# Genome-scale analysis of *Pichia pastoris* during recombinant protein production and methanol utilization

## Dissertation

zur Erlangung des Doktorgrades an der Universität für Bodenkultur

eingereicht von
**Dipl. Ing. Markus Buchetics**

**Betreuer:**
**Univ. Prof. DI Dr. Diethard Mattanovich**

Department für Biotechnologie
Universität für Bodenkultur Wien

Wien, im August 2016

## Danksagung

Ich möchte den folgenden Personen ausdrücklich für das hiermit Erreichte danken:

Univ. Prof. Dr. Diethard Mattanovich dafür, dass er mir in meiner Zeit in seiner Arbeitsgruppe ein guter Förderer und so liebenswerter und geduldiger Chef war. Er hat so vielen so viel ermöglicht – so auch mir.

Priv. Doz. Dr. Brigitte Gasser für ihre durchgehende, professionelle, freundliche und unermüdliche Unterstützung in sämtlichen Phasen meiner Arbeit. Ich habe ihr viel zu verdanken.

Dr. Minoska Valli für ihre herzliche Art und endlose Geduld im Aufbereiten endloser Excel-Dateien. Ich vermisse sie bereits sehr als Bürokollegin.

Corinna Rebnegger dafür, dass sie mir immer eine so gute Gesprächspartnerin für Fachsimpelei aller Art war und in so mancher Notsituation geholfen hat.

Gabriele Wilt dafür, dass wir so viel gemeinsam im Labor zustande gebracht haben. Mit ihr zu arbeiten hat immer viel mehr Freude gemacht als alleine.

Dr. Hannes Rußmayer, Dr. Roland Prielhofer, Dr. Martin Pfeffer, Dr. Hans Marx, Dr. Martin Dragosits, Dr. Lars Töllner, Dr. Lukas Marsalek und Jonas Hohlweg dafür, dass sie als Arbeitskollegen und Gesprächspartner meinen Alltag bereichert haben.

Meinen Eltern für ihr Vertrauen und ihre Unterstützung.

Institutioneller Dank für das Ermöglichen meines Bildungsweges gilt dem ACIB, meiner Universität sowie allen beteiligten Einrichtungen dieser Republik bzw. der Europäischen Union.

Ich werde diese prägende Zeit und ihre Protagonisten niemals vergessen und für die Erfahrungen immer dankbar sein.

# Contents

# 1 Abstract

The microbial host for recombinant protein production, the methylotrophic yeast *Pichia pastoris*, is able to deliver high product titers but efficiency of secretion is hampered by cellular bottlenecks associated with folding and secretion but presumably also with energy and precursor metabolism.

New insights into the methylotrophic lifestyle were gained by the analysis of the transcriptome, proteome and metabolome of cultures grown on methanol/glycerol in comparison to glucose. *P. pastoris* was found to assimilate methanol entirely in its peroxisomes (also using peroxisome-targeted isoforms of otherwise cytosolic enzymes) and to have a highly active translation apparatus and amino acid metabolism, which could be responsible for higher productivities during growth on methanol. The comparative analysis of the transcriptome and the proteome of different *P. pastoris* production strains showed that heterologous protein production has a lesser impact on the proteome than the carbon source. Production of an antibody Fab fragment with the glycosylated alpha mating factor leader attached to it resulted in an up-regulation of genes required for glycosylation as well as folding and quality control of glycoproteins.

A novel transcription factor (Aft1) for the enhanced secretion of a recombinant carboxylesterase (up to 2.5 fold in fed batch production) was found by browsing the genome for transcription factor binding sites in the promoter regions of genes that had previously been shown to enhance secretion. Unlike its homolog in *S. cerevisiae*, Aft1 was shown to play a role in carbon response rather than in iron regulation. For future strain engineering work, small scale cultivation protocols were developed in an educated down-scaling approach.

# Zusammenfassung

Die methylotrophe Hefe *Pichia pastoris* findet Anwendung als Wirtsorganismus für rekombinante Proteinproduktion. Obwohl bereits relativ hohe Produkttiter für diverse Proteine erreicht wurden, scheint eine weitere Effizienzsteigerung durch intrazelluläre Limitierungen im Sekretionsweg oder im Energiestoffwechsel behindert zu werden.

