta1, Ccp1) might be directly correlated with the lower abundance of proteins involved in fatty acid metabolism (Faa2 and Cat2) and respiratory proteins as many basic steps in energy metabolism involve transition metal ions and the generation of toxic oxygen species.[40]

Components of the protein folding and secretion machinery were also down-regulated at lower cultivation temperature. Chaperones Ssa4 and Ssb1 showed strongly reduced abundances, whereas two protein spots identified as fragments of these two heat shock proteins with a molecular weight of approximately 25 kDa showed significantly higher levels at 20 °C. A 50 kDa fragment of each of these two chaperones was also identified, but these 50 kDa fragments did not show a coherent up- or downward trend (Table 3). Both proteins belong to the HSP70 family and assist in the folding of newly synthesized proteins.[41] Ssa4 has previously been described as a potential secretion helper factor. Co-overexpression of Ssa4 resulted in an increase of specific productivity of a recombinant *P. pastoris* strain expressing an antibody fragment.[4] Interestingly, in the present chemostat study, Ssa4 was down-regulated under growth conditions that showed a 3-fold higher specific productivity. In contrast to Ssb1, which is a general high-abundance ribosome-associated chaperone, Ssa4 is highly induced upon stress in *S. cerevisiae*. Although positive effects on productivity and solubility of recombinant proteins are reported for chaperones of the HSP70 family and their bacterial homologues upon co-overexpression, for some recombinant proteins, also no or negative effects were reported. These contradictory results have been recently reviewed in detail by Gasser and co-workers.[42]

For the protein spots, which most likely represented the full-length proteins of Ssa4 and Ssb1, peptides spanning the whole protein from the N- to the C-terminal end were identified in the mass spectra; however, for the corresponding fragments, only N-terminal peptides were identified. This may lead to the assumption that decreased growth temperature led to a higher turnover of these chaperones.

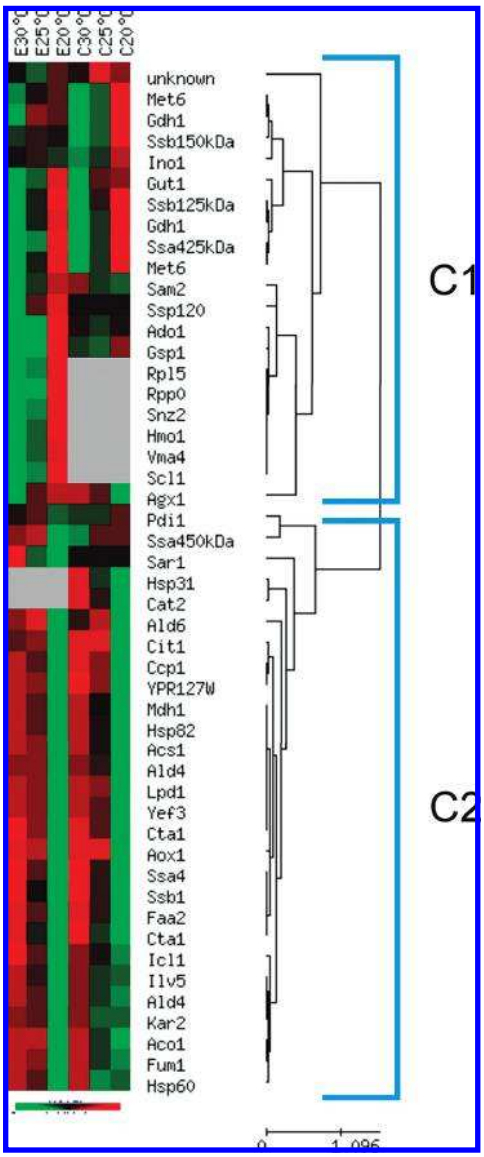


Figure 3. Hierarchical clustering of the proteins identified by mass spectrometry. Cluster of proteins with increased spot volumes at 20 °C (C1) and decreased spot volumes at 20 °C (C2). Production (E) and control strain (C). Red indicates increased spot volumes, green indicates decreased spot volumes (in comparison to the mean spot volume) and gray boxes indicate that no data were available for the corresponding protein spot.

Hsp82, a heat sensitive chaperone that is required for the correct folding of complex proteins[43] and Hsp60, involved in folding of mitochondrial proteins, showed the same trend as Ssa4 and Ssb1. The described chaperones behaved quite similarly in the production and the control strain. Additionally, we could also identify Hsp31 to have reduced protein levels during low-temperature growth in the control strain, whereas we could not identify the corresponding spot in the production strain due to lower gel image quality. Hsp31 has recently been identified to be involved in protection against oxidative stress,[44] which is in line with the lower abundance of Ccp1 and Cta1 at 20 °C.

We paid special attention to the behavior of the ER chaperone Kar2 (BiP), a key player in unfolded protein response (UPR)[45] and Pdi1, which is responsible for the formation of disulfide bonds. Therefore, these proteins were identified by

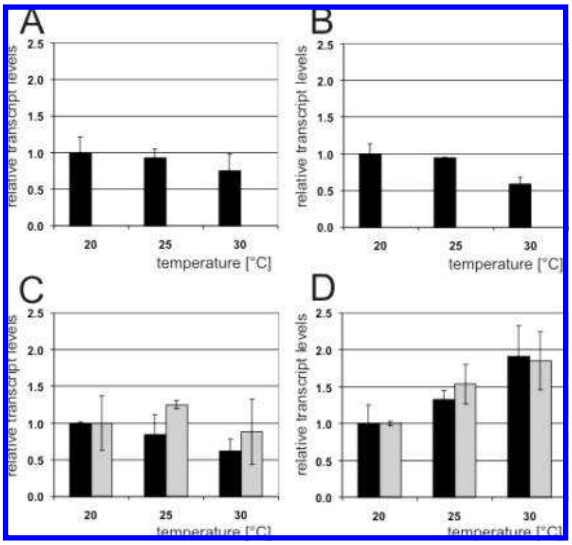


Figure 4. Results of Northern blot analysis of *P. pastoris* grown in aerobic chemostat cultures at 20, 25 and 30 °C. Relative mRNA levels to growth at 20 °C are shown for 3H6 Fab heavy chain (A), 3H6 Fab light chain (B), fumarase, FUM1 (C) and heat shock protein SSA4 (D). rRNA signal were used as a loading control. Black bars, 3H6 Fab expressing strain; gray bars, control strain. Error bars represent the standard error of the mean.

Western blotting and verified by mass spectrometry. Pdi1 has been reported to be a powerful candidate to enhance the secretion of heterologous proteins upon co-overexpression, since it participates in the formation of disulfide bonds, which are also present in antibodies and antibody fragments.[4] BiP levels decreased at lower temperatures in both strains (whereas the fold-change is only significant in the expressing strain), and Pdi1 levels seemed to be completely unaffected by growth temperature. This decreased level of BiP during 20 °C cultivation is concordant with transcript regulation data, obtained from a previous *P. pastoris* temperature experiment.[32] Since BiP participates in the UPR regulation,[46] it also influences the expression values of Pdi1. Therefore, it might be expected that these two proteins show a coordinate pattern, which was not the case in the current study. These data indicate a minor UPR induction under all culture conditions, although a decrease of cultivation temperature from 30 to 20 °C might help to relieve unfolded protein stress in the 3H6 Fab secreting strain, as BiP protein levels significantly decreased. It should be noted that, similarly to the cytosolic HSP70 protein family, inconsistent data about the effect on recombinant protein production upon co-overexpression of BiP and Pdi1 exist.[42]

We could also identify proteins involved in RNA transport and ribosome biogenesis, namely, Gsp1, Rpp0 and Rpl5. These proteins show higher abundance at 20 °C cultivation in the production strain, but were not identified in the control strain, most probably due to low gel image quality as these 3 proteins did not represent high intensity protein spots on the two-dimensional gels. The up-regulation of ribosomal proteins at suboptimal growth conditions is an already known phenomenon that has been reported for *S. cerevisiae* in anaerobic chemostat cultivations at low temperature on the transcript level.[11] In the current study, only 3 proteins were identified and they do not represent major components of the RNA and ribosome synthesis machinery. However, as the information that can be derived from 2D-DIGE and protein sequencing is limited, the up-regulation of few proteins may indicate changes on the cellular protein synthesis machinery, however, requiring further verification. Temperature influences the stability of RNA secondary structures, and at low cultivation temperature, translation initiation probably becomes a rate-limiting step in cell growth. Living cells are reported to show very constant levels of ribosomal protein and RNA in a wide temperature range[47] and the present study was performed in a yet physiological temperature range between 20 and 30 °C and not at cold-shock inducing temperatures. Therefore, only subtle changes of the cellular protein synthesis machinery might contribute to the improvement of q_P at lowered cultivation temperatures.

The levels of glutamate dehydrogenase (Gdh1) and homocysteine methyltransferase (Met6), both involved in amino acid metabolism, were increased in both strains during cultivation at 20 °C. The protein spot identified as Agx1 (responsible for the formation of glycine from glyoxylate) behaved differentially in the production and the control strain. Whereas Agx1 showed higher levels at 20 °C in the production strain, it decreased in its abundance in the control strain. Sam2, which is involved in methionine metabolism, also shows higher levels at low temperature in the production strain. On the other hand, Ilv5, a protein involved in branched-chain amino acid biosynthesis and mitochondrial genome maintenance,[48] showed decreased levels at low temperature in the production strain. Both proteins, Sam2 and Ilv5, did not show a significant change in the control strain (Table 3). Additionally, it has been shown by previous heterologous microarray experiments that Met6 and Sam2 belong to a tightly controlled set of genes in *P. pastoris*.[29] Met6 and Sam2 are also involved in the transfer of methyl-groups. The transfer of C1-units plays an important role in ribosomal subunit biogenesis, RNA processing and mRNA capping, thereby supporting the data on up-regulation of the cellular protein synthesis machinery, as it is indicated by increased levels of Gsp1, Rpp0 and Rpl5 at 20 °C.

There were also other cellular mechanisms that seemed to be affected by growth temperature on the proteome level, such as vitamin and adenosine metabolism. However, two protein spots, with no known function, were identified in this study (Table 3).

Northern Blot Analysis. Northern blots were performed to check whether differences in the secretion level of 3H6 Fab were related to different transcript levels. Transcript analysis showed that there is no difference in transcript levels for antibody 3H6 Fab heavy- and light chain between 20 and 25 °C cultivation. Also between 20 and 30 °C temperature setpoints, only minor differences in mRNA levels were visible. For the 3H6 light chain, a slightly significant lower transcript level at 30 °C, as compared to 20 °C, could be seen (Student's *t* test *p*-value = 0.038). Additionally, two other transcripts, *SSA4* and *FUM1*, were analyzed by Northern blotting. Ssa4 showed a significant response to growth temperature on the protein level in both strains, whereas Fum1 showed a strong response in the 3H6 Fab secreting strain and a weaker response in the control strain. *SSA4* mRNA levels appeared to be decreased at 20 °C in the expressing and the control strain (comparison of 30 and 20 °C Student's *t* test *p*-values of 0.022 and 0.028, respectively) and match the proteome data. The *FUM1* transcript level did not seem to be significantly affected by growth temperature, although there are differences among different cultivation temperatures on the proteome scale. The results of Northern blot analysis are summarized in Figure 4.

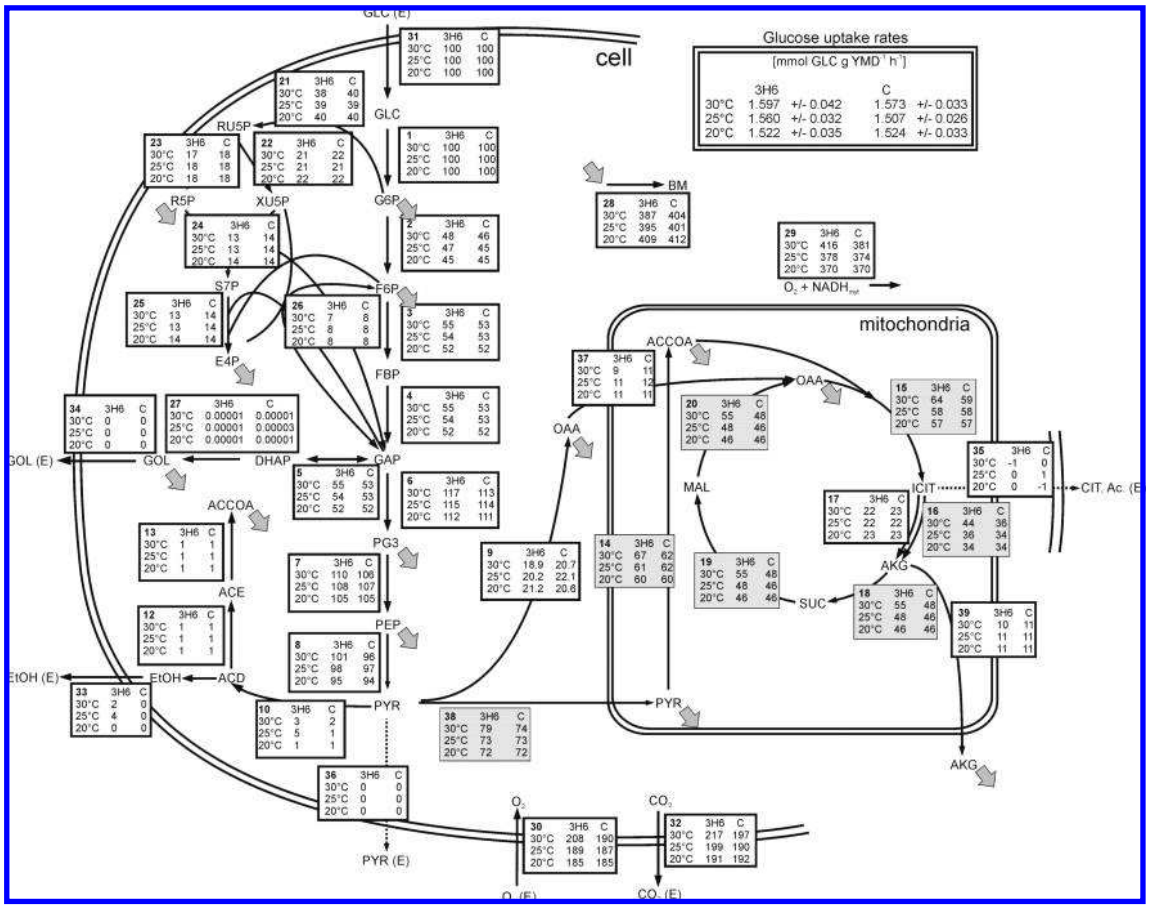


Figure 5. Central carbon fluxes of *P. pastoris* in aerobic chemostat cultures grown on glucose as sole carbon source at a dilution rate of $D = 0.1\ h^{-1}$. The fluxes are given as percentage of the total glucose utilization for the antibody fragment producing strain (3H6) and the control strain (C) at the three growth temperatures. Numbers in upper left corners represent the reaction numbers as given in Supplemental Data 1 in Supporting Information. Biomass producing reactions are illustrated with gray arrows. Fluxes affected most by growth temperature are highlighted in gray. Standard deviation of fluxes are provided in Supplemental Data 1 in Supporting Information.

Metabolic Flux Calculations. Two-dimensional gel electrophoresis and specific O_2 consumption and CO_2 production rates (Supplemental Data 1 in Supporting Information) indicated reduced TCA-cycle activity during 20 °C cultivation. From a thermodynamic point of view, this is interesting, as growth temperature influences the catalytic activity of enzymes,[39] but growth rate did not change in chemostat culture due to a constant dilution rate. Although many factors, such as allosteric effectors and isoenzymes with different catalytic properties, can help microorganisms to optimize their enzyme activity rates at lower temperatures, we would expect higher abundance of rate-limiting enzymes, such as citrate synthase (Cit1) at 20 °C as lower growth temperature generally results in a slowdown of enzymatic activity. However, in the present study, also this rate-limiting enzyme shows lower protein levels during growth at 20 °C.

To support the hypothesis that TCA-cycle activity decreased during growth at 20 °C, a metabolic network was built to calculate fluxes through the central carbon metabolism. A slightly but still significantly higher flux through glycolysis and especially the TCA-cycle at 30 °C compared with 20 °C was predicted by the present model for both the expressing and the control strain. The impact was higher in the antibody fragment producing strain (reactions 2−8, 14−20 and 38, respectively; Figure 5). These data correlate well with the results obtained from 2D-DIGE, where the average ratio for the proteins involved in energy metabolism were also higher in the 3H6 Fab expressing strain (Table 2). No significant changes were observed for the pentose-phosphate pathway.

Conclusions

So far, the general opinion toward increased productivity at decreased temperature was that enhanced product titers are most probably due to higher cell viability and lower proteolytic activity.[6,30] In this study, we demonstrated that this assumption is not the sole explanation for the increased product titers at decreased cultivation temperatures. Although a 3-fold increase in specific productivity was obtained by decreasing growth temperature from 30 to 20 °C, we could also show that the proteolytic activity did not change upon a temperature shift. These data indicate that there are other physiological parameters, besides viability and the release of cellular proteases, which are responsible for the enhancement of protein production at low-temperature cultivation.

It was shown by 2D-DIGE that a temperature reduction from 30 to 20 °C at steady-state culture conditions has a significant effect on important proteins of the *P. pastoris* metabolism (Figure 6).

In a recent study, Tai and co-workers reported on glycolytic fluxes in *S. cerevisiae* anaerobic chemostat cultures at normal

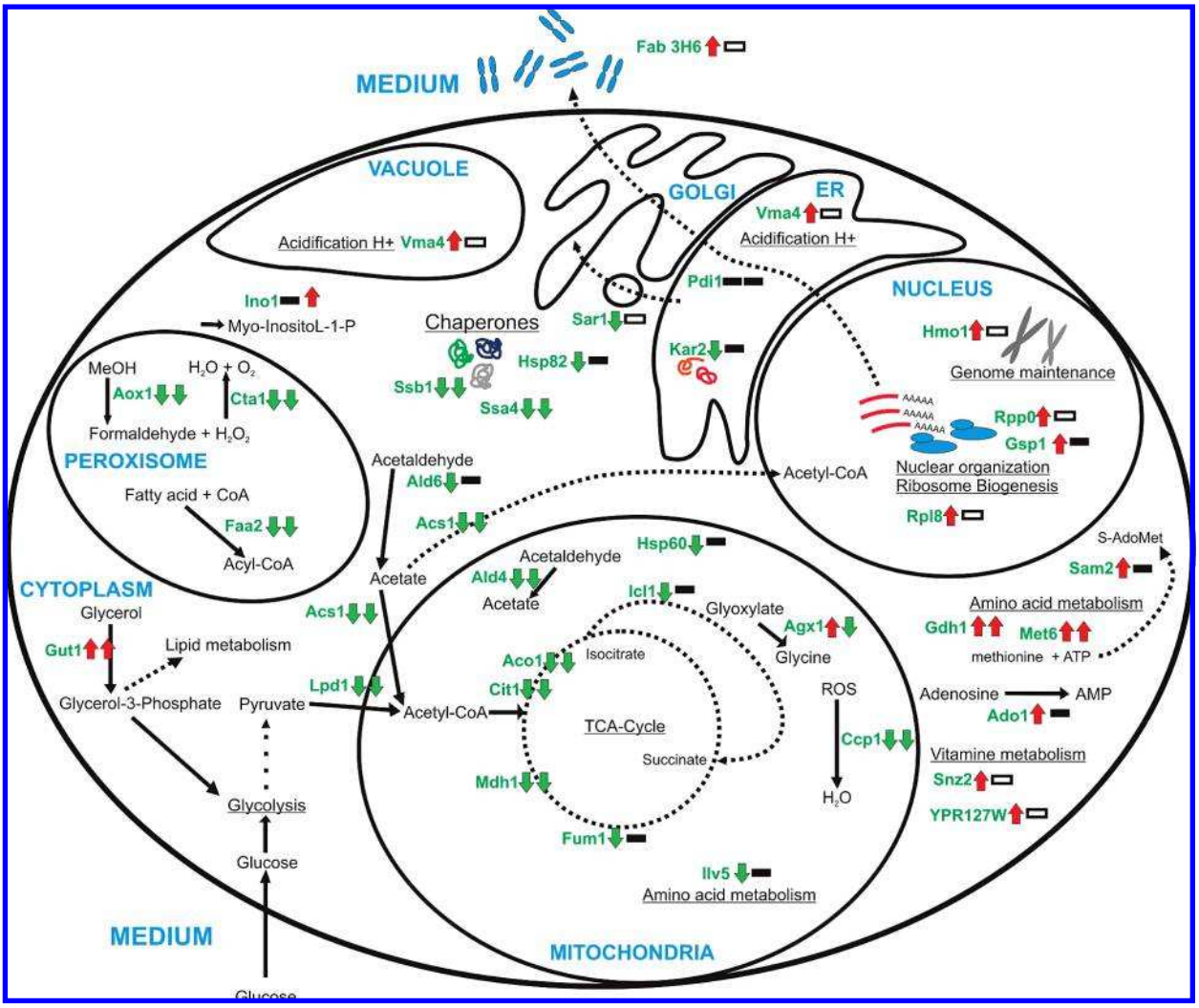


Figure 6. Main cellular changes in protein levels upon a temperature shift from 30 to 20 °C in chemostat cultures, as they were identified in the 3H6 Fab producing strain (first arrow next to protein short name) and the control strain (second arrow next to protein short name). Green downward arrows indicate reduced protein levels at 20 °C, red upward arrows indicate higher abundance at 20 °C, black blocks indicate no significant change in protein levels at different temperature setpoints and open blocks indicate missing data.

and suboptimal growth temperatures.[39] Major changes in the expression of glucose transporters occur during growth at suboptimal temperature, but there are only minor changes in gene expression of glycolytic enzymes. Nevertheless, they noticed massive changes in fermentative capacity and intracellular concentrations of nucleotides, indicating a major role of metabolic and post-transcriptional control of the central carbon pathways in *S. cerevisiae* chemostat cultures. Similar results were also obtained during cultivation on various carbon sources, wherein fluxes only partially correlate with gene expression levels.[49] Combining the results of 2D-DIGE, Northern blot analysis (especially Fum1 transcript levels) and flux calculations of the current study, these data also indicate important regulation patterns beyond the transcript level in *P. pastoris*.

It is most interesting that proteins involved in protein folding and secretion remain unchanged or show decreased levels during low temperature growth. Whereas lower mRNA levels could partially explain the reduced secretion of the 3H6 Fab fragment at 30 °C, other effects, such as mRNA stability, protein synthesis rates and protein folding related phenomena, could be responsible for the increased productivity at 25 and 20 °C. Molecular chaperones are essential catalysts for the production of heterologous proteins. It would be expected that increased levels of chaperones would lead to higher product titers. It is interesting that the 3-fold increase of q_P is accompanied by rather 'counterproductive' changes in the cellular protein folding machinery. Considering the narrow span of thermal stability of native proteins, the down-regulation of the protein folding machinery may also make sense. Higher growth temperature may result in increased denaturation and aggregation of native proteins, which consequently could lead to a higher demand of heat shock and stress proteins, including the chaperones identified in this study.[50] Therefore, cultivation at 20 °C may lead to generally higher stability of proteins and a reduced demand for protein refolding and/or degradation mechanisms, resulting in a higher secretion capacity as a positive 'side-effect'.

It is often reported that heterologous protein production poses a metabolic burden for the microbial host cell.[51,52] It can be speculated that heterologous protein production has an influence on the cellular folding and secretion machinery, leading to a higher demand for NADH and ATP than in a wild-type cell. At decreased temperature, reduced stress due to

reduced demand for protein (re)folding and degradation might also lead to lower energy demand. This assumption is in good compliance with carbon flux data and slightly increased biomass yields at 20 °C. Energy production was increased at 30 °C, but this increase was bigger in the production strain than in the control strain (reaction 29, Figure 5). Furthermore, it could explain the fact that the impact of growth temperature on proteins involved in TCA-cycle and energy production is more evident in the production strain than in the control strain.

The combination of 2D-DIGE proteome analysis and other approaches, such as DNA microarray analysis, might lead to an even more detailed picture of host cell physiology. Therefore, *P. pastoris* specific microarrays have been developed in our laboratory recently.[53] Beside the positive effect of decreased growth temperature on q_P, it has been reported that a reduction of oxygen supply can also lead to increased heterologous protein secretion in *P. pastoris*.[15] Hence, future work will include the analysis of the impact of other environmental factors, such as oxygenation and osmolarity, on host cell physiology under protein producing conditions, combining proteomic, transcriptomic and metabolomic methods.

Acknowledgment. This work has been supported by the European Science Foundation (ESF, program EuroSCOPE), the Austrian Science Fund (FWF), project no. I37, the Austrian Resarch Promotion Agency (Program FHplus), the Spanish program on Chemical Processes Technologies (project CTQ2007-60347/PPQ) and the Complementary Actions Plan supporting the ESF EuroSCOPE program (project BIO2005-23733-E), as well as the Generalitat de Catalunya (2005-SGR-00698). Special thanks to Martina Chang (Polymun Scientific) and Burghardt Scheibe (GE Healthcare) for their support and advice in establishing 2D-DIGE.

Supporting Information Available: Detailed information on the results of metabolite concentration determination, carbon flux calculations, the table of the reaction equations as well as the stoichiometric matrix are given in Supplemental Data 1; a complete list of spot volume ratios of the 2D-DIGE experiments is given in Supplemental Data 2; the total number of peptides identified by *de novo* sequencing and BLAST search for each protein and the corresponding amino acid sequences are listed in Supplemental Data 3. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Macauley-Patrick, S.; Fazenda, M.; McNeil, B.; Harvey, L. Heterologous protein production using the *Pichia pastoris* expression system. *Yeast* **2005**, *22* (4), 249–70.
(2) Gasser, B.; Maurer, M.; Gach, J.; Kunert, R.; Mattanovich, D. Engineering of *Pichia pastoris* for improved production of antibody fragments. *Biotechnol. Bioeng.* **2006**, *94* (2), 353–61.
(3) Smith, J.; Robinson, A. Overexpression of an archaeal protein in yeast: secretion bottleneck at the ER. *Biotechnol. Bioeng.* **2002**, *79* (7), 713–23.
(4) Gasser, B.; Sauer, M.; Maurer, M.; Stadlmayr, G.; Mattanovich, D. Transcriptomics-based identification of novel factors enhancing heterologous protein secretion in yeasts. *Appl. Environ. Microbiol.* **2007**, *73* (20), 6499–507.
(5) Blackwell, J. R.; Horgan, R. A novel strategy for production of a highly expressed recombinant protein in an active form. *FEBS Lett.* **1991**, 10–2.
(6) Shi, X.; Karkut, T.; Chamankhah, M.; Alting-Mees, M.; Hemmingsen, S.; Hegedus, D. Optimal conditions for the expression of a single-chain antibody (scFv) gene in *Pichia pastoris*. *Protein Expression Purif.* **2003**, *28* (2), 321–30.
(7) Sun, Z.; Zhou, R.; Liang, S.; McNeeley, K.; Sharfstein, S. Hyperosmotic stress in murine hybridoma cells: effects on antibody transcription, translation, posttranslational processing, and the cell cycle. *Biotechnol. Prog.* **2004**, *20* (2), 576–89.
(8) Mattanovich, D.; Gasser, B.; Hohenblum, H.; Sauer, M. Stress in recombinant protein producing yeasts. *J. Biotechnol.* **2004**, *113* (1−3), 121–35.
(9) Gasch, A.; Spellman, P.; Kao, C.; Carmel-Harel, O.; Eisen, M.; Storz, G.; Botstein, D.; Brown, P. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **2000**, *11* (12), 4241–57.
(10) Gasch, A. Comparative genomics of the environmental stress response in ascomycete fungi. *Yeast* **2007**, *24* (11), 961–76.
(11) Tai, S.; Daran-Lapujade, P.; Walsh, M.; Pronk, J.; Daran, J. Acclimation of *Saccharomyces cerevisiae* to low temperature: a chemostat-based transcriptome analysis. *Mol. Biol. Cell* **2007**, *18* (12), 5100–12.
(12) Gach, J.; Maurer, M.; Hahn, R.; Gasser, B.; Mattanovich, D.; Katinger, H.; Kunert, R. High level expression of a promising anti-idiotypic antibody fragment vaccine against HIV-1 in *Pichia pastoris*. *J. Biotechnol.* **2007**, *128* (4), 735–46.
(13) Hoskisson, P.; Hobbs, G. Continuous culture—making a comeback. *Microbiology* **2005**, *151* (Pt. 10), 3153–9.
(14) Regenberg, B.; Grotkjaer, T.; Winther, O.; Fausbøll, A.; Akesson, M.; Bro, C.; Hansen, L.; Brunak, S.; Nielsen, J. Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in *Saccharomyces cerevisiae*. *GenomeBiology* **2006**, *7* (11), R107.
(15) Baumann, K.; Maurer, M.; Dragosits, M.; Cos, O.; Ferrer, P.; Mattanovich, D. Hypoxic fed batch cultivation of Pichia pastoris increases specific and volumetric productivity of recombinant proteins. *Biotechnol. Bioeng.* **2008**, *100* (1), 177–83.
(16) Ferea, T.; Botstein, D.; Brown, P.; Rosenzweig, R. Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (17), 9721–6.
(17) Jansen, M.; Diderich, J.; Mashego, M.; Hassane, A.; de Winde, J.; Daran-Lapujade, P.; Pronk, J. Prolonged selection in aerobic, glucose-limited chemostat cultures of *Saccharomyces cerevisiae* causes a partial loss of glycolytic capacity. *Microbiology* **2005**, *151* (Pt. 5), 1657–69.
(18) Benthin, S.; Nielsen, J.; Villadsen, J. A simple and reliable method for the determination of cellular RNA content. *Biotechnol. Tech.* **1991**, *5* (1), 39–42.
(19) Verduyn, C.; Postma, E.; Scheffers, W.; van Dijken, J. Physiology of *Saccharomyces cerevisiae* in anaerobic glucose-limited chemostat cultures. *J. Gen. Microbiol.* **1990**, *136* (3), 395–403.
(20) Kolarich, D.; Altmann, F. N-Glycan analysis by matrix-assisted laser desorption/ionization mass spectrometry of electrophoretically separated nonmammalian proteins: application to peanut allergen Ara h 1 and olive pollen allergen Ole e 1. *Anal. Biochem.* **2000**, *285* (1), 64–75.
(21) Kolarich, D.; Altmann, F.; Sunderasan, E. Structural analysis of the glycoprotein allergen Hev b 4 from natural rubber latex by mass spectrometry. *Biochim. Biophys. Acta* **2006**, *1760* (4), 715–20.
(22) ERGO Bioinformatic software suite. http://ergo.integratedgenomics.com/ERGO/.
(23) Nissen, T.; Schulze, U.; Nielsen, J.; Villadsen, J. Flux distributions in anaerobic, glucose-limited continuous cultures of *Saccharomyces cerevisiae*. *Microbiology* **1997**, *143* (Pt. 1), 203–18.
(24) Pitkänen, J.; Aristidou, A.; Salusjärvi, L.; Ruohonen, L.; Penttilä, M. Metabolic flux analysis of xylose metabolism in recombinant *Saccharomyces cerevisiae* using continuous culture. *Metab. Eng.* **2003**, *5* (1), 16–31.
(25) Solà, A.; Maaheimo, H.; Ylönen, K.; Ferrer, P.; Szyperski, T. Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*. *Eur. J. Biochem.* **2004**, *271* (12), 2462–70.
(26) Solà, A.; Jouhten, P.; Maaheimo, H.; Sánchez-Ferrando, F.; Szyperski, T.; Ferrer, P. Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates. *Microbiology* **2007**, *153* (Pt 1), 281–90.
(27) Noorman, H.; Romein, B.; Luyben, K.; Heijnen, J. Classification, error detection, and reconciliation of process information in complex biochemical systems. *Biotechnol. Bioeng.* **1996**, *49* (4), 364–76.
(28) Stephanopoulos, G. N.; Aristidou, A. A.; Nielsen, J. *Metabolic Engineering. Principles and Methologies*; Academic Press: San Diego, CA, 1998.
(29) Sauer, M.; Branduardi, P.; Gasser, B.; Valli, M.; Maurer, M.; Porro, D.; Mattanovich, D. Differential gene expression in recombinant *Pichia pastoris* analysed by heterologous DNA microarray hybridisation. *Microb. Cell Fact.* **2004**, *3* (1), 17.
(30) Li, Z.; Xiong, F.; Lin, Q.; d'Anjou, M.; Daugulis, A.; Yang, D.; Hew, C. Low-temperature increases the yield of biologically active

herring antifreeze protein in *Pichia pastoris*. *Protein Expression Purif.* **2001**, *21* (3), 438–45.
(31) Jahic, M.; Wallberg, F.; Bollok, M.; Garcia, P.; Enfors, S. Temperature limited fed-batch technique for control of proteolysis in *Pichia pastoris* bioreactor cultures. *Microb. Cell Fact.* **2003**, *2* (1), 6.
(32) Gasser, B.; Maurer, M.; Rautio, J.; Sauer, M.; Bhattacharyya, A.; Saloheimo, M.; Penttilä, M.; Mattanovich, D. Monitoring of transcriptional regulation in *Pichia pastoris* under protein production conditions. *BMC Genomics* **2007**, *8*, 179.
(33) Lin, H.; Kim, T.; Xiong, F.; Yang, X. Enhancing the production of Fc fusion protein in fed-batch fermentation of *Pichia pastoris* by design of experiments. *Biotechnol. Prog.* **2007**, *23* (3), 621–5.
(34) Gygi, S.; Corthals, G.; Zhang, Y.; Rochon, Y.; Aebersold, R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (17), 9390–5.
(35) Pascoe, D.; Arnott, D.; Papoutsakis, E.; Miller, W.; Andersen, D. Proteome analysis of antibody-producing CHO cell lines with different metabolic profiles. *Biotechnol. Bioeng.* **2007**, *98* (2), 391–410.
(36) Grauslund, M.; Lopes, J.; Rønnow, B. Expression of GUT1, which encodes glycerol kinase in *Saccharomyces cerevisiae*, is controlled by the positive regulators Adr1p, Ino2p and Ino4p and the negative regulator Opi1p in a carbon source-dependent fashion. *Nucleic Acids Res.* **1999**, *27* (22), 4391–8.
(37) Fernández, E.; Moreno, F.; Rodicio, R. The ICL1 gene from *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **1992**, *204* (3), 983–90.
(38) Menendez, J.; Valdes, I.; Cabrera, N. The ICL1 gene of *Pichia pastoris*, transcriptional regulation and use of its promoter. *Yeast* **2003**, *20* (13), 1097–108.
(39) Tai, S.; Daran-Lapujade, P.; Luttik, M.; Walsh, M.; Diderich, J.; Krijger, G.; van Gulik, W.; Pronk, J.; Daran, J. Control of the glycolytic flux in *Saccharomyces cerevisia*e grown at low temperature: a multi-level analysis in anaerobic chemostat cultures. *J. Biol. Chem.* **2007**, *282* (14), 10243–51.
(40) Chang, T.; Chou, W.; Chang, G. Protein oxidation and turnover. *J Biomed Sci* **2000**, *7* (5), 357–63.
(41) Hartl, F. Molecular chaperones in cellular protein folding. *Nature* **1996**, *381* (6583), 571–9.
(42) Gasser, B.; Saloheimo, M.; Rinas, U.; Dragosits, M.; Rodríguez-Carmona, E.; Baumann, K.; Giuliani, M.; Parrilli, E.; Branduardi, P.; Lang, C.; Porro, D.; Ferrer, P.; Tutino, M.; Mattanovich, D.; Villaverde, A. Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microb. Cell Fact.* **2008**, *7*, 11.
(43) Nathan, D.; Vos, M.; Lindquist, S. In vivo functions of the *Saccharomyces cerevisiae* Hsp90 chaperone. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94* (24), 12949–56.
(44) Skoneczna, A.; Miciałkiewicz, A.; Skoneczny, M. *Saccharomyces cerevisiae* Hsp31p, a stress response protein conferring protection against reactive oxygen species. *Free Radical Biol. Med.* **2007**, *42* (9), 1409–20.
(45) Mori, K.; Sant, A.; Kohno, K.; Normington, K.; Gething, M.; Sambrook, J. A 22 bp cis-acting element is necessary and sufficient for the induction of the yeast KAR2 (BiP) gene by unfolded proteins. *EMBO J.* **1992**, *11* (7), 2583–93.
(46) Okamura, K.; Kimata, Y.; Higashio, H.; Tsuru, A.; Kohno, K. Dissociation of Kar2p/BiP from an ER sensory molecule, Ire1p, triggers the unfolded protein response in yeast. *Biochem. Biophys. Res. Commun.* **2000**, *279* (2), 445–50.
(47) Farewell, A.; Neidhardt, F. Effect of temperature on in vivo protein synthetic capacity in *Escherichia coli*. *J. Bacteriol.* **1998**, *180*, 4707–4710.
(48) Zelenaya-Troitskaya, O.; Perlman, P.; Butow, R. An enzyme in yeast mitochondria that catalyzes a step in branched-chain amino acid biosynthesis also functions in mitochondrial DNA stability. *EMBO J.* **1995**, *14* (13), 3268–76.
(49) Daran-Lapujade, P.; Rossell, S.; van Gulik, W.; Luttik, M.; de Groot, M.; Slijper, M.; Heck, A.; Daran, J.; de Winde, J.; Westerhoff, H.; Pronk, J.; Bakker, B. The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at post-transcriptional levels. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (40), 15753–8.
(50) Liberek, K.; Lewandowska, A.; Zietkiewicz, S. Chaperones in control of protein disaggregation. *EMBO J.* **2008**, *27* (2), 328–35.
(51) Glick, B. Metabolic load and heterologous gene expression. *Biotechnol. Adv.* **1995**, *13* (2), 247–61.
(52) Ramón, R.; Ferrer, P.; Valero, F. Sorbitol co-feeding reduces metabolic burden caused by the overexpression of a Rhizopus oryzae lipase in *Pichia pastoris*. *J. Biotechnol.* **2007**, *130* (1), 39–46.
(53) Graf, A.; Gasser, B.; Dragosits, M.; Sauer, M.; Leparc, G.; Tuechler, T.; Kreil, D.; Mattanovich, D. Novel insights into the unfolded protein response using *Pichia pastoris* specific DNA microarrays. *BMC Genomics* **2008**, *9* (1), 390.

PR8007623

The unfolded protein response is involved in osmotolerance of *Pichia pastoris*

Martin Dragosits[1], Johannes Stadlmann[2], Alexandra Graf[1, 3], Brigitte Gasser[1], Michael Maurer[3], Michael Sauer[3], David P. Kreil[4], Friedrich Altmann[2], and Diethard Mattanovich[1, 3*]

1 Department of Biotechnology, BOKU-University of Natural Resources and Applied Life Sciences, Vienna, Austria

2 Department of Chemistry, BOKU-University of Natural Resources and Applied Life Sciences, Vienna, Austria

3 School of Bioengineering, University of Applied Sciences FH-Campus Wien, Vienna, Austria

4 Chair of Bioinformatics, BOKU-University of Natural Resources and Applied Life Sciences, Vienna, Austria

e-mail addresses:

martin.dragosits@boku.ac.at
johannes.stadlmann@boku.ac.at
alexandra.graf@boku.ac.at
brigitte.gasser@boku.ac.at
michael.maurer@fh-campuswien.ac.at
michael.sauer@fh-campuswien.ac.at
david.kreil@boku.ac.at
friedrich.altmann@boku.ac.at
diethard.mattanovich@boku.ac.at

* To whom correspondence should be addressed. University of Natural Resources and Applied Life Sciences, Department of Biotechnology, Muthgasse 19, Vienna, Austria. Tel: (+43)1/36006-6569. Fax: (43) 1 3697615. E-mail: diethard.mattanovich@boku.ac.at.

Abstract

Background The effect of osmolarity on cellular physiology has been subject of investigation in many different species. Furthermore, several studies indicated that increased osmolarity of the growth medium can have a beneficial effect on recombinant protein production in different host organisms. Thus, the effect of osmolarity on the cellular physiology of the prominent host for recombinant protein production, *Pichia pastoris*, was studied in carbon limited chemostat cultures at three different media osmolarities. Transcriptome and proteome analysis were applied to assess differences upon growth at different osmolarities in both a non-expressing strain and a antibody Fab fragment expressing strain.

Results In contrast to the model yeast *Saccharomyces cerevisiae*, the main osmolyte in *P. pastoris* was arabitol rather than glycerol, demonstrating differences in energy metabolism, osmotic stress response and salt tolerance between these two yeast species. With the recently published genome sequence of *P. pastoris* at hand, 2D Fluorescence Difference Gel electrophoresis and microarray analysis were applied and demonstrated that processes such as the unfolded protein response, ribosome biogenesis and cell wall organization were affected by increased osmolarity. No activation of the high osmolarity glycerol (HOG) pathway was observed in these adapted conditions. While no changes in the specific productivity of recombinant Fab 3H6 were observed, transcriptome and proteome data indicated that less changes occurred upon growth at high osmolarity in recombinant *P. pastoris* than in the non-expressing strain.