Kulturen, die auf Methanol/Glycerin wuchsen, wurden verglichen mit solchen, die Glukose als einzige Kohlenstoffquelle nutzten, und analysiert auf Transkript-, Protein- und Metabolitenebene. Dadurch zeigte sich, dass in *P. pastoris* die Methanol-Aufnahme und Verwertung (bis zum Glyzerinaldehyd-3-Phosphat) zur Gänze in den Peroxisomen stattfindet. Eine aktivere Translationsmaschinerie sowie höherer Aminosäurenstoffwechsel wurden ebenfalls entdeckt. In Produktionsstämmen wurden weniger Auswirkungen auf Transkriptom und Proteom festgestellt. Bei der Produktion von Antikörper Fab Fragmenten, die mit dem Sekretionssignal des glykosylierten alpha-Paarungsfaktors (aus *S. cerevisiae*) versehen sind, wurde ein Hochregulieren von Faktoren zur Glykosylierung festgestellt.

Ein neuer Transkriptionsfaktor, Aft1, zur (bis zu 2,5-fach) verbesserten Sekretion einer rekombinanten Carboxylesterase konnte erstmalig beschrieben werden. Dazu wurde das *P. pastoris* Genom auf mögliche gemeinsame Transkriptionsfaktor-bindungsstellen untersucht. Die Suche konzentrierte sich dabei auf Promotorregionen von Genen, die zuvor in anderen Studien als erfolgreiche Sekretionshelfer durch begleitende Überexprimierung beschrieben worden waren. Anders als das Homolog in *S. cerevisiae* konnte Aft1 nicht der Regulierung des Eisenhaushalts zugeordnet werden. Eine Rolle in der zellulären Antwort auf die Verfügbarkeit einer Kohlenstoffquelle scheint wahrscheinlicher.

Für zukünftige Arbeiten zum Zweck von Sekretionsstudien wurden außerdem Kultivierungsprotokolle basierend auf physiologischen und technischen Erkenntnissen für den Kleinstmaßstab entwickelt.

# 2 Introduction

## 2.1 The demand for recombinant proteins

Numerous applications in the life sciences require specific proteins. DNA polymerases, reverse transcriptases, restriction endonucleases are the enzymatic work-horses in any molecular biology lab. Naturally occurring proteins, which are harvested from natural biological material, cannot provide all the needed proteins, both in terms of quantity and product characteristics. The biopharmaceutical market opts for recombinant therapeutic proteins with improved safety, reduced immunogenicity and of course customized specificity for their respective applications as e.g. enzymes, anticoagulants, protein scaffolds, hormones, blood or growth factors, interleukins, interferons, antibody-based drugs or Fc fusion proteins (Dimitrov 2012). In order to meet this demand, different production systems for recombinant proteins have been established and enhanced since the 1980s. *Escherichia coli* dominated the early market which gradually has expanded to alternative production hosts like yeasts, insect cells and especially mammalian cells -most notably Chinese Hamster Ovary (CHO) cell lines. Until now, about 400 recombinant protein drugs are on the market and other 1300 recombinant pharmaceuticals are under development of which one third is in clinical trials. Both prokaryotic and eukaryotic systems are constantly improving their properties as platforms for recombinant pharmaceutical production. With the well-established CHO system dominating the field in relative numbers, all production systems, both mammalian and non-mammalian, experienced increased use in absolute terms (Sanchez-Garcia et al. 2016).

### 2.1.1 The demand for suitable microbial production systems

Before the advent of recombinant expression systems, animal and human sources were indispensable for biopharmaceuticals. In the 1980s, with the development of recombinant insulin produced by *E.coli*, modern biotechnology and the use of microbial systems such as *E. coli* or yeast took off. By now, yeast is a competitor to mammalian cell culture systems in terms of the quality of the produced recombinant glycoproteins. Half of the world´s supply of insulin is produced in *S. cerevisiae*, with

*P. pastoris* also participating due to biosimilar product lines (Meehl and Stadheim 2014).

Both mammalian and yeast cells are capable of folding nascent polypeptides into active proteins along a secretory pathway to release them from the cell into the culture medium. The recovery of the active protein product which has been released to the extracellular space is a major advantage in terms of purification effort and costs. In this regard, cell lysis can pose a threat to extracellular product yields or protein activity as proteases, which might degrade the recombinant protein, are released from dead cells. In bacterial cultures, endotoxins can complicate protein purification as well. In contrast to mammalian cells, which are not designed by evolution to survive outside tissue compounds, microbial cells like bacteria or yeast are relatively robust in artificial environments, i.e. industrial-scale cultivation. They are less sensitive to external physical stress like shear forces or osmotic pressure. They replicate much faster, so fermentation times are only about the seventh part of e.g. Chinese hamster ovary (CHO) cells in a bioreactor. Apart from fast growth, the use of minimal media without the addition of complex molecules or amino acids is another key advantage in regard to cost-effectiveness and convenient handling during cultivation. The media for e.g. CHO cells can be ten times more expensive than microbial cell media.