Conclusion Taken together these data indicate that the unfolded protein response is involved in osmotic stress tolerance in *P. pastoris.* and the interrelation of environmental conditions such as high osmolarity and recombinant protein production.

Background

The response of cells to high osmotic pressure and increased salinity has been a subject of close investigation in many different organisms [1-4]. Depending on the intensity of the osmotic shock the immediate response to high osmolarity usually includes the activation of the environmental stress response (ESR) and of the high osmolarity glycerol (HOG) pathway to induce changes that are necessary to cope with the stressful environmental condition [5, 6]. In batch culture, osmotic shock usually implies a temporary growth arrest to adapt the metabolism [7]. Major adjustments of gene transcription in *Saccharomyces cerevisiae* include the induction of glycerol–3-phosphate dehydrogenase *GPD1* transcription [8], transcriptional repression of the plasma membrane glycerol efflux channel *FPS1* [2], but also the adjustment of ribosome biogenesis and the translation and protein folding machinery [9]. Glycerol production, but also the production of other small organic molecules, is induced in different yeast species to compensate variations of osmotic conditions [10]. Polyols, such as glycerol, pertain to a class of small molecules known as compatible solutes, which, in contrast to inorganic ions, can be safely accumulated and degraded in the cell without impairing cellular function or having a detrimental effect on protein and nucleic acid stability [11]. Furthermore, biomass yield is reduced upon exposure to high osmolarity because of higher maintenance energy in both, batch and chemostat cultures [7, 12]. However, it is known that after the immediate shock response, transcript levels of many stress responsive genes return to near basal levels after cells have adapted to the new environmental conditions [6].

The effect of osmolarity on cellular physiology is not only of particular interest for the basic research community. As environmental conditions can be changed very easily in an inexpensive manner, their interrelation with recombinant protein production is also of great interest for biotechnological applications. It has been reported for bacterial, yeast and mammalian host organisms that exposure to osmotic stress can have a beneficial effect on recombinant protein production by means of increased product yields or increased specific productivity [13-16]. Unfortunately, the positive effect of increased osmolarity on heterologous protein production is, at least in mammalian cells, often cell line specific [17]. *Pichia pastoris* represents a major fungal protein production host with key features that make it especially interesting for recombinant protein production. The genome sequence of *P. pastoris* has been recently published [18, 19] and with a publicly available sequence at hand thorough physiological investigations and characterization of this biotechnologically relevant organism becomes feasible.

In this context the effect of osmolarity on the physiology of *P. pastoris* was analyzed in both a control strain and a recombinant protein secreting strain. The secretion strain expressed the antibody Fab fragment 3H6 [20, 21] under the control of the constitutive glyceraldehyde–3-phosphate dehydrogenase (GAP) promoter. The effect of osmolarity was monitored in steady state by applying chemostat cultivation. Although chemostat cultivation differs from batch and fed batch systems, which are usually applied for large scale production of recombinant proteins, long term suboptimal growth conditions as they occur during batch and fed batch cultivation can also be applied in steady state chemostat conditions [22]. Furthermore, chemostat cultivation offers the advantage that growth rate related effects, which otherwise would interfere with high throughput protein and mRNA analytics, can be avoided [23].

To analyze the effect of increased osmolarity on host cell physiology, 2D Fluorescence Difference Gel Electrophoresis (2D-DIGE) and DNA microarray analysis were applied. These techniques have already been successfully applied to monitor the effect of environmental factors, such as temperature and osmolarity in yeasts [6, 24-26]. Furthermore, HPLC analysis was applied to analyze to intracellular polyol and trehalose content. The obtained data indicate a high importance of the unfolded protein response (UPR) for osmotic stress tolerance in *P. pastoris* and consequently differences in the response to increased osmolarity of a control strain and a recombinant protein secreting strain.

Results

General characteristics of cultures at different osmotic conditions Chemostat cultivations of *P. pastoris* were performed at three different osmotic conditions, which were achieved by different concentrations of KCl in the growth medium. This resulted in supernatant osmolarities of approximately 140, 850 and 1350mOsmol kg^{-1}, which will be named low, medium and high osmolarity thereafter. Samples were taken at steady-state, which means fully adapted cells were analyzed. The characteristics of chemostat cultures did not dramatically change upon cultivation at different osmotic conditions (Table 1). Biomass yield decreased with increasing osmolarity in the control strain and Fab 3H6 producing strain. However, the decrease in biomass yield was only statistically significant ($p \leq 0.05$) between cultivations at low and high osmolarity in both strains. The amount of total secreted protein did not change upon higher osmotic pressure, whereas the protein amount in the culture supernatant was higher in the Fab 3H6 production strain. Osmolarity of the growth medium

did not influence specific productivity (q_P) of the Fab 3H6. Generally, increased osmotic pressure poses a severe stress condition to cells [2]. Although the osmolarity was increased approximately 6–fold and ten–fold in the current study, no decrease of cell viability was observed. More than 97% of the cells in chemostat culture represented viable cells throughout all cultivations (Table 1). According to flow cytometry data, increased osmolarity resulted in a decrease of the mean cell size as indicated by a decrease of the mean forward scatter (Table 1).

Production of compatible solutes and trehalose in *P. pastoris* upon growth at different osmolarities In yeasts, glycerol is a very common solute but other polyols such as arabitol, mannitol and erythritol are also produced in some yeast species [10]. To analyze whether *P. pastoris* produces any of these substances, cell extracts were analyzed by HPLC.

It turned out that very low levels of mannitol and nearly no detectable amounts of erythritol were present in *P. pastoris* cells (Additional file 1). Intracellular glycerol levels were higher than mannitol and erythritol levels and a slight significant increase in the control strain from low to high and medium to high osmolarity occurred ($p \leq 0.05$), whereas no significant changes of glycerol content were observed in the Fab 3H6 expressing strain (Fig. 1A). Surprisingly, arabitol was the most abundant compound of the analyzed polyols in *P. pastoris* cells (five fold higher basal level then glycerol) and showed statistically significant increased levels ($p \leq 0.05$) when shifting growth conditions towards high osmolarity (Fig.1B). A 3-fold increase of intracellular arabitol levels was observed when comparing low and medium osmolarity conditions and a 4-fold increase when comparing cells grown at low and high osmolarity.

Furthermore, intracellular levels of trehalose were analyzed as trehalose is thought to be involved in relieving or impeding protein folding stress [27], which may also occur during salt stress [9]. Intracellular trehalose levels were in the same range as glycerol levels but showed a significant trend ($p \leq 0.05$) towards decreased concentrations at medium and high osmolarity growth conditions in the control strain but slightly missed the threshold p–value in the 3H6 Fab expressing strain (Fig.1C and Additional file 1)

The effect of osmolarity on the *P. pastoris* intracellular proteome As 2D-DIGE has already been successfully applied to monitor changes in the *P. pastoris* proteome upon growth at different temperatures [26], this method was also applied to track changes upon growth at elevated osmolarity. In the control strain of *P. pastoris* approximately 300 proteins passed the criteria (see experimental procedures section for details), whereas about 150 proteins passed these criteria in the recombinant protein producing strain. Most of the protein spots

represented low abundant proteins with too small quantities of proteins on the 2D gels to be confidently identified, resulting in 37 successfully identified proteins (Table 2). Figure 2 shows a representative gel image with the identified protein spots (a list of all identified protein spots with corresponding peptides, obtained from MS/MS analysis, is available in Additional file 2). As already reported in previous studies [26, 28], additionally to most likely full length proteins, protein fragments were identified according to the spot position on the gel (e.g Spot 4-Aco1, Spot 20 and 21-Ino1 and Spot 30-Ssb1; Table 2). Furthermore, protein levels showed the biggest changes when low and medium and low and high osmolarity cultivations were compared, whereas only minor changes occurred when comparing medium and high osmolarity setpoints (see Additional file 3 for a complete table of relative protein levels).

As can be seen in Figure 3, the major impact of osmolarity on the *P. pastoris* proteome was on proteins involved in energy metabolism and protein folding. Whereas protein levels of a major spot of aconitate hydratase (Aco1p) were increased at medium and high osmolarity in the control strain, they were not significantly affected by osmolarity in the Fab 3H6 producing strain. In contrast, in the Fab 3H6 expressing strain three minor isoforms or degradation products were significantly down-regulated at higher osmolarity, but showed no altered abundance in the control strain. Furthermore, citrate synthase (Cit1p) protein levels were decreased at high salt concentrations in the control strain but showed no significant change in the Fab producing strain. Formate dehydrogenase (Fdh1p), glycerol kinase (Gut1p), and isocitrate lyase (Icl1p) showed similar trends towards lower protein levels during medium and high osmolarity cultivations in both strains, whereas pyruvate kinase (Cdc19p) and phosphoglycerate kinase (Pgk1p) were generally up-regulated at higher osmolarity (it should be noted that Pgk1p levels returned to levels similar to low osmolarity cultivation in the production strain). Phosphoglucose isomerase (Pgi1p) showed lower levels at medium osmolarity in the production strain and a protein identified as Atp3p (a subunit of the mitochondrial F0F1 ATPase) was massively down-regulated at medium and high salt concentrations in the control strain. Furthermore, alcohol oxidase (Aox1p), a key enzyme in methanol utilization, was down-regulated at medium and high osmolarity in the control strain but did not show a significant change in the Fab 3H6 expressing strain. It should be pointed out that *AOX1* transcription is thought to be repressed during growth on glucose and that the current study was performed with glucose as carbon source to constitutively express the heterologous protein under the control of the GAP-promoter. However, it was shown

previously that basal levels of Aox1p were actually present during glucose limited growth of *P. pastoris* [26].

Similar discrepancies between the control and the recombinant protein expressing strain were observed for proteins involved in protein folding and secretion and folding stress response. Whereas the major ER chaperone and unfolded protein response (UPR) sensor Kar2p/BiP and the protein disulfide isomerase Pdi1p were up-regulated at medium and high osmolarity in the control strain, no changes of these two proteins were observed in the Fab producing strain. More prominently increased levels of cytosolic and mitochondrial chaperones Ssc1p, Sse1p, Ssz1p and Hsp60p were observed at medium and high osmolarity in the control strain than in the Fab 3H6 producing strain. The stress induced chaperone Ssa4p showed increased levels at medium salt concentrations in both strains but returns to below basal levels at high salt conditions in the Fab producing strain. Ino1p, a protein involved in synthesis of inositol phosphates and inositol-containing phospholipids and which is linked to the UPR [29], was also up-regulated at high salt concentrations in both strains analyzed.

Other protein spots that changed their abundance upon cultivation at increased salt concentrations were Agx1p and Gdh1p (both involved in amino acid synthesis). Both of them showed higher protein levels during growth at high osmolarity.

Figure 3 summarizes the osmolarity-induced effects observed on the proteome level of *P. pastoris*.

The effect of osmolarity on the *P. pastoris* transciptome To analyze the effect of different osmolarities on the *P. pastoris* transcriptome, *P. pastoris* specific microarrays were applied (Agilent platform). To support microarray analysis, real-time PCR was performed. Real-time PCR data proved to be consistent with microarray results (Additional file 4). In the *P. pastoris* control strain 521, 1024, and 6 open reading frames (ORFs) changed significantly when low osmolarity was compared to medium and high osmolarity and when medium osmolarity was compared to high osmolarity, whereas only 124, 177, and 43 ORFs were significantly regulated in the Fab 3H6 expressing strain. Although the number of regulated genes was lower in the heterologous protein producing strain, there was a high degree of overlap between the control and the Fab expressing strain among the significantly regulated ORFs (Table 3). Low p-values can result from a high technical variation within the replicates or reflect the biological truth within the samples. To determine if the lower amount of significantly regulated genes in the expressing strain is a technical artefact, the correlation, standard deviation and the coefficient of variation for the replicates of the control and expressing strain were compared. Correlation of intensity values was generally high between

all microarrays of one group (control/expressing red channel/green channel (see Additional file 5) with r^2 values between 0.95 and 0.97. The values for standard deviation and coefficient of variation (CV) indicated that the variance in replicates of the expressing strain was slightly but consistently higher than for the control strain (on average CV 0.18 for the control and CV 0.28 for the expressing strain, Additional file 5). Based on these results additional microarray experiments were performed to exclude any bias in the data. These additional data did not change the result or number of regulated genes, suggesting a true biological difference. To eliminate the possibility that the samples of the expressing strain vary more than the ones of the control strain, hierarchical cluster analysis (HCA) and gene set analysis (GSA) were performed on the fold change data and indicated that regulation was indeed different in the two strains analyzed (Additional file 5).

Because most of the genes that were regulated when comparing low to medium osmolarity were also regulated when comparing low to high osmolarity, the following data presentation and discussion will focus on the effects that were observed when low and high osmotic conditions were compared.

To get an overview of the general adaptations during steady-state cultivation, Fisher's exact test was performed to identify cellular processes, which were affected by different osmolarities on the transcript level. A total of 23 GO categories were either affected in both or at least in one of the analyzed strains (Additional file 5). Concordant with the mere number of regulated genes, there appeared more significantly affected cellular processes in the control strain than in the heterologous protein expressing strain. Only 3 GO categories occurred to be affected in both strains, namely GO:0006811 (ion transport), GO:0007047 (cell wall organization) and GO:0019725 (cellular homeostasis). Additionally, in the control strain the GO terms GO:0005975 (carbohydrate metabolism), GO:0006350 (transcription), GO:0006412 (translation) and GO:0042254 (ribosome biogenesis and assembly) were affected by increased extracellular osmolarity.

Figure 4 summarizes the important changes at the mRNA level of *P. pastoris,* grown in carbon–limited chemostat cultures when comparing high to low osmolarity, whereas microarray data for the discussed genes can be found in Additional file 6.

Regarding ion transport, uptake and metabolism, high osmolarity resulted in increased expression of the iron transporters *FTR1*, *SIT1* and the vacuolar iron reductase *FRE6* in both strains. Calcium ion homeostasis and calcium dependent signal transduction were obviously affected by high osmolarity in the control strain as the Ca^{2+} transporter *PMC1* and Calcineurin A (*CNA1*) were down–regulated at high osmolarity.

A major effect was apparent for genes involved in cell wall organization and its biogenesis. Whereas 21% of the genes belonging to this GO group were down-regulated at high osmolarity in the control strain, a similar effect, albeit with fewer significant genes, was observed in the 3H6 Fab secreting strain (Fig. 4A and 4B). Additionally, a putative extracellular or cell wall associated protein with homology to the *S. cerevisiae PRY1* gene was up-regulated in both strains at high osmolarity. Increased levels of a gene with homology to *S. cerevisiae PRY1* upon increased salinity have also been reported for the halotolerant yeast *Hortaea werneckii* previously [30]. However, no changes in the protein pattern, indicating higher protein levels, were observed by SDS-PAGE of the culture supernatant (data not shown).

A signaling cascade for sensing and adaptation to osmotic stress in *S. cerevisiae* has already been established based on the available data [2, 31] and genes with homology to the corresponding genes in *S. cerevisiae* were also identified in *P. pastoris.* None of the genes involved in osmotic stress sensing upstream of the mitogen activated protein kinase (MAPK) Hog1 showed significant regulation in the Fab 3H6 expressing strain, whereas *SHO1*, *SSK1* and *PTP3* were up–regulated and *STE50* was down–regulated at increased osmolarity in the control strain (Fig. 4A).

Several genes involved in energy metabolism and storage carbohydrate metabolism were affected by increased osmolarity. *FBA1*, a key enzyme in glycolysis and gluconeogenesis, was up–regulated in the heterologous protein expressing strain. The acetyl-coA synthetase *ACS2* was up-regulated in the control strain. Transcript levels of several genes involved in the tricarboxylic acid (TCA) cycle and the glyoxylate cycle, namely *ACO1*, *FUM1*, *MDH1*, *SFC1* and *ICL1* were reduced and subunits of the ATP synthase (*ATP5* and *ATP18*) were up-regulated during growth at high osmolarity in the *P. pastoris* control strain. In the Fab producing strain, *TKL1*, involved in the pentose phosphate (PP) pathway was significantly up-regulated. Glycogen synthesis was also affected by high osmolarity as *GLG1*, *GSY2* and *GLC3* showed decreased transcript levels during growth at high osmolarity. Decreased levels of *DGA1*, *GUT1* and *GTP2* indicated changes in glycerol and lipid metabolism. Additionally, a homologue to the *S. cerevisiae* putative passive glycerol channel *YFL054C* was down-regulated and the active glycerol importer *STL1* was up-regulated during growth at high osmotic conditions in the control strain. Furthermore, significant down-regulation of the *P. pastoris* alcohol oxidase *AOX1* was observed in the control strain at high osmolarity, whereas no significant regulation was observed in the Fab 3H6 expressing strain. These data are concordant with data on the proteome level (Table 2).

In the control strain approximately 7% of the genes involved in ribosome biogenesis and assembly were up–regulated during steady state cultivation at high osmolarity.

Salt tolerance of *Pichia pastoris* As no data on the salt tolerance of *P. pastoris* compared to *S. cerevisiae* were found in literature, a simple growth test on YPD agar plates, containing different amounts of NaCl or KCl, was performed. This growth test indicated higher tolerance to growth on NaCl and a less pronounced higher tolerance to growth on KCl of *P. pastoris* compared to *S. cerevisiae* (Figure 5).

Discussion

Production of Compatible Solutes To counterbalance the osmotic pressure by high or low salt or solute concentrations in the growth medium, microorganisms produce various compatible solutes. In *S. cerevisiae* and many other organisms glycerol is the main osmolyte accumulated during osmotic stress. However, we found that intracellular glycerol levels were rather low in *P. pastoris*. On the other hand, arabitol was more abundant than glycerol, even at low osmolarity, and it was accumulated in *P. pastoris* during growth at elevated osmotic pressure (Fig 1A and 1B). Glycerol production in *S. cerevisiae* depends on the increased expression of glycerol-3-phosphate dehydrogenase *GPD1* and glycerol-3-phosphatase *GPP2* [7, 25]. Nevertheless, we could not find neither of the two genes involved in glycerol metabolism, *GPD1* and *GPP2*, to be up–regulated on the transcript level and did not identify protein spots with altered expression which would match these two genes in *P. pastoris*. Furthermore, it was shown in *Debaryomyces hansenii* that NaCl stress lead to increased levels of proteins involved in the upper part of glycolysis and down–regulation of proteins involved in the TCA–cycle [24]. It was concluded that these changes may favour the accumulation of dihydroxyacetonephosphate and consequently the production of glycerol [32]. No changes related to the upper part of glycolysis were observed in the current study, although it is clear that the need for compatible solutes leads to a redirection of a part of the carbon source to alleviate the stress induced by increased osmolarity. This would make sense as arabitol obviously plays a more important role as compatible solute than glycerol in *P. pastoris*. However, the regulation of genes involved in glycerol uptake and efflux, such as the up-regulation of *STL1* and the down-regulation of *YFL054C* may be beneficial for *P. pastoris* as well. The loss of minor osmolytes may result in detrimental effects on cellular integrity at high KCl concentrations. No significant regulation was observed for other putative glycerol transporters of *P. pastoris* recently described by Mattanovich and co-workers [18]. Arabitol

synthesis is linked to the pentose phosphate (PP) pathway. However, no changes possibly linked to the PP pathway and arabitol synthesis were observed on the transcript or the proteome level. The regulation of arabitol synthesis might be mainly achieved on a post–transcriptional level by increased translation or by protein modification and changes of enzyme activity. Nevertheless, concordant with other studies, arabitol obviously has particular importance in the metabolism of *P. pastoris* as it is also secreted into the supernatant at certain growth conditions, such as low oxygenation [33].

Trehalose has been previously shown to play an important role in heat shock induced refolding of proteins in baker's yeast [34] and *in vitro* [27]. Furthermore, trehalose may also be involved in the response to temperature induced stress in *P. pastoris* as intracellular levels increased at elevated temperature (own unpublished data). However, as trehalose levels were lower during growth at high osmolarity and no changes of the stress induced cytosolic trehalase *NTH1* [35] were observed on the protein or mRNA level, trehalose may not be directly involved in the protection of proteins against osmotic induced protein denaturation or damage. It is more likely that, similar to *S. cerevisiae*, trehalose degradation may play a role during growth at elevated osmolarity [36], or that trehalose levels may be simply lower due to a redirection of carbon source to the production of arabitol rather then to the production of trehalose.

Effect on Energy Metabolism The differential response of Aco1p and the differences of transcript levels of genes involved the TCA cycle to different osmotic conditions between the control and the Fab 3H6 expressing strain lead to the conclusion that recombinant protein production influenced the osmo-dependent adaptation of the energy metabolism. Previous data already indicated a metabolic burden and influence of recombinant protein production on energy metabolism in *P. pastoris* [26, 37]. Furthermore, the key enzyme of methanol utilization, *AOX1*, was differently regulated in the two strains and indicated significant differences in the regulation of energy metabolism. Protein and transcript levels of the alcohol oxidase (*AOX1*) were significantly negatively affected by growth at high osmolarity in the control strain but not in the Fab 3H6 secreting strain. *P. pastoris* Aox1 seems to be tightly regulated upon exposure to various stresses and might represent an ideal candidate as a marker gene/protein to monitor diverse stresses in *P. pastoris*. Apart from these additional data supporting the idea of a metabolic burden during recombinant protein production in *P. pastoris*, no clear interpretation about the changes of energy metabolism upon growth at different osmolarities in chemostat cultures emerged. Further investigations using a different

approach to the one used in the current study will be necessary to elucidate the effect of osmolarity on the energy metabolism of *P. pastoris*.

Signaling pathways and the importance of the UPR Another major effect, connected to protein metabolism, was the massive increase of chaperones and UPR related proteins at high osmolarity. The UPR, including heat shock proteins and cellular chaperones, plays an essential role in the response to various stresses [38, 39]. Apart from its role in the ESR of unicellular organisms, the UPR is also of great importance in human disease as highlighted by its involvement in the development of several human maladies such as diabetes, neurodegenerative disorders and cancer [40, 41]. In the current study an induction of Hsp60 and Hsp70 chaperones (Ssc1p, Sse1p and Ssz1p) was observed on the proteome level but not on the transcript level; indicating once more substantial regulatory mechanisms beyond the level of transcription. The observation of increased levels of molecular chaperones during growth at high osmolarity is concordant with previous results for *Aspergillus nidulans* [1] and similar to results obtained for *D. hansenii* [24] and *S. cerevisiae* [25]. Furthermore, high osmotic pressure resulted in increased levels of Pdi1p and Kar2p, indicating activation of the UPR. Unlike *S. cerevisiae*, UPR induction has been reported to be a main event upon exposure to salt stress in the halotolerant yeast *Rhodotorula mucilaginosa* [9]. Generally, the induction of the UPR may not only be a result of high concentrations of ionic solutes such as salts but UPR is also triggered by high osmotic pressure induced by other substances such as sugar compounds. It has been reported for mammalian cells that low as well as high hexose concentrations can also lead to UPR induction [42, 43]. The fact that UPR was strongly induced upon growth at high osmolarity indicated that, similar to halotolerant yeast species like *R. mucilaginosa*, *P. pastoris* might use different mechanisms for gaining osmotic stress resistance than *S. cerevisiae*. According to the performed drop tests *P. pastoris* showed indeed higher resistance to increased salt concentrations in the growth medium than *S. cerevisiae* (Figure 5).

The obtained data indicated that other stress response mechanisms, such as response to oxidative stress, were also affected by high osmolarity. The interrelation of salt and oxidative stress is already established in plants [44] and the interrelation and cross–talk of the HOG pathway and other pathways such as protein kinase C (PKC) and calcineurin dependent signaling are also established in yeasts [2, 45, 46]. Changes in cell wall integrity signaling, which were evident by altered expression levels of cell wall components, may be directly related to the changes of the *CNA1* and *PMC1* transcripts, as some of these cell wall synthesis related genes are dependent on calcineurin signaling [46]. Although this effect was very

evident on the transcript level we were not able to monitor it on the proteome level. This may be simply due to the preparation of protein samples and the resulting absence of cell wall and membrane proteins, which are rather difficult to extract by standard protein preparation methods.

It is most interesting that induction of the UPR on the proteome level and induction of ribosome biogenesis and translation on the transcript level were apparent in the control strain but not in the heterologous protein producing strain. As indicated by proteome and microarray data, including HCA and GSA (Additional file 5), the Fab 3H6 expressing strain of *P. pastoris* showed a different response to altered osmolarity than the control strain. Concerning ribosome biogenesis and translation it has been reported for other yeasts, namely *H. werneckii* and *D. hansenii*, that genes involved in protein synthesis were up–regulated during osmotic stress [30, 47]. Furthermore, studies on brewing strains of *S. cerevisiae* also concluded that the faster adaption to higher salt concentration compared to a laboratory strain was achieved by higher expression levels of genes involved in protein synthesis [48]. Similar to other environmental factors, such as temperature [49], translation might become a rate-limiting factor during growth at high osmolarities because of stress caused by decreased intracellular water availability. Generally, it is well established that UPR is induced by the expression of heterologous proteins [50, 51]. Furthermore, it was shown by a previous microarray study that over–expression of the UPR key transcription factor *HAC1* [52] in *P. pastoris* resulted in increased expression of genes involved in mRNA translation and to a massive increase of genes involved in ribosome biogenesis and assembly [53]. The up-regulation of ribosome biogenesis, translation and additional most likely co-regulated processes at high osmolarity was obviously not necessary in a heterologous protein producing strain as these changes have already been triggered by the recombinant protein induced UPR response.

Conclusion

Although the central ESR pathways are well conserved among fungi, the up- and downstream elements can be significantly different among species to satisfy niche–specific requirements [54]. Most notably, the presented data demonstrate a very high similarity and/or cross-talk of recombinant protein induced stress and the response to elevated osmolarity via the involvement of the UPR in *P. pastoris*. The recombinant Fab 3H6 secreting *P. pastoris* strain was less prone to osmotic induced stress. Distinct differences, especially in central carbon metabolism and UPR existed between the control and the 3H6 Fab producing strain.

Although in the current study elevated osmolarity did not result in increased productivity for recombinant Fab 3H6, the obtained data might be useful to explain the results of other research groups. It has been reported previously that osmotic stress applied prior to induction of protein secretion resulted in higher levels of scFv antibody in *P. pastoris* [13]. Because osmotic stress obviously induces UPR in *P. pastoris* it seems plausible that cells may be prepared for recombinant protein production because unfolded protein stress may be relieved or the folding competence of the host cells may be increased compared to non pre–conditioned cells.

According to the current data post-translational control mechanisms play an essential role in *P. pastoris*, especially during chemostat cultivation. Other proteomic methods such as the analysis of the phosphoproteome [55] might be very useful to gain detailed insights into these yet unestablished mechanisms. However, the current data represent a first step towards a systems wide approach to assess the response to environmental stresses, as well as their interrelation with recombinant protein production, in *P. pastoris*.

Methods

Materials All chemicals for yeast cultivations were molecular biology grade and were purchased from Roth, Germany. All chemical reagents for two-dimensional gel electrophoresis were high purity grade and were purchased from Sigma, unless stated otherwise.

Yeast Strains Two strains, which have been described recently [26], have been used in this study. For secreting the Fab 3H6, both the light and the heavy chain of the Fab fragment were expressed under the control of the constitutive GAP-promoter using the pGAPZαA vector. Secretion was mediated by the *S. cerevisiae* α–mating factor secretion signal. For the non-expressing control strain, *P. pastoris* X-33 was transformed with an empty pGAPZαA vector as described by Gasser and co–workers [56]

Chemostat cultivation For chemostat cultivations a 3.5 L bench-top bioreactor (MBR, Switzerland) was used at a working volume of 1.5 L. A 1000 mL shake flask containing 150 mL YPG medium (2% (w/v) peptone, 1% (w/v) yeast extract, 1% (w/v) glycerol) was inoculated with 1 mL cryostock of the respective *P. pastoris* clones. The cultures were grown for approximately 24 h at 28°C and shaking at 170 rpm, before they were used to inoculate the bioreactor to an optical density (OD_{600}) of 1.0. After a batch phase of approximately 24 hours the continuous culture was started at a dilution rate of $D = 0.1\ h^{-1}$ (growth medium flow

rate of 150 g h^{-1}). pH was controlled at 5.0 with 25% ammonium hydroxide (w/w). Gas flow rate was kept constant at 1.5vvm (volume gas per volume medium and minute) and dissolved oxygen was kept at 20% by controlling the stirrer speed. Three chemostat media, with different osmolarities, were used.

Batch medium contained per liter: 39.9 g glycerol, 1.8 g citric acid, 12.6 g $(NH_4)_2HPO_4$, 0.022 g $CaCl_2 \cdot 2H_2O$, 0.9 g KCl, 0.5 g $MgSO_4 \cdot 7H_2O$, 2 mL Biotin (0.2 g L^{-1}), 4.6 mL trace salts stock solution. The pH was set to 5.0 with 25% (w/w) HCl. Osmolarity of the growth medium was controlled by KCl concentration. Chemostat medium contained per liter: 50 g glucose $\cdot 1H_2O$, 0.9 g citric acid, 4.35 g $(NH_4)2HPO_4$, 0.01 g $CaCl_2 \cdot 2H_2O$, 1.7 (low) or 29.9 (medium) or 48.5 (high) g KCl, 0.65 g$MgSO_4$ $7H_2O$, 1mL Biotin (0.2 g L^{-1}), and 1.6 mL trace salts stock solution. The pH was set to 5.0 with 25% (w/w) HCl. Trace salts stock solution contained per liter: 6.0 g $CuSO_4 \cdot 5H_2O$, 0.08 g NaI, 3.0 g $MnSO_4 \cdot H_2O$, 0.2 g $Na_2MoO_4 \cdot 2H_2O$, 0.02 g H_3BO_3, 0.5 g $CoCl_2$, 20.0 g $ZnCl_2$, 5.0 g $FeSO_4 \cdot 7H_2O$, and 5.0 mL H_2SO_4 (95–98% w/w).

Three chemostat cultivations were performed, where the cultivation regime was different for each cultivation to avoid adaptive evolution effects and sample bias due to long term cultivation [57]. Samples were taken at steady state after 8 residence times after a switch of culture medium. Biomass was determined by drying duplicates of 10mL chemostat culture to constant weight at 105°C in pre-weight beakers. Samples for 2D-DIGE and DNA microarray analysis were taken from the chemostat and immediately frozen at -80°C until use, whereat the samples for transcript analysis were fixed with 5% (v/v) phenol / ethanol prior to freezing. Osmolarity determinations of culture supernatants were performed on a Semi-Microosmometer (Knaur). Viability of cells was determined immediately after samples were taken from the chemostat on a FACSCalibur flow cytometer (BD Biosciences) and a cell viability kit (BD Biosciences) as described previously [58].

Analysis of intracellular Polyols and Trehalose To quantify intracellular levels of glycerol, arabitol, mannitol, erythritol and trehalose, heat extraction was performed as described by Philips and co-workers [59]. Cell pellets were resuspended in 0.5M TrisCl pH 7.5, heated to 95°C for 10min and centrifuged for 10min to remove cell debris. Supernatants were kept for analysis via HPLC. Isocratic conditions, using 4mM H_2SO_4 as solvent and a flow rate of 0.6mL min^{-1} on a Aminex HPX-87H column (Biorad) at 40°C and a Biologic DuoFlow (Biorad) combined with a Smartline RI Detector 2300 (Knauer) were applied to separate and analyze substances (Additional file 1). Concentrations were determined by external standard solutions. Solute concentrations were correlated with biomass.

2D Fluorescence Difference in Gel Electrophoresis (2D-DIGE) and protein identification 2D-DIGE and protein identification were essentially performed as described previously [26]. After adequate sample preparation, cleaning, quantification and Cy-dye labeling, proteins were separated on IPG DryStrips pH 3-11NL (GE Healthcare) on a IPGPhor for a total of 65kVh. 2nd dimension separation was performed by SDS polyacrylamide gel electrophoresis on 12% polyacrylamide gels. Fluorescence gel images were taken at a resolution of 100µm on a Typhoon 9400 Fluorescence scanner. The DeCyder Software package v.5 (GE Healthcare) was used to analyze the obtained gel images. Significantly regulated protein spots (fold- change ≥ 1.5, 1-way ANOVA ≤ 0.05 in at least one comparison of cultivation conditions and present on at least 80% of the spot maps) were picked from Coomassie stained gels and after a tryptic digest subjected to reversed phase capillary chromatography (BioBasic C18, 5 µ, 100 x 0.18 mm, Thermo) and ESI-MS/MS on a quadrupole time-of-flight (Q-TOF) Ultima Global (Waters Micromass) mass spectrometer. Mass spectra were analyzed either by using the Protein Lynx Global Server 2.1 software (Waters) or X!Tandem (http://www.thegpm.org/tandem/). Only proteins identified by at least 2 peptides were considered to represent confident hits, except for Pdi1, which was verified via Western blotting [26].

DNA microarray analysis DNA microarray analysis was performed using *P. pastoris* specific microarrays (Agilent) as described by Graf and co-workers [53]. RNA was extracted from Ethanol/Phenol fixed cell samples. Reverse Transcription and synthesis of Cy3/5 labeled cRNA was done using the Low RNA Input Two-Color Amplification Kit (Agilent). cRNAs were purified via RNeasy Mini spin colums (Qiagen). Quality of total RNA and labeled cRNA was confirmed on an Agilent Bioanalyzer 2100 and the RNA Nano 6000 Assay Kit (Agilent). RNA concentrations were determined on a ND-1000 (Nanodrop). After hybridization at 65°C for 17h, slides were scanned on an Agilent MicroArray scanner and raw data were extracted using Feature Extraction v.9.1 (Agilent). Normalization steps and statistical analysis of microarray data, including Hierarchical cluster analysis, Fisher's exact test and Gene Set Analysis (GSA), were done using the R software package (www.**r**-project.org). For identifying differentially expressed genes, the False Discovery Rate was controlled strongly less than 5% ($q < 0.05$) using a Benjamini-Yekutieli correction for multiple testing. For Fisher's exact test and GSA, 63 gene ontology terms were considered. This list of terms was compiled based on the GOslim annotation of the *Saccharomyces* genome database (www.yeastgenome.org), where some of the larger categories were resolved at a finer gene ontology level. A threshold of $p \leq 0.05$ was chosen to be appropriate to identify

significantly regulated GO categories. Microarray data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress).

Real–Time PCR To support microarray data, Real-time PCR was performed. Total RNA was reverse-transcribed using a Superscript III cDNA synthesis kit (Invitrogen). Quantity of cDNA was determined on a ND-1000 (Nanodrop). Real time PCR was performed using the SensiMix Plus PCR premix (GenXpress) on a Rotorgene 6000 (Corbett Life Sciences). The following target genes were selected for Real-time PCR analysis: *ACT1*, *AOX1*, *DGA1*, *GLG1*, *SIT1*, *PDI1*, *3H6 Fab HC* and *3H6 Fab LC* (Additional file 4). Data were analyzed via the Rotorgene Software package and Microsoft Excel. *ACT1* was chosen as reference to determine relative mRNA levels of the other genes.

Growth tests on different salt concentrations *P. pastoris* X-33 and *S. cerevisiae* HA232 (http://www.biotec.boku.ac.at/acbr.html) were grown in YPD medium (2% (w/v) peptone, 1% (w/v) yeast extract, 2% (w/v) glucose) at 28°C on a shaker at 170rpm over night. Cultures were diluted to an OD of 0.1 in sterile PBS and 1:10 serially diluted in sterile PBS. 3µL were spotted onto YPD agar plates (2% (w/v) peptone, 1% (w/v) yeast extract, 1% (w/v) agar, 2% (w/v) glucose) containing 0, 0.6, 1.2, 1.4 and 1.6M NaCl or KCl. Plates were incubated at 28°C for 4 to 6 days.

3H6 Fab quantification To analyze the 3H6 Fab produced during chemostat cultivation, a sandwich ELISA was performed as described in previous studies [56].

Author's Contributions

MD performed chemostat cultivation, microarray analysis, Real-time PCR, 2D-DIGE, HPLC analysis and growth tests. JS performed protein identifications by LC-ESI-MS/MS. AG performed the statistical evaluation of the microarray experiments. MM set up media recipes for cultivation and assisted in chemostat cultivations. BG, MS and DM contributed to the design of the study and data interpretation. DPK contributed to the design and advised on the analysis of the microarray experiments. FA supervised MS/MS analytics. MD, BG and DM drafted the manuscript. DM conceived of the study.

Acknowledgements

This work has been supported by the European Science Foundation (ESF, program EuroSCOPE), the Austrian Science Fund (FWF), project no. I37 and the Austrian Resarch Promotion Agency (Program FHplus) and is part of the Genophys research project. DPK acknowledges funding by the the Vienna Science and Technology Fund (WWTF), Baxter

AG, Austrian Research Centres Seibersdorf, and the Austrian Centre of Biopharmaceutical Technology. Thanks to Martina Chang (Polymun Scientific) and Burghardt Scheibe (GE Healthcare) for their support and advice in 2D-DIGE, Hans Marx (School of Bioengineering, University of Applied Sciences FH-Campus Wien) for his advice in HPLC analytics and Astrid Mecklenbräuker (Department of Biotechnology, University of Natural Resources and Applied Life Sciences) for her advice in real-time PCR.

References

1. Kim Y, Nandakumar M, Marten M: **Proteome map of Aspergillus nidulans during osmoadaptation.** *Fungal Genet Biol* 2007, **44**(9):886-895.
2. Mager W, Siderius M: **Novel insights into the osmotic stress response of yeast.** *FEMS Yeast Res* 2002, **2**(3):251-257.
3. Shen D, Sharfstein S: **Genome-wide analysis of the transcriptional response of murine hybridomas to osmotic shock.** *Biotechnol Bioeng* 2006, **93**(1):132-145.
4. Zeller G, Henz S, Widmer C, Sachsenberg T, Rätsch G, Weigel D, Laubinger S: **Stress-induced changes in the Arabidopsis thaliana transcriptome analyzed using whole-genome tiling arrays.** *Plant J* 2009, **58**(6):1068-1082.
5. O'Rourke S, Herskowitz I: **Unique and redundant roles for HOG MAPK pathway components as revealed by whole-genome expression analysis.** *Mol Biol Cell* 2004, **15**(2):532-542.
6. Gasch A, Spellman P, Kao C, Carmel-Harel O, Eisen M, Storz G, Botstein D, Brown P: **Genomic expression programs in the response of yeast cells to environmental changes.** *Mol Biol Cell* 2000, **11**(12):4241-4257.
7. Norbeck J, Blomberg A: **Metabolic and regulatory changes associated with growth of Saccharomyces cerevisiae in 1.4 M NaCl. Evidence for osmotic induction of glycerol dissimilation via the dihydroxyacetone pathway.** *J Biol Chem* 1997, **272**(9):5544-5554.
8. Larsson K, Ansell R, Eriksson P, Adler L: **A gene encoding sn-glycerol 3-phosphate dehydrogenase (NAD+) complements an osmosensitive mutant of Saccharomyces cerevisiae.** *Mol Microbiol* 1993, **10**(5):1101-1111.
9. Lahav R, Nejidat A, Abeliovich A: **Alterations in protein synthesis and levels of heat shock 70 proteins in response to salt stress of the halotolerant yeast Rhodotorula mucilaginosa.** *Antonie Van Leeuwenhoek* 2004, **85**(4):259-269.
10. Kayingo G, Kilian S, Prior B: **Conservation and release of osmolytes by yeasts during hypo-osmotic stress.** *Arch Microbiol* 2001, **177**(1):29-35.
11. Yancey P: **Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses.** *J Exp Biol* 2005, **208**(Pt 15):2819-2830.
12. Olz R, Larsson K, Adler L, Gustafsson L: **Energy flux and osmoregulation of Saccharomyces cerevisiae grown in chemostats under NaCl stress.** *J Bacteriol* 1993, **175**(8):2205-2213.
13. Shi X, Karkut T, Chamankhah M, Alting-Mees M, Hemmingsen S, Hegedus D: **Optimal conditions for the expression of a single-chain antibody (scFv) gene in Pichia pastoris.** *Protein Expr Purif* 2003, **28**(2):321-330.
14. Blackwell JR, Horgan R: **A novel strategy for production of a highly expressed recombinant protein in an active form**. In. FEBS Lett.; 1991: 10-12.
15. Kim N, Lee G: **Response of recombinant Chinese hamster ovary cells to hyperosmotic pressure: effect of Bcl-2 overexpression.** *J Biotechnol* 2002, **95**(3):237-248.
16. Wu M, Dimopoulos G, Mantalaris A, Varley J: **The effect of hyperosmotic pressure on antibody production and gene expression in the GS-NS0 cell line.** *Biotechnol Appl Biochem* 2004, **40**(Pt 1):41-46.
17. Park S, Lee G: **Enhancement of monoclonal antibody production by immobilized hybridoma cell culture with hyperosmolar medium.** *Biotechnol Bioeng* 1995, **48**(6):699-705.

18. Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, Kleinheinz M, Sauer M, Altmann F, Gasser B: **Genome, secretome and glucose transport highlight unique features of the protein production host Pichia pastoris.** *Microb Cell Fact* 2009, **8**:29.
19. De Schutter K, Lin Y, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouzé P, Van de Peer Y, Callewaert N: **Genome sequence of the recombinant protein production host Pichia pastoris.** *Nat Biotechnol* 2009, **27**(6):561-566.
20. Gach J, Maurer M, Hahn R, Gasser B, Mattanovich D, Katinger H, Kunert R: **High level expression of a promising anti-idiotypic antibody fragment vaccine against HIV-1 in Pichia pastoris.** *J Biotechnol* 2007, **128**(4):735-746.
21. Gach J, Quendler H, Strobach S, Katinger H, Kunert R: **Structural analysis and in vivo administration of an anti-idiotypic antibody against mAb 2F5.** *Mol Immunol* 2008, **45**(4):1027-1034.
22. Mattanovich D, Gasser B, Hohenblum H, Sauer M: **Stress in recombinant protein producing yeasts.** *J Biotechnol* 2004, **113**(1-3):121-135.
23. Regenberg B, Grotkjaer T, Winther O, Fausbøll A, Akesson M, Bro C, Hansen L, Brunak S, Nielsen J: **Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in Saccharomyces cerevisiae.** *Genome Biol* 2006, **7**(11):R107.
24. Gori K, Hébraud M, Chambon C, Mortensen H, Arneborg N, Jespersen L: **Proteomic changes in Debaryomyces hansenii upon exposure to NaCl stress.** *FEMS Yeast Res* 2007, **7**(2):293-303.
25. Blomberg A: **Global changes in protein synthesis during adaptation of the yeast Saccharomyces cerevisiae to 0.7 M NaCl.** *J Bacteriol* 1995, **177**(12):3563-3572.
26. Dragosits M, Stadlmann J, Albiol J, Baumann K, Maurer M, Gasser B, Sauer M, Altmann F, Ferrer P, Mattanovich D: **The Effect of Temperature on the Proteome of Recombinant Pichia pastoris.** *J Proteome Res* 2009, **8**(3):1380-1392.
27. Zancan P, Sola-Penna M: **Trehalose and glycerol stabilize and renature yeast inorganic pyrophosphatase inactivated by very high temperatures.** *Arch Biochem Biophys* 2005, **444**(1):52-60.
28. Pascoe D, Arnott D, Papoutsakis E, Miller W, Andersen D: **Proteome analysis of antibody-producing CHO cell lines with different metabolic profiles.** *Biotechnol Bioeng* 2007, **98**(2):391-410.
29. Chang H, Jones E, Henry S: **Role of the unfolded protein response pathway in regulation of INO1 and in the sec14 bypass mechanism in Saccharomyces cerevisiae**. *Genetics* 2002, **162**(1):29-43.
30. Vaupotic T, Plemenitas A: **Differential gene expression and Hog1 interaction with osmoresponsive genes in the extremely halotolerant black yeast Hortaea werneckii.** *BMC Genomics* 2007, **8**:280.
31. Gat-Viks I, Shamir R: **Refinement and expansion of signaling pathways: the osmotic response network in yeast.** *Genome Res* 2007, **17**(3):358-367.
32. Neves M, Oliveira R, Lucas C: **Metabolic flux response to salt-induced stress in the halotolerant yeast Debaryomyces hansenii.** *Microbiology* 1997, **143 (Pt 4)**:1133-1139.
33. Carnicer M, Baumann K, Töplitz I, Sánchez-Ferrando F, Mattanovich D, Ferrer P, Albiol J: **Macromolecular and elemental composition analysis and extracellular metabolite balances of Pichia pastoris growing at different oxygen levels**.
34. Simola M, Hänninen A, Stranius S, Makarow M: **Trehalose is required for conformational repair of heat-denatured proteins in the yeast endoplasmic reticulum but not for maintenance of membrane traffic functions after severe heat stress.** *Mol Microbiol* 2000, **37**(1):42-53.

35. Zähringer H, Burgert M, Holzer H, Nwaka S: **Neutral trehalase Nth1p of Saccharomyces cerevisiae encoded by the NTH1 gene is a multiple stress responsive protein**. *FEBS Lett* 1997, **412**(3):615-620.
36. Garre E, Pérez-Torrado R, Gimeno-Alcañiz J, Matallana E: **Acid trehalase is involved in intracellular trehalose mobilization during postdiauxic growth and severe saline stress in Saccharomyces cerevisiae.** *FEMS Yeast Res* 2009, **9**(1):52-62.
37. Ramón R, Ferrer P, Valero F: **Sorbitol co-feeding reduces metabolic burden caused by the overexpression of a Rhizopus oryzae lipase in Pichia pastoris.** *J Biotechnol* 2007, **130**(1):39-46.
38. Hohmann S: **Osmotic stress signaling and osmoadaptation in yeasts.** *Microbiol Mol Biol Rev* 2002, **66**(2):300-372.
39. Welch W: **How cells respond to stress.** *Sci Am* 1993, **268**(5):56-64.
40. Marciniak S, Ron D: **Endoplasmic reticulum stress signaling in disease.** *Physiol Rev* 2006, **86**(4):1133-1149.
41. Hetz C: **The UPR as a survival factor of cancer cells: More than folding proteins?** *Leuk Res* 2009, **33**(7):880-882.
42. Fonseca S, Fukuma M, Lipson K, Nguyen L, Allen J, Oka Y, Urano F: **WFS1 is a novel component of the unfolded protein response and maintains homeostasis of the endoplasmic reticulum in pancreatic beta-cells.** *J Biol Chem* 2005, **280**(47):39609-39615.
43. Mulhern M, Madson C, Danford A, Ikesugi K, Kador P, Shinohara T: **The unfolded protein response in lens epithelial cells from galactosemic rat lenses.** *Invest Ophthalmol Vis Sci* 2006, **47**(9):3951-3959.
44. Gueta-Dahan Y, Yaniv Z, Zilinskas B, Ben-Hayyim G: **Salt and oxidative stress: similar and specific responses and their relation to salt tolerance in citrus.** *Planta* 1997, **203**(4):460-469.
45. García-Rodríguez L, Valle R, Durán A, Roncero C: **Cell integrity signaling activation in response to hyperosmotic shock in yeast.** *FEBS Lett* 2005, **579**(27):6186-6190.
46. Cyert M: **Calcineurin signaling in Saccharomyces cerevisiae: how yeast go crazy in response to stress.** *Biochem Biophys Res Commun* 2003, **311**(4):1143-1150.
47. Gonzalez N, Vázquez A, Ortiz Zuazaga H, Sen A, Olvera H, Peña de Ortiz S, Govind N: **Genome-wide expression profiling of the osmoadaptation response of Debaryomyces hansenii.** *Yeast* 2009, **26**(2):111-124.
48. Hirasawa T, Nakakura Y, Yoshikawa K, Ashitani K, Nagahisa K, Furusawa C, Katakura Y, Shimizu H, Shioya S: **Comparative analysis of transcriptional responses to saline stress in the laboratory and brewing strains of Saccharomyces cerevisiae with DNA microarray.** *Appl Microbiol Biotechnol* 2006, **70**(3):346-357.
49. Farewell A, Neidhardt F: **Effect of Temperature on In Vivo Protein Synthetic Capacity in Escherichia coli**. In., vol. 180. J Bacteriology; 1998: 4707-4710.
50. Kauffman K, Pridgen E, Doyle Fr, Dhurjati P, Robinson A: **Decreased protein expression and intermittent recoveries in BiP levels result from cellular stress during heterologous protein expression in Saccharomyces cerevisiae.** *Biotechnol Prog* 2002, **18**(5):942-950.
51. Xu P, Raden D, Doyle Fr, Robinson A: **Analysis of unfolded protein response during single-chain antibody expression in Saccaromyces cerevisiae reveals different roles for BiP and PDI in folding.** *Metab Eng* 2005, **7**(4):269-279.
52. Mori K, Kawahara T, Yoshida H, Yanagi H, Yura T: **Signalling from endoplasmic reticulum to nucleus: transcription factor with a basic-leucine zipper motif is required for the unfolded protein-response pathway.** *Genes Cells* 1996, **1**(9):803-817.

53. Graf A, Gasser B, Dragosits M, Sauer M, Leparc G, Tüchler T, Kreil D, Mattanovich D: **Novel insights into the unfolded protein response using Pichia pastoris specific DNA microarrays.** *BMC Genomics* 2008, **9**:390.
54. Nikolaou E, Agrafioti I, Stumpf M, Quinn J, Stansfield I, Brown A: **Phylogenetic diversity of stress signalling pathways in fungi.** *BMC Evol Biol* 2009, **9**:44.
55. Stasyk T, Morandell S, Bakry R, Feuerstein I, Huck C, Stecher G, Bonn G, Huber L: **Quantitative detection of phosphoproteins by combination of two-dimensional difference gel electrophoresis and phosphospecific fluorescent staining.** *Electrophoresis* 2005, **26**(14):2850-2854.
56. Gasser B, Maurer M, Gach J, Kunert R, Mattanovich D: **Engineering of Pichia pastoris for improved production of antibody fragments.** *Biotechnol Bioeng* 2006, **94**(2):353-361.
57. Ferea T, Botstein D, Brown P, Rosenzweig R: **Systematic changes in gene expression patterns following adaptive evolution in yeast.** *Proc Natl Acad Sci U S A* 1999, **96**(17):9721-9726.
58. Hohenblum H, Borth N, Mattanovich D: **Assessing viability and cell-associated product of recombinant protein producing Pichia pastoris with flow cytometry.** *J Biotechnol* 2003, **102**(3):281-290.
59. Philips J, Herskowitz I: **Osmotic balance regulates cell fusion during mating in Saccharomyces cerevisiae**. In., vol. 5. J Cell Biol; 1997: 961-974.

Figure 1 - Intracellular glycerol (A), arabitol (B) and trehalose (C) levels during carbon limited steady state cultivation of *P. pastoris* at different osmolarities (low, medium and high).

Fab 3H6 – recombinant protein secreting strain. VC – vector control strain. Error bars represent the standard error of the mean.

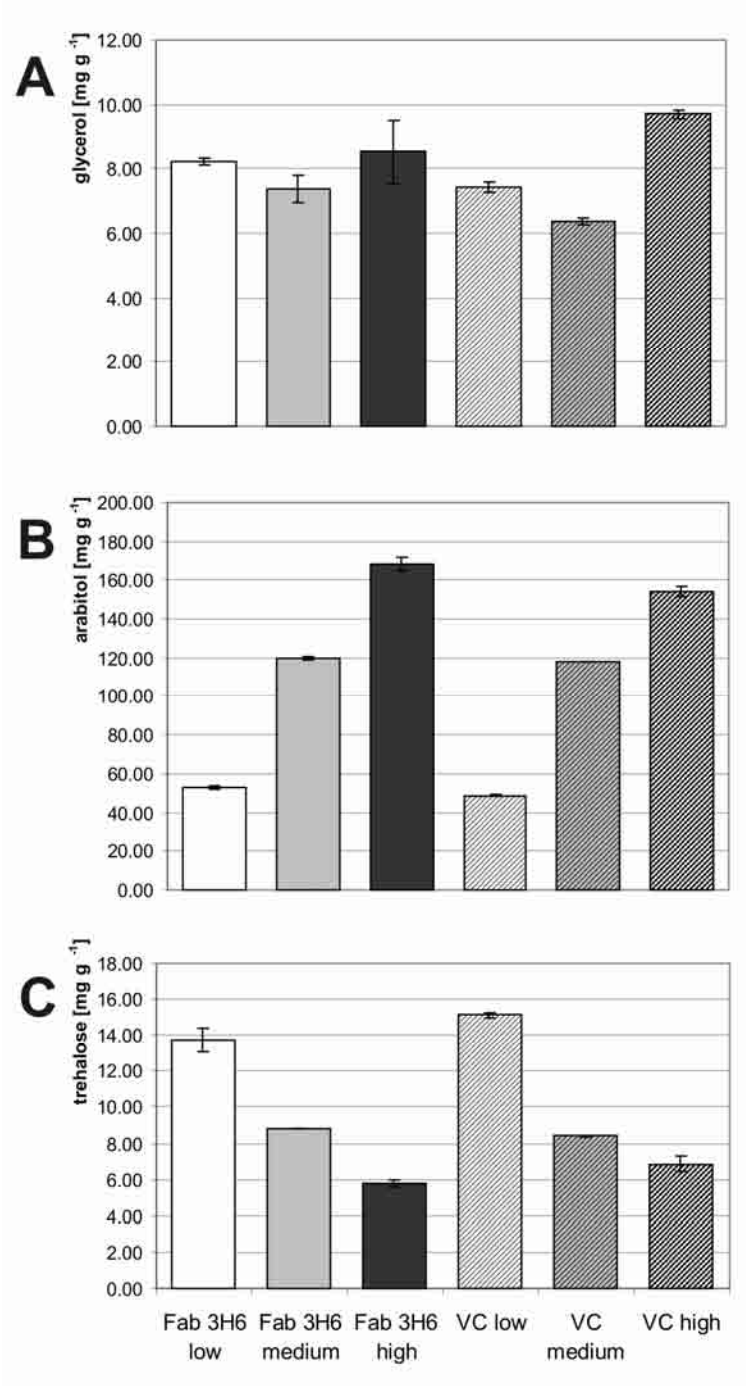


Figure 2 - Representative gel image of the 2D-DIGE experiment.

Identified spots are marked with an arrow and a number corresponding to Table 2.

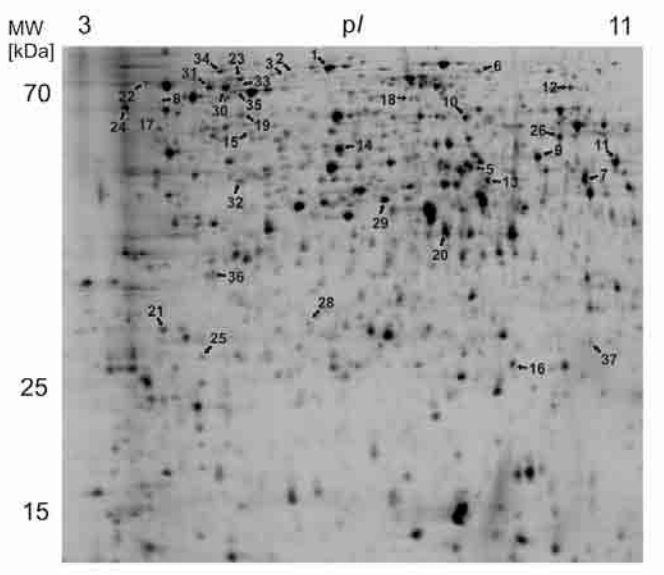


Figure 3 - Schematic representation of differences on the proteome level between low and high osmolarity cultivation of *P. pastoris* in carbon limited chemostats.

Left arrow represents changes in control strain of *P. pastoris*, and the right arrow represents changes in the Fab 3H6 secreting strain. Red upward arrow – up-regulation at high osmolartiy; Blue downward arrow – down-regulation at high osmolarity; black bar - no regulation.

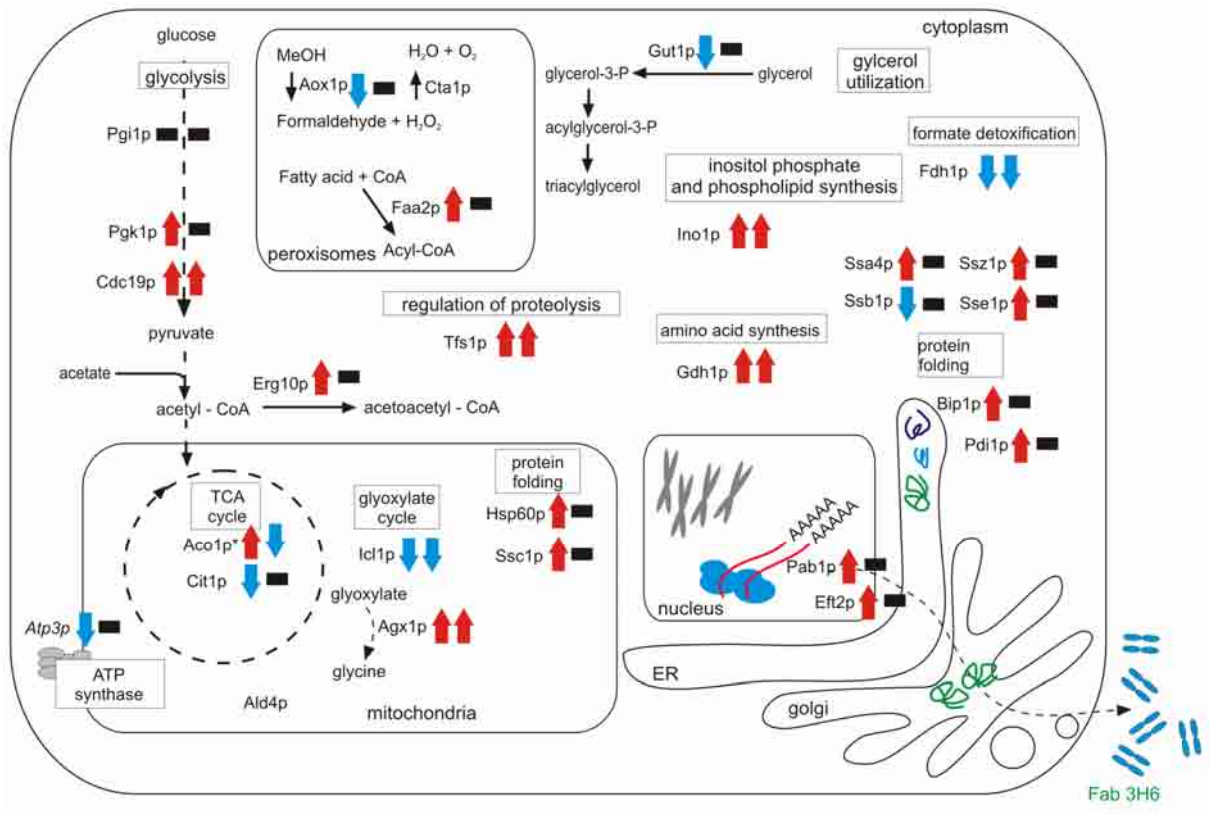


Figure 4 - Schematic presentation of changes at the transcript level of carbon limited grown *P. pastoris* X-33 when high and low osmolarity growth were compared.

Control strain (A). 3H6 Fab expressing strain (B). Blue downward arrows indicate down-regulation of genes at high osmolarity. Red upward arrows indicate up-regulated genes at high osmolarity. No arrow indicates no significant regulation of the genes or gene groups.

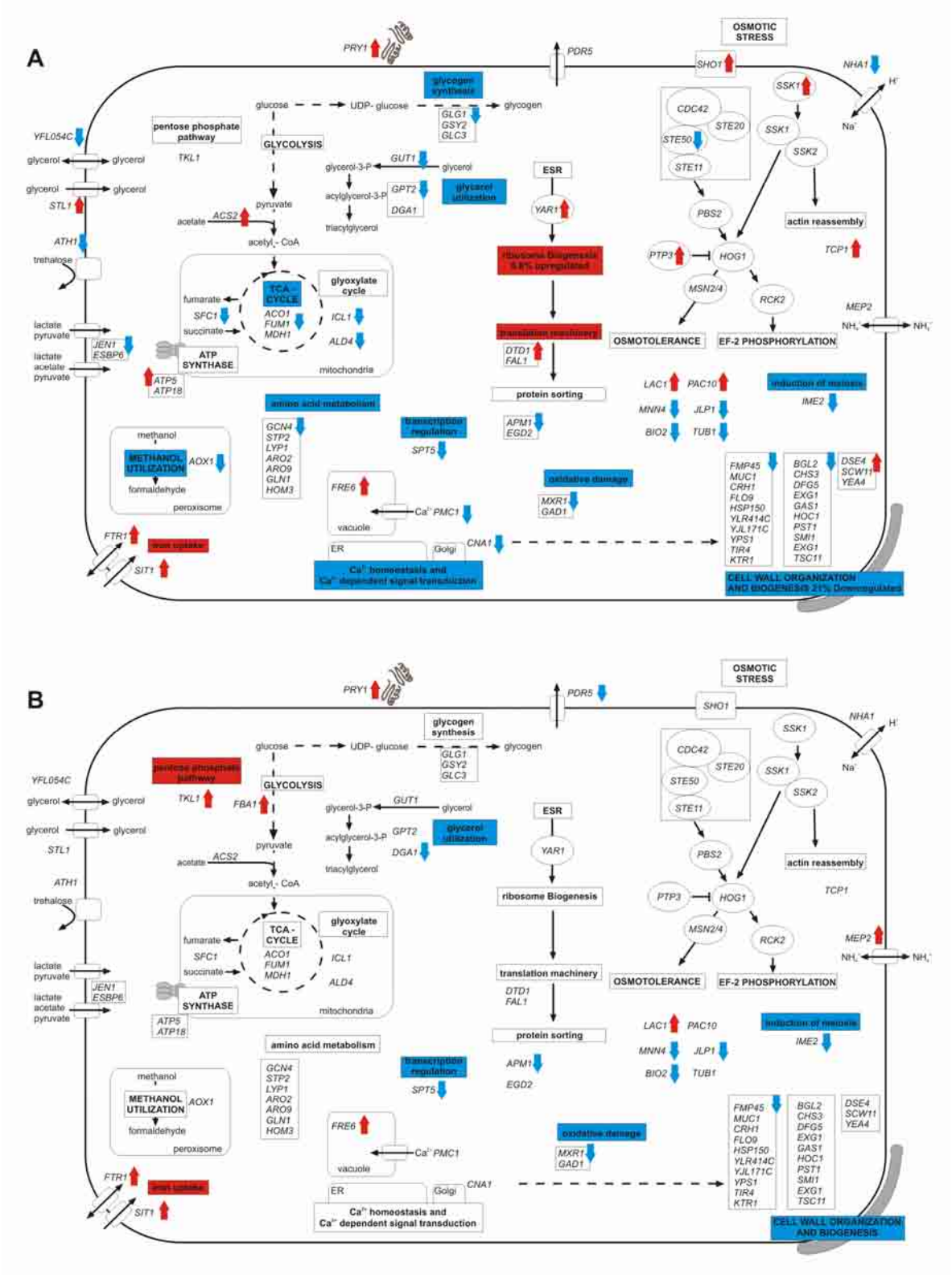


Figure 5 - Growth of *S. cerevisiae* HA232 (A) and *P. pastoris* X-33 (B) on YPD agar with different concentrations of NaCl or KCl.

YPD agar plates and plates containing 0.6M NaCl or KCl were incubated at 28°C for 4 days, whereas all other plates were incubated at 28°C for 6 days.

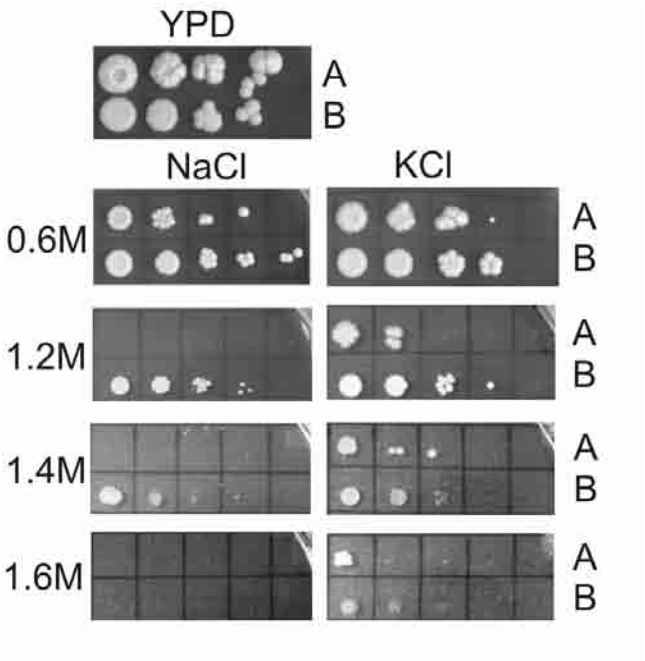


Table 1 - Characteristics of *P. pastoris* X-33 grown in carbon limited chemostat cultures at different osmolarities.

clone	Osmolarity [mOs kg^{-1}]	YDM [g L^{-1}]	Total Protein Supernatant [mg L^{-1}]	viability [%]	forward scatter	qP FabYDM$^{-1}h^{-1}$
Fab	135.2 +/- 3.3	27.8 +/- 0.2	0.47 +/- 0.04	97.2 +/- 0.4	355 +/- 60	0.039 +/- 0.004
Fab	857.3 +/- 8.5	26.8 +/- 0.4	0.44 +/- 0.05	97.9 +/- 0.6	225 +/- 30	0.042 +/- 0.002
Fab	1352 +/- 10,1	24.9 +/- 0.3	0.45 +/- 0.05	97.8 +/- 0.5	206 +/- 12	0.047 +/- 0.006
VC	149 +/- 11,4	27.7 +/- 0.3	0.38 +/- 0.01	97.5 +/- 0.3	443 +/- 95	-
VC	865.7 +/- 3,2	27.3 +/- 0.2	0.36 +/- 0.02	98.8 +/- 0.2	258 +/- 15	-
VC	1351.3 +/- 2,0	26.0 +/- 0.3	0.38 +/- 0.00	98.1 +/- 0.2	297 +/- 4	-

Fab represents the recombinant protein producing strain. VC represents the control strain. +/- represents the standard error of the mean. - not applicable.

Table 2 - Proteins that were affected by growth at different osmolarities in carbon limited chemostat cultures of *P. pastoris* X-33.

				Fab low / high		C low / high	
Spot no	**Protein**	**description**	**MW / p*I***	**Av ratio**	**1-ANOVA**	**Av ratio**	**1-ANOVA**
1	Aco1	aconitase	84.5 / 5.93	1.01	**7.60E-04**	-2.25	**3.70E-04**
2	Aco1	aconitase	84.5 / 5.93	1.54	**4.40E-05**	-1.13	5.80E-02
3	Aco1	aconitase	84.5 / 5.93	1.49	**5.10E-04**	-1.23	8.40E-02
4	Aco1	aconitase	84.5 / 5.93	1.51	**2.30E-04**	1.10	2.10E-01
5	Agx1	alanine:glyoxylate aminotransferase	31.0 / 6.36	-1.43	**8.80E-03**	-1.79	**1.00E-04**
6	Aox1	alcohol oxidase	73.8 / 6.41	-1.36	1.80E-01	2.34	**1.90E-02**
7	Atp3	F1F0 ATPase subunit	31.6 / 7.74	1.26	4.20E-01	3.28	**3.20E-06**
8	Cdc19	pyruvate kinase	49.6 / 6.24	-1.26	**9.10E-03**	-1.81	**3.40E-03**
9	Cit1	citrate synthase	51.9 / 8.32	-1.04	1.30E-01	3.23	**2.90E-04**
10	Eft2	Elongation Factor 2	93.6 / 6.29	1.14	6.30E-02	-1.93	**3.40E-04**
11	Erg10	acetyl CoA acetyltransferase	41.7 / 6.10	-1.20	5.30E-02	-1.83	**7.00E-05**
12	Faa2	long chain fatty acyl-CoA synth.	25.4 / 6.73	1.16	5.10E-01	-1.65	**3.40E-03**
13	Fdh1	formate dehydrogenase	40.3 / 6.61	1.35	**3.20E-03**	1.79	**1.30E-03**
14	Gdh1	glutamate dehydrogenase	49.3 / 5.67	-1.45	**2.20E-04**	-2.38	**3.40E-02**
15	Gut1	glycerol kinase	68.2 / 5.33	1.18	**1.20E-03**	1.31	**5.90E-03**
16	Hbn1	nitroreductase (similar to bacterial)	21.8 / 6.30	1.36	**3.00E-08**	1.12	2.10E-01
17	Hsp60	heat shock protein 60	60.2 / 5.08	-1.13	**1.40E-04**	-1.83	**1.20E-05**
18	Icl1	isocitrate lyase	61.5 / 6.15	1.56	**1.90E-05**	1.41	**2.10E-03**
19	Ino1	inositol-1-P synthase	58.4 / 5.26	-1.88	**1.30E-04**	-2.61	**1.30E-05**
20	Ino1	inositol-1-P synthase	58.4 / 5.26	-1.11	**1.60E-04**	-1.78	**1.80E-05**
21	Ino1	inositol-1-P synthase	58.4 / 5.26	1.41	**9.00E-04**	-1.06	3.60E-01
22	Kar2	BiP	74.2 / 4.79	-1.05	5.90E-01	-2.95	**4.70E-05**
23	Pab1	poly A binding protein	68.6 / 5.07	-1.03	**3.70E-05**	-1.77	**5.00E-04**
24	Pdi1	protein disulfide isomerase	57.8 / 4.63	1.08	**1.00E-02**	-1.66	**2.10E-05**
25	Pgi1	phosphoglucose isomerase	61.9 / 5.83	1.07	**7.00E-04**	1.12	1.00E-01
26	Pgk1	phosphoglycerate kinase	44.1 / 7.77	-1.03	**2.40E-02**	-1.97	**3.30E-05**
27	Pil1	Primary component of eisosomes	35.3 / 5.03	1.12	**6.20E-03**	1.35	**2.60E-02**
28	Rib3	DHBP synthase / riboflavin	22.9 / 5.09	-1.16	7.60E-02	-1.84	**7.80E-06**
29	Sor2	similar to sorbitol dehydrogenase	38.6 / 5.76	1.24	**2.50E-03**	1.34	**3.50E-02**
30	Ssa4	heat shock protein	70.3 / 5.12	1.15	**1.10E-05**	-1.20	**4.90E-02**
31	Ssb1	heat shock protein	66.5 / 5.12	1.17	**1.70E-03**	-1.02	8.70E-01
32	Ssb1	heat shock protein	66.5 / 5.12	1.03	**1.50E-05**	1.58	**1.20E-03**
33	Ssc1	mitochondrial matrix ATPase	69.7 / 5.71	-1.20	**5.40E-05**	-4.7	**3.80E-03**
34	Sse1	hsp70 family ATPase	78.7 / 5.11	-1.14	**3.70E-04**	-2.79	**5.80E-05**
35	Ssz1	hsp70 family ATPase	57.9 / 4.83	-1.21	**4.30E-05**	-1.81	**8.00E-06**
36	Tfs1	carboxypeptidase Y inhibitor	24.2 / 4.92	-1.39	**2.00E-02**	-2.00	**3.90E-04**
37	Ymr090W	unknown function	25.1 / 6.91	-1.23	**1.30E-04**	-1.41	**3.40E-05**

Analysis was performed by 2D-DIGE and subsequent LC ESI–MS/MS identification. Protein standard name (according to the SGD, www.yeastgenome.org), protein functional description, theoretical molecular weight and theoretical isoelectric point (p*I*), average expression values and 1-ANOVA values are shown. Significant ANOVA values are indicated in bold letters.

Table 3 - Number of regulated annotated genes (up- and down–regulated) in the control strain and the Fab expressing strain at different osmolarities during carbon limited chemostat cultivation.

strain	low/high up	low/high down	low/medium up	low/medium down	medium/high up	medium/high down
control	226	165	153	50	1	1
Fab	27	13	13	10	8	6
common genes	22	7	11	4	0	0

The number of genes that are similarly regulated in both strains are listed as common genes.

Additional data files

Additional file 1

Name additional_file_1

Format pdf-file

File Name **Determination of intracellular polyol and trehalose content in *P. pastoris* upon growth at different osmolarities.**

Description contains data on methodology of HPLC measurements, retention times of analytes and analyte concentrations with corresponding standard errors of the mean.

Additional file 2

Name additional_file_2

Format Excel spreadsheet

File Name **Table of peptides of interesting proteins identified by 2D-DIGE and LC-ESI-MS/MS.**

Description contains a list of all peptides assigned to the proteins identified by 2D-DIGE and LC-ESI-MS/MS and described in the manuscript (table 2). Additionally, scores and scoring schemes are indicated.

Additional file 3

Name additional_file_3

Format pdf-file

File Name **Fold-change and one-way ANOVA data for all contrasts of the 2D-DIGE experiment.**

Description contains 2D-DIGE data of all comparisons (low to medium, low to high and medium to high) of both strains. Tables contain protein master numbers, short protein names, protein descriptions, fold-changes and corresponding one-way ANOVA values.

Additional file 4

Name additional_file_14

Format pdf-file

File Name **Real-time PCR results of *P. pastoris* grown at different osmolarities**.

Description contains detailed data on real-time PCR. Primers sequences, PCR conditions as well as result diagrams are included.

Additional file 5

Name additional_file_5

Format pdf-file

File Name **Quality of microarray experiments and statistical test.**

Description contains supplemental data on microarray analysis: Signal intensity plots, correlation of intensities, standard deviations and variations of the microarray experiment. Additionally, results of Hierarchical Cluster Analysis (HCA), Gene Set Analysis (GSA) and Fisher's exact test are included.

Additional file 6

Name additional_file_6

Format Excel spreadsheet

File Name **Microarray results of interesting genes**

Description contains microarray experiment results of interesting genes, which are described and discussed in the manuscript.

Comparative analysis of antibody fragment production in diverse microbial host cells

Martin Dragosits[1§], Gianni Frascotti[2§], Lise Bernard-Granger[3§], Felícitas Vázquez[4,5§], Maria Giuliani[6§], Kristin Baumann[7§], Escarlata Rodríguez-Carmona[4,5], Jaana Tokkanen[3], Ermenegilda Parrilli[6], Marilyn G. Wiebe[3], Renate Kunert[1], Michael Maurer[1,8], Brigitte Gasser[1], Michael Sauer[1,8], Paola Branduardi[2], Tiina Pakula[3], Markku Saloheimo[3], Merja Penttilä[3], Pau Ferrer[7], Maria Luisa Tutino[6], Antonio Villaverde[4,5], Danilo Porro[2] and Diethard Mattanovich[1,8*]

1 Department of Biotechnology, University of Natural Resources and Applied Life Sciences, Vienna, Austria

2 Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milan, Italy

3 VTT Technical Research Centre, Espoo, Finland

4 Institute of Biotechnology and Biomedicine, Universitat Autònoma de Barcelona, Spain

5 CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, 08193 Barcelona Spain.

6 Department of Organic Chemistry and Biochemistry, Università di Napoli Federico II, Naples, Italy.

7 Department of Chemical Engineering, Escola Tècnica Superior d'Enginyeria, Universitat Autònoma de Barcelona, Spain

8 School of Bioengineering, University of Applied Sciences FH-Campus Wien, Austria

§ Equal contributors

To whom correspondence should be addressed. University of Natural Resources and Applied Life Sciences, Department of Biotechnology, Muthgasse 19, Vienna, Austria. Tel: +43 136006-6569. Fax: +43 13697615. E-mail: diethard.mattanovich@boku.ac.at

Running title: Comparison of Fab production in microbial hosts

Abstract

While efficient heterologous protein production has proven to be feasible in general terms, the production of complex proteins has turned out to have a low success rate. A lot of molecular, biochemical and biological information about different host cells has been progressively accumulated, but the majority of such data has been gained through a variety of non-comparable approaches, leading to a poor base for generalisation. This study therefore aimed at a comparative analysis of five different host species with proven capacities for protein expression.

Expression of a heterodimeric antibody Fab fragment was assessed in five microbial host systems, namely *Saccharomyces cerevisiae*, *Pichia pastoris*, *Trichoderma reesei*, *Escherichia coli* and *Pseudoalteromonas haloplanktis*. The comparative analysis of the specific production and the effect of cultivation temperature and oxygenation as examples for production relevant environmental parameters highlight fundamental differences but also common features of these protein production platforms. Especially *P. haloplanktis* is a novel strong potential candidate for the expression of antibody Fab fragments. Furthermore, we could show that lowering the cultivation temperature had a positive effect on recombinant antibody fragment Fab secretion in 3 out of 4 analyzed microorganisms.

Introduction

Microbial host cells, most notably *Escherichia coli* and *Saccharomyces cerevisiae*, represent the most commonly used model and host organisms for basic research and recombinant protein production. Of the few protein based biopharmaceuticals on the market, the majority are still produced in these two host organisms (10). Apart from these two work horses of biotechnological research, many different organisms are currently used or tested for the production of technically and therapeutically relevant proteins, including bacterial and fungal host cells, insect and mammalian cell lines and transgenic plants. Since post-translational modifications (PTMs) such as glycosylation (6) play a major role for product quality and suitability for human administration, higher eukaryotic expression systems are commonly applied. They can often produce human or human-like PTMs and generally achieve high product titers. Furthermore, the production of large proteins is often difficult in microbial, especially bacterial hosts (10). Nevertheless, detailed research has enabled substantial progress in producing PTMs such as glycosylation in bacterial cells (54) and humanized glycosylation patterns in the fungal host *Pichia pastoris* (18, 20, 29, 42). Large proteins such as full length antibodies have also already been successfully produced in microbial hosts (36, 43, 49). In addition to these advances and the strong molecular biology knowledge background of microbial hosts, there are also the benefits of simple growth requirements on defined mineral growth media, high growth rates and consequently short fermentation and process times. Thus there is motivation to further develop and improve microbial host systems.

In the current study, different microbial host systems for recombinant protein production were compared. A complex model protein, the Fab 3H6 (12, 26), was expressed and secreted either into the periplasm or into the culture broth in *S. cerevisiae*. *Pichia pastoris*, *Trichoderma reesei*, *E. coli* and *Pseudoalteromonas haloplanktis*. These organisms represent well

established platforms for heterologous protein production in academia and industry or highly prospective new host systems, as in the case of *P. haloplanktis* (40). The general characteristics of these organisms are summarized in Table. 1. There are large differences in the capability to express specific proteins even among yeast species, indicating the importance of identifying differences as well as common features also in closely related species. For each species a commonly used vector system for recombinant protein production, containing either an inducible or a constitutive promoter, in combination with chemostat cultivation, was applied. Generally, secretion of the recombinant product into the periplasm or culture supernatant results in relatively pure protein and facilitates downstream processing. On the other hand periplasmic secretion of recombinant polypeptides in bacteria such as *E. coli* may result in relatively low protein concentrations while filamentous fungi such as *T. reesei* secrete a vast amount of host protein into the culture supernatant (33). Codon optimization for the specific host organism may also be applied to increase the product yield (7). In the current study non-codon optimized 3H6 Fab was used to facilitate comparison of the organisms and because codon optimization would not have been applicable to all organisms.

It is often still unpredictable which host systems will be the most suitable for the production of a particular protein, thus a trial and error approach is necessary to find the best suited host cell. The results of this study confirm that there is a great variability in the production of Fab 3H6 among different host organisms. Furthermore, microorganisms encounter diverse stress conditions during fermentation processes, which may interfere with process performance (34). These environmental stresses and the physiological reactions they trigger, including changes in the protein folding/secretion machinery, are highly interrelated (16), making the continued investigation of environmental factors influencing protein expression and secretion very important. Of all the possible stresses, temperature, oxidative stress, osmolarity and pH appear particularly important. Therefore, the impact of temperature and oxygenation on recombinant protein production was analysed in steady-state using chemostat cultivation. As

106 microorganisms generally encounter long-term suboptimal conditions during production
107 processes rather than short-term perturbations, the analysis of environmental factors in steady-
108 state conditions was the method of choice. Furthermore growth rate related effects could be
109 avoided by applying chemostat cultivation (46).
110 The results obtained in the current study highlight common features and differences in the
111 response to temperature and oxygenation in different microbial species.
112
113 **Materials and Methods**
114
115 Strain construction
116
117 *Saccharomyces cerevisiae* – The 3H6 heavy and light chain (12) coding sequences were both
118 integrated on a pYX integrative expression vector. For constitutive expression the *S.*
119 *cerevisiae TPI* (triose phosphate isomerase) promoter was used and the *S. cerevisiae* α-factor
120 leader sequence served as secretion signal (4). A polyA sequence served as terminator. HIS3
121 and URA3 were used as selection markers and expression vectors were integrated into the
122 genome of *S. cerevisiae* CEN.PK strain (53).
123
124 *Pichia pastoris* – Heavy and light chain coding sequences were both inserted into a single
125 pGAPzαA expression vector. Both chains were expressed under the control of the *GAP*
126 (glyceraldehyde-3-phosphate dehydrogenase) promoter and the AOX terminator (2). The *S.*
127 *cerevisiae* α-factor secretion signal was used to target the protein into the supernatant. The
128 plasmid was linearized by restriction digest and integrated into the genomic GAP promoter
129 locus of *P. pastoris* X-33.
130

Trichoderma reesei – Two expression cassettes were constructed, where the 3H6 light and heavy chain fragments were located between the *cbh1* promoter and terminator. The c*bh1* signal sequence was fused with the antibody chain sequences. The expression cassettes were subsequently cloned to a final expression plasmid where the light and heavy chain cassettes were in tandem, followed by the acetamidase selection marker gene and a *cbh1* 3´ flanking region fragment. The fragment carrying these elements was transformed into the *T. reesei* strain RutC-30 (37). The Fab production strain studied in this work had the Fab fragment expression cassettes and the acetamidase gene integrated into the *cbh1* locus.

Escherichia coli – 3H6 Fab heavy and light chain expression cassettes were present on the pET27b vector, using T7 promotor and terminator sequences and the pelB secretion signal for secretion into the periplasm. *E. coli* HMS174(DE3) was transformed with the obtained plasmid.