### 2.1.2 Recombinant protein yields depend on the specific protein

Although predictions on how well a recombinant protein might be expressed seem difficult some protein characteristics were observed to influence yields of soluble and functionally active proteins after purification. Eukaryotic cells can produce proteins 200-250 kDa large proteins, a necessity when it comes to the production of e.g. monoclonal antibodies. Another factor is complexity, i.e. the number of protein subunits: Multidomain proteins often lack activity in *E. coli*. Eukaryotic cells, however, are better in secreting fully assembled multimeric products such as monoclonal antibodies consisting of four subunits. In general, the more subunits a protein has, the harder it is to get high yields of a functional product.

A protein in solution is much easier to recover than from aggregates, which can form due to suboptimal expression or purification conditions. When using the bacterial

system *E. coli*, inclusion body aggregates may form when high amounts of heterologous protein (esp. when of eukaryotic origin) accumulate intracellularly. It is not always possible to solubilize those inclusion bodies and even if it is feasible, refolding into a functionally active form is often unlikely.

Protein glycosylation motifs are specific for the production host species and they not only influence protein activity, but also functionality in the target environment via antigenicity or specific functions. Glycans are essential for pharmacological properties as stability, solubility, in vivo activity, pharmacokinetics, and immunogenicity. Glycosylation of e.g. antibody Fc is necessary for receptor-mediated activity. Small molecules are often removed rapidly from circulation due to their small size when they lack larger oligosaccharide patterns (Li and d'Anjou 2009). By nature, only mammalian cells are able to create glycosylation patterns similar to humans, but they are also hard to control (Ferrer-Miralles et al. 2009). In recent years, yeast species, in particular *P. pastoris*, have been genetically engineered to mimic human glycosylation. The proclaimed complete humanization of the glycosylation pathway in *P. pastoris* by Gerngross et al. is one of the most complex cellular engineering work achieved to date.

### 2.1.3 From invention to production: A challenge for manufacturers

Manufacturers of recombinant products make high demands on the production process itself: Scalability from pilot to industrial scale is desirable for process development. The process has to deliver consistent specific protein yields, the overall productivity should be reproducible and stable. This includes consistent protein recovery as protein purification is a crucial cost factor. Especially for biopharmaceuticals, the recombinant biologic needs to display the specified structure, comprising primary amino acid sequence, folded state and post translational modifications which are crucial for (non)immunogenicity. Significant differences need to be scientifically justified and examined in preclinical studies and clinical trials, usually taking place in three phases with each phase more complex and expensive than the previous one.

Unlike chemical process development, bioprocess development is very time-consuming and costly. Trial and error approaches to find the ideal parameters and operation points are often a matter of staff resources and competence, especially at

the interface between product and process development as well as between lab- and industrial scale production. Apart from the initial invention during the so-called discovery phase, achieved usually by scientists, the actual manufacturability is crucial for a marketable product. According to process developers, mostly with engineering background, there is a lack of awareness among product developers for the complexity of process upscaling.

Neubauer et al. emphasize that, especially in the biotech industry, new product development is highly interdisciplinary and has to combine many different kinds of tasks that all require experts in the respective fields. The different educational backgrounds of the product developers are the reason why the biotech industry is a prime example for so-called open innovation, where the innovation itself is generated by an interacting and dynamic network of partners. From a technical point of view, the development of a new product can be divided in screening phase, scale-up phase and manufacturing phase. However, these phases can overlap to a certain extent (Neubauer et al. 2013).

It is generally understood, that the cell as a biological system is very complex. Bioprocess engineering has to deal with gradual cellular adaptation to the environment. Thus, a cell´s constitution is influenced by the growth mode, its history and the cycle time of environmental variations. It is suggested by Noorman et al. that process development should pay tribute to the complex task of transferring the initial discovery to the industrial scale. A top to bottom approach starts the process development phase (and maybe even the product development phase) already from the perspective of the industrial scale (Noorman 2011).