Pseudoalteromonas haloplanktis – An artificial operon was constructed for 3H6 heavy and light chain recombinant expression in *P. haloplanktis* strain TAC125. The operon was cloned in the psychrophilic vector pUCRP containing a strong L-malate inducible promoter (39) and each gene was fused to a psychrophilic signal peptide encoding sequence for periplasmic secretion. 6xHis tag was fused at the C-terminus of recombinant 3H6 heavy chain.

Secretion / Periplasmic expression of the 3H6 Fab fragment was verified by Western Blotting and quantified by ELISA (see section Analytical procedures for details) in all the hosts.

157 Growth conditions

158

159 Generally, all cultivations were performed in triplicate In experiments involving more than
160 one physiological condition, i.e. changes in temperature or oxygenation, the sequence in
161 which either temperature or oxygenation was changed when more than one conditions was
162 included in the same cultivation was taken into account to reduce the effects of evolutionary
163 adaption. Samples were taken during steady state, indicated either by constant RQ and
164 biomass yield or as in the case of *T. reesei* also by monitoring the expression of 13 genes
165 (*rpl16a*, *hen6*, *hsp70*, *rps16b*, *gcn4*, *chs1*, *acs1*, *bgl2*, *bga1*, *egl1*, *vpa1*, antibody light chain,
166 antibody heavy chain, and *nth1*) using TRAC analysis (45).
167 Detailed media recipes for all model organisms of this study can be found in Supplemental
168 data 1.

169

170 *Saccharomyces cerevisiae* – Shaking flask cultivations of *S. cerevisiae* were performed in
171 mineral medium with 2% (w/v) glucose as carbon source (batch medium, Supplemental data
172 1). Chemostat cultivations were performed in a BIOSTAT B (B. Braun Biotech International
173 GmbH) bioreactor as follows: after a batch phase of approximately 24 hours (initial OD660
174 was 20), chemostat cultivation was performed on 2% (w/w) glucose and 0.05% ethanol (w/w)
175 as carbon sources under fully aerobic conditions at pH 5.0 and at a dilution rate of $D = 0.1h^{-1}$
176 and a working volume of 1.25L with an expected yeast dry mass of $10g\ L^{-1}$. Ethanol was
177 added to avoid culture oscillation (41). The total gas flow was kept constant for all
178 experiments at 1 vvm (volume gas per volume batch medium and minute) and the agitation
179 was set at 1000 rpm. Chemostat cultivation was performed at 3 different temperatures: 23, 26
180 and 30°C. For cultivation in different oxygenation conditions, chemostat medium was used,
181 with 2% (w/w) glucose, at a dilution rate of $D = 0.1h^{-1}$. No ethanol was added in the oxygen
182 experiment because no culture oscillations were observed without ethanol in the growth

183 medium. The temperature was set at 26°C and the air concentration in the inlet gas stream
184 was changed in a stepwise manner by partial replacement with an equivalent volume of
185 nitrogen. The corresponding oxygen concentrations in the inlet air were 20.97 %, 5% and 2
186 %, resulting in normoxic, oxygen limited and respiro-fermentative (hypoxic) conditions,
187 respectively. The dissolved oxygen tension (DOT) in oxygen limited condition was about 2-
188 5%, in hypoxic conditions it was zero, whereas in fully aerobic conditions it was about 80%.
189
190 *Pichia pastoris* – Shake flask cultivations of *P. pastoris* were performed in YPD medium (2%
191 (w/v) peptone, 1% (w/v) yeast extract, 2% (w/v) glucose). Bioreactor cultures were performed
192 in a 3L Bioreactor (MBR) at a working volume of 1.5L as follows. After a batch phase of
193 approximately 24 hours on mineral medium with glycerol as carbon source, chemostat
194 cultivation was performed on mineral medium with glucose as carbon source under fully
195 aerobic conditions (20% dissolved oxygen, DO), pH 5.0 and at a dilution rate of D = $0.1h^{-1}$
196 with an expected yeast dry mass of 25g L^{-1}. Chemostat cultivation was performed at 20, 25
197 and 30°C, respectively. For cultivation in different oxygenation conditions a 2L Biostat B
198 bioreactor (B. Braun Biotech International GmbH) at a working volume of 1L was used. The
199 chemostat medium (Supplemental data 1) was used, the temperature was set to 25 °C and the
200 oxygen concentrations in the inlet air were: 20.97% (normoxic condition), 10.91% (oxygen
201 limitation) and 8.39% (hypoxic condition). The total gas flow was kept constant for all
202 experiments at 1.5 vvm with a headspace overpressure of 0.2 bar, 700 rpm and 25 °C. In both
203 limited and hypoxic conditions the dissolved oxygen tension (DOT) was zero, whereas in
204 fully aerobic conditions DOT was about 45%.
205
206 *Trichoderma reesei* – *T. reesei* was grown in Biostat CT2 bioreactors (B. Braun Biotech
207 International GmbH) with a working volume of 2 L, with 0.5 vvm aeration at 800 rpm and pH
208 4.8 ± 0.1. Cultures were inoculated with 200 mL (10% final culture volume) filamentous pre-

cultures inoculated with 1 x 10^6 conidia mL^{-1} and grown for approximately 72 h at 28°C in 50 mL volumes in 250 mL flasks which were shaken at 200 rpm. Nitrogen, oxygen and carbon dioxide were monitored online with an OmniStar mass spectrometer (Pfeiffer Vacuum, Germany). Chemostats were maintained at D = 0.03 h^{-1}. Cultivations were carried out at 28, 24 and 21.5°C.

Escherichia coli – Chemostat cultivations were performed in a Biostat B bioreactor (B. Braun Biotech International GmbH) as follows. During the batch phase, the temperature was set to 33ºC and the dissolved oxygen concentration was maintained above 20% saturation by cascade automatic control of the stirrer speed between 700 and 1000 rpm and the air-flow between 1.5 and 4 L min^{-1}. pH was controlled at 7.0. After approximately 10 h of batch growth, continuous medium flow was started at a dilution rate of D = 0.1 h^{-1}. Induction was performed by addition of IPTG after the continuous process began. IPTG concentration was maintained constant at 25 μM throughout the process. Chemostat cultivations were carried out in glucose limited cultures at 37, 33 and 30 °C. Fully aerobic conditions were maintained by means of constant aeration and stirring.

Pseudoalteromonas haloplanktis – Shaking flask cultivations of *P. haloplanktis* were performed in complex TYP medium (1.6% (w/v) bacto-tryptone, 1.6% (w/v) yeast extract, 1% (w/v) NaCl pH 7.5). Fermentation was performed in a STR 3L Bioreactor (Applikon) connected to an ADI 1030 Bio Controller (Applikon) in SCHATZ mineral medium (39) containing L-leucine 0.5% (w/v), L-isoleucine 0.5% (w/v) and L-valine 1% (w/v) as carbon sources and supplemented with ampicillin 100μg mL^{-1} for plasmid selection and L-malate 0.4% (w/v) for promoter induction. After a batch phase of approximately 48 hours, chemostat cultivation was performed on same medium under fully aerobic conditions (≥30% DOT) at a

234 dilution rate of D = 0.05h^{-1} and a working volume of 1.0L. Chemostat cultivation was
235 performed at 15°C (µmax=100%).
236
237 Analytical procedures
238
239 *Biomass determination* - For biomass determination adequate sample volumes were washed in
240 demineralised water and either dried to constant weight in pre-weight beakers at 105°C until
241 constant weight (2) or collected on pre-weighed filter discs and dried at 110°C until constant
242 weight (44).
243
244 *Periplasmic protein preparation for bacterial hosts* - Bacterial pellets were resuspended in
245 1/20 of culture volume of borate buffer ($Na_2B_4O_7$ 200mM, NaCl 130mM, EDTA 5mM, pH
246 8) and incubated 18h at 4°C. The suspension was centrifuged at 8000rpm for 15min at 4°C
247 and the supernatant used for ELISA.
248
249 *Product quantification by ELISA* – A sandwich ELISA was performed as described in
250 previous studies (12, 14).
251
252 *Western Blot Analysis* – For Western Blot analysis, samples were subjected to standard
253 polyacrylamide gel electrophoresis (SDS-PAGE) under non-reducing conditions to avoid
254 reduction of disulfide-bonds and blotted onto a nitrocellulose membrane (Biorad). After
255 blocking the membrane in PBS 2% (w/v) dry milk for at least 2 hours, 3H6 Fab was detected
256 using either anti-human IgG (light chains) or anti-human Fab specific alkaline phosphatase
257 conjugate (Sigma). Blots were developed using the AP-substrate kit (Biorad).
258
259

Results

Fab 3H6 could be successfully produced in all bacterial and fungal host cells. A band of approximately 50 kDa in size was obtained in all host organisms and indicated full length Fab 3H6 (Supplemental data 2). Additionally, it can be seen that besides the correctly assembled heterodimeric Fab 3H6, free unassembled light chains of approximately 25 kDa in size were released in all host organisms, to various extents with the exception of *P. haloplanktis*. Furthermore, a small fraction of degradation products was detectable in *P. pastoris* and *E. coli* cultures. Moreover, during *E. coli* chemostat cultivation, presence of the 3H6 Fab was not only observed in the periplasmic space but also in the culture broth (Supplemental data 2).

Production at temperatures optimal for growth For comparative purposes, the standard operating conditions for each microorganism were defined as an aerobic carbon limited continuous culture with a temperature setpoint at which μ_{max} in batch culture would be 100%. However, it should be noted that chemostat cultivations were not performed at a dilution rate corresponding to μ_{max} but at a dilution rate that corresponds to $\leq 60\ \%\ \mu_{max}$ depending on the host organism. Cultivations were not performed at μ_{max} to allow subsequent cultivation at different temperatures that would not support growth at μ_{max} and to enable the analysis of the impact of growth temperature on specific production in a growth rate independent manner. Final product titers differed between different host organisms, but also specific production varied between the expression hosts (Table 2). *P. haloplanktis* and *P. pastoris* were able to produce the 3H6 Fab very effectively and showed the highest specific production, followed by *S. cerevisiae* and *E. coli*. For *T. reesei* product yields were low compared to the other 4 organisms.

285 *The effect of temperature on biomass yield and specific production* Temperatures below
286 the optimal growth value have already been reported to have a beneficial effect on
287 heterologous protein production, and specifically on the Fab 3H6 secretion in *P. pastoris* (3,
288 9, 15, 27, 48). Therefore, we investigated whether decreased temperature had an impact on
289 3H6 Fab production in the other host species as well. Continuous cultivations were performed
290 at 3 different temperatures and samples were collected after steady-state was reached. For
291 each host organism temperature setpoints corresponding to those at which μ in a batch culture
292 would be 60, 80 and 100% of μ_{max}, were chosen. The impact of temperature was analysed for
293 *S. cerevisiae, P. pastoris*, *T. reesei* and *E. coli*, whereas no data are available for *P.*
294 *haloplanktis*.
295 Interestingly biomass yield during chemostat cultivation was affected by growth temperature
296 in some of the host organisms (Table 3). *T. reesei* and *P. pastoris* had a higher biomass yield
297 at lower growth temperature, whereas *S. cerevisiae* had a lower biomass yield at the lower
298 temperature. No changes were observed for *E. coli*.
299 Growth temperature had an effect on specific production during chemostat cultivation in 3 out
300 of 4 host organisms. As can be seen in Figure 1, *E. coli* showed a 2.4-fold increase while *P.*
301 *pastoris* and *T. reesei* showed a 2-fold increase in specific production when the growth
302 temperature was decreased to that at which μ would be 80% of μ_{max} in batch culture. At the
303 temperature at which μ would be 60% of μ_{max} in batch culture, these 3 organisms showed an
304 even higher increase in specific production (3 fold for *P. pastoris* and *E. coli* and 2.5 fold for
305 *T. reesei* compared to the optimum temperature). In contrast, *S. cerevisiae* showed no change
306 of specific production with temperature.
307 Additionally an interesting effect was observed in *E. coli*. As already mentioned, during *E.*
308 *coli* cultivations product leakage into the culture broth occurred. 3H6 Fab was found both in
309 the periplasm and in the culture medium. It was observed that the amounts found in both
310 fractions varied with temperature and hence with the amount of Fab produced (data not

shown). At the lowest Fab 3H6 concentration, more protein was found in the periplasm but as the amount of Fab increased, it was mainly found in the culture broth.

As mentioned above for *P. pastoris* and *E. coli* a small amount of apparently degraded Fab was detectable on Western Blots. However, the amount of degraded Fab was not affected by temperature (data not shown).

The Effect of Oxygenation on biomass yield and specific productivity It has already been reported in a previous study that decreased oxygen supply during chemostat cultivation of *P. pastoris* can lead to increased specific productivity (2). A 2.5 fold increase in specific productivity was achieved by shifting from normoxic to hypoxic growth conditions. Based on these data, an oxygen limited fed batch strategy was developed (2). This shift was also accompanied by a 2-fold decrease in biomass yield and increased ethanol formation.

To check whether this is a general effect that also occurs in other yeasts, the same experiment was performed with *S. cerevisiae*. The data obtained showed no positive effect on specific production with decreasing oxygen availability in *S. cerevisiae* (Table 4). As for *P. pastoris* a reduction of oxygen supply resulted in a nearly 2-fold reduction of biomass yield and increased ethanol accumulation in *S. cerevisiae*, but the specific Fab production did not change in any of the three oxygenation conditions.

No data are available for *E. coli* and *T. reesei*. For *E. coli* no stable cultivations could be established at decreased oxygen concentration because of excessive foaming at oxygen-limited conditions. Hypoxic conditions could not be applied due to the fact that media such as the one used in this study, do not support anaerobic cultivations. Standard media for aerobic cultivation of *E. coli* lack the co-factors selenium, nickel and molybdenum which are necessary for the formate hydrogen lyase complex (FLC). The FLC itself is induced under anaerobic conditions and serves to remove formate, which is toxic to cells at high concentrations (50).

Discussion

A complex protein, the antibody fragment Fab 3H6 was produced in 5 microbial host organisms with different degrees of success. The results demonstrate variation in the specific production between different expression hosts and platforms (Table 2). *P. haloplanktis* showed the highest specific production, demonstrating its potential as novel high yield protein expression host. Nevertheless, further optimization is required to improve *P. haloplanktis* growth rate and biomass yield and also to reduce medium complexity. Without these optimizations elaborate effort might be necessary to perform large scale cultivations.

The methylotrophic yeast *P. pastoris*, which showed the second highest specific production, has already been used for a long time as a well suited host for heterologous protein production. High level production of antibody fragments have been reported using both the methanol based AOX as well as the glucose based GAP system (12, 13, 57).

S. cerevisiae showed rather moderate product levels, which may be partially explained by the pYX integrative vector systems, which generally result in lower gene copy number / mRNA level, than the commonly used episomal vector systems. Furthermore, *S. cerevisiae* is generally known as a non high performance secretor. Additionally, the 3H6 Fab is a glycosylated protein (12) and as *P. pastoris* and *S. cerevisiae* differ in the extent to which they glycosylate proteins (Table 1), this may also contribute to their differential secretory behaviour.

E. coli also showed low levels of heterologous Fab, similar to the amounts produced in *S. cerevisiae*, but it is well established that large proteins with an MW of >50kDa are difficult to express in bacterial host systems without extensive host/vector fine tuning. Specific production in *E. coli* might be increased by strictly controlling the ratio of heavy and light chain mRNA levels (49). In contrast, in the current study no strain engineering was applied

362 and the heavy and light chain mRNA ratio was expected to be near 1:1 as they were both
363 expressed under the control of the same promoter.
364 Interestingly *T. reesei* showed the lowest specific production although other proteins could be
365 produced at high yields in this filamentous fungus (24, 25). TRAC analysis for *T. reesei*
366 revealed that mRNA levels for both chains were very low (data not shown). Therefore, a
367 bottleneck at the transcript level might exist. Low levels of Fab secretion (< 1 mg L^{-1}) in *T.*
368 *reesei* have been reported previously. A more than a 50-fold increase in Fab production was
369 achieved by fusing the heavy chain to *T. reesei* cellobiohydrolase CBHI, a strategy commonly
370 applied for heterologous protein production in filamentous fungi (38). This fusion-strategy
371 was not applied in the present study to enable the comparison with the other host organisms,
372 but not producing a fusion protein most likely contributed to the low secretion levels of the
373 recombinant Fab in *T. reesei*.
374 Generally, codon optimization might also increase product yields. Codon optimization has
375 already been successfully applied in various host organisms to increase production of
376 heterologous proteins (5, 7, 11, 31, 52). However, it has been shown previously that codon
377 optimization had no effect on the production of Fab 3H6 in *P. pastoris* (12). Furthermore,
378 codon usage analysis revealed that there are no significant differences in the general codon
379 preferences of *P. pastoris* and *S. cerevisiae* (Supplemental data 3), indicating that for *S.*
380 *cerevisiae* no improvement can be expected by codon optimization. In contrast codon usage
381 analysis for *E. coli* revealed that many codons of the 3H6 Fab heavy and light expression
382 cassettes represent less frequent codons in this bacterial host (Supplemental data 3). A great
383 effect of codon optimization on productivity has been observed for *E. coli* (7), including our
384 own unpublished data on Fab 3H6 expression, which showed that the productivity was
385 increased by a factor of >10 when codon optimized sequences were used. With no data at
386 hand for *T. reesei* and *P. haloplanktis* it remains unclear whether codon optimization can have
387 a positive effect on Fab 3H6 secretion in these hosts.

Product formation and secretion correlates with cellular growth (19). Generally, higher growth rates may increase the specific production in all organisms due to higher translational activity. In *S. cerevisiae* and *P. pastoris*, systems which use glycolytic promoters, increased specific production may be also achieved at higher growth rate due to increased transcript levels of the heterologous protein. However, as mentioned before in the current study chemostat cultivation with μ smaller than μ_{max} has been applied to ensure equal growth rates for all experimental setups. In contrast, less product would have been expected if *T. reesei* had been grown at a higher specific growth rate because the CBHI promoter has optimal expression at D = 0.03 h^{-1}.

Reducing the cultivation temperature had a positive effect on specific production in 3 out of 4 species analysed, namely *E. coli*, *P. pastoris* and *T. reesei* but had no significant effect on specific production in *S. cerevisiae*. Increase cell viability and reduced proteases in the culture broth in lower compared to higher growth temperatures may be responsible for the increased product yield in yeasts (21, 30). However, these effects were obtained in batch and fed batch cultures and not in controlled chemostat cultures at constant growth rate. For example, a temperature controlled methanol fed-batch strategy was developed for *P. pastoris* to increase product yield by reducing cellular lysis and contamination with intracellular proteases (21). As we have reported previously, no significant proteolytic activity or changes thereof were measured in glucose limited chemostat cultures (9, 35). We observed increased specific production and increased biomass yield at decreased temperature for *T. reesei* and *P. pastoris*, whereas no such increase of biomass yield was observed in *E. coli* and *S. cerevisiae*. In *S. cerevisiae* a decrease of temperature resulted in a decrease of biomass yield, although a positive effect of a reduction of growth temperature below the so-called optimum growth temperature has already been reported in several studies for a wide variety of bacteria and yeasts (23, 27, 32, 56).

413 Reduced protein folding stress and therefore lower maintenance energy may be responsible
414 for the increase in biomass yield at low temperatures. This conclusion is in good agreement
415 with transcript data and proteome data available for *S. cerevisiae* and *P. pastoris*.
416 Environmental stress response (ESR) gene transcription is reduced and ribosome biogenesis
417 related gene transcription increased at low compared with high temperature during steady
418 state growth of *S. cerevisiae* (50). Transcriptome and proteome data for *P. pastoris* have
419 shown that transcript and protein levels for stress induced chaperones such as SSA4 were
420 reduced during chemostat cultivation at 20°C compared to 25 and 30°C, whereas no change in
421 mRNA levels for recombinant Fab 3H6 was observed (8, 15). Thus, it seems plausible that
422 physiological changes related to protein folding and energy demand lead to increased specific
423 production and / or increased biomass yield in *T. reesei*, *P. pastoris* and *E.coli*.
424 Differences in protein folding and processing between *S. cerevisiae* compared to *T. reesei* (1)
425 or *P. pastoris* (17) have been reported, which may also account for the different behaviour of
426 *S. cerevisiae* observed in this study. The behaviour of *S. cerevisiae* may also be simply
427 dependent on the low production level of Fab 3H6 resulting in no protein folding / secretion
428 bottleneck that can be relieved by decreased temperature like in the other species.
429 *E. coli* additionally responded with increased leakage of recombinant Fab into the culture
430 broth at lower temperature. This effect has already been described in previous studies (47),
431 but it remains unclear whether this effect was caused by increased recombinant protein
432 amount itself or if low temperature triggered physiological changes that enhanced product
433 leakage into the culture broth.
434 Nevertheless, this work demonstrated that producing recombinant proteins at temperatures
435 below the optimal growth temperature improves protein production in many if not all hosts,
436 although the reasons for this are not well understood and are beyond the scope of this
437 investigation.
438

Although oxygen availability and its strong effect on cellular physiology have been analysed for several species (22, 28, 44, 55), its potential impact on heterologous protein production and secretion had not been reported until recently (2). In oxygen-limited conditions metabolism is respirofermentative rather than purely respirative, as reflected in increased ethanol levels in *S. cerevisiae* and *P. pastoris* (Table 4). Increased transcript levels of the 3H6 Fab heavy and light chain during hypoxic cultivation conditions in *P. pastoris* (unpublished data) might contribute to the increased productivity as during respirofermentative growth the flux and transcript levels for the glycolytic machinery increased and may also have boosted heterologous protein mRNA levels under the control of the GAP promoter. Still, this has to be further investigated, especially as no improvement of specific production in *S. cerevisiae* occurred, although similar physiological events happen in this yeast during growth at depleted oxygenation conditions (22). As in *P. pastoris*, the flux through glycolysis increases as the amount of oxygen available for the cellular metabolism is reduced. However, in *S. cerevisiae* the mRNAs of genes involved in glycolysis were not increased in oxygen limited conditions (55). The different response of these two yeast species may highlight substantial differences in cellular control regimes, indicating massive regulation at the transcript level *in P. pastoris* and largely post-transcriptional regulation in *S. cerevisiae*. Nevertheless, it is still unclear whether recombinant protein transcript levels would be the sole explanation for increased production of Fab 3H6 in *P. pastoris*.

The presented data show that common features as well as differences exist in the well-established protein production factories used in this study. Although eukaryotic systems are thought to be better equipped for the production of complex, disulfide bonded proteins such as Fab fragments we showed that also in prokaryotic systems such as *P. haloplanktis* Fab fragments can be produced with great success without elaborate strain and vector engineering. In contrast, the filamentous fungus *T. reesei* showed only a weak performance in the secretion

of the 3H6 Fab, without application of the commonly used method of producing a fusion protein to enhance heterologous protein production. Nonetheless, to analyse the host / platform dependencies of protein production more closely, other secreted model proteins, various vector systems and strains should be included in future studies. Additionally, with affordable high throughput transcript and proteome analysis now available, proteome and transcript profiling can be applied to more fully understand the influence of environmental factors on heterologous protein secretion and to shed more light onto the differences that appeared among different host organisms.

Acknowledgements

This work has been supported by the European Science Foundation (ESF, program EuroSCOPE), the Austrian Science Fund (FWF), project no. I37, the Austrian Research Promotion Agency (Program FHplus), the Spanish program on Chemical Processes Technologies (project CTQ2007- 60347/PPQ), the Complementary Actions Plan supporting the ESF EuroSCOPE program (project BIO2005-23733-E), the integrated action HU2005-0001 of the Spanish Ministry of Science and Innovation, the Generalitat de Catalunya (2009SGR-108) and by the Academy of Finland (grant no. 112391). This work was partially supported by F.A.R. 2008, 12-1-5140000-27, to DP. AV has been granted with an ICREA ACADEMIA award (from ICREA, Catalonia, Spain).

The authors thank Polymun GmbH (Vienna, Austria) for providing the genes for the 3H6 Fab fragment. Furthermore, the authors also thank Elena Vitale, Valerio Mezzasalma, Rosaria Varlese and Hanna Kuusinen for technical contributions.

Author's contributions

MD performed cloning procedures for *P. pastoris* and *E. coli*. MD and KB performed cultivations for *P. pastoris*. FV and ER planned and performed cultivations for *E. coli*. PB performed cloning procedures for the *S. cerevisiae* studies. GF planed and performed cultivations for *S. cerevisiae*. JT constructed the *T. reesei* production strains, and LB, MGW and TP carried out the *T. reesei* cultivations and their analysis. MM assisted in media design and fermentation for *S. cerevisiae*, *P. pastoris* and *E. coli*. MG performed cloning procedures and cultivations for *P. haloplanktis*. EP assisted in media design for *P. haloplanktis*. RK developed Fab 3H6 and provided vectors, ELISA and initial data on *P. pastoris* fermentations. BG, MiS, TP, MaS, AV and PB participated in the design of the study. AV, PF, DP, MLT, MP and DM conceived of the study.

References

1. **Arvas, M., T. Pakula, K. Lanthaler, M. Saloheimo, M. Valkonen, T. Suortti, G. Robson, and M. Penttilä. 2006. Common features and interesting differences in transcriptional responses to secretion stress in the fungi Trichoderma reesei and Saccharomyces cerevisiae. BMC Genomics 7:32.**
2. **Baumann, K., M. Maurer, M. Dragosits, O. Cos, P. Ferrer, and D. Mattanovich. 2008. Hypoxic fed-batch cultivation of Pichia pastoris increases specific and volumetric productivity of recombinant proteins. Biotechnol Bioeng 100:177-83.**
3. **Blackwell, J., and R. Horgan. 1991. A novel strategy for production of a highly expressed recombinant protein in an active form. FEBS Lett 295:10-2.**
4. **Brake, A., J. Merryweather, D. Coit, U. Heberlein, F. Masiarz, G. Mullenbach, M. Urdea, P. Valenzuela, and P. Barr. 1984. Alpha-factor-directed synthesis and secretion of mature foreign proteins in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 81:4642-6.**
5. **Brat, D., E. Boles, and B. Wiedemann. 2009. Functional expression of a bacterial xylose isomerase in Saccharomyces cerevisiae. Appl Environ Microbiol 75:2304-11.**
6. **Brooks, S. 2004. Appropriate glycosylation of recombinant proteins for human use: implications of choice of expression system. Mol Biotechnol 28:241-55.**
7. **Burgess-Brown, N., S. Sharma, F. Sobott, C. Loenarz, U. Oppermann, and O. Gileadi. 2008. Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study. Protein Expr Purif 59:94-102.**
8. **Dragosits, M., J. Stadlmann, J. Albiol, K. Baumann, M. Maurer, B. Gasser, M. Sauer, F. Altmann, P. Ferrer, and D. Mattanovich. 2008. The Effect of**

Temperature on the Proteome of Recombinant Pichia pastoris. Journal of Proteome Research submitted.

9. **Dragosits, M., J. Stadlmann, J. Albiol, K. Baumann, M. Maurer, B. Gasser, M. Sauer, F. Altmann, P. Ferrer, and D. Mattanovich. 2009. The Effect of Temperature on the Proteome of Recombinant Pichia pastoris. J Proteome Res 8:1380-92.**

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

11. **Fletcher, S., M. Muto, and S. Mayfield. 2007. Optimization of recombinant protein expression in the chloroplasts of green algae. Adv Exp Med Biol 616:90-8.**

12. **Gach, J., M. Maurer, R. Hahn, B. Gasser, D. Mattanovich, H. Katinger, and R. Kunert. 2007. High level expression of a promising anti-idiotypic antibody fragment vaccine against HIV-1 in Pichia pastoris. J Biotechnol 128:735-46.**

13. **Gasser, B., and D. Mattanovich. 2007. Antibody production with yeasts and filamentous fungi: on the road to large scale? Biotechnol Lett 29:201-12.**

14. **Gasser, B., M. Maurer, J. Gach, R. Kunert, and D. Mattanovich. 2006. Engineering of Pichia pastoris for improved production of antibody fragments. Biotechnol Bioeng 94:353-61.**

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

16. **Gasser, B., M. Saloheimo, U. Rinas, M. Dragosits, E. Rodríguez-Carmona, K. Baumann, M. Giuliani, E. Parrilli, P. Branduardi, C. Lang, D. Porro, P. Ferrer, M. Tutino, D. Mattanovich, and A. Villaverde. 2008. Protein folding and**

554 **conformational stress in microbial cells producing recombinant proteins: a host**
555 **comparative overview. Microb Cell Fact 7:11.**
556 **17. Graf, A., B. Gasser, M. Dragosits, M. Sauer, G. Leparc, T. Tüchler, D. Kreil, and**
557 **D. Mattanovich. 2008. Novel insights into the unfolded protein response using**
558 **Pichia pastoris specific DNA microarrays. BMC Genomics 9:390.**
559 **18. Hamilton, S., R. Davidson, N. Sethuraman, J. Nett, Y. Jiang, S. Rios, P.**
560 **Bobrowicz, T. Stadheim, H. Li, B. Choi, D. Hopkins, H. Wischnewski, J. Roser,**
561 **T. Mitchell, R. Strawbridge, J. Hoopes, S. Wildt, and T. Gerngross. 2006.**
562 **Humanization of yeast to produce complex terminally sialylated glycoproteins.**
563 **Science 313:1441-3.**
564 **19. Hensing, M., R. Rouwenhorst, J. Heijnen, J. van Dijken, and J. Pronk. 1995.**
565 **Physiological and technological aspects of large-scale heterologous-protein**
566 **production with yeasts. Antonie Van Leeuwenhoek 67:261-79.**
567 **20. Jacobs, P., S. Geysens, W. Vervecken, R. Contreras, and N. Callewaert. 2009.**
568 **Engineering complex-type N-glycosylation in Pichia pastoris using GlycoSwitch**
569 **technology. Nat Protoc 4:58-70.**
570 **21. Jahic, M., F. Wallberg, M. Bollok, P. Garcia, and S. Enfors. 2003. Temperature**
571 **limited fed-batch technique for control of proteolysis in Pichia pastoris**
572 **bioreactor cultures. Microb Cell Fact 2:6.**
573 **22. Jouhten, P., E. Rintala, A. Huuskonen, A. Tamminen, M. Toivari, M. Wiebe, L.**
574 **Ruohonen, M. Penttilä, and H. Maaheimo. 2008. Oxygen dependence of**
575 **metabolic fluxes and energy generation of Saccharomyces cerevisiae**
576 **CEN.PK113-1A. BMC Syst Biol 2:60.**
577 **23. Kasemets, K., I. Nisamedtinov, T. Laht, K. Abner, and T. Paalme. 2007. Growth**
578 **characteristics of Saccharomyces cerevisiae S288C in changing environmental**
579 **conditions: auxo-accelerostat study. Antonie Van Leeuwenhoek 92:109-28.**

24. Kiiskinen, L., K. Kruus, M. Bailey, E. Ylösmäki, M. Siika-Aho, and M. Saloheimo. 2004. Expression of Melanocarpus albomyces laccase in Trichoderma reesei and characterization of the purified enzyme. Microbiology 150:3065-74.

25. Kontkanen, H., A. Westerholm-Parvinen, M. Saloheimo, M. Bailey, M. Rättö, I. Mattila, M. Mohsina, N. Kalkkinen, T. Nakari-Setälä, and J. Buchert. 2009. Novel Coprinopsis cinerea polyesterase that hydrolyzes cutin and suberin. Appl Environ Microbiol 75:2148-57.

26. Kunert, R., R. Weik, B. Ferko, G. Stiegler, and H. Katinger. 2002. Anti-idiotypic antibody Ab2/3H6 mimics the epitope of the neutralizing anti-HIV-1 monoclonal antibody 2F5. AIDS 16:667-8.

27. Lee, J., Y. Jeoin, D. Kim, J. Seo, and B. Hur. 2006. Eicosapentaenoic acid (EPA) biosynthetic gene cluster of Shewanella oneidensis MR-1: Cloning, heterologous expression, and effects of temperature and glucose on the production of EPA in Escherichia coli. Biotechnol. Bioprocess Eng 11:510-515.

28. Levanon, S., K. San, and G. Bennett. 2005. Effect of oxygen on the Escherichia coli ArcA and FNR regulation systems and metabolic responses. Biotechnol Bioeng 89:556-64.

29. Li, H., N. Sethuraman, T. Stadheim, D. Zha, B. Prinz, N. Ballew, P. Bobrowicz, B. Choi, W. Cook, M. Cukan, N. Houston-Cummings, R. Davidson, B. Gong, S. Hamilton, J. Hoopes, Y. Jiang, N. Kim, R. Mansfield, J. Nett, S. Rios, R. Strawbridge, S. Wildt, and T. Gerngross. 2006. Optimization of humanized IgGs in glycoengineered Pichia pastoris. Nat Biotechnol 24:210-5.

30. Li, Z., F. Xiong, Q. Lin, M. d'Anjou, A. Daugulis, D. Yang, and C. Hew. 2001. Low-temperature increases the yield of biologically active herring antifreeze protein in Pichia pastoris. Protein Expr Purif 21:438-45.

605 **31. Lim, L., H. Li, N. Cheong, B. Lee, and K. Chua. 2004. High-level expression of a**
606 **codon optimized recombinant dust mite allergen, Blo t 5, in Chinese hamster**
607 **ovary cells. Biochem Biophys Res Commun 316:991-6.**
608 **32. Margesin, R. 2009. Effect of temperature on growth parameters of psychrophilic**
609 **bacteria and yeasts. Extremophiles 13:257-62.**
610 **33. Martinez, D., R. Berka, B. Henrissat, M. Saloheimo, M. Arvas, S. Baker, J.**
611 **Chapman, O. Chertkov, P. Coutinho, D. Cullen, E. Danchin, I. Grigoriev, P.**
612 **Harris, M. Jackson, C. Kubicek, C. Han, I. Ho, L. Larrondo, A. de Leon, J.**
613 **Magnuson, S. Merino, M. Misra, B. Nelson, N. Putnam, B. Robbertse, A.**
614 **Salamov, M. Schmoll, A. Terry, N. Thayer, A. Westerholm-Parvinen, C. Schoch,**
615 **J. Yao, R. Barabote, R. Barbote, M. Nelson, C. Detter, D. Bruce, C. Kuske, G.**
616 **Xie, P. Richardson, D. Rokhsar, S. Lucas, E. Rubin, N. Dunn-Coleman, M.**
617 **Ward, and T. Brettin. 2008. Genome sequencing and analysis of the biomass-**
618 **degrading fungus Trichoderma reesei (syn. Hypocrea jecorina). Nat Biotechnol**
619 **26:553-60.**
620 **34. Mattanovich, D., B. Gasser, H. Hohenblum, and M. Sauer. 2004. Stress in**
621 **recombinant protein producing yeasts. J Biotechnol 113:121-35.**
622 **35. Mattanovich, D., A. Graf, J. Stadlmann, M. Dragosits, A. Redl, M. Maurer, M.**
623 **Kleinheinz, M. Sauer, F. Altmann, and B. Gasser. 2009. Genome, secretome and**
624 **glucose transport highlight unique features of the protein production host Pichia**
625 **pastoris. Microb Cell Fact 8:29.**
626 **36. Mazor, Y., T. Van Blarcom, B. Iverson, and G. Georgiou. 2009. Isolation of full-**
627 **length IgG antibodies from combinatorial libraries expressed in Escherichia coli.**
628 **Methods Mol Biol 525:217-39, xiv.**

37. Montenecourt , B., and D. Eveleigh. 1977. Selective screening methods for the isolation of high yielding cellulase mutants of Trichoderma reesei on various carbon sources, p. 289-301, Adv Chem Ser.

38. Nyyssönen, E., and S. Keränen. 1995. Multiple roles of the cellulase CBHI in enhancing production of fusion antibodies by the filamentous fungus Trichoderma reesei. Curr Genet 28:71-9.

39. Papa, R., V. Rippa, G. Sannia, G. Marino, and A. Duilio. 2007. An effective cold inducible expression system developed in Pseudoalteromonas haloplanktis TAC125. J Biotechnol 127:199-210.

40. Parrilli, E., A. Duilio, and M. Tutino. 2008. Heterologous protein expression in psychrophilic hosts. Springer-Verlag Berlin Heidelberg.

41. Porro, D., E. Martegani, B. Ranzi, and L. Alberghina. 1988. Oscillations in continuous cultures of budding yeast: A segregated parameter analysis. Biotechnol Bioeng 32:411-7.

42. Potgieter, T., M. Cukan, J. Drummond, N. Houston-Cummings, Y. Jiang, F. Li, H. Lynaugh, M. Mallem, T. McKelvey, T. Mitchell, A. Nylen, A. Rittenhour, T. Stadheim, D. Zha, and M. d'Anjou. 2009. Production of monoclonal antibodies by glycoengineered Pichia pastoris. J Biotechnol 139:318-25.

43. Rakestraw, J., S. Sazinsky, A. Piatesi, E. Antipov, Wittrup, and KD. 2009. Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in Saccharomyces cerevisiae, Biotechnol Bioeng.

44. Rautio, J., M. Bailey, T. Kivioja, H. Söderlund, M. Penttilä, and M. Saloheimo. 2006. Transcriptional monitoring of steady state and effects of anaerobic phases in chemostat cultures of the filamentous fungus Trichoderma reesei, p. 247, BMC Genomics.

45. Rautio, J., K. Kataja, R. Satokari, M. Penttilä, H. Söderlund, and M. Saloheimo. 2006. Rapid and multiplexed transcript analysis of microbial cultures using capillary electophoresis-detectable oligonucleotide probe pools. J Microbiol Methods 65:404-16.

46. Regenberg, B., T. Grotkjaer, O. Winther, A. Fausbøll, M. Akesson, C. Bro, L. Hansen, S. Brunak, and J. Nielsen. 2006. Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in Saccharomyces cerevisiae. Genome Biol 7:R107.

47. Rinas, U., and F. Hoffmann. Selective leakage of host-cell proteins during high-cell-density cultivation of recombinant and non-recombinant Escherichia coli. Biotechnol Prog 20:679-87.

48. Shi, X., T. Karkut, M. Chamankhah, M. Alting-Mees, S. Hemmingsen, and D. Hegedus. 2003. Optimal conditions for the expression of a single-chain antibody (scFv) gene in Pichia pastoris. Protein Expr Purif 28:321-30.

49. Simmons, L., D. Reilly, L. Klimowski, T. Raju, G. Meng, P. Sims, K. Hong, R. Shields, L. Damico, P. Rancatore, and D. Yansura. 2002. Expression of full-length immunoglobulins in Escherichia coli: rapid and efficient production of aglycosylated antibodies. J Immunol Methods 263:133-47.

50. Soini, J., K. Ukkonen, and P. Neubauer. 2008. High cell density media for Escherichia coli are generally designed for aerobic cultivations - consequences for large-scale bioprocesses and shake flask cultures. Microb Cell Fact 7:26.

51. Tai, S., P. Daran-Lapujade, M. Walsh, J. Pronk, and J. Daran. 2007. Acclimation of Saccharomyces cerevisiae to low temperature: a chemostat-based transcriptome analysis. Mol Biol Cell 18:5100-12.