A process engineer in the field of recombinant protein production is aware that the choice of feed medium not only has implications on the productivity of the system but on the process and its feasibilty. Growth kinetics, biomass concentrations, induction of protein production, etc. can cause manifold problems during the bioprocess. For example, in the industrial scale the power input achieved by stirrers can be a bottleneck for aeration of the culture. Insufficient mixing of the microbial culture broth results in inhomogeneities and media concentrations gradients (Neubauer et al. 2013).

## 2.2 *Pichia pastoris*: A workhorse of recombinant protein production

For yeasts, successful production of several recombinant proteins in the gram per liter range has been reported throughout the past decades, as exemplarily shown in Table 1.

**Table 1. Product titer achieved by *P. pastoris* (Paradise et al. 2011).**

| Yeast host | Product | Product titer |
|---|---|---|
| ***S. cerevisiae*** | *A. niger* glucose oxidase | 9 g/L (Demain and Vaishnav 2009) |
| ***P. pastoris*** | Mammalian recombinant protein, insulin precursor | 1.5 g/L (Wang et al. 2001) |
| ***P. pastoris*** | Intracellular Interleukin 2 | 30% protein by total DW (Cregg et al. 1993) |
| ***P. pastoris*** | Human serum albumin | 4 g/L (Cregg et al. 1993) |
| ***P. pastoris*** | Tumor necrosis factor | 6 g/L (Sreekrishana et al. 1999)<br>10 g/L (Sreekrishna et al. 1989) |
| ***P. pastoris*** | Recombinant proteins | 20-30 g/L (Morrow 2007) |

The methylotrophic yeast *Komagataella spp*, long known only under the established name *Pichia pastoris*, entered the industrial field in the early 1970s when Phillips Petroleum used it for the production of single cell protein (SCP) as animal feed additive. When the oil crisis made methanol (a by-product of oil-refinery) too expensive, it rendered it uneconomical for SCP production. Ten years later, the group of James Cregg made it popular again for heterologous protein expression using the strong, tightly regulated methanol-inducible *AOX1* promoter. Common fermentation strategies separate growth from production (with less growth) phase, enabling growth to high cell densities without the negative implications due to cellular stress that come with recombinant protein production. When the *AOX1* promoter is chosen for expression of the recombinant product gene, constitutive promoters are usually used to express selection markers. Promoters are still a considerable aspect for strain engineering. Varying promoter strength fitting to the growth and induction strategy (constitutively active, de-repressible, inducible) could not only further fine-

tune the expression of the recombinant product but also of helper proteins from metabolic or secretory pathways or the PTM machinery. Methanol-free induction strategies are also pursued by looking for alternative inducible promoters or by engineering the *AOX1* promoter itself (Takagi et al. 2008) Even stronger *AOX1* promoters variants (up to 160% of the native activity) were generated by (Hartner et al. 2008)) using synthetic promoter libraries.

Constitutive promoters such as the glyceraldehyde-3-phosphate promoter (GAP promoter) for the expression of the product gene simplify fed batch operations and avoid the toxic and flammable methanol. As glycolytically active promoter, the GAP promoter ensures product formation on a variety of carbon sources with glucose (highest activity), glycerol (half the activity relative to glucose) and even methanol (two thirds the activity relative to glucose) being the most commonly used substrates (Ahmad et al. 2014).

Recombinant genes are permanently integrated into the host genome by homologous recombination. This is achieved by using rather long homologous regions flanking the promoter-gene-terminator cassette in order to avoid non-homologous end-joining (Näätsaari et al. 2012). Secretion is mediated by the α-MF signal sequence, consisting of a pre-region for ER targeting and a pro-region for ER to Golgi transport (Ahmad et al. 2014).

It has been shown that *P. pastoris* clones with more than one copy of the recombinant gene expression cassette are usually able to produce more recombinant protein product (Marx et al. 2009), at least until limiting effects of protein secretion start to predominate further downstream of the secretory path, like the unfolded protein response (Hohenblum et al. 2004). Unwanted homologous recombination events can be stimulated by transcription, making DNA more susceptible to DNA damage and clones might get unstable. Elevated stress due to overexpression of the foreign gene might also result in a loss of copy numbers (Ohi et al. 1998). Multiple integrations of the foreign gene expression cassette are a rather rare events compared to integration of only one single copy. Thus, many clones will have to be screened to obtain multi-copy clones (Lin-Cereghino et al. 2008).

### 2.2.1 Main advantages of *P. pastoris*

Unlike *E. coli*, eukaryotic fungi like the yeast *P. pastoris* are capable of protein glycosylation and secretion. Proteins rich of disulfide bonds can also be accommodated by yeasts. For recombinant products such as antibodies (for therapy or diagnostics) *P. pastoris* is more preferable than a prokaryotic system. However, full immunological functionality is often impaired due to the lack of mammalian glycan structures. In biopharmaceutical production, safety concerns are a major contributor to overall costs. Yeast systems are able to deliver endotoxin- and virus-free product batches. Removal of pathogenic contaminants requires several (product yield diminishing) chromatography steps and is highly time- and cost-consuming. Other main advantages of *P. pastoris* are strain stability and the ability to grow to high cell density, coupled with high productivity (Paradise et al. 2011).

### 2.2.2 Products

In recent years, more and more products (some on the way to the market) were produced with the *P. pastoris* expression system. Research Corporation Technologies (RCT) licenses a commercial *P. pastoris* expression system. RCT claims that there are more than 70 products on the market or in a late stage of development. Table 2 shows some of the recombinant proteins produced with *P. pastoris*.

**Table 2. Marketed biopharmaceuticals produced in *Pichia pastoris* (adapted from RCT homepage, 10.07.2016, http://www.pichia.com/science-center/commercialized-products/)**

| Product | Company | Use |
|---|---|---|
| **Kalbitor® (DX-88 ecallantide, a recombinant kallikrein inhibitor protein)** | Dyax (Cambridge, MA) | Hereditary angioedema treatment |
| **Insugen® (recombinant human insulin)** | Biocon (India) | Diabetes therapy |
| **Medway (recombinant human serum albumin)** | Mitsubishi Tanabe Pharma (Japan) | Blood volume expansion |
| **Shanvac ™ (recombinant hepatitis B vaccine)** | Shantha/Sanofi (India) | Hepatitis B prevention |
| **Shanferon™ (recombinant interferon-alpha 2b)** | Shantha/Sanofi (India) | Hepatitis C & Cancer treatment |
| **Ocriplasmin (recombinant microplasmin)** | ThromboGenics (Belgium) | Vitreomacular adhesion (VMA) treatment |
| **Nanobody® ALX-0061 (recombinant anti-IL6 receptor single domain antibody fragment)** | Ablynx (Belgium) | Rheumatoid arthritis treatment |

| | | |
|---|---|---|
| **Nanobody® ALX00171 (recombinant anti-RSV single domain antibody fragment)** | Ablynx (Belgium) | Respiratory syncytial virus (RSV) infection treatment |
| **Heparin-binding EGF-like growth factor (HB-EGF)** | Trillium (Canada) | Treatment of interstitial cystitis/bladder pain syndrome (IC/BPS) treatment |
| **Purifine (recombinant phospholipase C)** | Verenium/DSM (San Diego, CA/Netherlands) | Degumming of high phosphorus oils |
| **Recombinant trypsin** | Roche Applied Science (Germany) | Digestion of proteins |
| **Recombinant collagen** | Fibrogen (San Francisco, CA) | Medical research reagents/dermal filler |
| **AQUAVAC IPN (recombinant infectious pancreatic necrosis virus capsid proteins)** | Merck/Schering Plough Animal Health (Summit, NJ) | Vaccines for infectious pancreatic necrosis in salmon |
| **Recombinant phytase** | Phytex, LLC (Sheridan, IN) | Animal feed additive |
| **Superior Stock recombinant nitrate reductase** | The Nitrate Elimination Co. (Lake Linden, MI) | Enzyme-based products for water testing and water treatment |
| **Recombinant human cystatin C** | Scipac (United Kingdom) | Research reagent |

In addition to industrial enzymes and biopharmaceuticals, the production of membrane proteins and of small antimicrobial peptides in *P. pastoris* might have the potential to replace chemical synthesis. Yet, is of utmost importance to improve protein secretion by strain and bioprocess engineeering, because product yields can vary massively depending on the specific recombinant protein to be secreted (Ahmad et al. 2014). Bioprocess optimization efforts can lead to a significant reduction of costs per dose of the biopharmaceutical produced with *P. pastoris* (Paradise et al. 2011).