679 **52. Tokuoka, M., M. Tanaka, K. Ono, S. Takagi, T. Shintani, and K. Gomi. 2008.**
680 **Codon optimization increases steady-state mRNA levels in Aspergillus oryzae**
681 **heterologous gene expression. Appl Environ Microbiol 74:6538-46.**
682 **53. van Dijken JP, J. Bauer, L. Brambilla, P. Duboc, J. Francois, C. Gancedo, M.**
683 **Giuseppin, J. Heijnen, M. Hoare, H. Lange, E. Madden, P. Niederberger, J.**
684 **Nielsen, J. Parrou, T. Petit, D. Porro, M. Reuss, van Riel N, M. Rizzi, H.**
685 **Steensma, C. Verrips, J. Vindeløv, and J. Pronk. 2000. An interlaboratory**
686 **comparison of physiological and genetic properties of four Saccharomyces**
687 **cerevisiae strains. Enzyme Microb Technol 26:706-714.**
688 **54. Wacker, M., D. Linton, P. Hitchen, M. Nita-Lazar, S. Haslam, S. North, M.**
689 **Panico, H. Morris, A. Dell, B. Wren, and M. Aebi. 2002. N-linked glycosylation in**
690 **Campylobacter jejuni and its functional transfer into E. coli. Science 298:1790-3.**
691 **55. Wiebe, M., E. Rintala, A. Tamminen, H. Simolin, L. Salusjärvi, M. Toivari, J.**
692 **Kokkonen, J. Kiuru, R. Ketola, P. Jouhten, A. Huuskonen, H. Maaheimo, L.**
693 **Ruohonen, and M. Penttilä. 2008. Central carbon metabolism of Saccharomyces**
694 **cerevisiae in anaerobic, oxygen-limited and fully aerobic steady-state conditions**
695 **and following a shift to anaerobic conditions. FEMS Yeast Res 8:140-54.**
696 **56. Wittmann, C., J. Weber, E. Betiku, J. Krömer, D. Böhm, and U. Rinas. 2007.**
697 **Response of fluxome and metabolome to temperature-induced recombinant**
698 **protein synthesis in Escherichia coli. J Biotechnol 132:375-84.**
699 **57. Zhao, H., Q. He, C. Xue, B. Sun, X. Yao, and Z. Liu. 2009. Secretory expression**
700 **of glycosylated and aglycosylated mutein of onconase from Pichia pastoris using**
701 **different secretion signals and their purification and characterization. FEMS**
702 **Yeast Res 9:591-9.**
703

704

host organism	cytoplasmic production	secretory production	periplasmic production	disulfide bond formation	glycosylation
S. cerevisiae	Yes – may result in poor product titers	Yes - generally low titers	*na*	Yes	Yes – protein often hyperglycosylated
P. pastoris	Yes	Yes – generally lower titers than for intracellulary production	*na*	Yes	Yes – smaller glycan-structures than *S. cerevisiae*
T. reesei	Yes	Yes	*na*	Yes	Yes
E. coli	Yes – often highly pure protein in the form of inclusion bodies (IBs)	Yes – but inefficient and the underlying mechanisms are poorly understood	Yes	Not for intracellulary produced proteins, but can occur in the periplasm	No
P. haloplanktis	Yes	Yes	Yes	Can occur in the periplasm	No

705

706 Table 1. Host organisms used in this study and their general protein production

707 characteristics. *na* - not applicable

708

host organism	**growth temperature** [°C]	**D**[h^{-1}]	**product titer** [mg L^{-1}]	**$Y_{P/X}$** [mg g biomass $^{-1}$]
S. cerevisiae	30	0.10	0.22 +/- 0.05	0.02 +/- 0.005
P. pastoris	30	0.10	5.4 +/- 0.2	0.21 +-/ 0.01
T. reesei	28	0.03	$3.3\ 10^{-3}$ +/- $2\ 10^{-4}$	$9.4\ 10^{-4}$ +/- $7\ 10^{-4}$
E. coli *	37	0.10	0.21 +/-0.043	0.017+/- 0.003
P. haloplanktis	15	0.05	4.14 +/- 0.16	0.93 +/- 0.05

709

710 Table 2. Final product titers and specific Fab production ($Y_{P/X}$) at optimal growth
711 temperature, grown in carbon limited chemostat cultures at constant dilution rates (D). * For
712 *E. coli* the combined (periplasmic + extracellular product) values are shown. +/- represents
713 the standard error of the mean.

714

	$Y_{x/s}$		
Host organism	**T = 100% μ_{max}**	**T = 80% μ_{max}**	**T = 60% μ_{max}**
S. cerevisiae	0.48 +/- 0.006	0.46 +/- 0.012	0.41 +/- 0.010
P. pastoris	0.52 +/- 0.01	0.53 +/- 0.01	0.55 +/- 0.01
T. reesei	0.36 +/- 0.006	0.38 +/- 0.006	0.44 +/- 0.018
E. coli	0.37 +/- 0.004	0.38 +/- 0.012	0.37 +/-0.015
P. haloplanktis	0.22 +/- 0.00	*nd*	*nd*

715

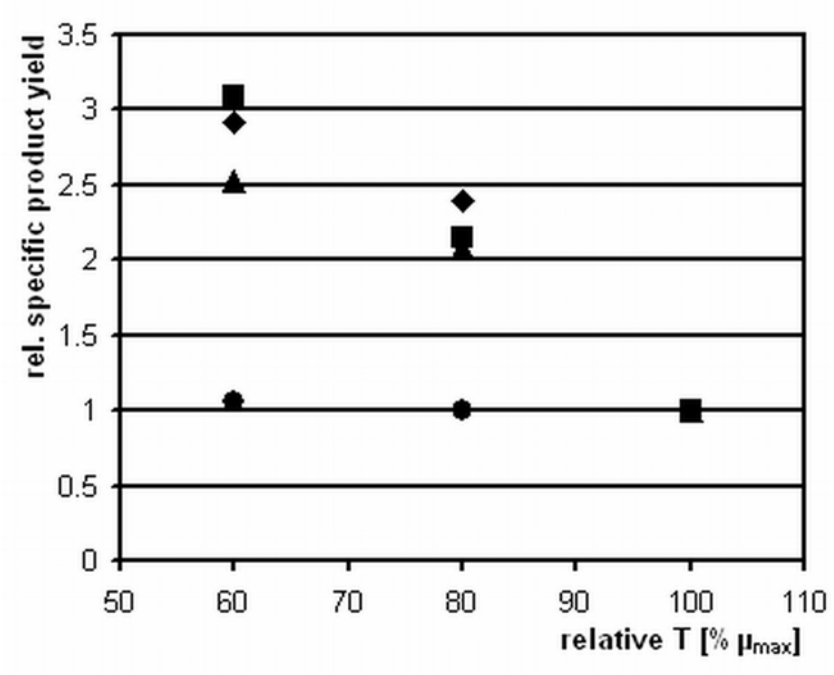
716 Table 3: Biomass yield coefficient $Y_{X/S}$ [g g^{-1} biomass per carbon source] in carbon limited
717 chemostat cultures at constant dilution rate at different growth temperatures. T represents the
718 temperature at which the indicated percentage of μ_{max} could be reached in batch culture. +/-
719 represents the standard error of the mean. *nd* no data available.

720

organism	oxygenation	$Y_{X/S}$ [g g carbon source $^{-1}$]	$Y_{P/X}$ [mg g biomass $^{-1}$]	ethanol in culture broth [g L^{-1}]
S. cerevisiae	N	0.46 +/- 0.01	0.027 +/- 0.0046	0.004 +/- 0.0004
	L	0.45 +/- 0.00	0.029 +/- 0.0046	0.008 +/- 0.0004
	H	0.31 +/- 0.01	0.023 +/- 0.0039	2.16 +/- 0.11
P. pastoris	N	0.48 +/- 0.01	0.27 +/- 0.02	0.31 +/- 0.17
	L	0.45 +/- 0.02	0.36 +/- 0.06	0.71 +/- 0.19
	H	0.25 +/- 0.01	0.69 +/- 0.02	5.72 +/- 0.14

721

722 Table 4. The effect of oxygen provision on biomass yield, Fab 3H6 yield and ethanol
723 production of recombinant *S. cerevisiae* and *P. pastoris (2)*, grown in glucose limited
724 chemostat cultures at D = $0.1h^{-1}$. N – normoxic conditions, L – limited oxygenation, H –
725 hypoxic conditions. +/- represents the standard error of the mean.



726

727 Fig.1 Relative specific production in carbon-limited chemostat cultures ($D \leq 60\% \mu_{max}$) at
728 different temperatures for *S. cerevisiae* (●), *T. reesei* (▲), *P. pastoris* (■) and *E. coli* (♦).The
729 relative temperature values indicate the temperatures at which μ_{max} in batch culture would be
730 reduced, relative to growth at the optimal temperature, by the percentages indicated.

731

Microbial Cell Factories

Research

Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*

Diethard Mattanovich*[1,2], Alexandra Graf[1,2], Johannes Stadlmann[3], Martin Dragosits[1], Andreas Redl[1,2], Michael Maurer[1,2], Martin Kleinheinz[1], Michael Sauer[1,2], Friedrich Altmann[3] and Brigitte Gasser[1]

Address: [1]Department of Biotechnology, University of Natural Resources and Applied Life Sciences, Vienna, Austria, [2]School of Bioengineering, University of Applied Sciences FH-Campus Wien, Vienna, Austria and [3]Department of Chemistry, University of Natural Resources and Applied Life Sciences, Vienna, Austria

Email: Diethard Mattanovich* - diethard.mattanovich@boku.ac.at; Alexandra Graf - alexandra.graf@boku.ac.at; Johannes Stadlmann - johannes.stadlmann@boku.ac.at; Martin Dragosits - martin.dragosits@boku.ac.at; Andreas Redl - andreas.redl@boku.ac.at; Michael Maurer - michael.maurer@boku.ac.at; Martin Kleinheinz - martin.kleinheinz@boku.ac.at; Michael Sauer - michael.sauer@fh-campuswien.ac.at; Friedrich Altmann - friedrich.altmann@boku.ac.at; Brigitte Gasser - brigitte.gasser@boku.ac.at

* Corresponding author

Published: 2 June 2009

Microbial Cell Factories 2009, **8**:29 doi:10.1186/1475-2859-8-29

Received: 13 May 2009
Accepted: 2 June 2009

This article is available from: http://www.microbialcellfactories.com/content/8/1/29

Abstract

Background: *Pichia pastoris* is widely used as a production platform for heterologous proteins and model organism for organelle proliferation. Without a published genome sequence available, strain and process development relied mainly on analogies to other, well studied yeasts like *Saccharomyces cerevisiae*.

Results: To investigate specific features of growth and protein secretion, we have sequenced the 9.4 Mb genome of the type strain DSMZ 70382 and analyzed the secretome and the sugar transporters. The computationally predicted secretome consists of 88 ORFs. When grown on glucose, only 20 proteins were actually secreted at detectable levels. These data highlight one major feature of *P. pastoris*, namely the low contamination of heterologous proteins with host cell protein, when applying glucose based expression systems. Putative sugar transporters were identified and compared to those of related yeast species. The genome comprises 2 homologs to *S. cerevisiae* low affinity transporters and 2 to high affinity transporters of other Crabtree negative yeasts. Contrary to other yeasts, *P. pastoris* possesses 4 H^+/glycerol transporters.

Conclusion: This work highlights significant advantages of using the *P. pastoris* system with glucose based expression and fermentation strategies. As only few proteins and no proteases are actually secreted on glucose, it becomes evident that cell lysis is the relevant cause of proteolytic degradation of secreted proteins. The endowment with hexose transporters, dominantly of the high affinity type, limits glucose uptake rates and thus overflow metabolism as observed in *S. cerevisiae*. The presence of 4 genes for glycerol transporters explains the high specific growth rates on this substrate and underlines the suitability of a glycerol/glucose based fermentation strategy. Furthermore, we present an open access web based genome browser http://www.pichiagenome.org.

Background

Yeasts have attracted renewed interest in the last few decades as production hosts for biopharmaceutical proteins as well as for bulk chemicals. The methylotrophic yeast *Pichia pastoris* (Guillermond) Phaff (1956) is well reputed for efficient secretion of heterologous proteins [1], and has come into focus for metabolic engineering applications recently. Especially reengineering of the N-glycosylation pathway has enabled the production of heterologous proteins with human-like N-glycan structures [2-4]. While protein production is the major application of *P. pastoris*, production of metabolites has come into research focus recently too [5,6]. Apart from these biotechnological applications, it is widely used as a model for peroxisome [7] and secretory organelle research [8]. *P. pastoris* has recently been reclassified into a new genus, *Komagataella* [9], and split into three species, *K. pastoris*, *K. phaffii*, and *K. pseudopastoris* [10]. Strains used for biotechnological applications belong to two proposed species, *K. pastoris* and *K. phaffii*. The strains GS115 and X-33 are *K. phaffii*, while the SMD series of protease deficient strains (most popular SMD1168) is classified into the type species, *K. pastoris*. Apart from these strains which have been made available by Invitrogen, research labs and industry use different other strains belonging to either of these two species, and no trend towards a superior expression level of one of the two species has been observed. In order to provide a common information basis across the different strains, we have performed this work with the type strain (DSMZ 70382) of the type species *K. pastoris*, which is the reference strain for all the available *P. pastoris* strains. In coherence with the published literature, we name all strains *P. pastoris*, which thus stands for the entire genus *Komagataella*. As other strains, DSMZ 70382 was isolated from tree exudate, in this case from the chestnut tree.

The majority of *P. pastoris* processes described so far utilize methanol as substrate and inducer for heterologous protein production. While tight gene regulation and high product titers can be achieved with this strategy, the disadvantages as large scale use of a flammable substrate, high heat production and oxygen consumption, and significant cell lysis have been reported. Apart from technological challenges in large scale fermentation, this leads to significant contamination of culture supernatants with intracellular proteins including proteases [11]. *P. pastoris* has been described to secrete some heterologous proteins like human serum albumin [12] or as recently reported glycoengineered antibodies [13] in the g L^{-1} range, while naturally secreted proteins account only for low amounts [14], which supports the easy production of highly pure proteins. However, several secreted *P. pastoris* proteins are observed as contaminants in culture supernatants, requiring elaborate product purification and analytical effort. A detailed characterization of the secretome would significantly improve production and quality control of biopharmaceuticals produced with this expression system. The secretomes of few yeasts and filamentous fungi have been analyzed experimentally. Computational analyses of yeast genomes predicted approximately 200 potentially secreted proteins [15,16]. Secretomes of filamentous fungi contain numerous enzymes for degradation of starch, cellulose, lignin and similar plant polymers [17-19]. However, these predictions suffer from some limitations. As certain targeting sequences are not recognized, the predictions may contain proteins which are retained in cellular organelles. Most cell wall associated proteins can be predicted, but due to the fluctuating nature of the cell wall during growth and budding a fraction of these may be released from the cell wall structure and add to the secretome. Finally the actual composition of the secretome will depend on growth conditions and the actual expression of the genes encoding potentially secreted proteins. Therefore the extracellular proteome of *P. pastoris* was analyzed here and compared to the predicted secretome.

Substrate uptake kinetics determine growth kinetics and the characteristics of biotechnological processes. *P. pastoris* is described as a Crabtree-negative yeast, featuring respiratory metabolism under glucose surplus [20]. A major reason for the easy growth to high biomass concentrations is assumed in the endowment with hexose transporters and their features. We report here the determination and analysis of the *P. pastoris* draft genome sequence and its application in correlating *in silico* and mass spectrometric analysis of the extracellular proteome. Furthermore, a comparative analysis of hexose transporters allows drawing conclusions towards glucose uptake kinetics, a major determinant of growth and bioprocess characteristics in relation to substrate supply. Additionally, a web based database with search functions and annotation data for analysis of the genome sequence is reported.

Results

Sequencing

The genome of *P. pastoris* was sequenced using two next generation sequencing methods. First a Roche GS-FLX run was used to take advantage of the longer reads (400 nts) of this method, which was then complemented by a paired end run with the short read method of Illumina Genome Analyzer (36 nts) to improve the quality of the sequence. The combined result of both methods was a draft genome of 326 assembled contigs of which 93 were larger than 10 kb and 60 between 1 and 10 kb. The longest contig comprised 419,475 nts and the shortest 128 nts. 125 of the 326 contigs could be aggregated into 38 supercontigs. Overall 9,405,451 bases were sequenced with a coverage of 22× with Roche GS-FLX and 60× with Illu-

mina GA. Key statistical data of the draft genome are presented in table 1.

Gene prediction

We initially predicted 7,935 open reading frames using two different gene finders. Manual curation reduced this number to 5,450 ORFs. The eukaryotic gene finder Augustus has been pre-trained on a number of datasets including various yeast species. Of these, *Candida guilliermondii*, *Debariomyces hansenii* and *Pichia stipitis* were selected for their relatively close relation to *P. pastoris* (based on sequence similarity), and *Saccharomyces cerevisiae* as a reference yeast species with the best sequence annotation. In addition the prokaryotic gene finder Glimmer3 was applied since many eukaryotic gene finders overpredict intron containing genes. As yeast genomes are generally compact a large amount of intron containing genes was not expected. All putative ORFs < 100 nts or comprising a starting codon other than ATG were excluded from the set except for genes on contig borders. 194 of the predicted genes are truncated because they crossed contig borders. Ribosomal RNAs were annotated by homology to *S. cerevisiae* rRNAs. Contrary to *S. cerevisiae*, the 5S rRNA is not part of the cluster containing 18S, 26S and 5.8S rRNA but spread across the genome. 149 transfer RNAs were identified using tRNA Scan, which is lower than the average number of tRNAs identified in other yeasts (216 on average).

Table 1: Genome statistics overview

Sequencing Data:	
Total DNA bases after Roche GS FLX	9,408,251
Average coverage Roche GS FLX	22
Total DNA bases after Illumina GA	9,405,451
Average coverage Illumina GA	60
Number of reads Roche FLX	562,515
Number of reads Illumina GA	15,761,520
Number of contigs	326
Contigs > 1 kbp	153
Largest contig	419,475
Smallest contig	128
Average contig size	28,906
GC content	41.34%
Gene Prediction Data:	
Predicted ORFs	7,935
Manually curated number of ORFs	5,450
Thereof ORFs with introns	741
Truncated ORFs	194
ORFs with annotation	4,257
GC content coding regions	41.90%
RNA Prediction Data:	
tRNA genes	149
5S rRNA	14
rRNA cluster (18S, 26S, 5.8S rRNA)	1

Functional Annotation

Functional annotation was performed computationally with a reciprocal best hit (RBH) strategy, using BLAST [21] searches against a selected dataset of the subphylum *Saccharomycotina* to which *P. pastoris* also belongs, and the Uniprot database. All *P. pastoris* genes and proteins that were publicly available at the NCBI (National Center for Biotechnology Information) were manually compared against our predictions. The native genes and proteins were present in our set. The average identity between these genes deposited in NCBI and their homologs in the present genome sequence was 95%. For all proteins that were predicted to be secreted and all others that are discussed here the functional annotation was manually curated. The distribution in GO functional terms of all functionally annotated ORFs was compared to *S. cerevisiae* (figure 1). The distribution is rather similar with differences observed mainly in the groups organelle organization, protein modification, lipid, amino acid and cofactor metabolism.

Secretome

To validate the secretome prediction pipeline (see Materials and Methods) used for *P. pastoris*, it was applied to the *S. cerevisiae* proteome beforehand. The majority of proteins which were described to be extracellular in the Saccharomyces genome database SGD [22] were found in the secreted dataset, for the rest a GPI-anchor signal was predicted. Due to the good performance of the prediction pipeline with *S. cerevisiae* and the successful application of similar methods for *K. lactis* [15] and *C. albicans* [16] respectively, a high accuracy for the secretome predictions was expected for *P. pastoris* as well. The predicted secretome of *P. pastoris* comprises 88 putative proteins of which 55 could be functionally annotated. Additionally, 172 ORFs were predicted to encode proteins entering the general secretion pathway but being localized in different cellular compartments (for the complete list see additional file 1). Obviously the secretome prediction cannot easily discriminate between ER/Golgi localized and secreted proteins, as the chaperone Kar2 and protein disulfide isomerise (Pdi1) appear among the predictions. Therefore the experimental determination of the extracellular proteins is essential for an assessment.

To identify the extracellular secretome of *P. pastoris*, the strain DSMZ 70382 was grown in chemostat culture on glucose as limiting carbon source, reaching 26.4 ± 0.1 g L^{-1} dry biomass (YDM). The supernatants contained 407 mg L^{-1} total protein. Analysis by SDS-PAGE indicated that approximately 15 distinct protein bands, ranging from 12 kDa to 170 kDa, were present in the culture supernatant (figure 2a). On 2D gels, 28 protein spots were visible at higher abundance, at least 7 thereof being obviously isoforms of other protein spots with identical MW but differ-

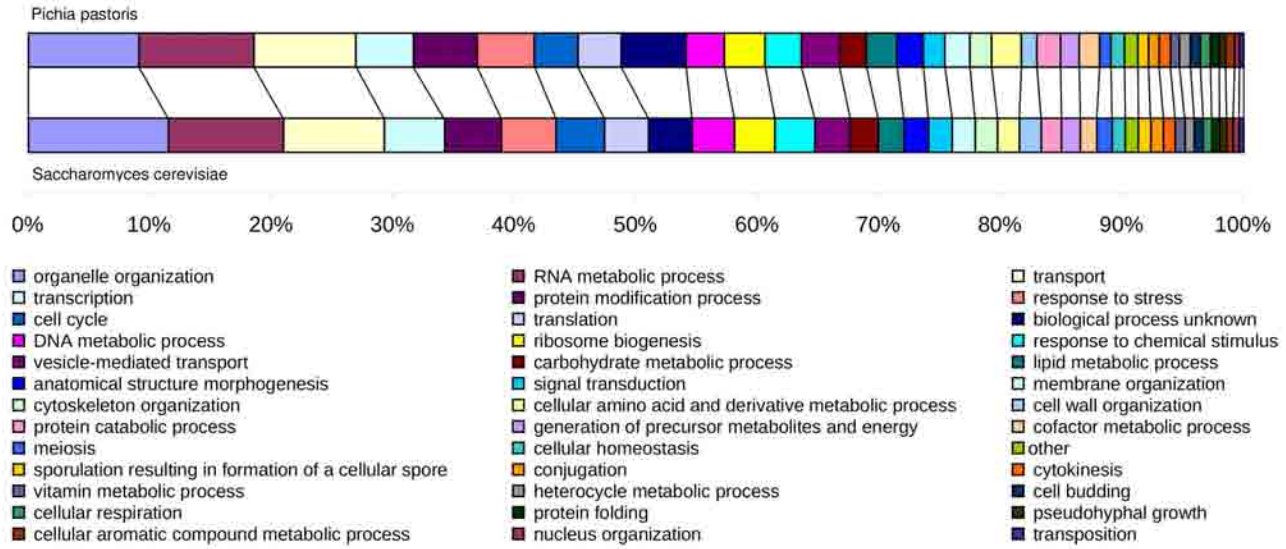


Figure 1
Categorization of the *P. pastoris* annotated genome compared to *S. cerevisiae*. The GO functional groups are displayed based on their relative representation with annotated ORFs.

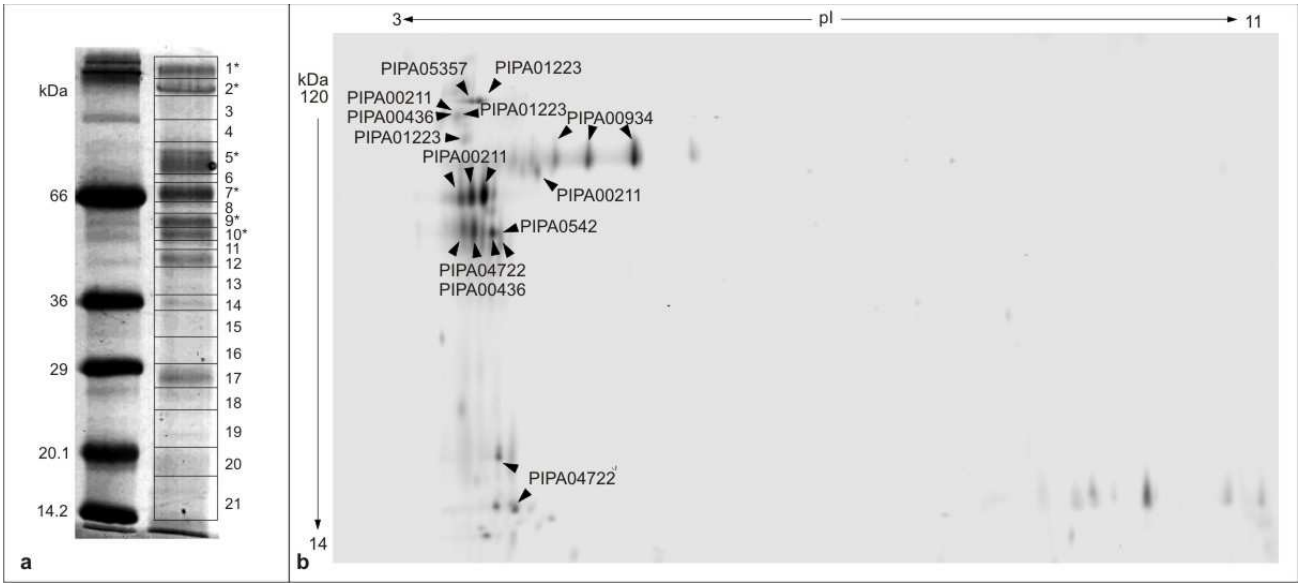


Figure 2
Secretome of *P. pastoris*. (a) SDS-polyacrylamide gel. Left lane: molecular weight marker, right lane: supernatant of *P. pastoris* chemostat culture. Boxes indicate the gel slices used for LC-MS protein identification. Bands corresponding to glycoproteins are marked with an asterisk. (b) 2D electrophoresis gel of *P. pastoris* culture supernatants. Proteins identified by LC-MS are indicated.

Table 2: Secreted proteins of *P. pastoris*

PIPA ID	Predicted function	theoretical pI/MW [kDa]	Predicted N-glycosylation sites	Predicted localization
PIPA00211	Covalently-bound cell wall protein of unknown function	5.01/45.73	1	secreted
PIPA00246	hypothetical fungal hexokinase	5.98/24.92	1	no SP
PIPA00436	Cell wall protein related to glucanases	4.83/36.07	0	secreted
PIPA00545	Cell wall protein related to glucanases	4.33/45.02	2	secreted
PIPA00748	O-glycosylated protein required for cell wall stability	4.22/31.86	1	secreted
PIPA00934	SCP-domain family protein, unknown function, extracellular	5.55/31.72	0	secreted
PIPA00956	60S ribosomal protein L18A	9.92/21.82	1	no SP
PIPA01008	GAS1; Beta-1,3-glucanosyltransferase	3.98/57.20	4	secreted
PIPA01010	GAS1; Beta-1,3-glucanosyltransferase	3.99/58.37	5	secreted
PIPA01223	potential cell wall glucanase	4.34/49.39	0	secreted
PIPA01958	Endo-beta-1,3-glucanase	4.03/33.76	1	secreted
PIPA02332	no similarity found	6.01/23.64	2	no SP
PIPA02510	Glyceraldehyde-3-phosphate dehydrogenase	6.24/35.74	1	no clear SP
PIPA02524	glucan 1,3-beta-glucosidase similar to *S. cerevisiae* EXG1 (YLR300W)	4.51/46.22	1	secreted
PIPA02544	aldehyde dehydrogenase, Adh2p [*S. cerevisiae*]	6.00/36.86	0	no SP
PIPA03955	endo-1,3-beta-glucanase [*P. stipitis* CBS 6054], Dse4p [*S. cerevisiae*]	4.70/109.45	5	secreted
PIPA04722	Cell wall protein with similarity to glucanases	5.18/32.95	0	secreted
PIPA05357	no similarity found	4.25/66.46	1	no SP, 2 TM
PIPA05673	YLR286Cp-like protein [*S. cerevisiae*], endochitinase	4.05/71.87	1	no clear SP
PIPA05771	Chitin deacetylase, Cda2p [*S. cerevisiae*]	5.25/34.66	2	secreted, lower probability

List of identified secreted proteins, with theoretical pI and theoretical MW, and information on the predicted localization (SP = signal peptide, TM = transmembrane domain).

ent pI (figure 2b). Almost all highly abundant proteins ran at low pI values between 3 and 5.5. As the cellular viability was 99% throughout the cultivation, and total DNA content of the supernatants was 1.12 ± 0.03 μg mL^{-1}, a maximum of 1% lysed cells was estimated, accounting for maximally 10% of total protein in the supernatant. Therefore, the potential contamination by intracellular protein was assumed to be minor. A 1D SDS PAGE gel was cut into 21 slices and analyzed by LC-ESI-MS/MS. Detailed data on protein identification are found in additional file 2. Twenty different proteins were identified (table 2), 12 of which appeared in more than one gel slice (additional file 2). Nine proteins ran at higher molecular weight than predicted from the sequence. Eight out of these proteins contained potential N-glycosylation sites (table 2 and additional file 2) and corresponded to detected glycoproteins (figure 2a). Apparently 6 of these proteins were subject to proteolysis. However, the proteolytic activity in the supernatants was very low (equivalent to 11 ± 0.9 ng mL^{-1} trypsin), and in contrast to other yeast secretomes, no protein with putative proteolytic activity was identified. Fourteen of the proteins identified by homology are obviously secreted or cell wall bound, 6 of them with homology to glucanases. The other proteins with extracellular localization comprise 7 cell wall modifying enzymes and 1 secreted protein of unknown function. Four proteins are homologous to intracellular proteins (including glyceraldehyde phosphate dehydrogenase which has been described to be also located at the cell wall in *S. cerevisiae* [23], and for 2 no similarity was found. The putative intracellular proteins mainly comprise glycolytic enzymes and ribosomal proteins which are highly abundant on glucose [24]. A comparison of predicted to identified secretome reveals a good correlation of prediction, putative function, and experimentally determined localization (table 2). All proteins homologous to intracellular proteins were predicted to be intracellular, and only for 2 of the 14 putatively secreted proteins the prediction was unclear or slightly below threshold.

Hexose transporters

Fourteen putative sugar transporters all belonging to the major facilitator superfamily (MFS) were identified by sequence similarity. All *P. pastoris* sugar transporters feature the classical 12 transmembrane domains, and contain the PESP motif and at least one of the two sugar transporter signature sequences. Contrary to *S. cerevisiae*, which comprises 20 isogenes for low and high affinity hexose transport, only two putative transporters with sequence similarity to *S. cerevisiae* transporters are present in the *P. pastoris* genome. While PIPA00236 possesses more than 60% identity to *S. cerevisiae* HXT-family proteins, and the low-affinity transporters of *Kluyveromyces lactis* Rag1 [25] and *Hansenula polymorpha* Hxt1 [26] on the amino acid level, PIPA08653 shows only low similarity (max. 37% identity/58% positives) to these proteins as well as to other *P. pastoris* sugar transporters. Although all 5 conserved amino acids that have been postulated to be required for high affinity transporters in *S. cerevisiae* Hxt2 [27] are present also in the respective translated protein sequence of *P. pastoris* gene PIPA00236, disruption of the gene led to impaired growth on high concentrations of glucose (2%). Disruption of PIPA08653 did not show a distinct growth phenotype. This indicates that PIPA00236 encodes the major low affinity glucose transporter in *P. pastoris*.

For high affinity transport, two *P. pastoris* proteins (PIPA02561 and PIPA00372) with high sequence similarity (>65% identity) to *K. lactis* high affinity glucose transporter Hgt1 were identified (see figure 3). The potential transporter-like hexose sensor is encoded by PIPA01691, and lacks the C-terminal "glucose sensor domain" as do the respective orthologous sensors in *H. polymorpha* (Hxt1) and *Candida albicans* [26]. Additionally a gene with similarity to quinate permease of *P. stipitis* and filamentous fungi was identified, which has putative orthologs in many other yeast species, but is missing in *S. cerevisiae*. According to Barnett et al. [28] *P. pastoris* cannot utilize quinate as a carbon source, although some of the genes required for the utilization of quinate are part of the shikimate pathway leading to the production of aromatic amino acids, and are present as part of the pentafunctional AROM protein. However, regulatory proteins of the quinate pathway are missing in the genome of *P. pastoris*. Interestingly, *P. pastoris* possesses four transporters that are highly similar to putative glycerol transporters from *K. lactis* (KLLA0A03223g) and *Yarrowia lipolytica* (YALI0F06776g), and weakly similar to the *S. cerevisiae* glycerol transporter Slt1. Sequence similarities of the proteins discussed above to their respective orthologs in *S. cerevisiae*, *P. stipitis*, *H. polymorpha*, *K. lactis*, and *Emericella nidulans* are illustrated in figure 3.

Database, genome browser

To make the genomic data accessible it was loaded into a relational database. For visualization a genome browser was installed on a web server and connected to the database.

The genome browser of *P. pastoris* is publicly available at http://www.pichiagenome.org [29].

The draft genome sequence data are deposited at EMBL-EBI, accession number CABH01000001 – CABH01000326.

Discussion

The predicted size of the haploid genome of *P. pastoris* [30] was confirmed here to comprise 9.4 Mb, which is

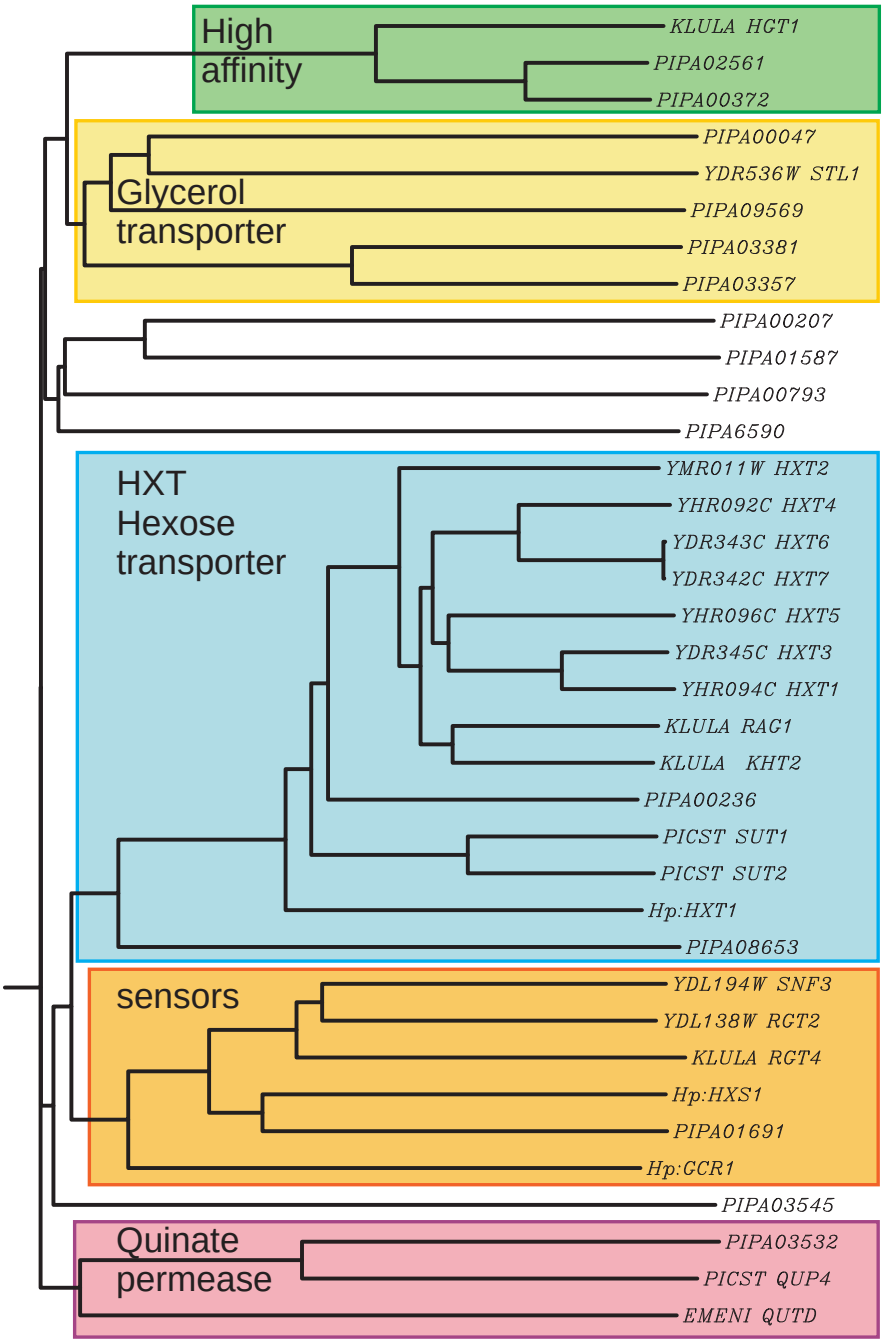


Figure 3
Branch length dendrogram of sugar transporters and related proteins of different yeasts. Putative hexose transporters and sensors and related proteins were aligned with ClustalW, and clusters of functional categories are highlighted. High affinity = high affinity glucose transporters; glycerol transporters = H^+/glycerol symporter; HXT = low affinity *S. cerevisiae* hexose transporter family; sensors = transporter-like glucose sensors; quinate permease = homologs to fungal quinate permeases. ORF IDs relate to: PIPA = *P. pastoris*; Ynnnnnn = *S. cerevisiae*; KLULA = *K. lactis*; PICST = *P. stipitis*; Hp = *H. polymorpha*; EMENI = *Emericella nidulans*. ORFs not highlighted are homologous to other substrate transporters with sequence similarity to hexose transporters.

smaller than the genomes of other yeasts, spanning from 10–20 Mb [31]. Nevertheless the number of functionally annotated genes is comparable to other yeasts, which can be attributed to the fact that *P. pastoris* contains fewer genome redundancies compared e.g. to S. *cerevisiae* and *D. hansenii*, which have undergone genome duplications followed by partial genome losses during evolution [32]. While *P. pastoris* contains specific subclasses of genes for methanol metabolism and peroxisome synthesis, structure and degradation which are present only in methylotrophic yeasts, most metabolic enzymes are present only in single copies, and the number of secreted proteins is low. To verify the quality of gene prediction, all 173 *P. pastoris* genes and 245 proteins currently deposited in NCBI were BLAST searched among the predicted gene list. All of the *P. pastoris* specific genes were present, indicating a high quality of gene prediction.

The secretomes of *K. lactis* and *C. albicans* have been predicted computationally [15,16], yielding 178 ORFs of *K. lactis* and 283 of *C. albicans*. The *C. albicans* secretome apparently is more complex and contains numerous lipases, proteases and agglutinin-like proteins, while both for *K. lactis* and *P. pastoris* only few enzymes apart from glucanases and chitin modifying enzymes appear. As *P. pastoris* utilizes only few carbon sources [28] it appears obvious that neither proteolytic, lipolytic or saccharolytic activities are secreted for substrate utilization. Yeast glucanases and chitinases are required for cell wall plasticity during cell growth and division [33]. While these enzymes are commonly regarded to be cell wall associated, it is plausible that they reach the culture supernatant during cell wall remodelling, indicating that a distinct border cannot be drawn between cell wall and the exterior space.

Fourteen of the 20 proteins identified in the culture supernatant of *P. pastoris* were homologous to proteins implicated in cell wall or extracellular functions. No other secretory enzyme homologs were identified, further indicating that cell wall associated proteins are the essential constitutents of the secretome of glucose grown *P. pastoris*. The computationally predicted secretome contains all secreted proteins plus mainly soluble cellular proteins containing a signal peptide but no transmembrane domains. Thus these predictions obviously overestimate secretory proteomes (figure 4). The culture supernatant of *K. lactis* contained significantly more (82) proteins [15] of which 34 were predicted to be secreted or cell wall bound, and the rest were assumed to be localized either to the ER or the cytosol. The latter group of proteins indicates a significant release of intracellular proteins in this study, probably by cell lysis due to the culture conditions.

The low concentration, together with the small number of actually secreted proteins from *P. pastoris* highlights a major advantage of this protein production system, as secreted products are much less contaminated with host cell protein. Jahic et al. [34] have shown that host cell protein released from *P. pastoris* grown on methanol mainly derives from cell lysis, which occurs to a much lower extent upon growth on glucose. Combined with the fact that strong promoters for use on glucose are available [34,35], these data provide convincing arguments for a reconsideration of methanol based protein production with *P. pastoris*. The toxicity of methanol and several of its metabolites is the main reason for cell lysis and consequently also protease leakage to the culture supernatant. Additionally other host cell proteins are released, leading to significant contamination of protein products. A common approach to reduce product proteolysis is the knock out of protease genes. However, multiple protease knockout strains tend to be growth retarded, so that it appears reasonable to employ a production strategy based on glucose media which avoids the detrimental effects of methanol at all. Detailed knowledge of the secreted host cell proteins, as presented here, can have a strong positive effect on product purification and quality control, as specific assays can be developed. Additionally a knock out of major secreted proteins can reduce the host cell protein load significantly [36].

Substrate uptake kinetics determines growth kinetics and the characteristics of biotechnological processes. The fermentative (Crabtree-positive) yeast *S. cerevisiae* consumes glucose at high rates when supplied with high concentrations. This exceptionally high glucose uptake rate is attributed to high abundance of hexose transporters, encoded by more than 10 isogenes [37]. Respiratory (Crabtree-negative) yeasts limit glucose uptake, as they contain few hexose transporter genes, encoding energy dependent symporters with high affinity to glucose [38]. The endowment of *P. pastoris* with hexose transporters is in good accordance to other respiratory yeasts such as *K. lactis*, *H. polymorpha* and *P. stipitis*, all having a reduced number of hexose transporters in comparison to *S. cerevisiae*. Moreover, Crabtree-negative yeasts usually exhibit K_m values in the micromolar range for glucose [37], due to their very high-affinity transporters such as *K. lactis* Hgt1, which is an ortholog of *P. pastoris* PIPA02561 and PIPA00372. While K_m values for *P. pastoris* specific transporters remain to be determined in future, conclusions to glucose uptake behavior can be drawn. Accordingly, specific glucose uptake rate is limited to q_{Smax} = 0.35 g g^{-1} YDM h^{-1} (at growth rates near μ_{max} = 0.193 h^{-1}) in *P. pastoris* chemostat cultivations [39], in comparison to q_{Smax} = 2.88 g g^{-1} YDM h^{-1} in fully aerobic *S. cerevisiae* [40]. The limited glucose uptake prevents Crabtree-negative yeasts such as *P. pastoris* from extensive overflow metabolism, which leads to the aerobic formation of ethanol and a reduced biomass yield at high external glucose concentrations in *S. cerevisiae*.