Glycosylated full length antibodies are not mandatory for antigen recognition as Fv and Fab regions are the ones responsible for that. Fabs are able to penetrate tissue easier, but a missing Fc domain reduces stability. Consequently, microbial hosts are highly suitable for Fv and Fab products as they can be easier engineered and cultivated (Spadiut et al. 2014).

To this date, there are numerous *P. pastoris*-based biopharmaceuticals in clinical development, e.g. nanobody-based therapeutics, or have been approved by the FDA recently, e.g. ocrisplasmin, a protease for the treatment of vitreomacular adhesion, a disease affecting human vision (Meehl and Stadheim 2014).

### 2.2.3 Host strain engineering

Product titers and yields are still the main requirements for the economical production of a biopharmaceutical. When protein folding seems to be a limiting factor, e.g. for an antibody fragment, co-overexpression of chaperones like protein disulfide isomerase (Pdi) can lead to a significant increase in protein folding, hence titers (Damasceno et al. 2007). Overexpression of PDI or the transcription factor Hac1 (triggering the unfolded protein response) increased protein secretion for other products as well (Gasser et al. 2006; Gasser et al. 2007; Guerfal et al. 2010).

Recent and current optimization work deals with different engineering targets, ranging from enhancing/replacing secretion signals or the secretion leader sequence to the investigation of alternative promoters for methanol-free induction of protein expression. Genetic engineering could make the host even more efficient by removing rather redundant or yield hampering functions such as proteolytic decay (Ahmad et al. 2014). However, the crucial step during secretion, the release of recombinant protein from the cell, remains largely a mystery to the community and it is still to be found out why the budding yeast *P. pastoris* produces and secretes some proteins or protein classes better than others.

Many efforts have been taken in recent years to increase recombinant protein secretion in the yeast *P. pastoris*. A pre-requisite for this work was the open-access availability of genome sequences of *P. pastoris* wild-type strains (the original SCP production strain CBS7435 (Küberl et al. 2011), the first host strain developed for heterologous protein expression GS115 (De Schutter et al. 2009), as well as of the type strain *P. pastoris* DSMZ 70382 (Mattanovich et al. 2009b). In recent years, metabolic models have been published for *P. pastoris* as well (Caspeta et al. 2012; Chung et al. 2013; Chung et al. 2010; Irani et al. 2016; Sohn et al. 2010).

Unlike *S. cerevisiae*, *P. pastoris* shows less fermentative growth. Thus, cultures (when not oxygen limited) can reach very high cell densities, e.g. 160 g/L cell dry weight. For that reason, mixed-feed fed-batch cultivations are an advantage of *P. pastoris* over *S. cerevisiae* as this feed strategy can help to significantly reduce oxygen consumption, which translates to an easier aeration of the culture especially in large-scale.

*P. pastoris* might become a viable alternative expression system for the ever-growing antibody market when the previously addressed problems of immunologically detrimental post-translational modifications can be overcome. A startling approach is glycoengineering of the strains in order to get rid of non-human-like, i.e. immunogenic, glycosylations (Li et al. 2006; Vervecken et al. 2007; Vervecken et al. 2004).

In nature, yeast produce hyper-mannosylated glycans because – unlike in mammalian cells – the mannose sugar structures are not removed in the Golgi but rather further extended. This problem was solved by removal of the mannosyltransferase (*OCH1*) responsible for the initial mannose sugar branch extension. Combinatorial library approaches resulted in the successful introduction of several oligosaccharide transferases, finalized by the transfer of sialic acid onto the terminal ends of the glycoproteins (sialic acid ensures half-life extension of the glycoprotein). In this context, the addition of sugar and sialic acid transporters was necessary. It also needed adequate yeast localization signals and leader sequences fused to the newly introduced enzymes for the engineered glycosylation machinery (Hamilton and Gerngross 2007). Glycoengineered *P. pastoris* were also shown to work in surface display and cell sorting approaches: Different Fab (fragment of antigen binding) molecules with mammalian mannose-type $Man_5GlcNAc_2$ N-linked glycans was successfully displayed on the surface using glycoengineered *P. pastoris*. Subsequent cell sorting enriched the cell variants displaying higher amounts of the Fab or with higher affinity for the antigen (Lin et al. 2012).

## 2.3 Omics-driven large-scale studies

For many years, genetic research relied on the individual investigation of genes or sets of genes. The advent of the often called “Omics”-methods enabled larger-scale studies of cellular processes and has led to new insights in global regulatory networks. Once more, the already well-studied yeast *S. cerevisiae* was an early work-horse for systems biology research also because of its relatively small genome and well-established genetic manipulation techniques. In order to elucidate regulatory networks, high throughput protein-protein interaction studies - initially with yeast two-hybrid studies, then with subsequent affinity purification of tagged proteins and mass

spectrometry – and smaller-scale protein microarrays were the first (but method-biased and limited) efforts. DNA microarrays, and nowadays more and more RNA sequencing approaches (lacking cross hybridization problems) are powerful tools for expression profiling under different conditions, strains, growth stages or other experimental setups. Transcription factor binding sites could be identified by looking for common sequence motifs in promoter regions of mutually regulated genes.

The throughput in Systems Biology has increased more and more, both in terms of data collecting (a biologist´s task) as well as scoring of results and modeling of interactions or pathways (requiring a computational background). The outcome of Systems Biology research, such as a proposed model, can then again be further tested experimentally. More detailed knowledge of a production host´s genetics, biochemistry, physiology, production performance can be acquired faster and faster with high-throughput systems biology tools. A rapid characterization of cellular behavior in different situations is crucial for developments in the field of metabolic engineering. Industrial applications can be developed and brought to the market more likely (because faster) when it is possible to quickly characterize a microbial system in sufficient detail in order to assess its potential for different engineering strategies (Nielsen and Jewett 2008; Petranovic and Vemuri 2009).

Apart from uncovering biological occurrences for the purposes of industrial biotechnology, a major future application for global analyses of molecular components will be in the field of diagnostics and disease and treatment monitoring (Snyder and Gallagher 2009).

### 2.3.1 Gene expression analysis with DNA microarrays

For the parallel analysis of DNA homology and RNA concentrations for a whole genome with thousands of genes, DNA microarrays have been used over the past years with increasing efficiency and quality. So far, many microorganisms were subjected to gene expression analysis or the investigation of genome differences in order to better understand their metabolism and physiology. This development was accompanied by an increasing availability of complete microbial genomes (Ye et al. 2001).

Different probe types are used in different microarray platforms. At first, cDNA libraries were common, but due to a more reliable annotation, identity and hybridization characteristics of 50- to 70-mer oligonucleotides, oligonucleotide microarrays became more popular: Short (e.g. 60-mer probes in Agilent design) and specific DNA oligonucleotides are printed onto the array as probes (sometimes referred to as reporters or oligos) in picomole ($10^{-12}$ mol) amounts. The DNA sequences are synthesized directly onto the array surface. A gene is often represented by a probe set consisting of several probes to specifically match different regions of the gene. The arrays are then subjected to hybridization with labeled cDNA or cRNA (anti-sense RNA) sample (Woo et al. 2004). Oligonucleotide probes provide higher target specificity, experimental control of hybridization properties, and better discrimination of highly similar targets such as splice variants or gene families (Kreil et al. 2006).

Two-color microarrays offer a main advantage over one-color arrays: Hybridization takes place with cDNA prepared from two different samples, which therefore can be compared to each other simultaneously on the same array. This requires labeling with two different fluorescent dyes, such as Cy3 (fluorescence emission wavelength of 570 nm) and Cy5 (670 nm). After hybridization, a microarray scanner visualizes the fluorescence of both fluorophores after excitation with a laser beam of a defined wavelength (Shalon et al. 1996). An experimental design that incorporates dye swaps as well as a common reference is suggested for microarray data mining (Zahurak et al. 2007). Oligonucleotide microarrays often carry control probes designed to hybridize with RNA spike-ins. The extense of hybridization between the spike-ins and the control probes is used to normalize the hybridization measurements for the target probes (Tang et al. 2007). Before up-or down-regulated genes can be identified, the rough fluorescence intensities of the spots undergo normalization, which means to adjust microarray data for technical effects such as background noise or varying dye balance due to varying spot intensities or spatial position on the array (Smyth and Speed 2003). Local regression (LOESS) normalization without background subtraction delivers more reliable results than other methods (Zahurak et al. 2007).

### 2.3.2 Analysis of the proteome

It is well documented that mRNA levels are often unreliable for a consistent prediction of corresponding levels of active protein. Apart from technological limitations, biology itself contributes to experimental errors and difficult data interpretation. Among the main biological reasons for the deviations are mRNA splicing events (undetectable via microarray hybridization but with RNA-Seq), post translational modifications or allosteric protein interactions (Hegde et al. 2004).