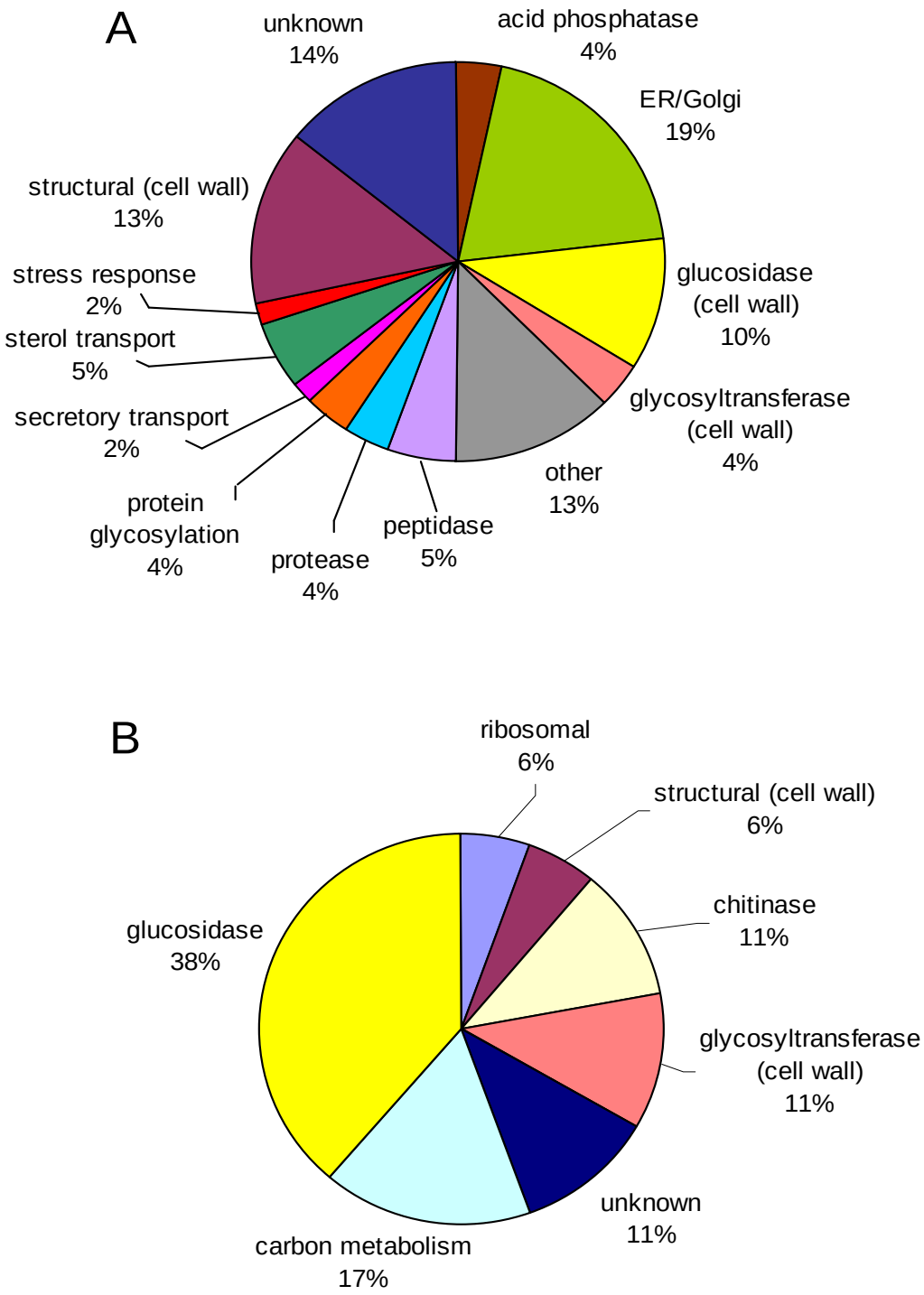


Figure 4
Categorization of *P. pastoris* secretome. (a) predicted and (b) detected secretome based on GO terms. Proteins without *S. cerevisiae* homologs are classified as "unknown".

This difference is also reflected in the very high biomass concentrations (more than 100 g l^{-1}) that can be achieved in *P. pastoris* cultivations. For heterologous protein production, aerobic ethanol formation is a substantial problem, because it lowers the yield of the desired product due to a lower biomass concentration.

Interestingly, *P. pastoris* contains four genes encoding putative H^+/glycerol symporters, contrary to all other sequenced yeasts up-to-date. Consequently, the maximum glycerol uptake rate of *P. pastoris* is $q_{Glycerol\_max}$ = 0.37 g g^{-1} YDM h^{-1}. This is substantially higher than the uptake rates reported for *S. cerevisiae* ($q_{Glycerol\_max}$ = 0.046 g g^{-1} YDM h^{-1}) and many other yeast species [41]. The ability to grow on glycerol as a single carbon and energy source – a mode of cultivation widely applied for generation of biomass with *P. pastoris* prior to methanol induction or glucose fed batch – is dependent on the activity of a constitutive salt-independent active glycerol transport by the H^+/glycerol symport and has also been reported for *Pichia sorbitophila* and *Pichia jadinii* [41]. Specific growth rates of these yeasts on glycerol are similar to the specific growth rates that can be obtained on glucose (e.g. for *P. pastoris* on mineral media $\mu_{Glycerol\_max}$ = 0.26 h^{-1}, $\mu_{Glucose\_max}$ = 0.19 h^{-1}), whereas yeasts lacking the activity of such a type of carrier have significantly reduced growth rates on glycerol. The high specific glycerol uptake rate, enabled by the exceptional endowment with specific transporters emphasizes the suitability of glycerol as a substrate for biomass growth.

Conclusion

The availability of genome data has become an essential tool for cell and metabolic engineering of biotechnological production organisms. This work highlights major advantages of *P. pastoris* as a protein production platform and the benefits of glycerol/glucose based production technology. Apart from lower heat production and oxygen demand compared to methanol based processes, glucose grown cultures display higher viability and essentially no protease release to the culture supernatant. Furthermore detailed insights into the sugar transport will enable rational modulation of substrate fluxes, especially for efficient metabolite production.

Material and methods

Strain

The *P. pastoris* type strain (DSMZ 70382 = CBS704) was selected as the source of genomic DNA, and used for all experimental work. Genomic DNA was prepared as described in Hohenblum et al. using the Qiagen Genomic G-20 kit [42].

Sequencing

Genomic DNA was sequenced by GATC Biotech AG, Konstanz (Germany) with a Roche GS FLX-Titanium Series complemented by an Illumina Genome Analyzer paired end run. The reads were assembled with SeqMan NGen by DNASTAR. To verify the sequencing quality all *P. pastoris* gene and protein sequences available at NCBI were downloaded and the sequences were compared using BLAST searches.

Gene prediction and annotation

Gene prediction was performed with the eukaryotic gene finder Augustus [43] using the option for overlapping genes as well as the prokaryotic gene finder Glimmer3 [44]. Predicted open reading frames were kept if they were longer than 100 nucleotides and started with ATG, except for genes predicted on contig boarders. The ORF sets were merged and made non redundant using the clustering program cd-hit-est [45] with a similarity cut-off of 95%.

Annotation was done by a reciprocal protein BLAST against a dataset consisting of the publicly available *Saccharomycotina* species and the UNIPROT protein database with an E-value threshold of 10^{-10}. All *P. pastoris* proteins and genes available at NCBI, all proteins that were predicted to be secreted and all sugar transporters were manually curated. Gene Ontology annotation was done for all proteins with a homolog in *S. cerevisiae*.

Ribosomal RNA annotation was done through homology with *S. cerevisiae* using nucleotide BLAST against the *P. pastoris* contigs, and the results were manually analyzed. tRNAs were localized using the program tRNAscan-SE [46]. Gene predictions were manually curated using BLASTx.

In silico *secretome prediction*

A similar method was used as described to predict the secretomes of *K. lactis* [15] and *C. albicans* [16], respectively. The prediction pipeline included SignalP 3.0 [47,48] to identify the N-terminal signal peptide, Phobius [49] to predict the transmembrane topology, GPI-SOM [50] and the fungal version of big-PI [51] for GPI anchor prediction, TargetP [52] to exclude all proteins with predicted mitochondrial localization. Additionally WoLF PSORT [53] was used for general localization prediction.

Proteins were considered to be secreted when an N-terminal signal peptide existed but neither a transmembrane domain (except one within the first 40 residues), nor a GPI-anchor, nor any localization signal to other organelles were identified.

The prediction pipeline was tested on an *S. cerevisiae* dataset of 5,884 proteins which was downloaded from the Saccharomyces Genome Database SGD [22].

Experimental secretome analysis

P. pastoris DSMZ 70382 was grown in fully aerobic chemostat cultures on minimal medium with glucose as carbon source until steady state (biomass yield and RQ constant for at least 2 residence times). Detailed data on media compositions, fermentation data and the analysis of culture supernatant can be found in additional file 3. Culture supernatants were concentrated by acetone precipitation and subjected to 1D SDS-PAGE on a 12% PAA gel and 2D-DIGE, respectively. For 2D-DIGE supernatant protein was Cy5 labelled and separated on a IPGDryStrip (3-11NL) in the first dimension, followed by SDS-PAGE on a 12% PAA gel as described in Dragosits et al. [24]. 1D gel lanes were cut into 21 slices, and protein spots from CBB stained 2D gels were picked. After tryptic digest, samples were analyzed by reversed-phase chromatography (UltiMate 3000 Capillary LC-system, Dionex) coupled with ESI MS/MS analysis (Q-TOF Ultima Global, Waters). The obtained mass spectra were subsequently analysed using X!Tandem 2008.12.01 [54]. The identified proteins had to meet the following criteria: protein score e-value ≤ 10^{-5} with at least 2 peptides per protein. Glycoproteins were detected by SDS-PAGE and blotting of proteins onto a nitrocellulose membrane followed by detection via Concanavalin A and Horseradish peroxidase. Putative N-glycosylation sites were identified with NetNGlyc 1.0 server [55].

Analysis of hexose transporters

P. pastoris ORFs encoding putative sugar transporters were identified by sequence similarity using BLAST. Multiple sequence alignment of the respective protein sequences to previously identified hexose transporters and sensors from other yeasts was performed by ClustalW [56] using BLOSUM weight matrix, and a dendrogram with branch length was generated. Additionally an integrated search in PROSITE [57], Pfam, PRINTS and other family and domain databases was performed with InterProScan [58] for all these protein sequences.

Disruption cassettes for PIPA00236 and PIPA08653 were generated by PCR (primers: PIPA08653FW: ATGGCAGGTATTAAAGTTGGATC; PIPA08653BW: TACTGCCATCTGCTTCTTTC; PIPA00236FW: GCAGGAGAATAGTCCAGTTTAC; PIPA00236BW: TTCATAGCCTCGTCGACTCTG). 200–300 bp each up- and downstream of the start codon were exchanged for the Zeocin resistance cassette. These cassettes were introduced into the genome of *P. pastoris* DSMZ 70382 by electroporation, and clones were selected on YP plates containing 1% yeast extract, 2% peptone, 2% agar-agar, 2% glycerol and 25 μg mL^{-1} Zeocin. Positively growing clones were then analyzed for their growth behavior on YP plates containing either 2% glycerol, 2% glucose or 0.01% glucose for 48 h at 28°C.

Genome Database

The gene predictions were parsed into GFF file format and loaded into a Chado [59] database which is designed especially to hold a wide variety of biological data.

Gbrowse [60], the Generic Genome Browser, was installed on a web server in the latest stable version (1.69) and configured to display the genomic data from the Chado database.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DM initiated and coordinated the *P. pastoris* genome project. AG and AR were responsible for genome annotation and analysis. AG predicted the secreted proteins. MD performed the chemostat cultivations and 2D-gel electrophoresis. AR developed the genome database. JS performed the MS identification of the secreted proteins. FA coordinated and supervised proteomics. MM, MK and MS contributed to annotation. BG carried out the analysis of the hexose transporters and contributed to gene annotation. DM, AG, MD, MM and BG wrote the final text of the manuscript.

Additional material

Additional file 1

***Predicted secretome of* P. pastoris**. *Predicted localization of all genes containing a predicted signal peptide. The output of the prediction pipeline is given, as well as ORF and gene names and descriptions of* S. cerevisiae *homologs, if available.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1475-2859-8-29-S1.xls]

Additional file 2

Summary of identified proteins. *List of mass spectrometry identified proteins on both 1D and 2D gels, including protein scores and all individual peptides with corresponding peptide scores.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1475-2859-8-29-S2.xls]

Additional file 3

Chemostat cultivation data. *Detailed chemostat cultivation data including culture medium composition and evaluation of DNA, RNA and protein content of the supernatant.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1475-2859-8-29-S3.xls]

Acknowledgements

This work has been supported by the European Science Foundation (ESF, program EuroSCOPE), the Austrian Science Fund (FWF), project no. 137, and the Austrian Resarch Promotion Agency (Program FHplus). Special thanks to Harald Pichler (TU Graz) for critically reviewing the manuscript.

Addendum

During revision of this manuscript, De Schutter et al. have published the genome sequence of *K. phaffii* (*P. pastoris*) strain GS115 (Nat. Biotechnol. doi:10.1038/nbt.1544).

References

1. Cereghino G, Cereghino J, Ilgen C, Cregg J: **Production of recombinant proteins in fermenter cultures of the yeast Pichia pastoris.** *Curr Opin Biotechnol* 2002, **13(4):**329-332.
2. Hamilton S, Davidson R, Sethuraman N, Nett J, Jiang Y, Rios S, Bobrowicz P, Stadheim T, Li H, Choi B, *et al.*: **Humanization of yeast to produce complex terminally sialylated glycoproteins.** *Science* 2006, **313(5792):**1441-1443.
3. Hamilton S, Gerngross T: **Glycosylation engineering in yeast: the advent of fully humanized yeast.** *Curr Opin Biotechnol* 2007, **18(5):**387-392.
4. Jacobs P, Geysens S, Vervecken W, Contreras R, Callewaert N: **Engineering complex-type N-glycosylation in Pichia pastoris using GlycoSwitch technology.** *Nat Protoc* 2009, **4(1):**58-70.
5. Marx H, Mattanovich D, Sauer M: **Overexpression of the riboflavin biosynthetic pathway in Pichia pastoris.** *Microb Cell Fact* 2008, **7:**23.
6. He J, Deng J, Zheng Y, Gu J: **A synergistic effect on the production of S-adenosyl-L-methionine in Pichia pastoris by knocking in of S-adenosyl-L-methionine synthase and knocking out of cystathionine-beta synthase.** *J Biotechnol* 2006, **126(4):**519-527.
7. Dunn WJ, Cregg J, Kiel J, Klei I van der, Oku M, Sakai Y, Sibirny A, Stasyk O, Veenhuis M: **Pexophagy: the selective autophagy of peroxisomes.** *Autophagy* 2005, **1(2):**75-83.
8. Payne W, Kaiser C, Bevis B, Soderholm J, Fu D, Sears I, Glick B: **Isolation of Pichia pastoris genes involved in ER-to-Golgi transport.** *Yeast* 2000, **16(11):**979-993.
9. Yamada Y, Matsuda M, Maeda K, Mikata K: **The phylogenetic relationships of methanol-assimilating yeasts based on the partial sequences of 18S and 26S ribosomal RNAs: the proposal of Komagataella gen. nov. (Saccharomycetaceae).** *Biosci Biotechnol Biochem* 1995, **59(3):**439-444.
10. Kurtzman C: **Description of Komagataella phaffii sp. nov. and the transfer of Pichia pseudopastoris to the methylotrophic yeast genus Komagataella.** *Int J Syst Evol Microbiol* 2005, **55(Pt 2):**973-976.
11. Curvers S, Brixius P, Klauser T, Thömmes J, Weuster-Botz D, Takors R, Wandrey C: **Human chymotrypsinogen B production with Pichia pastoris by integrated development of fermentation and downstream processing. Part 1. Fermentation.** *Biotechnol Prog* **17(3):**495-502.
12. Kobayashi K, Kuwae S, Ohya T, Ohda T, Ohyama M, Ohi H, Tomomitsu K, Ohmura T: **High-level expression of recombinant human serum albumin from the methylotrophic yeast Pichia pastoris with minimal protease production and activation.** *J Biosci Bioeng* 2000, **89(1):**55-61.
13. Potgieter T, Cukan M, Drummond J, Houston-Cummings N, Jiang Y, Li F, Lynaugh H, Mallem M, McKelvey T, Mitchell T, *et al.*: **Production of monoclonal antibodies by glycoengineered Pichia pastoris.** *J Biotechnol* 2009, **139(4):**318-325.
14. Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM: **Heterologous protein production using the Pichia pastoris expression system.** *Yeast* 2005, **22(4):**249-270.
15. Swaim C, Anton B, Sharma S, Taron C, Benner J: **Physical and computational analysis of the yeast Kluyveromyces lactis secreted proteome.** *Proteomics* 2008, **8(13):**2714-2723.
16. Lee S, Wormsley S, Kamoun S, Lee A, Joiner K, Wong B: **An analysis of the Candida albicans genome database for soluble secreted proteins using computer-based prediction algorithms.** *Yeast* 2003, **20(7):**595-610.
17. Martinez D, Berka R, Henrissat B, Saloheimo M, Arvas M, Baker S, Chapman J, Chertkov O, Coutinho P, Cullen D, *et al.*: **Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina).** *Nat Biotechnol* 2008, **26(5):**553-560.
18. Ravalason H, Jan G, Mollé D, Pasco M, Coutinho P, Lapierre C, Pollet B, Bertaud F, Petit-Conil M, Grisel S, *et al.*: **Secretome analysis of Phanerochaete chrysosporium strain CIRM-BRFM41 grown on softwood.** *Appl Microbiol Biotechnol* 2008, **80(4):**719-733.
19. Oda K, Kakizono D, Yamada O, Iefuji H, Akita O, Iwashita K: **Proteomic analysis of extracellular proteins from Aspergillus oryzae grown under submerged and solid-state culture conditions.** *Appl Environ Microbiol* 2006, **72(5):**3448-3457.
20. Porro D, Sauer M, Branduardi P, Mattanovich D: **Recombinant protein production in yeasts.** *Mol Biotechnol* 2005, **31(3):**245-259.
21. Altschul S, Gish W, Miller W, Myers E, Lipman D: **Basic local alignment search tool.** *J Mol Biol* 1990, **215(3):**403-410.
22. **Saccharomyces Genome Database** [http://www.yeastgenome.org]
23. Delgado M, O'Connor J, Azorín I, Renau-Piqueras J, Gil M, Gozalbo D: **The glyceraldehyde-3-phosphate dehydrogenase polypeptides encoded by the Saccharomyces cerevisiae TDH1, TDH2 and TDH3 genes are also cell wall proteins.** *Microbiology* 2001, **147(Pt 2):**411-417.
24. Dragosits M, Stadlmann J, Albiol J, Baumann K, Maurer M, Gasser B, Sauer M, Altmann F, Ferrer P, Mattanovich D: **The Effect of Temperature on the Proteome of Recombinant Pichia pastoris.** *J Proteome Res* 2009 in press.
25. Wesolowski-Louvel M, Goffrini P, Ferrero I, Fukuhara H: **Glucose transport in the yeast Kluyveromyces lactis. I. Properties of an inducible low-affinity glucose transporter gene.** *Mol Gen Genet* 1992, **233:**1-2.
26. Stasyk OG, Maidan MM, Stasyk OV, Van Dijck P, Thevelein JM, Sibirny AA: **Identification of hexose transporter-like sensor HXS1 and functional hexose transporter HXT1 in the methylotrophic yeast Hansenula polymorpha.** *Eukaryot Cell* 2008, **7(4):**735-746.
27. Kasahara T, Maeda M, Ishiguro M, Kasahara M: **Identification by comprehensive chimeric analysis of a key residue responsible for high affinity glucose transport by yeast HXT2.** *J Biol Chem* 2007, **282(18):**13146-13150.
28. Barnett JA, Payne RW, Yarrow D: **Yeasts: Characteristics and Identification.** *Volume 3*. 3rd edition. Cambridge, UK: Cambridge University Press; 2000.
29. **Pichia pastoris genome browser** [http://www.pichiagenome.org]
30. Ohi H, Okazaki N, Uno S, Miura M, Hiramatsu R: **Chromosomal DNA patterns and gene stability of Pichia pastoris.** *Yeast* 1998, **14(10):**895-903.
31. Jeffries T, Grigoriev I, Grimwood J, Laplaza J, Aerts A, Salamov A, Schmutz J, Lindquist E, Dehal P, Shapiro H, *et al.*: **Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast Pichia stipitis.** *Nat Biotechnol* 2007, **25(3):**319-326.
32. Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuvéglise C, Talla E, *et al.*: **Genome evolution in yeasts.** *Nature* 2004, **430(6995):**35-44.
33. Adams DJ: **Fungal cell wall chitinases and glucanases.** *Microbiology* 2004, **150(Pt 7):**2029-2035.
34. Jahic M, Wallberg F, Bollok M, Garcia P, Enfors S: **Temperature limited fed-batch technique for control of proteolysis in Pichia pastoris bioreactor cultures.** *Microb Cell Fact* 2003, **2(1):**6.
35. Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM: **Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter.** *Gene* 1997, **186(1):**37-44.
36. Nombela C, Gil C, Chaffin WL: **Non-conventional protein secretion in yeast.** *Trends Microbiol* 2006, **14(1):**15-21.
37. Boles E, Hollenberg CP: **The molecular genetics of hexose transport in yeasts.** *FEMS Microbiol Rev* 1997, **21(1):**85-111.
38. van Urk H, Postma E, Scheffers W, van Dijken J: **Glucose transport in crabtree-positive and crabtree-negative yeasts.** *J Gen Microbiol* 1989, **135(9):**2399-2406.
39. Maurer M, Kuehleitner M, Gasser B, Mattanovich D: **Versatile modeling and optimization of fed batch processes for the produc-**

tion of secreted heterologous proteins with Pichia pastoris. *Microb Cell Fact* 2006, **5:**37.
40. Otterstedt K, Larsson C, Bill RM, Stahlberg A, Boles E, Hohmann S, Gustafsson L: **Switching the mode of metabolism in the yeast Saccharomyces cerevisiae.** *EMBO Rep* 2004, **5(5):**532-537.
41. Lages F, Silva-Graca M, Lucas C: **Active glycerol uptake is a mechanism underlying halotolerance in yeasts: a study of 42 species.** *Microbiology* 1999, **145(Pt 9):**2577-2585.
42. Hohenblum H, Gasser B, Maurer M, Borth N, Mattanovich D: **Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant Pichia pastoris.** *Biotechnol Bioeng* 2004, **85(4):**367-375.
43. Stanke M, Waack S: **Gene prediction with a hidden Markov model and a new intron submodel.** *Bioinformatics* 2003, **19(Suppl 2):**ii215-225.
44. Delcher A, Bratke K, Powers E, Salzberg S: **Identifying bacterial genes and endosymbiont DNA with Glimmer.** *Bioinformatics* 2007, **23(6):**673-679.
45. Li W, Godzik A: **Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide.** *Bioinformatics* 2006, **22(13):**1658-1659.
46. Lowe T, Eddy S: **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.** *Nucleic Acids Res* 1997, **25(5):**955-964.
47. Bendtsen J, Nielsen H, von Heijne G, Brunak S: **Improved prediction of signal peptides: SignalP 3.0.** *J Mol Biol* 2004, **340(4):**783-795.
48. Nielsen H, Engelbrecht J, Brunak S, von Heijne G: **Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.** *Protein Eng* 1997, **10(1):**1-6.
49. Käll L, Krogh A, Sonnhammer E: **Advantages of combined transmembrane topology and signal peptide prediction – the Phobius web server.** *Nucleic Acids Res* 2007:W429-432.
50. Fankhauser N, Mäser P: **Identification of GPI anchor attachment signals by a Kohonen self-organizing map.** *Bioinformatics* 2005, **21(9):**1846-1852.
51. Eisenhaber B, Schneider G, Wildpaner M, Eisenhaber F: **A sensitive predictor for potential GPI lipid modification sites in fungal protein sequences and its application to genome-wide studies for Aspergillus nidulans, Candida albicans, Neurospora crassa, Saccharomyces cerevisiae and Schizosaccharomyces pombe.** *J Mol Biol* 2004, **337(2):**243-253.
52. Emanuelsson O, Nielsen H, Brunak S, von Heijne G: **Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.** *J Mol Biol* 2000, **300(4):**1005-1016.
53. Horton P, Park K, Obayashi T, Fujita N, Harada H, Adams-Collier C, Nakai K: **WoLF PSORT: protein localization predictor.** *Nucleic Acids Res* 2007:W585-587.
54. **X!Tandem** [http://www.thegpm.org/tandem/]
55. **NetNGlyc 1.0 server** [http://www.cbs.dtu.dk/services/NetNGlyc/]
56. **ClustalW** [http://align.genome.jp]
57. Bairoch A: **PROSITE: a dictionary of sites and patterns in proteins.** *Nucleic Acids Res* 1992, **20(Suppl):**2013-2018.
58. **InterProScan** [http://www.ebi.ac.uk/Tools/InterProScan/]
59. Mungall C, Emmert D: **A Chado case study: an ontology-based modular schema for representing genome-associated biological information.** *Bioinformatics* 2007, **23(13):**i337-346.
60. Stein L, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich J, Harris T, Arva A, *et al.*: **The generic genome browser: a building block for a model organism system database.** *Genome Res* 2002, **12(10):**1599-1610.

BMC Genomics

BioMed Central

Research article

Open Access

Novel insights into the unfolded protein response using *Pichia pastoris* specific DNA microarrays

Alexandra Graf†[1], Brigitte Gasser†[1], Martin Dragosits[1], Michael Sauer[2], Germán G Leparc[3], Thomas Tüchler[3], David P Kreil[3] and Diethard Mattanovich*[1,2]

Address: [1]Institute of Applied Microbiology, Department of Biotechnology, University of Natural Resources and Applied Life Sciences Vienna, Muthgasse 18, 1190 Vienna, Austria, [2]School of Bioengineering, University of Applied Sciences FH Campus Vienna, Muthgasse 18, 1190 Vienna, Austria and [3]Vienna Science Chair of Bioinformatics, Department of Biotechnology, University of Natural Resources and Applied Life Sciences Vienna, Muthgasse 18, 1190 Vienna, Austria

Email: Alexandra Graf - alexandra.graf@boku.ac.at; Brigitte Gasser - brigitte.gasser@boku.ac.at; Martin Dragosits - martin.dragosits@boku.ac.at; Michael Sauer - michael.sauer@fh-campuswien.ac.at; Germán G Leparc - german.leparc@boku.ac.at; Thomas Tüchler - thomas.tuechler@boku.ac.at; David P Kreil - pichia08@kreil.org; Diethard Mattanovich* - diethard.mattanovich@boku.ac.at

* Corresponding author †Equal contributors

Published: 19 August 2008

BMC Genomics 2008, **9**:390 doi:10.1186/1471-2164-9-390

Received: 8 February 2008
Accepted: 19 August 2008

This article is available from: http://www.biomedcentral.com/1471-2164/9/390

Abstract

Background: DNA Microarrays are regarded as a valuable tool for basic and applied research in microbiology. However, for many industrially important microorganisms the lack of commercially available microarrays still hampers physiological research. Exemplarily, our understanding of protein folding and secretion in the yeast *Pichia pastoris* is presently widely dependent on conclusions drawn from analogies to *Saccharomyces cerevisiae*. To close this gap for a yeast species employed for its high capacity to produce heterologous proteins, we developed full genome DNA microarrays for *P. pastoris* and analyzed the unfolded protein response (UPR) in this yeast species, as compared to *S. cerevisiae*.

Results: By combining the partially annotated gene list of *P. pastoris* with *de novo* gene finding a list of putative open reading frames was generated for which an oligonucleotide probe set was designed using the probe design tool TherMODO (a thermodynamic model-based oligoset design optimizer). To evaluate the performance of the novel array design, microarrays carrying the oligo set were hybridized with samples from treatments with dithiothreitol (DTT) or a strain overexpressing the UPR transcription factor *HAC1*, both compared with a wild type strain in normal medium as untreated control. DTT treatment was compared with literature data for *S. cerevisiae*, and revealed similarities, but also important differences between the two yeast species. Overexpression of *HAC1*, the most direct control for UPR genes, resulted in significant new understanding of this important regulatory pathway in *P. pastoris*, and generally in yeasts.

Conclusion: The differences observed between *P. pastoris* and *S. cerevisiae* underline the importance of DNA microarrays for industrial production strains. *P. pastoris* reacts to DTT treatment mainly by the regulation of genes related to chemical stimulus, electron transport and respiration, while the overexpression of *HAC1* induced many genes involved in translation, ribosome biogenesis, and organelle biosynthesis, indicating that the regulatory events triggered by DTT treatment only partially overlap with the reactions to overexpression of *HAC1*. The high reproducibility of the results achieved with two different oligo sets is a good indication for their robustness, and underlines the importance of less stringent selection of regulated features, in order to avoid a large number of false negative results.

Background

Transcriptomics, the parallel quantification of many, or all transcripts of an organism in given conditions, has become a favorite tool for basic research [1]. Messenger-RNA regulation patterns of model organisms under many different conditions have become available during the last years. However, these methods are still not applicable for many industrially important organisms, mainly due to the lack of DNA microarrays targeting these organisms. A typical example is the yeast *Pichia pastoris*, which is widely applied for the production of recombinant proteins. Several approaches have been taken to derive transcriptomic data without specific microarrays. Sauer et al. [2] have applied heterologous hybridization of *P. pastoris* samples to *Saccharomyces cerevisiae* microarrays. Alternative methodological concepts like Transcript Analysis with the Aid of Affinity Capture (TRAC) [3] may be applied preferentially to subsets of the transcriptome [4], provided that genome sequence data are available. If this is not the case, total cDNA may be utilized as a source of probes, either by applying expressed sequence tags to microarrays [5] or employing RNA fingerprinting like cDNA-amplified fragment length polymorphism (cDNA-AFLP) [6], which has recently been applied to *Trichoderma reesei* [7]. These unannotated methods bear of course the disadvantage that specific hits may only be identified after sequencing their respective probes.

Therefore oligonucleotide microarrays have become the method of choice for many applications, although their design depends on the availability of a genomic sequence with good gene identification and annotation. The genome sequence of *P. pastoris* is not published yet. The data available from Integrated Genomics (IG, Chicago, IL, USA; [8]) contain a partial gene identification and annotation, so that additional effort in this direction was a first step necessary towards development of comprehensive DNA microarrays for this yeast species. There is a wide choice of computational gene finders available at the moment which can be classified into intrinsic and extrinsic prediction programs. Intrinsic or *de novo* gene finder only use information from the sequences to be studied, building statistical models to distinguish between coding and non-coding regions of the genome on the basis of biological sequence patterns [9-11]. Extrinsic gene finder utilize homology search to determine where protein coding regions are in the genome. Their applicability is therefore limited to organisms that have homologs in current databases that are correctly annotated. Because of this limitation it is common to integrate homology search with *de novo* prediction [12]. Most state of the art gene finders use a form of Hidden Markov Model (HMM) differing in the implementation and complexity of the model as well as the ease in which users can adapt the application to their needs [13].

It is well known that cross-hybridization can confound microarray results rendering good probe design an essential requirement for accurate microarray analyses. The specificity of oligonucleotides is determined by the Gibbs free energy (ΔG) of the hybridization reaction between potential binding partners. Highly specific probes will bind their target transcript much more strongly than any other transcript. Considering that microarray experiments are non-equilibrium measurements, it is desirable that microarray probes exhibit uniform thermodynamic properties, which many probe design tools aim to achieve by demanding a narrow distribution of the probe-target melting temperature T_m. Ideally, probes should have a uniform binding free energy at the hybridization temperature T_{hyb} [14].

Previous studies have demonstrated that industrial production strains may behave quite differently to laboratory strains and model organisms [15], which emphasizes the importance of analytical tools for industrially relevant strains and species. As an example, the unfolded protein response (UPR), a regulation circuit of high relevance for heterologous protein production in eukaryotic cells [16], has been shown to be differentially regulated in *P. pastoris* [4] compared to *S. cerevisiae* [17], which is the typical model species for hemiascomycete yeasts. The development of specific microarrays for *P. pastoris* was intended to allow a detailed analysis of UPR regulation in *P. pastoris*. As in previous transcriptomics work with *S. cerevisiae* the induction of UPR was either accomplished by addition of dithiothreitol (DTT) or tunicamycin, this work aimed at a comparison of DTT induced gene regulation in *P. pastoris* to that in *S. cerevisiae* published by Travers et al. [17]. Finally we aimed at the comparison of DTT induced regulation to the regulatory response to overexpression of *HAC1*, the transcription factor controlling the UPR. Transcriptional regulation of *HAC1* overexpression has not been studied for yeasts so far, so that we expected valuable data to better define the core UPR regulated transcriptome.

Results and Discussion

Gene prediction and Oligo Design

To evaluate available gene finders for their performance on yeast genomes, three *de novo* gene finders (GeneMark, Glimmer3, GlimmerHMM) were tested on the genome sequence of *S. cerevisiae*. GeneMark and Glimmer3 work with a prokaryotic Hidden Markov Model (HMM) whereas GlimmerHMM employs a eukaryotic gene model. GeneMark was trained with coding and non-coding sequences of *S. cerevisiae*, building an HMM transition probability matrix of the 7th order. Glimmer3 and GlimmerHMM could be trained directly on the genome in question without specifying coding and non-coding regions. In Lomsadze et al. [18] and Besemer and Boro-

dovsky [9] the difficulty of eukaryotic gene finders in the prediction of genes for organisms with few introns is discussed and linked to a lack of data for representative exon – intron models. Our results confirmed that a gene finder written for eukaryotes (GlimmerHMM) could not be trained well on yeast genomes, introducing far too many introns into the predicted genes. Both prokaryotic versions performed much better, with GeneMark predicting less false negatives but more false positives than Glimmer3 (Table 1). Even though the positive prediction value was somewhat lower with GeneMark it was more important not to miss true positives than to achieve a lower rate of false positives. A further improvement could be achieved by a GeneMark model for lower eukaryotes, in which the prokaryotic algorithm is modified to use Kozak start sites instead of prokaryotic ribosomal binding sites. *P. pastoris* genes were predicted using this version of GeneMark with the lowest possible threshold (probability score t = 0.05) so that filter conditions could be better controlled at a later state. The prediction yielded a total of 26,471 putative genes for the genome of *P. pastoris*.

In a WU-BLASTN search against *S. cerevisiae*, 6,374 sequences that were predicted by GeneMark, and 3,964 of the IG predictions produced hits with *S. cerevisiae* using an *E* value (Expectation value, [19]) of $< 10^{-4}$, a hit length > 100 nucleotides and an identity of >50%. To reduce the redundancy within the data set the predicted genes were clustered into groups sharing more than 90% similarity using cd-hit [20]. From a total of 31,896 candidate sequences (GeneMark and IG predictions), 22,020 cd-hit groups were obtained. From the cluster file it was clear that some of the clusters had to be analyzed further before selecting target sequences for the oligo design. After the removal of all sequences that had a short length and a low prediction value, complex clusters were defined as clusters for which the minimum relative length of all sequences was smaller than 0.9. A total of 2,612 clusters fell into this category and were excluded at a first design stage.

Finally 19,508 predicted target sequences remained to be tested in the first microarray experiments. OligoArray 2.1 [21] was able to design oligonucleotide probes for 17,161 sequences ranging in length from 57 to 60 nucleotides.

Validation arrays for the first list of predicted transcript sequences (Same-Same experiment)

With these probes 4 × 44 K slides were produced on the Agilent microarray platform and employed for an initial validation of the predicted transcript sequences by hybridization with the Pool samples of *P. pastoris* (for preparation of Pool samples see Material and Methods). One slide had to be discarded because of quality issues. For the remaining 12 arrays the number of probes showing a signal varied between 10,708 and 15,598. Of these, 7,980 had a signal on all 12 arrays, and only 951 probes showed no hybridization on all 12 arrays.

Second, curated list of predicted target sequences and second oligo design

The results of the initial validation arrays were utilized to adapt the list of predicted genes, keeping all predictions for which a hybridization signal could be observed for all arrays plus all predictions with significant sequence similarity to annotated genes as well as all sequences with an average gene prediction score > 0.5. This approach allows for the fact that not all genes will have been actively expressed in the target samples. Additionally, predicted transcripts resulting from a subsequent analysis of the complex clusters were included at this stage. Of the 2,612 complex cluster that were not included in the design for the first batch of arrays, only 223 contained more than 2 sequences and for a further 14 no subsequence match of at least 60 nucleotides could be found within the last 1000 bases at the 3'-end. These 237 clusters were manually curated while the rest could be automatically reduced to one sequence. To make full use of the 15,208 features available on the Agilent microarray platform, it was decided to also include predicted sequences with somewhat lower gene prediction score that showed a hybridization signal in at least 8 of the 12 arrays. Finally, a selected set of 15,253 predicted transcript sequences was used as targets for probe design of a comprehensive *P. pastoris* microarray. While it is obvious that this list is larger than the expected number of open reading frames (6,000–7,000), as judged in comparison to other yeast species [22], we intentionally included more putative transcript sequences, as false positives with a distinct sequence will not negatively affect microarray design or

Table 1: Comparison of gene finder performance on yeast genomic sequence data

Gene finder	True positives	Partly	False positives	False negatives	Sensitivity (%)	Positive prediction value (%)
Glimmer3	75	13	21	31	73.9	68.8
GlimmerHMM	1	3	68(234)	115	3.2	1.4
GeneMark	81	16	32	22	81.5	62.7

Three different gene finders were tested on the genome sequence of *S. cerevisiae* chromosome 1 to evaluate the quality of gene prediction. Sensitivity = TP/(TP + FN), positive prediction value = TP/(TP + FP); For Glimmer HMM the column False Positives contains the number of genes and in brackets the number of exons.

experiments, in contrast to the damage of falsely excluding a potential transcript target.

Oligonucleotide probes were designed using a probe design tool developed in-house, a thermodynamic model-based oligoset optimizer ('TherMODO', [23]). TherMODO designed probes for 15,035 sequences, of which only 665 were predicted as having cross-hybridization potential. The TherMODO design was compared to probe design with eArray [24]. The distributions of ΔG and T_m of both designs are shown in additional file 1. Clearly the TherMODO designed probes are more uniform in respect to the Gibbs free energy ΔG, indicating a superior hybridization performance [14].

The final probe design was manufactured on 8 × 15 K slides by Agilent, and evaluated for reproducibility and biological meaningfulness. Pool samples were applied to 2 arrays on 2 slides each, including dye swap. The scatterplots show uniformly high correlations > 97% both within and between arrays, both on same and different slides, indicating high reproducibility of hybridization signals between identical samples. Exemplarily, a scatterplot of signal intensities derived from the same samples (wild type strain untreated) is shown in Figure 1. For the final gene list the annotation was improved in addition to the annotation provided by IG. This resulted in 3954 annotated ORFs, of which 2989 had an IG annotation. 965 newly annotated ORFs were found, and the annotation of 288 hypothetical proteins was confirmed. All annotated genes are listed in Additional file 2.

Biological evaluation of the new microarrays

The performance of the new arrays was examined by a hybridization experiment using samples, for which transcript regulation data have been obtained before [4]. The biological question evaluated was the regulatory response of *P. pastoris* to constitutive overexpression of the active form of *S. cerevisiae HAC1*, the transcription factor controlling UPR target genes. By this approach, the regulation of 52 genes which have been studied before using TRAC [3] could be verified, with 80% of these genes showing the same regulation pattern for both methods (genes highlighted in bold in Additional file 2). This correlation is statistically significant based on calculating the regression ($p = 8.8 \cdot 10^{-6}$).

The similarities and differences of UPR induction and reaction to DTT stress have been discussed before [4,25,26]. To achieve further insight into this technologically relevant issue, we compared the gene regulation patterns of a *HAC1* overexpressing strain *vs* wildtype control with the regulation pattern of the wildtype treated with DTT for 60 min *vs* the untreated control. Genes were qualified as significantly regulated with a *p*-value < 0.05 (adjusted for multiple testing). 11,262 of all features on the microarrays appeared as differentially regulated either upon DTT treatment or *HAC1* overexpression, or both. 8,480 reacted to *HAC1*, and 6,870 to DTT, with an overlap of 4,088. Considering only the 3,954 annotated genes, a similar pattern is observed with roughly half of the regulated genes overlapping between DTT and *HAC1*, and another half being typical only for either of the treatments

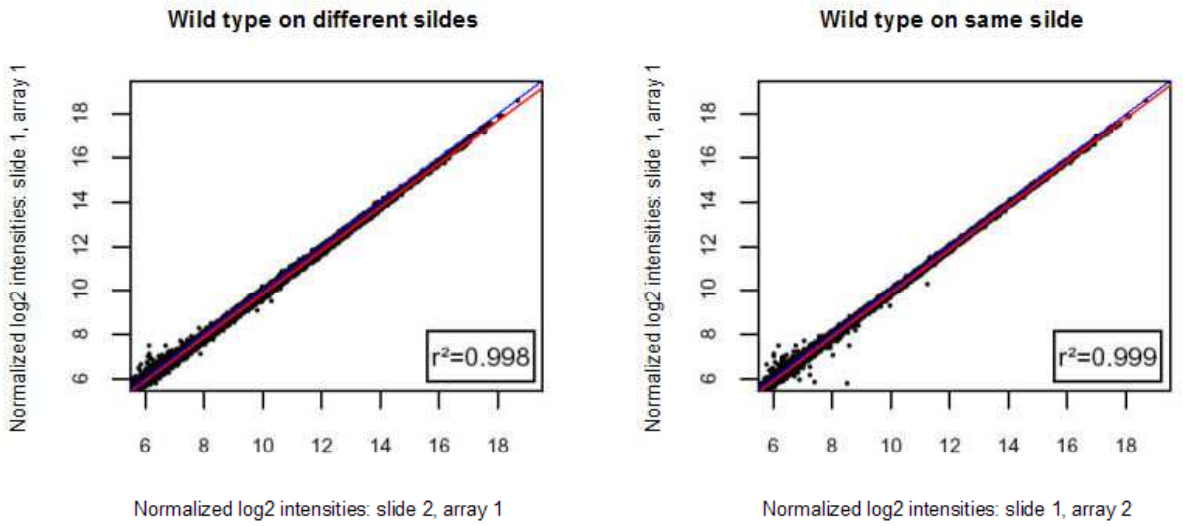


Figure 1
Correlation of signal intensities. Scatterplots of untreated wild type strain samples on (A) different arrays of the same slide; (B) different arrays on different slides. Red line: linear regression of the data; blue line: theoretical perfect correlation.

(Figure 2). Accordingly, the correlation of log fold changes of the two treatments is apparent but rather weak (Figure 3). While DTT treatment is widely accepted as a standard inducer of UPR, these observations indicate that the gene regulation pattern triggered by the UPR transcription factor Hac1 differs to a significant extent from that exerted by DTT.

As previous research on transcriptome regulation upon UPR induction usually employs a fold change (FC) cut-off to highlight the strongly regulated genes, we decided to introduce FC > 1.5 as a second criterion to identify more strongly regulated genes for further detailed analysis (Volcano plots visualizing the two criteria are provided in Additional file 3). Although the introduction of a FC cut-off alters the absolute number of regulated genes, it does not alter the relative distribution of regulated genes categorized into functional groups (GO slim biological process), as can be seen in Figure 4 and Additional file 4.

Comparison of UPR induction by DTT in P. pastoris *and* S. cerevisiae

In order to compare the effects of DTT treatment in *S. cerevisiae* with those in *P. pastoris*, the data published by Travers et al. [17] for 60 min treatment of *S. cerevisiae* with DTT were evaluated alongside with our results for *P. pastoris*. All genes of *S. cerevisiae* which were listed in [17] and for which homologs in *P. pastoris* were identified were

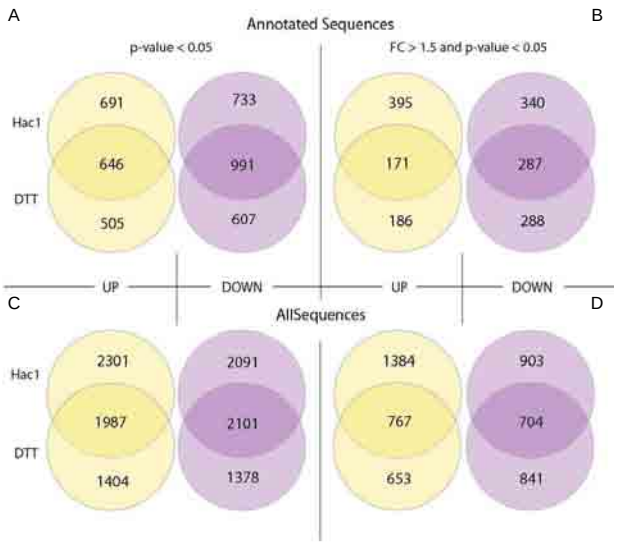


Figure 2
Venn diagrams of differentially expressed genes upon DTT treatment or *HAC1* overexpression. (A, B) Regulated hits with annotation; (C, D) all regulated features; (A, C) cut-off adjusted *p*-value < 0.05; (B, D) cut-off adjusted *p*-value < 0.05 and FC > 1.5.

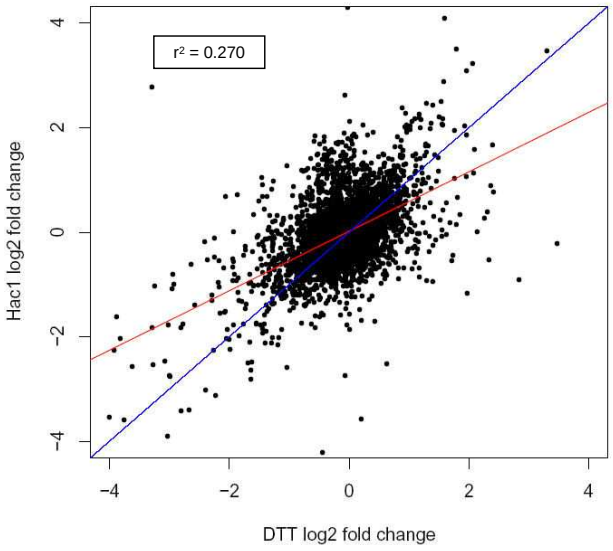


Figure 3
Comparison of expression changes induced by DTT treatment and *HAC1* overexpression, respectively. Log_2 values of expression changes (log_2 FC) caused by DTT (DTT treated wildtype *vs* untreated wildtype) and by Hac1 (*HAC1* overexpression *vs* wildtype) are compared. The correlation coefficient r^2 is indicated. Red line: linear regression of the data; blue line: theoretical perfect correlation.

classified as upregulated, downregulated or unregulated. In order to compare the two data sets, a cutoff of 1.5 fold differential expression was set in both to define regulated genes. A significance threshold on *p*-values could not be employed, as these data were not provided for *S. cerevisiae*. 48% of these genes defined as regulated or unregulated reacted in *P. pastoris* just as in *S. cerevisiae*.

A closer evaluation revealed that certain GO groups were regulated very similarly in both yeast species, while others showed only a low degree of similarity (Table 2). Fisher's exact test was performed to evaluate the significance of groups with low similarity. Especially the GO groups 'translocation', 'protein folding', 'protein degradation', and to some extent 'glycosylation' and 'transport' showed high degrees of similarity. In some GO groups, only some subgroups reacted similarly while others behaved differently in the two yeasts. Of the 'glycosylation' group, core oligosaccharide synthesis and glycosyltransferase genes behaved very similarly, while glycoprotein processing, GPI anchoring and O-glycosylation related genes were regulated significantly different ($p < 0.05$). In the 'protein degradation' group, more similarity was observed for ERAD genes than for ubiquitin/proteasome related genes. Among the 'transport' gene group, budding, fusion and retrieval of ER to Golgi showed a high degree of similar regulation, contrary to the subgroup distal secretion. Low

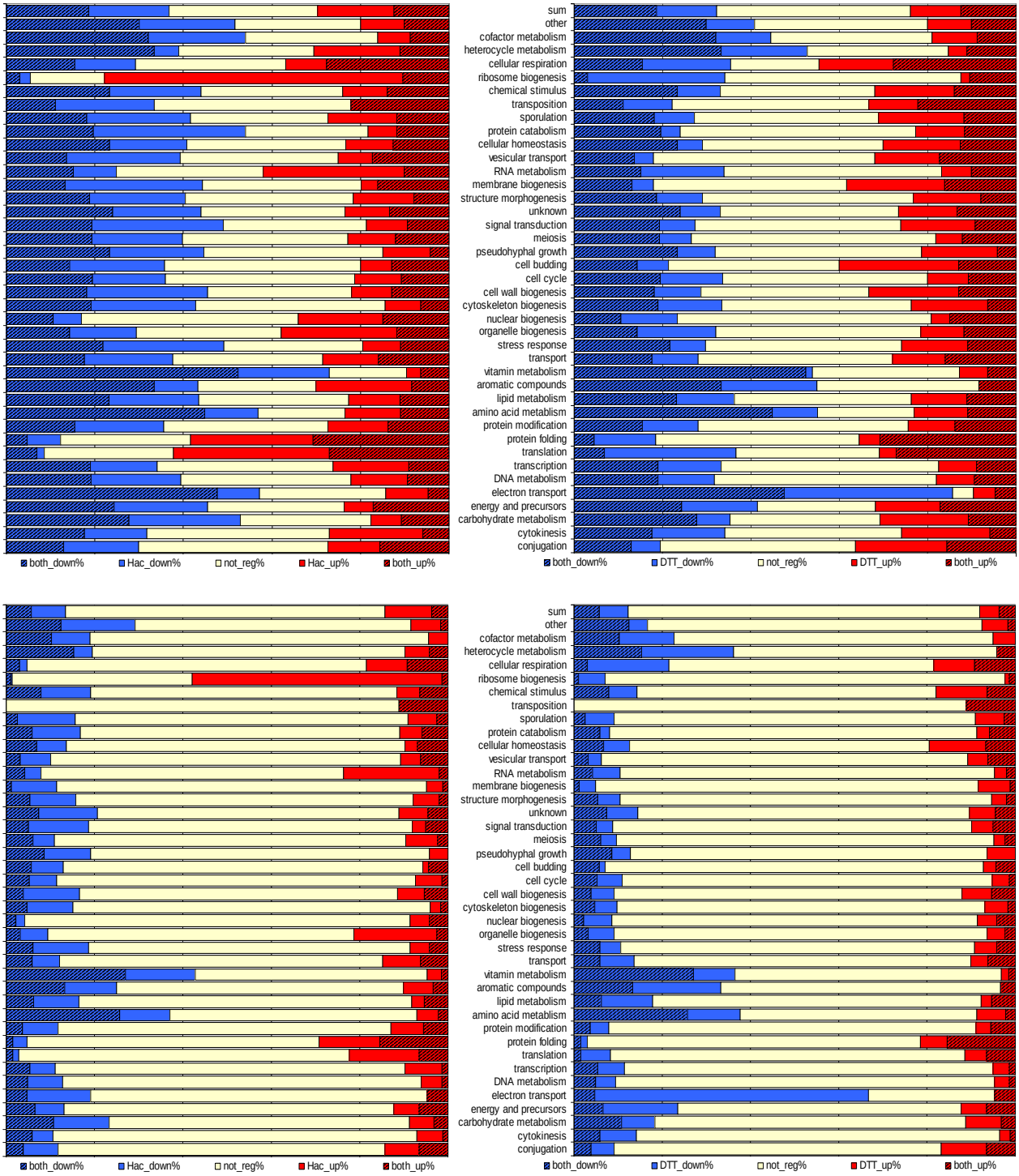


Figure 4

Fractions of up- and downregulated genes in functional groups. Relative numbers of upregulated (red), downregulated (blue) and unregulated (yellow) genes categorized in GO biological process terms upon *HAC1* overexpression (left panel) and DTT treatment (right panel). Shaded in black: regulated in both treatments. Upper panels: cut-off *p*-value < 0.05, lower panel cut-off *p*-value < 0.05 and FC > 1.5. The results of significance testing (Fisher's exact test) are given in additional file 4.

Table 2: Similarity of gene regulation between *P. pastoris* and *S. cerevisiae* upon DTT treatment

Function	Subfunction	No. of similarly regulated/total	% similar regulation
Translocation	total	4/6	67
Glycosylation	total	11/22	50
	Core oligosaccharide synthesis	3/4	75
	Oligosaccharyltransferase	4/4	100
	Glycoprotein processing	1/5	20
	GPI anchoring	1/4	25
	Golgi/O-linked	2/5	40
Protein Folding	total	5/8	63
	Chaperones	3/5	60
	Disulfide bond formation	2/3	67
Protein Degradation	total	4/5	80
	ERAD	3/3	100
	Ubiquitin/Proteasome	1/2	50
Transport	total	11/20	55
	Budding (ER-Golgi)	4/7	57
	Fusion (ER-Golgi)	1/1	100
	Retrieval (ER-Golgi)	4/5	80
	Distal secretion	2/7	29
Lipid Metabolism	total	5/18	28
	Fatty acid metabolism	0/4	0
	Heme biosynthesis	2/5	40
	Phospholipid biosynthesis	2/6	33
	Sphingolipid biosynthesis	0/1	0
	Sterol metabolism	1/2	50
Vacuolar Protein Sorting	total	1/4	25
Cell Wall Biogenesis	total	4/10	40

All genes that were indicated in [17] as core UPR genes in *S. cerevisiae* and having an annotation in *P. pastoris* were grouped by their GO process functions. Similar regulation of a gene means upregulated, downregulated or below cut-off, respectively, in both yeasts.

similarities were observed for 'lipid metabolism', 'vacuolar protein sorting' and 'cell wall biogenesis' genes. It becomes obvious that core UPR genes related to protein translocation, folding and ER transport, as well as core N-glycosylation react similarly to DTT treatment in *P. pastoris* as compared to *S. cerevisiae*, while genes involved in processes which are more distal from ER protein folding behave more differently, indicating that those processes (like functions in the Golgi, [27]) differ significantly between the two yeasts.

Overexpression of Hac1 triggers a different regulation pattern compared to DTT treatment

In most previous studies of the UPR in lower eukaryotic cells, treatment with DTT or tunicamycin, or heterologous protein expression has been employed to trigger the UPR. This study clearly indicates that the set of regulatory events triggered by DTT analysis only partially overlaps with the reactions to constitutive expression of the activated form of the UPR transcription factor Hac1 (see Figures 2 and 3). Interestingly, both treatments resulted in the same amount of genes being down-regulated as being up-regulated, a fact that has been neglected to some extent in the existing literature.

Those genes appearing beyond the threshold (p-value < 0.05 and FC >1.5) were subjected to a more detailed comparison between the effects of DTT treatment and Hac1 induced regulation. The relative numbers of up- and down-regulated genes in each GO biological process term based on the SGD GO slim tool [28] are depicted in Figure 4.

A pattern common to both treatments is the down-regulation of major metabolic processes like carbohydrate, amino acid and lipid metabolism, as well as that of vitamins, cofactors and aromatic and heterocyclic compounds. This makes it obvious that the UPR has a major impact on decreasing both catabolic and anabolic processes. On the other side, both treatments lead to up-regulation of protein folding and vesicular transport. These effects are in line with the published literature, indicating the cellular reaction towards alleviation of the UPR [4,25,26,17].

As expected, the genes coding for classical UPR targets are induced both in Hac1 overproducing and in DTT stressed cells, and genes underlined in the following paragraphs have been identified as UPR targets in previous studies. Especially the ER folding catalysts *PDI1* and *ERO1*, the

DnaJ homologs *JEM1* and *SCJ1*, the ER resident chaperones *CNE1* (calnexin), *KAR2*/BiP and *LHS1* and the mitochondrial chaperones *HSP60* and *SSC1* are significantly up-regulated in both conditions. Among the functional group of 'protein modification' the majority of up-regulated genes belong to the core oligosaccharide synthesis (*DPM1*, *DIE2*), oligosaccharyltransferase complex (*OST1*, *OST2*, *OST3*, *SWP1*, *STT3*, *WBP1*), glycoprotein processing (*ALG2*, *ALG7*, *SEC53*), GPI anchor biosynthesis (*GPI2*, *GPI14*, *PSA1*) and Golgi/O-linked glycosylation (*PMT1*, *PMT2*, *PMT4*, *PMT6*). Besides these, several genes coding for the translocon pore complex (*SEC61*, *SEC62*, *SEC63*, *SEC72*, *SSS1*), which aid the translocation of nascent polypeptides into the ER, are induced. Higashio and Kohno [29] describe the stimulation of ER-to-Golgi transport through the UPR by inducing COPII vesicle formation. In this context, we see *SEC23*, *SEC24*, *SFB2*, *YIP3*, and *ERV2* upregulated. However, also proteins building the COPI coatomer, which are required for retrograde Golgi-to-ER transport, show increased transcription levels upon ER stress in our experiments (*COP1*, *RET2*, *SEC21*, *SEC27*).

While we cannot give any information on ERAD regulation, as *HRD1* is the only annotated gene of this protein degradation process (up-regulated in the Hac1 strain), we observed the down-regulation of some components involved in the assembly of the 20 S core of the 26 S proteasome (*ADD66*, *PRE1*, *PRE4*, *SCL1*) and ubiquitin *UBI4* upon constitutive UPR activation. In this context, Shaffer et al. [30] describe reduced degradation of newly synthesized proteins in XBP1-overexpressing human Raji cells.

Induction of genes encoding cytosolic chaperones (Cns1, Jjj3, Hsp82, Ssa1, Ssa2, Sse1, Ydj1, Zuo1) can only be seen in the Hac1-overproducing strain. Additionally, the ER-resident Pdi homolog Mpd1 and two members of the PPIases (*FPR4* and *CPR6*) are only up-regulated in the engineered strain, but not upon DTT addition.

One of the most striking patterns is the significant up-regulation of a large number of genes with functions in ribosomal biogenesis (233 genes assigned to the GO-categories 'ribosome biogenesis and assembly' and 'RNA metabolic process'). Most of these genes are contributing to rRNA processing (RRP family) and ribosome subunit nuclear export and assembly, while the ribosomal proteins (RPS and RPL families) themselves are not among the regulated genes for *P. pastoris* (see Additional file 2). No genes with a function in mRNA decay show increased transcription levels. The induction of the above functional categories came as a surprise, as translational down-regulation of proteins involved in ribosomal biogenesis was recently reported when *S. cerevisiae* cell were treated with DTT [31]. In contrast, the transcription levels of 9 out of the 16 mRNAs listed by these authors are enhanced in our study. Transcriptional down-regulation of ribosomal proteins during ER stress conditions was also revealed when reanalysing the raw data provided by Travers et al. [17]. However, Shaffer et al. [30] describe an increase in total protein synthesis as well as in the number of assembled ribosomes upon the overexpression of the mammalian Hac1 homolog XBP1 in Raji cells, but did not observe upregulation of genes related to ribosome biogenesis. A similar effect was observed after XBP1 overexpression in CHO-K1 cells [32]. These results may be an indication that the positive effect of overexpression of the UPR transcription factor on heterologous protein production [33,34,16,35] results not just from stimulation of folding and secretion of proteins but also their synthesis. The induction of protein folding related genes upon Hac1 overexpression is in line with the literature on UPR effects, while an impact on organelle biosynthesis other than ER and Golgi has so far only been described for mammalian cells.

The stimulatory effects of XBP1 induction on ribosomes and organelle synthesis in mammalian cells like lymphocytes have been attributed to their function as dedicated protein factories. On the other hand the UPR in lower eukaryotes should rather serve to alleviate the load of unfolded, aggregation prone protein. It will be of interest in the future to investigate whether Hac1 stimulates ribosome biogenesis in other yeasts and fungi as well, and whether this leads to increased translation.

In this context, it is worthwhile to mention the induction of two pathways leading to the unusual post-translationally modified amino acid derivatives diphthamide and hypusine which are exclusively found in eukaryotic translation elongation factors 2 (eEF2) and 5 (eEF5), respectively [36,37]. As these biosynthetic pathways are rather complex, and outstanding in the otherwise downregulated group of 'amino acid biosynthesis', this induction underlines the increased demand for protein synthesis.

Furthermore, we observe that ER stress leads to increased transcription of genes coding for the large and small subunits of the mitochondrial ribosomes (*MRPS*, *RSM* and *MRPL* families), mitochondrial translation initiation and elongation factors (*IFM1*, *MEF1*, *MEF2*) and mitochondrial DNA polymerase (*MIP1*). Several essential constituents of the mitochondrial inner membrane presequence translocase (*TIM* family) are also up-regulated, indicating increased necessity for protein import into the mitochondria. Similarly, XBP1 was shown to increase mitochondrial mass and function in two types of mammalian cells [30].

While previous studies analysing UPR regulation mainly focus on up-regulated genes [17], more than half of the genes identified in our study to be regulated are strongly down-regulated (at least 1.5 fold). As can be seen in Figure 4, anabolic processes such as vitamin production, amino

acid and aromatic compound biosynthesis, heterocycle metabolic processes, carbohydrate, lipid and cofactor metabolism are among the most prominent repressed classes in both DTT-treated as well as Hac1-overproducing cells. The down-regulation of energy consuming biosynthetic pathways emerges as a general picture during ER stress conditions. However, it becomes obvious that the response to the folding perturbation agent DTT strongly differs from constitutive UPR induction by Hac1-overproduction. Especially the prominent down-regulation of genes belonging to 'electron transport' and 'cellular respiration' can easily be explained by the strong reducing capacities of DTT. Prominent members of the mitochondrial inner membrane electron transport chain such as subunits of the cytochrome c oxidase (*COX4*, *COX4*, *COX5A*, *COX13*) and the ubiquinol cytochrome-c reductase complex (*COR1*, *QRC6*, *QRC7*, *QRC9*, *RIP1*) are significantly repressed upon DTT treatment. Additionally, cytochrome c (*CYC1*), cytochrome c1 (*CYT1*) and cytochrome c heme lyase (*CYC3*) are only under DTT-dependent repression (GO: 'generation of precursor metabolites and energy'). The reducing features of DTT are most probably also the reason for the up-regulation of genes involved in the upkeeping of 'cellular homeostasis' and clearly, addition of DTT is provoking a 'response to a chemical stimulus'.

Down-regulated genes appearing in both Hac1 and DTT in the 'protein modification' group focus on protein kinases (*CDC5*, *CDH1*, *DBF2*) and components of the ubiquitinylation complex (*BUL1*, *CUL3*) involved in cell cycle regulation driving the cells towards mitotic exit (*CDC5*, *CDH1*, *MOB1*). These effects are even more pronounced in the Hac1-strain, where several more histone modifying enzymes as well as cycline-dependent protein kinases and components of the protein kinase C signalling pathway show reduced transcription levels compared to the wild type. Unlike reported for the filamentous fungi *T. reesei* [7] and *A. nidulans* [26], genes encoding the histones H2A, H2B, H3 and H4 appear to be down-regulated upon secretion stress in *P. pastoris*.

No clear picture emerges regarding the regulation of 'lipid metabolism': While sterol and ergosterol biosynthesis tend to be inhibited, the production of sphingolipid precursor substances is enhanced. On the other hand, a down-regulation of the major cell wall constituents (β-1,3 glucanases *BGL2* and *EXG1*, cell wall mannoproteins *CCW12*, *CWP2* and *TPI1*, GPI-glycoproteins *GAS1* and *SED1*, *PST1*) and genes coding for proteins required for the transport of cell wall components to the cell surface (*SBE22*) is manifest. Taken together, these results indicate a significant remodelling process regarding the *P. pastoris* cell envelope during ER stress conditions.

Interestingly, the major groups of metabolic genes were down-regulated upon Hac1 overexpression, indicating a decrease of the supply of metabolites. However, it should be noted that no reduction of the specific growth rate was observed as compared to the wild type strain ($\mu = 0.37$ and $0.39\ h^{-1}$, respectively). A reduction of metabolic processes, and amino acid synthesis in particular, is contradictory to translation stimulation. Further research will be needed to elucidate the overall regulatory pattern of UPR in respect to protein synthesis.

Conclusion

Additional gene finding and annotation added to the available data for *P. pastoris* lead to a list of approximately 4,000 genes with a putative identification of their function, and 11,000 more potential open reading frames. An oligonucleotide probe set was designed, the hybridization results were evaluated for reproducibility, and results from a biologically relevant analysis were tested for meaningfulness. In a direct comparison to *S. cerevisiae* employing DTT treatment for UPR induction, 45 out of 93 genes reacted similarly. The differences thus observed between *P. pastoris* and *S. cerevisiae* underline the importance of DNA microarrays for industrial production strains. *HAC1* overexpression in *P. pastoris* obviously leads to induction of many genes involved in translation: most genes of ribosome biogenesis, as well as many related to RNA metabolism and translation were up-regulated, an effect that has never been observed in yeasts and filamentous fungi so far.

The upregulation of ribosomal biogenesis, RNA metabolism, translation, and organelle biosynthesis is specific for *HAC1* overexpression and not observed with DTT treatment, while the latter leads specifically to the upregulation of genes related to chemical stimulus, and the downregulation in the groups electron transport and respiration, so that these reactions have to be regarded as specific for the treatment with a reducing agent rather than UPR regulated.

Methods

Gene Prediction and Sequence Selection

Gene prediction and the selection of sequences for oligonucleotide probes were based on sequenced contigs of the *P. pastoris* genome including predictions of protein coding genes, available through Integrated Genomics [8]. The number of predicted genes was 5,425 of which 3,680 had an assigned function. The ORFs were made up of experimentally identified genes, as well as ORFs predicted by a proprietary gene finder [38].

To validate and possibly improve these predictions, *de novo* gene finding was conducted. First three *de novo* gene

finder (GeneMark, Glimmer3, GlimmerHMM) were tested on the genome sequence of *S. cerevisiae* (data from BioMart, [39]) to evaluate their performance on yeast genomes. As described in Results and Discussion, GeneMark [40] was selected for further gene prediction on the *P. pastoris* genome sequence. To run the gene prediction it was necessary to train GeneMark on *S. cerevisiae* by building a matrix with transition probabilities for coding and non-coding regions used by the Hidden Markov Model (HMM) of the program. With the amount of data available we were able to generate a matrix of the 7th order. The genes of *P. pastoris* were predicted using the *S. cerevisiae* matrix and the lowest possible probability score cut-off (t = 0.05). In the initial stage of the microarray design the aim was to predict as many putative ORFs as possible. In this context a higher false positives rate was accepted in order to keep the false negatives rate as low as possible.

The predicted sequences were merged with data from IG and clustered by running cd-hit [20] with a similarity cut-off of 90%. For all of the resulting sequences a BLASTX search was done against *S. cerevisiae* using WU-BLAST [19]. Blast data was further filtered for length (cutoff 55 bp) and low prediction score. Clusters comprised of more than one gene were represented by the longest sequence, or curated manually, if appropriate.

From this first gene list (PpaV1) microarrays were analyzed as described below. Spots with a positive signal were determined using the mean plus one standard deviation of the negative control probes as a cut-off. Sequences were selected if they were positive in at least 8 out of 12 arrays. This criterion was chosen to fill the array capacity. Additionally all sequences with a probability score higher than 0.5 or having an annotation were kept for the second set of sequences (PpaV2).

Annotation

For the PpaV2 sequence set the program cd-hit-est [20] was used to find all ORFs that had a global identity of > 80% with *S. cerevisiae*. WU-BLASTX and WU-TBLASTN searches were conducted against *S. cerevisiae*, using a low complexity filter and $E < 10^{-7}$. For all the sequences that did not have a match with *S. cerevisiae* under these conditions the two BLAST searches were repeated against the SwissProt/TrEMBL [41] database. A perl script was developed to summarize and compare the BLAST results.

Oligo Design and Array platform

Oligos for the PpaV1 sequences were designed with the Program OligoArray 2.1 [21] to match the melting temperature distribution of Agilent's *S. cerevisiae* oligos on the Yeast Oligo Microarray (V2), design number 013384.

The oligo-set for the PpaV2 sequence set was designed using the thermodynamic model-based oligoset optimizer 'TherMODO'. This tool incorporates advanced quantitative models for probe-target binding region accessibility and position-dependent target labelling efficiency, and replaces the common greedy search algorithm by a global set optimization step, achieving high discrimination power for particularly uniform probe sets [23]. Probes for Agilent arrays are limited to a maximum length of 60 nucleotides by the manufacturing process. For increased flexibility in the probe design, the oligoset design optimization considered probes ranging in length from 57 to 60 nucleotides.

These arrays were produced on Agilent 60 mer oligonucleotide high density arrays 4 × 44 K (with 42,034 available features) for PpaV1 and 8 × 15 K (with 15,208 available features) for PpaV2.

Experimental Design

For the first batch of arrays a same-same design was used, employing six replicates each of Pool 1 and of Pool 2. The aim of this experiment was to determine which of the probes hybridize to *P. pastoris* targets. For the second batch of arrays a two-state comparison set up was chosen with 6 replicates for each experiment of which 3 were dye swapped.

Strains und Cultures

For the first batch of arrays the aim was to determine which of the predicted probes hybridize with targets from *P. pastoris*. To make sure that many genes were active it was important to pool samples from various conditions of the cells. Samples were taken from two different *P. pastoris* strains, X-33 and CBS2612, grown on different media and taken at both exponential and stationary growth phase. The media were YP Medium (1% yeast extract, 2% peptone and either 2% glucose, 2% glycerol or 0.5% methanol as carbon source), Buffered Minimal Medium (1.34% yeast nitrogen base, 4×10^{-5}% biotin, 100 mM potassium phosphate pH 6.0 and either 2% glucose, 2% glycerol or 0.5% methanol as carbon source), and Buffered Minimal Medium described above supplemented with amino acids (0.005% of L-glutamic acid, L-methionine, L-lysine, L-leucine and L-isoleucine). The samples were combined into two pools with Pool 1 containing 18 samples from the exponential growth phase and Pool 2 containing 18 samples from the stationary phase. Both pools additionally contained seven chemostat samples of the strain X-33 3H6Fab, grown as in [42].

For the UPR experiments, strains GS115 HAC1, constitutively overproducing the activated form of *S. cerevisiae* Hac1, as described in Gasser et al. [33,4], as well as GS115

transformed with the empty vector pGAPHIS (a histidine prototrophic isogenic strain of GS115) were cultivated in YPD (YP as above with glucose) at 28°C. After growing the cultures to an OD_{600} = 5.7, dithiothreitol (2.5 mM) was added where appropriate. After 1 more hour of cultivation, 1 ml culture was added to 0.5 ml precooled phenol solution (5% in absolute ethanol) and centrifuged immediately for 30 sec at 13.000 rpm. After discarding the supernatants the pellets were frozen at -80°C.

RNA Isolation

All samples were resuspended with 1 mL TRI Reagent (Sigma). Cells were disrupted after addition of 500 μL glass-beads with a Thermo Savant Fastprep FP120 Ribolyzer by treatments of 2 × 20 sec at 6.5 ms^{-1}. RNA was extracted with chloroform, precipitated with isopropanol, washed with 75% ethanol and dissolved with diethylpyrocarbonate treated water. The extracted RNAs were quantified via absorption at 260 and 280 nm. The quality of the RNA samples was verified with the Agilent Bioanalyzer 2100 and RNA 6000 Nano Assay kit (Agilent Technologies, California).

Labeling and Hybridization

Hybridization targets for *P. pastoris* microarrays were prepared according to Agilent's Two-Color Microarray-Based Gene Expression Analysis protocol (Version 5.5, February 2007). Purification of the labelled and amplified RNA was conducted using RNeasy mini spin columns (Qiagen). The quality of labelled cRNA was evaluated on the Agilent Bioanalyzer 2100 and quantified using a ND-1000 NanoDrop Spectrophotometer. Fragmented cRNA samples were applied to the individual arrays. The slides were placed into Agilent hybridization oven and hybridized for 17 h, at 65°C and 10 rpm.

Microarray Analysis

Slides were scanned with an Agilent MicroArray Scanner and intensities were extracted using Agilent's Feature Extraction software (version 9.1). The resulting data was imported into R where data pre-processing and normalization was performed. In the pre-processing step all outliers and saturated spots were given the weight zero. After plotting the data we decided to refrain from background correction since it has the tendency to add more noise to the data [43]. The data were normalized using locally weighted MA-scatterplot smoothing (LOESS) followed by a between array scale normalization. Both functions are available within the limma package of R [44]. For the selection of differentially expressed genes linear models were fitted to the log-ratios of the expression data separately for each gene. An empirical Bayes approach was used to shrink the probe-wise sample variances towards a common value yielding a moderated *t*-statistic per gene [45]. *P*-values were corrected for multiple testing using Holm's method [46]. Features were defined as differentially expressed if they had a *p*-value < 0.05. For the identification of stronger regulatory effects an additional cut-off for the fold change (FC) of 1.5 > FC > 1/1.5 was applied. Description of the platform, array, raw data as well as processed data were deposited at ArrayExpress [47] under the accession numbers A-MEXP-1157.

All annotated *P. pastoris* genes were categorized into GO biological process terms using the SGD GO slim tool [28], whereby *P. pastoris* specific genes were included into the term 'other'. The significance of a deviation of the number of up- or downregulated genes in each group from the average was verified with a Fisher test (Additional file 4).

Authors' contributions

AG performed gene finding and annotation, statistical data analysis, supported data evaluation, and drafted part of the manuscript. BG performed data evaluation, supported annotation, study design, array design and drafted part of the manuscript. MD performed the cultivations and hybridizations. MS contributed to study design, annotation and array design. GGL developed the employed probe design tool and supported the array design. TT developed the quantitative model for the position-dependent target labelling efficiency and adapted it for the relevant end-primed labelling protocol. DPK supervised gene identification and annotation, supervised and contributed to the development of the employed probe design tool, and contributed to the manuscript. DM conceived of the study, and participated in data evaluation and manuscript drafting. All authors read and approved the final manuscript.

Additional material

Additional File 1

Thermodynamic properties of the TherMODO probe design compared to probes designed through Agilent's eArray. *Distribution of Gibbs free energy ΔG (A) and the probe-target melting temperature T_m (B) of the oligo sets. The upper row (1) shows the oligos designed through eArray and the lower row (2) the oligos designed with TherMODO. PpaV2 is the name of the second set of sequences as described in the Materials and Methods section.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-9-390-S1.pdf]

Additional File 2

Differential expression values of all annotated genes upon DTT treatment and* HAC1 *overexpression. *Differential expression values and adjusted* p*-values of all annotated genes of* P. pastoris*, denominated with the gene name of their respective* S. cerevisiae *homolog. Genes that were tested with TRAC as a different method for transcript quantification are highlighted in bold letters. Legend of headers: id – internal unique identifier of sequence; sequ_id – ERGO identifier (RPPA.) or gene finder identifier (orf.) respectively; DTT_logFC – log_2 fold change of DTT treatment compared to control;* HAC1 *logFC – log_2 fold change of* HAC1 *overexpression compared to control; Gene name – Standard gene name or if missing systematic ORF name according to* S. cerevisiae *nomenclature; GO – Gene Ontology term (for descriptions see additional file 4). If a gene is present in a certain GO group it has a 1 in the respective column, if not it has a 0.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-390-S2.xls]

Additional File 3

Volcano plots of fold change vs. adjusted* p*-values. *(A) DTT treatment; (B)* HAC1 *overexpression. Blue line:* p*-value cut-off* p > *0.05; red lines: optional fold change cut-off FC > 1.5.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-390-S3.pdf]

Additional File 4

Fisher's exact test of the up/down regulated gene groups upon DTT treatment and* HAC1 *overexpression. *Fisher's exact test was applied to test significance of the up- and downregulated gene groups displayed in figure 4.* p_{adj} *values are given for each GO group. Legend of headers: group – Gene Ontology term; Description – Gene Ontology description; odds.ratio – measure of independence between variables; adj.p – Holm adjusted* p*-value;* HAC1 *up/down – up/down regulated in* HAC1 *overexpression experiment; DTT up/down – up/down regulated in DTT experiment. The first work sheet represents results using only a* p*-value cut-off* p > *0.05, the second work sheet represents results using a* p*-value cut-off* p > *0.05 and a fold change cut-off FC > 1.5.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-390-S4.xls]

Acknowledgements

This work was supported by the Austrian Science Fund (project No. I37-B03), the European Science Foundation (programme EuroSCOPE), and the Austrian Research Promotion Agency (programme FHplus).

The Vienna Science Chair of Bioinformatics gratefully acknowledges support by the Vienna Science and Technology Fund (WWTF), Baxter AG, Austrian Research Centres (ARC) Seibersdorf, and Austrian Centre of Biopharmaceutical Technology (ACBT). TT acknowledges partial funding by the GenAu BIN-II PhD programme.

References

1. Ye RW, Wang T, Bedzyk L, Croker KM: **Applications of DNA microarrays in microbial systems.** *J Microbiol Methods* 2001, **47(3):**257-272.
2. Sauer M, Branduardi P, Gasser B, Valli M, Maurer M, Porro D, Mattanovich D: **Differential gene expression in recombinant Pichia pastoris analysed by heterologous DNA microarray hybridisation.** *Microb Cell Fact* 2004, **3(1):**17.
3. Rautio JJ, Kataja K, Satokari R, Penttila M, Soderlund H, Saloheimo M: **Rapid and multiplexed transcript analysis of microbial cultures using capillary electophoresis-detectable oligonucleotide probe pools.** *J Microbiol Methods* 2006, **65(3):**404-416.
4. Gasser B, Maurer M, Rautio J, Sauer M, Bhattacharyya A, Saloheimo M, Penttilä M, Mattanovich D: **Monitoring of transcriptional regulation in *Pichia pastoris* under protein production conditions.** *BMC Genomics* 2007, **8:**179.
5. Sims AH, Robson GD, Hoyle DC, Oliver SG, Turner G, Prade RA, Russell HH, Dunn-Coleman NS, Gent ME: **Use of expressed sequence tag analysis and cDNA microarrays of the filamentous fungus *Aspergillus nidulans*.** *Fungal Genet Biol* 2004, **41(2):**199-212.
6. Bachem CW, van der Hoeven RS, de Bruijn SM, Vreugdenhil D, Zabeau M, Visser RG: **Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development.** *Plant J* 1996, **9(5):**745-753.
7. Arvas M, Pakula T, Lanthaler K, Saloheimo M, Valkonen M, Suortti T, Robson G, Penttila M: **Common features and interesting differences in transcriptional responses to secretion stress in the fungi *Trichoderma reesei* and *Saccharomyces cerevisiae*.** *BMC Genomics* 2006, **7:**32.
8. Genomics I: **ERGO bioinformatics suite.** [http://ergo.integratedgenomics.com/ERGO/].
9. Besemer J, Borodovsky M: **GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses.** *Nucleic Acids Res* 2005, **33(Web Server issue):**W451-4.
10. Majoros WH, Pertea M, Salzberg SL: **TigrScan and Glimmer-HMM: two open source ab initio eukaryotic gene-finders.** *Bioinformatics* 2004, **20(16):**2878-2879.
11. Mathé C, Sagot MF, Schiex T, Rouzé P: **Current methods of gene prediction, their strengths and weaknesses.** *Nucleic Acids Res* 2002, **30(19):**4103-4117.
12. Majoros WH, Pertea M, Salzberg SL: **Efficient implementation of a generalized pair hidden Markov model for comparative gene finding.** *Bioinformatics* 2005, **21(9):**1782-1788.
13. Pedersen JS, Hein J: **Gene finding with a hidden Markov model of genome structure and evolution.** *Bioinformatics* 2003, **19(2):**219-227.
14. Kreil DP, Russell RR, Russell S: **Microarray oligonucleotide probes.** *Methods Enzymol* 2006, **410:**73-98.
15. Erasmus DJ, van der Merwe GK, van Vuuren HJ: **Genome-wide expression analyses: Metabolic adaptation of *Saccharomyces cerevisiae* to high sugar stress.** *FEMS Yeast Res* 2003, **3(4):**375-399.
16. Valkonen M, Penttilä M, Saloheimo M: **Effects of inactivation and constitutive expression of the unfolded- protein response pathway on protein production in the yeast Saccharomyces cerevisiae.** *Appl Environ Microbiol* 2003, **69(4):**2065-2072.
17. Travers KJ, Patil CK, Wodicka L, Lockhart DJ, Weissman JS, Walter P: **Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation.** *Cell* 2000, **101(3):**249-258.
18. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M: **Gene identification in novel eukaryotic genomes by self-training algorithm.** *Nucleic Acids Res* 2005, **33(20):**6494-6506.
19. University W: **WU-BLAST.** [http://blast.wustl.edu/].
20. Li W, Godzik A: **Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences.** *Bioinformatics* 2006, **22(13):**1658-1659.
21. Rouillard JM, Zuker M, Gulari E: **OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach.** *Nucleic Acids Res* 2003, **31(12):**3057-3062.
22. Arvas M, Kivioja T, Mitchell A, Saloheimo M, Ussery D, Penttila M, Oliver S: **Comparison of protein coding gene contents of the fungal phyla Pezizomycotina and Saccharomycotina.** *BMC Genomics* 2007, **8:**325.
23. Leparc GG, Tuechler T, Striedner G, Bayer K, Sykacek P, Hofacker I, Kreil DP: **Model based probe set optimization for high-performance microarrays.** *Nucleic Acids Research* 2008, **submitted:**.
24. Agilent: **eArray.** [https://earray.chem.agilent.com/earray/].
25. Guillemette T, van Peij NN, Goosen T, Lanthaler K, Robson GD, van den Hondel CA, Stam H, Archer DB: **Genomic analysis of the**

secretion stress response in the enzyme-producing cell factory *Aspergillus niger*. *BMC Genomics* 2007, **8:**158.
26. Sims AH, Gent ME, Lanthaler K, Dunn-Coleman NS, Oliver SG, Robson GD: **Transcriptome analysis of recombinant protein secretion by *Aspergillus nidulans* and the unfolded-protein response *in vivo*.** *Appl Environ Microbiol* 2005, **71(5):**2737-2747.
27. Rossanese OW, Soderholm J, Bevis BJ, Sears IB, O'Connor J, Williamson EK, Glick BS: **Golgi structure correlates with transitional endoplasmic reticulum organization in *Pichia pastoris* and *Saccharomyces cerevisiae*.** *J Cell Biol* 1999, **145(1):**69-81.
28. SGD: **SGD Gene Ontology Slim Mapper.** [http://db.yeastgenome.org/cgi-bin/GO/goSlimMapper.pl].
29. Higashio H, Kohno K: **A genetic link between the unfolded protein response and vesicle formation from the endoplasmic reticulum.** *Biochem Biophys Res Commun* 2002/08/15 edition. 2002, **296(3):**568-574.
30. Shaffer AL, Shapiro-Shelef M, Iwakoshi NN, Lee AH, Qian SB, Zhao H, Yu X, Yang L, Tan BK, Rosenwald A, Hurt EM, Petroulakis E, Sonenberg N, Yewdell JW, Calame K, Glimcher LH, Staudt LM: **XBP1, downstream of Blimp-1, expands the secretory apparatus and other organelles, and increases protein synthesis in plasma cell differentiation.** *Immunity* 2004, **21(1):**81-93.
31. Payne T, Hanfrey C, Bishop AL, Michael AJ, Avery SV, Archer DB: **Transcript-specific translational regulation in the unfolded protein response of *Saccharomyces cerevisiae*.** *FEBS Lett* 2008.
32. Tigges M, Fussenegger M: **Xbp1-based engineering of secretory capacity enhances the productivity of Chinese hamster ovary cells.** *Metab Eng* 2006.
33. Gasser B, Maurer M, Gach J, Kunert R, Mattanovich D: **Engineering of Pichia pastoris for improved production of antibody fragments.** *Biotechnol Bioeng* 2006, **94(2):**353-361.
34. Gasser B, Sauer M, Maurer M, Stadlmayr G, Mattanovich D: **Transcriptomics-based identification of novel factors enhancing heterologous protein secretion in yeasts.** *Appl Environ Microbiol* 2007/09/04 edition. 2007, **73(20):**6499-6507.
35. Valkonen M, Ward M, Wang H, Penttilä M, Saloheimo M: **Improvement of foreign-protein production in *Aspergillus niger* var. awamori by constitutive induction of the unfolded-protein response.** *Appl Environ Microbiol* 2003, **69(12):**6979-6986.
36. Mattheakis LC, Sor F, Collier RJ: **Diphthamide synthesis in *Saccharomyces cerevisiae*: structure of the DPH2 gene.** *Gene* 1993, **132:**149-154.
37. Wolff EC, Kang KR, Kim YS, Park MH: **Posttranslational synthesis of hypusine: Evolutionary progression and specificity of the hypusine modification.** *Amino Acids* 2007, **33(2):**341-350.
38. Overbeek R, Larsen N, Walunas T, D'Souza M, Pusch G, Selkov E Jr., Liolios K, Joukov V, Kaznadzey D, Anderson I, Bhattacharyya A, Burd H, Gardner W, Hanke P, Kapatral V, Mikhailova N, Vasieva O, Osterman A, Vonstein V, Fonstein M, Ivanova N, Kyrpides N: **The ERGO genome analysis and discovery system.** *Nucleic Acids Res* 2003, **31(1):**164-171.
39. (EBI) EBI, (CSHL) CSHL: **BioMart.** [http://www.biomart.org/biomart/martview/].
40. Technology GI: **GeneMark.** [http://exon.gatech.edu/GeneMark/].
41. (SIB) SIB: **Swiss-Prot/TrEMBL.** [http://www.expasy.ch/sprot/].
42. Baumann K, Maurer M, Dragosits M, Cos O, Ferrer P, Mattanovich D: **Hypoxic fed batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins.** *Biotechnol Bioeng* 2008, **100(1):**177-83.
43. Zahurak M, Parmigiani G, Yu W, Scharpf RB, Berman D, Schaeffer E, Shabbeer S, Cope L: **Pre-processing Agilent microarray data.** *BMC Bioinformatics* 2007, **8:**142.
44. Smyth GK, Speed T: **Normalization of cDNA microarray data.** *Methods* 2003, **31(4):**265-273.
45. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3:**Article3.
46. Holm S: **A Simple Sequentially Rejective Bonferroni Test.** *Scandinavian Journal of Statistics* 1979:65 -670.
47. EMBL-EBI: **ArrayExpress.** [http://www.ebi.ac.uk/microarray-as/aer/].

COMMUNICATION TO THE EDITOR

BIOTECHNOLOGY and BIOENGINEERING

Hypoxic Fed-Batch Cultivation of *Pichia pastoris* Increases Specific and Volumetric Productivity of Recombinant Proteins

Kristin Baumann,[1] Michael Maurer,[2,3] Martin Dragosits,[2] Oriol Cos,[1] Pau Ferrer,[1] Diethard Mattanovich[2,3]

[1]Departament d'Enginyeria Química, Escola Tècnica Superior d'Enginyeria, Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès), Spain

[2]Department of Biotechnology, Institute of Applied Microbiology, University of Natural Resources and Applied Life Sciences, Vienna, Austria; telephone: +431360066569; fax: +4313697615; e-mail: diethard.mattanovich@boku.ac.at

[3]School of Bioengineering, University of Applied Sciences FH Campus Vienna, Austria

Received 26 October 2007; revision received 13 November 2007; accepted 26 November 2007

Published online 13 December 2007 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.21763

ABSTRACT: High cell density cultivation of *Pichia pastoris* has to cope with several technical limitations, most importantly the transfer of oxygen. By applying hypoxic conditions to chemostat cultivations of *P. pastoris* expressing an antibody Fab fragment under the GAP promoter, a 2.5-fold increase of the specific productivity q_P at low oxygen supply was observed. At the same time the biomass decreased and ethanol was produced, indicating a shift from oxidative to oxidofermentative conditions. Based on these results we designed a feedback control for enhanced productivity in fed batch processes, where the concentration of ethanol in the culture was kept constant at approximately 1.0% (vv^{-1}) by a regulated addition of feed medium. This strategy was tested successfully with three different protein producing strains, leading to a three- to sixfold increase of the q_P and threefold reduced fed batch times. Taken together the volumetric productivity Q_P increased 2.3-fold.

Biotechnol. Bioeng. 2008;100: 177–183.

KEYWORDS: oxygen limitation; *Pichia pastoris*; antibody fragment; trypsinogen; productivity

Introduction

The design of bioprocesses has to meet both economical and technical requirements. One of the most limiting parameters, especially during large scale high cell density cultivations of microorganisms, is the maximal oxygen transfer rate (Porro et al., 2005). This is also the case for the methylotrophic yeast *Pichia pastoris*, a popular host for the production of recombinant protein which enables the utilization of different metabolic pathways. In the methanol pathway, the oxidation of methanol with molecular oxygen is the first step of both energy production and carbon assimilation. Surplus of oxygen may lead to accumulation of formaldehyde, especially during the induction phase on methanol. It was assumed by most authors that high oxygen supply and a high level of dissolved oxygen tension (DOT) are needed during the production phase, and often pure oxygen is employed for aeration (Lee et al., 2003). Charoenrat et al. (2005) have shown that oxygen limitation during fed batch may be employed to limit the growth, thus maintaining a higher methanol concentration, which may further induce the AOX1 promoter used to express the heterologous gene. However, this approach still runs under fully aerobic conditions (which are required by the methanol metabolism), and the oxygen-limited situation leads to very low specific growth rates and long process times.

Alternatively, the glycolytic glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter is used to produce proteins on glucose. When using this expression system, q_P increases asymptotically to a maximum value with increasing μ (Maurer et al., 2006). An analogous behavior has been also reported for the AOX1 promoter for *P. pastoris* cells growing on methanol (Ohya et al., 2005). Hence the ratio of q_P to μ, which corresponds to the yield of product per biomass (Y_{PX})

K. Baumann and M. Maurer contributed equally to this work.

Correspondence to: D. Mattanovich

Contract grant sponsor: European Science Foundation (Program EuroSCOPE)
Contract grant sponsor: Complementary Actions Plan and Project and Integrated Actions Plan of the Spanish Ministry of Science and Education
Contract grant number: BIO 2005-23733-E and HU2005-0001, respectively
Contract grant sponsor: Ministry of Innovation and Universities of the Generalitat de Catalunya
Contract grant number: 2005-SGR-00698
Contract grant sponsor: Xarxa de Referència en Biotecnologia and doctoral fellowship for K.B.
Contract grant sponsor: Austrian Science Fund
Contract grant number: I37-B03
Contract grant sponsor: Austrian Exchange Service

decreases with increasing μ, so that low product titers are achieved at higher growth rates. Maurer et al. (2006) have developed a model to describe growth and product formation, and to optimize the feeding profile of glucose limited fed batch cultures to increase volumetric productivity. However, this model is based on fully aerobic conditions, and growth is solely controlled by a feed forward strategy of the fed batch medium, following a predefined feed rate profile.

The impact of oxygen supply on the production of heterologous proteins and disulfide bridge formation has not been investigated rigorously yet. Gross et al. (2006) describe insights into mechanisms for maintaining disulfide bond formation under anaerobic conditions in the endoplasmic reticulum by regenerating oxidized Ero1p, a key enzyme in the disulfide bond formation pathway in eukaryotic cells in both aerobic and anaerobic environments. Oxygen limitation in *Escherichia coli* cultures leads to a stress response (Schweder et al., 1999) and by-product formation such as acetate, which inhibit recombinant protein production (Bauer et al., 1990). Wiebe et al. (2007) investigated the impact of oxygen depletion in *Saccharomyces cerevisiae*. They provide a general survey of the changes in metabolite levels as well as transcript levels pointing out that it is not possible to linearly interpolate between aerobic and anaerobic conditions. Also, the impact of oxygen limitation on recombinant protein production was not approached.

Growth on methanol is an aerobic process, so that oxygen limitation leads to growth limitation irrespective of methanol supply (Charoenrat et al., 2005). On glucose, however, *P. pastoris* shows a facultative anaerobic behavior so that oxygen limitation leads to by-product (mainly ethanol) fermentation. Therefore oxygen limitation on glucose does not limit substrate consumption but represents a switch to an alternative metabolic pathway. Based on the obtained results, we further developed a feedback controlled fed batch process for enhanced productivity which was tested successfully with three different protein producing strains.

Materials and Methods

Strains

For the construction of the *P. pastoris* strain expressing the human monoclonal antibody 3H6 Fab fragment, the light and heavy chain genes were prepared as described by Gach et al. (2007), cloned under control of the constitutive GAP promoter with the *S. cerevisiae* alpha-mating factor signal sequence for secretion, and transformed into the strain X-33, as described in Gasser et al. (2006). The *P. pastoris* strain producing human trypsinogen was described by Hohenblum et al. (2003), and the strain for porcine trypsinogen was prepared accordingly, both utilizing the GAP promoter and the alpha-mating factor signal sequence.

Cultivation

Chemostat

The chemostat cultivations were performed at a working volume of 1 L in a 2 L bench-top bioreactor (Biostat B, Braun Biotech, Melsungen, Germany). One liter shake flask containing 300 mL of YPG medium (per liter: 10 g yeast extract, 20 g peptone, 10 g glycerol) was inoculated with a 1.0 mL cryostock of the recombinant *P. pastoris* cell bank. The culture was grown for approximately 24 h at 30°C with shaking at 200 rpm, and used to inoculate the reactor to an optical density (OD_{600}) of 1.0. After a batch period of approximately 24 h the continuous culture started at a dilution rate of $D = 0.1\ h^{-1}$ by adjusting the flow rate to $100\ g\,h^{-1}$.

The cultivation parameters were set to 25°C and 700 rpm, and the pH was controlled at 5.0 with ammonium hydroxide. The total gas flow was kept constant for all experiments at 1.5 vvm (volume gas per volume batch medium and minute), however, for cultures that received less than 20.97% O_2 in the gas stream, air was partially replaced with an equivalent volume of nitrogen to create oxygen limiting and respiro-fermentative (hypoxic) conditions. The set points were changed from high to low air concentrations, resulting in a stepwise reduction of the oxygen concentration in the inlet air from 20.97% to 10.91%, 8.39% and 5.87% (vv^{-1}). In three independent experiments, the continuous cultivation was performed for at least five residence times τ before taking samples and shifting the inlet air-flow to the next set point, except for the hypoxic conditions (5.87% or 8.39% O_2) where we observed wash out after 3.5 residence times and therefore implemented the sampling step after at least three residence times.

The media were as follows:

If not otherwise stated all chemicals were purchased from Merck Eurolab (Darmstadt, Germany).

Batch medium contained per liter: 39.9 g glycerol, 1.8 g citric acid, 12.6 g $(NH_4)_2HPO_4$, 0.022 g $CaCl_2 \cdot 2H_2O$, 0.9 g KCl, 0.5 g $MgSO_4 \cdot 7H_2O$, 2 mL Biotin ($0.2\ g\,L^{-1}$; Sigma, Munich, Germany), 4.6 mL PTM1 trace salts stock solution. The pH was set to 5.0 with 25% HCl.

Chemostat medium contained per liter: 50 g glucose $\cdot$ $1H_2O$, 0.9 g citric acid, 4.35 g $(NH_4)_2HPO_4$, 0.01 g $CaCl_2 \cdot 2H_2O$, 1.7 g KCl, 0.65 g $MgSO_4 \cdot 7H_2O$, 1 mL Biotin ($0.2\ g\,L^{-1}$), and 1.6 mL PTM1 trace salts stock solution. The pH was set to 5.0 with 25% HCl.

PTM1 trace salts stock solution contained per liter: 6.0 g $CuSO_4 \cdot 5H_2O$, 0.08 g NaI, 3.0 g $MnSO_4 \cdot H_2O$, 0.2 g $Na_2MoO_4 \cdot 2H_2O$, 0.02 g H_3BO_3, 0.5 g $CoCl_2$, 20.0 g $ZnCl_2$, 65.0 g $FeSO_4 \cdot 7H_2O$, and 5.0 mL H_2SO_4 (95–98%). The chemicals for PTM1 trace salts stock solution were from

Riedel-de Haën (Seelze, Germany), except for H_2SO_4 (Merck Eurolab).

Fed Batch Cultivation

The fed batch cultivations were carried out in a 40 L stirred tank bioreactor (MBR, Wetzikon, Switzerland) with a computer based process control (ISE, Vienna, Austria). The temperature was controlled at 25°C and the pH was controlled at 5.0 with addition of 25% ammonium hydroxide. The concentration of ethanol in the outlet air was continuously analyzed using a gas analyzer type GS 2/6 (Vogelbusch, Vienna, Austria), which was linked to the process control unit to control the feed during hypoxic fed batch cultivation. The ethanol signal was calibrated (at process parameter settings: stirrer speed 1,200 rpm; aeration rate 1 vvm; and temperature 25°C) by addition of aliquots of ethanol to the reactor in 17 L water.

The media were as follows:

The batch medium was the same as used in the chemostat experiments.

Glucose fed batch solution contained per liter: 550 g glucose $\cdot H_2O$, 10 g KCl, 6.45 g $MgSO_4 \cdot 7H_2O$, 0.35 g $CaCl_2 \cdot 2H_2O$ and 12 mL PTM1 trace salts stock solution.

One liter pre-culture was prepared as described above, harvested by centrifugation at 3,000*g* for 10 min and resuspended in 500 mL sterile batch medium. This inoculum was used to inoculate the starting volume (17.0 L) of the bioreactor to an optical density (OD_{600}) of 1.0.

After approximately 24 h the batch was finished and—depending on the cultivation strategy—the feed was started.

The standard cultivation strategy was a fed batch with a constant feed, this means that the batch phase was followed by the glucose fed batch with a feed rate $F = 161.7\ \mathrm{g\,h^{-1}}$. During these cultivations the dissolved oxygen concentration was maintained above 20% saturation by controlling the stirrer speed between 600 and 1,200 rpm, whereas the air-flow was kept at 1 vvm. The cultivation was terminated at an approximate fed batch time $t = 90$ h. Samples were taken frequently and processed as described below.

The hypoxic cultivation strategy (O_2—limited fed batch) consists of different phases to adjust the desired phenomena. During the hypoxic cultivation the dissolved oxygen concentration was controlled by a cascade regulation between 600 and 1,200 rpm, but the maximum oxygen transfer rate was limited by the air-flow rate of $1{,}200\ \mathrm{L\,h^{-1}}$.

The batch phase was followed by a pre-programmed exponential feed phase, where Equation (1) was used to calculate the feed profile.

$$F = \frac{\mu}{\Delta s Y_{X/S}} X_0 e^{\mu t} \qquad (1)$$

For the calculation of the flow rate F a growth rate μ of $0.2\ \mathrm{h^{-1}}$, a utilizable substrate concentration Δs of $0.5\ \mathrm{g\,g^{-1}}$ and a substrate yield coefficient of $0.54\ \mathrm{g\,g^{-1}}$ was used. The exponential feed was kept until an oxygen limit (0% dissolved oxygen) and an ethanol concentration of 1.0% ($\mathrm{vv^{-1}}$) in the culture broth was achieved. This was the starting point of the third phase where a feedback control based on the ethanol concentration in the culture regulates the flow rate. An ethanol concentration of approximately 1.0% ($\mathrm{vv^{-1}}$) was kept constant by regulated addition of feed medium. At actual levels of ethanol below the set point the PI controller increased the flow rate, while at higher ethanol concentrations the flow rate was decreased. The end point of the cultivation is the total addition of 14.5 kg feed medium. Samples were also taken frequently and processed as described below.

Analytics

Optical Density

The cell concentration was measured by diluting the samples in ddH_2O up to 1:500 to measure the OD at 600 nm.

Biomass Determination

Culture samples (2×5 mL) were collected by centrifugation and the supernatants frozen for further analysis. The pellets were washed in 10 mL ddH_2O, centrifuged again, resuspended in 10 mL ddH_2O, transferred to a weighed beaker, and dried at 105°C until constant weight.

Quantification of Product Titers

To determine the Fab content, 96-well microtiter plates (MaxiSorb, Nunc, Denmark) were coated with anti-hIgG (Fab specific, 1:1,000 in PBS, pH 7.4) overnight at room temperature (RT, 22–25°C), before serially diluted supernatants of cultivation samples containing 3H6 Fab were applied and incubated for 2 h at RT. A papain digested purified normal human Fab/kappa from IgG (Bethyl, Montgomery, TX) was used as a standard protein at a starting concentration of $200\ \mathrm{ng\,mL^{-1}}$. After each incubation step the plates were washed three times with PBS containing 1% Tween-20 adjusted to pH 7.4. 100 μL of anti-kappa light chain—alkaline phosphatase conjugate as secondary antibody (1:1,000 in PBS/Tween + 2% bovine serum albumin) were added to each well, and incubated for 1 h at RT. After washing, the plates were stained with pNPP ($1\ \mathrm{mg\,mL^{-1}}$ *p*-nitrophenyl phosphate in coating buffer, 0.1 N $Na_2CO_3/NaHCO_3$; pH 9.6) and read at 405 nm (reference wavelength 620 nm).

The trypsinogen quantification was performed as described before (Hohenblum et al., 2004b) with a trypsin kinetic assay using *p*-tosyl-L-arginine methyl ester (TAME) after activation with bovine enterokinase.

Alcohol and Substrate

Glucose and ethanol were analyzed by HPLC (Hewlett Packard 1050, Boeblingen, Germany) analysis using an ion exchange column Aminex HPX-87H (Bio Rad, Barcelona, Spain). The mobile phase was 15 mM sulfuric acid.

Cellular viability was determined by propidium iodide staining, measured by flow cytometry as described by Hohenblum et al. (2003).

Results

Chemostat

The chemostat cultivations were performed in triplicates and showed a good reproducibility. The mean standard deviation of all measured parameters (biomass and ethanol concentration, and q_P) between the biological replicas was 12%. The oxygen concentrations in the inlet air represent following conditions: fully aerobic (20.97%), limited aerobic (10.91%), and hypoxic (8.39%, 5.87%). In both limited and hypoxic conditions the DOT was zero, whereas in fully aerobic conditions DOT was about 45%. Limited aerobic conditions still allow fully aerobic glucose metabolism while no oxygen remains in the culture, whereas hypoxic conditions lead to partially fermentative metabolism. Biomass concentration, specific productivity q_P and ethanol concentration in the culture supernatant, plotted against the oxygen supply in the inlet air, are illustrated in Figure 1. *P. pastoris* cells showed a fully respiratory metabolism under oxygen excess and limiting conditions, but when hypoxic conditions were reached, biomass decreased twofold and ethanol was produced at significant levels ($5.82 \pm 0.33\ g\,L^{-1}$), that is, indicating a shift to an oxidofermentative metabolism. The q_P shows no significant change in all aerobic conditions, whether oxygen was in excess or limited, but, interestingly, increases 2.5-fold upon shifting to hypoxic conditions. It should be noted that steady state could only be maintained for three residence times under hypoxic conditions. A possible reason for this is the sterol and unsaturated fatty acids depletion under hypoxic conditions. However, according to the findings in Wiebe et al. (2007), where the level of intracellular metabolites was shown to stabilize after 30 h at $D = 0.1$, three residence times will be sufficient to reach a physiological equilibrium.

Based on the unexpected increase of q_P, we decided to design and test a fed batch protocol under hypoxic conditions as a potential strategy to increase process productivities.

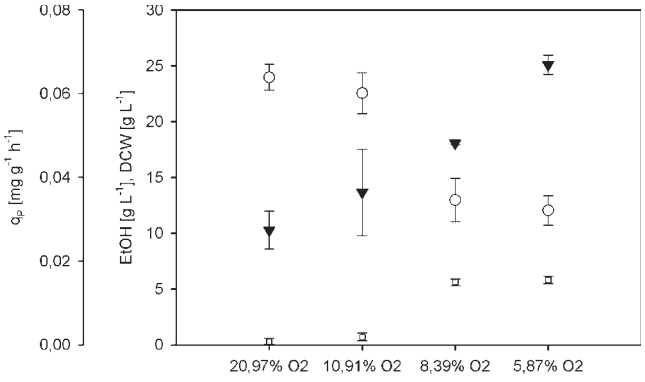


Figure 1. Comparison of dry cell weight (DCW, ○), q_P (▼) and ethanol concentrations (□) at different oxygen supply levels.

Fed Batch

Two cultivation strategies were compared, the standard strategy with constant feed resulting in carbon limited conditions, and a hypoxic strategy based on a feedback feed control ensuring a permanent low level of fermentative metabolism controlled by the equilibrium of oxygen transfer rate and glucose feed rate. Both cultivation strategies had the same termination criterion, the maximal volume of the reactor, so that the same amount of feed medium was applied in both cases. Kinetics of growth and product secretion are shown in Figure 2. The standard fed batch shows a nearly linear increase in biomass and product, while specific growth rate and specific productivity q_P decrease, as typically observed with constant feed protocols. The long feed phase of 85–100 h is characteristic for standard *P. pastoris* cultures. On the contrary, the hypoxic feed control led to a drastically reduced feed time of 29.9 h with a 13% higher product titer, while the biomass concentration reached only 71% of the standard cultivation. The specific growth rate did not differ significantly from the standard cultivation during the same time period of 29.9 h, but q_P was about twofold higher, which is in accordance with the results obtained in the chemostat (Fig. 2B). The ethanol concentration varied between 4.1 and 14.5 $g\,L^{-1}$ (0.52–1.8% vv^{-1}, average 1.16% vv^{-1}; Fig. 2A). Glucose was not detectable in any of the samples.

For the evaluation of a new process strategy it is essential to prove the concept with more than one heterologous protein, as expression characteristics may vary significantly. Therefore two further mammalian proteins, porcine and human trypsinogen, were produced with both the standard process and the hypoxic feed strategy, respectively. The results of these cultivations were highly comparable to the 3H6 Fab results, with 2.9-fold lower feed time, 11% higher product titer and 72% of the biomass concentration, as compared to the standard fed batch (Table I).

The calculations of trends, specific rates and yield coefficients were performed by using suitable smoothing routines (Matlab 7.1 Curvefit Toolbox, The Mathworks, Inc., Matlab, Natick, MA) and mass balances that permitted to calculate complete data sets based on coincident off-line data, as described recently (Cos et al., 2005).

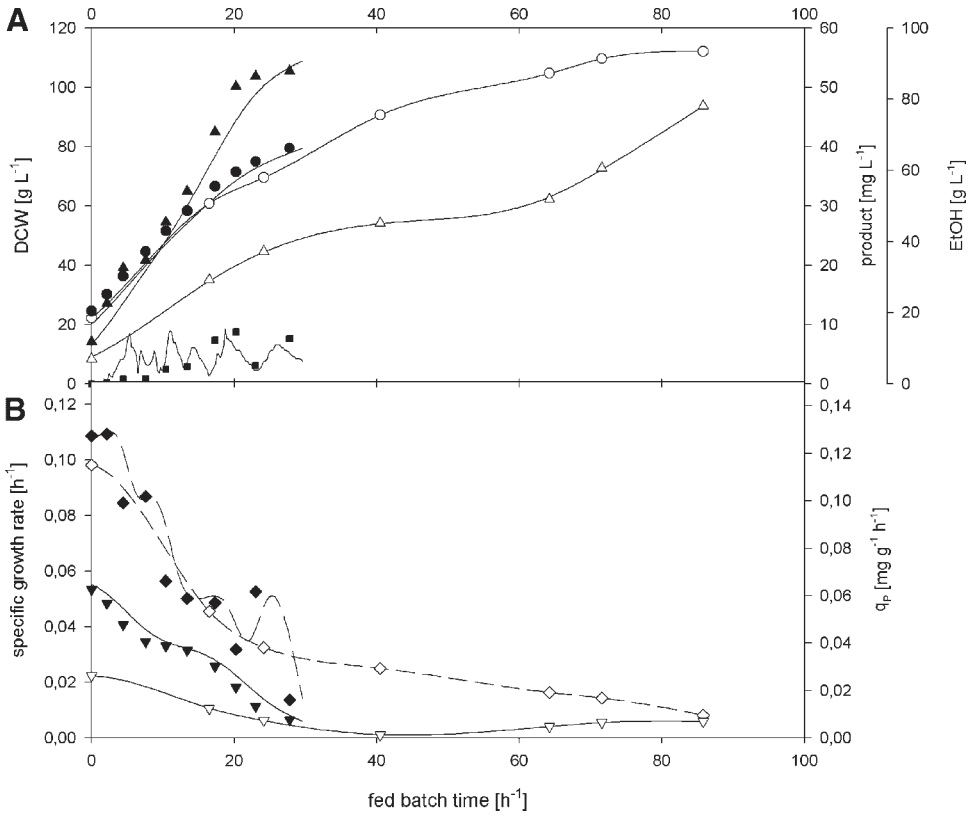


Figure 2. Comparison of the standard fed batch (open symbols) and hypoxic fed batch (closed symbols) of the 3H6 Fab fragment. **A**: Dry cell weight (DCW, ○), 3H6 Fab concentration (Δ), ethanol concentration in the supernatant (■) and ethanol concentration in the off gas (thick line). **B**: Specific growth rate μ (◇) and specific production rate q_P (▽).

Discussion

A strong positive effect of hypoxic conditions on specific productivity was observed, while fully aerobic but oxygen limited conditions did not have a significant effect. These hypoxic conditions differ from approaches with limited oxygen supply as described by Charoenrat et al. (2005), Khatri and Hoffmann (2006), and Trentmann et al. (2004), where the aim was to limit methanol metabolism by a kind of substrate limitation (with oxygen as the limited substrate), so that growth could be limited but a higher methanol concentration maintained for strong induction of the AOX1 promoter. As the glycolytic flux is higher in the hypoxic cultivations, it can be assumed that glycolytic genes like GAP will be upregulated. Thus it is tempting to speculate that heterologous genes controlled by the GAP promoter are upregulated in the hypoxic conditions as well. However, previous work indicates that the secretion of proteins with complex structures is rather limited at the level of secretion than transcription (Gasser et al., 2006, 2007; Hohenblum et al., 2004a).

The results obtained in chemostat cultures clearly indicate that a fed batch cultivation at the limit of aerobic to fermentative metabolism could be beneficial for the production of heterologous proteins, at least in the configuration of a constitutive promoter on glucose. While based on different metabolic principles, the process control requirements resemble the classical baker's yeast process, which should run at the boarder to fermentation to induce the ethanol pathway (Verduyn et al., 1992). Thus a control strategy known from baker's yeast production was employed, namely controlling the feed with an alcohol sensor in order to avoid extensive fermentation but maintaining the oxidofermentative conditions. However, it should be noted that *P. pastoris* is a Crabtree negative yeast, so that fermentation will only be induced when oxygen is severely limited. Consequently oxygenation was limited, so that the feed control was based on the equilibrium of oxygen transfer rate and glucose feed rate.

The major advantage of this new feed control is the drastically decreased time period of the feed phase. While the specific growth rate kinetics resembles that of the standard fed batch, the specific productivity is significantly higher than in fully aerobic conditions, resembling the data obtained in the chemostat cultures. It becomes obvious that the secretion of heterologous proteins benefits from initially

Table I. Summary of final product titers, final biomass concentrations, mean specific growth rates μ, mean as well as maximum specific production rates q_P, volumetric productivities Q_P, substrate yields $Y'_{X/glucose}$, and feed times during standard and hypoxic fed batch cultivations, performed with three different model proteins, namely the antibody fragment 3H6 Fab, human trypsinogen, and porcine trypsinogen.

Recombinant protein	Fermentation strategy	Max. Fab concentration [mg L^{-1}] trypsin concentration [U L^{-1}]	Max. yeast dry matter [g L^{-1}]	μ mean [h^{-1}]	Mean q_P [mg Fab g^{-1} h^{-1}] [U trypsin g^{-1} h^{-1}]	Max q_P [mg Fab g^{-1} h^{-1}] [U trypsin g^{-1} h^{-1}]	Q_P [mg Fab L^{-1} h^{-1}] [U trypsin L^{-1} h^{-1}]	$Y'_{X/glucose}$ [g g^{-1}]	Fed batch time [h]
3H6 Fab	Standard fed batch	46.85	112.07	0.0317	0.0079	0.0263	0.43	0.459	85.80
3H6 Fab	Hypoxic fed batch	52.75	79.46	0.0603	0.0330	0.0627	0.94	0.338	29.90
Human trypsinogen	Standard fed batch	20.08	96.99	0.0178	0.0026	0.0041	0.17	0.433	93.75
Human trypsinogen	Hypoxic fed batch	22.34	79.35	0.0675	0.0166	0.0240	0.43	0.351	25.50
Porcine trypsinogen	Standard fed batch	79.20	95.23	0.0201	0.0227	0.0295	0.79	0.428	97.30
Porcine trypsinogen	Hypoxic fed batch	86.95	59.76	0.0428	0.0671	0.1438	1.82	0.265	29.65

high growth, followed by decreasing specific growth rates, as demonstrated by optimization of the growth, and feed profile of fed batch (Maurer et al., 2006). While the growth profile over time does not differ significantly between hypoxic and standard fed batch during the first 30 h, the standard process continues for much longer time at low growth, so that the average specific growth rate of the hypoxic cultivations was 1.9-fold higher than that of the standard process. The mean q_P of all three proteins was 4.5 ± 1.7-fold higher in the hypoxic processes, and, most importantly the volumetric productivities Q_P (total product per total volume and time) in all cases were 2.36 ± 0.16-fold higher with the hypoxic fed batch strategy.

Additionally, the partially fermentative metabolism leads to a decreased biomass yield coefficient, so that less biomass is accumulated during the process, resulting in facilitated cell removal, which is a major challenge with high cell density cultures (Wang et al., 2006). As a low dissolved oxygen set point with limited oxygen concentration in the culture results in an increased oxygen transfer rate due to a higher concentration gradient, lower air-flow and stirrer speed are required to supply enough oxygen to the culture. Taken together, the lower aeration rate required for the hypoxic cultivation protocol represents another technological advantage, as aeration (oxygen transfer) is another limit of large scale high density cultivation.

Conclusions

Based on the observation that hypoxic conditions lead to an increased specific productivity of protein secreting *P. pastoris*, a hypoxic fed batch protocol was developed. It differs from the previously described oxygen limited fed batch in the sense that growth is still controlled by carbon limitation, while oxygen limitation is applied to modulate metabolism and heterologous protein productivity. It could be shown in all three cases tested so far that this hypoxic fed batch leads to increased specific productivity, but more importantly to twofold increased volumetric productivity. Additional technological benefits are lower aeration rates and lower final biomass concentrations. In summary, hypoxic fed batch represents an interesting alternative production strategy for heterologous proteins with several important technological advantages over the standard fed batch strategy.

This work has been supported by the European Science Foundation (program EuroSCOPE), the Project BIO2005-23733-E of the Complementary Actions Plan and Project HU2005-0001 of the Integrated Actions Plan of the Spanish Ministry of Science and Education, as well as the Ministry of Innovation and Universities of the Generalitat de Catalunya (2005-SGR-00698, Xarxa de Referència en Biotecnologia, and pre-doctoral fellowship for KB), the Austrian Science Fund (FWF, project I37-B03) and the Austrian Exchange Service (ÖAD). The authors thank Karola Vorauer-Uhl for the HPLC analysis.

References

Bauer KA, Ben-Bassat A, Dawson M, de la Puente VT, Neway JO. 1990. Improved expression of human interleukin-2 in high-cell-density fermentor cultures of Escherichia coli K-12 by a phosphotransacetylase mutant. Appl Environ Microbiol 56(5):1296–1302.

Charoenrat T, Ketudat-Cairns M, Stendahl-Andersen H, Jahic M, Enfors SO. 2005. Oxygen-limited fed-batch process: An alternative control for Pichia pastoris recombinant protein processes. Bioprocess Biosyst Eng 27(6):399–406.

Cos O, Serrano A, Montesinos JL, Ferrer P, Cregg JM, Valero F. 2005. Combined effect of the methanol utilization (Mut) phenotype and gene dosage on recombinant protein production in Pichia pastoris fed-batch cultures. J Biotechnol 116(4):321–335.

Gach JS, Maurer M, Hahn R, Gasser B, Mattanovich D, Katinger H, Kunert R. 2007. High level expression of a promising anti-idiotypic antibody fragment vaccine against HIV-1 in Pichia pastoris. J Biotechnol 128(4):735–746.

Gasser B, Maurer M, Gach J, Kunert R, Mattanovich D. 2006. Engineering of Pichia pastoris for improved production of antibody fragments. Biotechnol Bioeng 94(2):353–361.

Gasser B, Sauer M, Maurer M, Stadlmayr G, Mattanovich D. 2007. Transcriptomics-based identification of novel factors enhancing heterologous protein secretion in yeasts. Appl Environ Microbiol 73(20):6499–6507.

Gross E, Sevier CS, Heldman N, Vitu E, Bentzur M, Kaiser CA, Thorpe C, Fass D. 2006. Generating disulfides enzymatically: Reaction products and electron acceptors of the endoplasmic reticulum thiol oxidase Ero1p. Proc Natl Acad Sci USA 103(2):299–304.

Hohenblum H, Borth N, Mattanovich D. 2003. Assessing viability and cell-associated product of recombinant protein producing Pichia pastoris with flow cytometry. J Biotechnol 102(3):281–290.

Hohenblum H, Gasser B, Maurer M, Borth N, Mattanovich D. 2004a. Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant Pichia pastoris. Biotechnol Bioeng 85(4):367–375.

Hohenblum H, Vorauer-Uhl K, Katinger H, Mattanovich D. 2004b. Bacterial expression and refolding of human trypsinogen. J Biotechnol 109(1–2):3–11.

Khatri NK, Hoffmann F. 2006. Oxygen-limited control of methanol uptake for improved production of a single-chain antibody fragment with recombinant Pichia pastoris. Appl Microbiol Biotechnol 72(3):492–498.

Lee CY, Lee SJ, Jung KH, Katoh S, Lee EK. 2003. High dissolved oxygen tension enhances heterologous protein expression by recombinant Pichia pastoris. Process Biochem 38(8):1147–1154.

Maurer M, Kühleitner M, Gasser B, Mattanovich D. 2006. Versatile modeling and optimization of fed batch processes for the production of secreted heterologous proteins with Pichia pastoris. Microb Cell Fact 5:37.

Ohya T, Ohyama M, Kobayashi K. 2005. Optimization of human serum albumin production in methylotrophic yeast Pichia pastoris by repeated fed-batch fermentation. Biotechnol Bioeng 90(7):876–887.

Porro D, Sauer M, Branduardi P, Mattanovich D. 2005. Recombinant protein production in yeasts. Mol Biotechnol 31(3):245–259.

Schweder T, Kruger E, Xu B, Jurgen B, Blomsten G, Enfors SO, Hecker M. 1999. Monitoring of genes that respond to process-related stress in large-scale bioprocesses. Biotechnol Bioeng 65(2):151–159.

Trentmann O, Khatri NK, Hoffmann F. 2004. Reduced oxygen supply increases process stability and product yield with recombinant Pichia pastoris. Biotechnol Prog 20(6):1766–1775.

Verduyn C, Postma E, Scheffers WA, Van Dijken JP. 1992. Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. Yeast 8(7):501–517.

Wang A, Lewus R, Rathore AS. 2006. Comparison of different options for harvest of a therapeutic protein product from high cell density yeast fermentation broth. Biotechnol Bioeng 94(1):91–104.

Wiebe MG, Rintala E, Tamminen A, Simolin H, Salusjarvi L, Toivari M, Kokkonen JT, Kiuru J, Ketola RA, Jouhten P, Huuskonen A, Maaheimo H, Ruohonen L, Penttilä M. 2007. Central carbon metabolism of Saccharomyces cerevisiae in anaerobic, oxygen-limited and fully aerobic steady-state conditions and following a shift to anaerobic conditions. FEMS Yeast Res.

7. Curriculum Vitae

Dragosits Martin

Grafenschachen 301, 7423 Austria

E-mail martin.dragosits@gmail.com

Date of birth 29.08.1982

Education

2006-2009	PhD Thesis 'The Interrelation of Environmental Factors and Heterologous Protein Production in *Pichia pastoris* – A Systems Biology Approach' University of Natural Resources and Applied Life Sciences Vienna, Department of Biotechnology Supervisor: Ao Prof. DI Dr Diethard Mattanovich.
2005	Diploma (MSc) Graduation with honors Diploma Thesis 'Detection of microbial contamination of foods via optical nanosensors.' University of Vienna, Department of Biochemistry and Attophotonics Biosciences GmbH. Supervisor: Ao Prof. Mag Dr Thomas Schalkhammer
2001-2005	Studies of Biology (Microbiology and Genetics) at the University of Vienna, Austria
2000-2001	Military service at the JgB19, Pinkafeld, Austria
2000	General qualification for university entrance. Graduation with honors
1993-2000	Secondary School. Gymnasium Oberschützen, Austria
1990-1993	Primary School, Grafenschachen, Austria

Languages

German	mother tongue
English	fluently, both written and spoken
French	highschool level

Publications

Dragosits M, Stadlmann J, Albiol J, Baumann K, Maurer M, Gasser B, Sauer M, Altmann F, Ferrer P, Mattanovich D. The effect of temperature on the proteome of recombinant *Pichia pastoris*. J Proteome Res. 2009 Mar;8(3):1380-92.

Dragosits M, Stadlmann J, Graf A, Maurer M, Gasser B, Sauer M, Kreil DP, Altmann F, Mattanovich D. The unfolded protein response is involved in osmotolerance of *Pichia pastoris*. **submitted manuscript**

Dragosits M, Frascotti G, Bernard-Granger L, Vázquez F, Giuliani M, Baumann K, Rodríguez-Carmona E, Tokkanen J, Parrilli E, Wiebe MG, Kunert R, Maurer M, Gasser B, Sauer M, Branduardi P, Pakula T, Saloheimo M, Penttilä M, Ferrer P, Tutino ML, Villaverde A, Porro D, Mattanovich D. Comparative analysis of antibody fragment production in diverse microbial host cells**. in preparation**

Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, Kleinheinz M, Sauer M, Altmann F, Gasser B. Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*. Microb Cell Fact. 2009 Jun 2;8:29.

Graf A, Dragosits M, Gasser B, Mattanovich D. Yeast systems biotechnology for the production of heterologous proteins. FEMS Yeast Res. 2009 May;9(3):335-48. Epub 2009 Apr 1. Review.

Graf A, Gasser B, Dragosits M, Sauer M, Leparc GG, Tüchler T, Kreil DP, Mattanovich D. Novel insights into the unfolded protein response using *Pichia pastoris* specific DNA microarrays. BMC Genomics. 2008 Aug 19;9:390.

Gasser B, Saloheimo M, Rinas U, Dragosits M, Rodríguez-Carmona E, Baumann K, Giuliani M, Parrilli E, Branduardi P, Lang C, Porro D, Ferrer P, Tutino ML, Mattanovich D, Villaverde A. Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microb Cell Fact. 2008 Apr 4;7:11.

Baumann K, Maurer M, Dragosits M, Cos O, Ferrer P, Mattanovich D. Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins. Biotechnol Bioeng. 2008 May 1;100(1):177-83.

Patent Applications

Gasser B, Marx H, Dragosits M, Mattanovich D, Maurer M, Sauer M. 2009. Expression sequences. European patent application EP09161403.2

Professional Membership

Member of the European Federation of Biotechnology (EFB)