At present, proteomics studies are capable of quantitative measurements of proteins. This development was made possible by the ongoing progress in the field of mass spectrometry and associated sample preparation, separation and data analysis. Modern HPLC-MS systems with state-of-the-art mass spectrometers/analyzers (e.g. quadrupole time-of-flight) exhibit higher sensitivity and data acquisition speed, being capable of identifying and quantifying 5000-10000 proteins from a given proteome. The sheer amount of generated data (often “noisy” or incomplete) has spurred the co-improvement of software tools for data handling and statistical treatment. Concerning peptide separation, one major trend is the use of (currently mainly performed off-line) two-dimensional LC separations in order to detect peptides with PTMs or to reduce matrix interference. Sample complexity is reduced and isotope signals overlap less often. Another very popular LC separation approach nowadays is to couple a reversed-phase chromatography system to the MS. The high backpressure of the relatively long columns requires standard HPLC or even ultra-HPLC systems (Bantscheff et al. 2012). The joint analyzation of transcriptomic and proteomic data is a complex task with different outcomes, answers and problems depending on the selected approach. Haider and Pal (Haider and Pal 2013) categorized the used methods in literature into eight prime types. The most obvious integration type is the union of both datasets to create a reference data set. In one example, the reference dataset helped to find a significant number of bacterial metabolic enzymes that were not present in the dataset from proteomic study alone (Delmotte et al. 2010).

### 2.3.3 Combined analysis of transcript and protein abundances

Maintaining or shifting protein abundance is highly dynamic, achieved by a balance of many processes. Throughout all examined organisms so far (esp. eukaryotic), steady-state transcript levels were shown to generally correlate with protein levels, but only partially. A squared Pearson correlation coefficient of ~0.4 implies that about 40% of the variations in protein concentrations can by directly linked to corresponding variations in mRNA concentrations. The non-correlating rest can be attributed to post-transcriptional regulation, if not measurement noise. A pre-requisite for the investigation of the regulation of cellular protein abundances were the recent advances in large-scale quantification of the proteome, i.e. mass spectrometry. As a rule of thumb, the lower the detected mRNA levels are, the less likely the corresponding polypeptide can be detected using current proteomics technology. Vice versa, above a certain mRNA concentration, the likelihood of detecting the protein remains the same. As indicated in Figure 1, the correlation of mRNA abundance and the probability to even detect the protein (here in yeast; also observed by other groups in bacteria and human cell cultures) graphically results in a 'lazy-step-function'. (Vogel and Marcotte 2012) take this as a strong hint for RNA expression to act more in a switch-like fashion: If the cell wants to make sure that a certain protein is present, mRNA levels should reach a minimum threshold first.

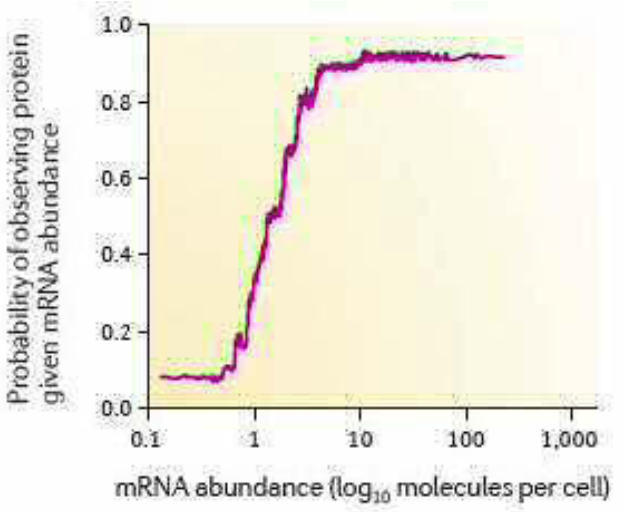


**Figure 1. The correlation of mRNA abundance and the probability to detect the protein** (adapted from (Vogel and Marcotte 2012)

Of course, the various modes and velocities of mRNA and protein turnover are important features of the cell and its evolutionary-acquired regulatory efforts to maintain life. Both methods, transcriptomics and proteomics, combined can deliver additional insights on regulations: For example, unaltered mRNA levels with altered protein abundance at the same time indicates post-transcriptional regulation events. Figure 2 illustrates different levels of protein abundance regulation.

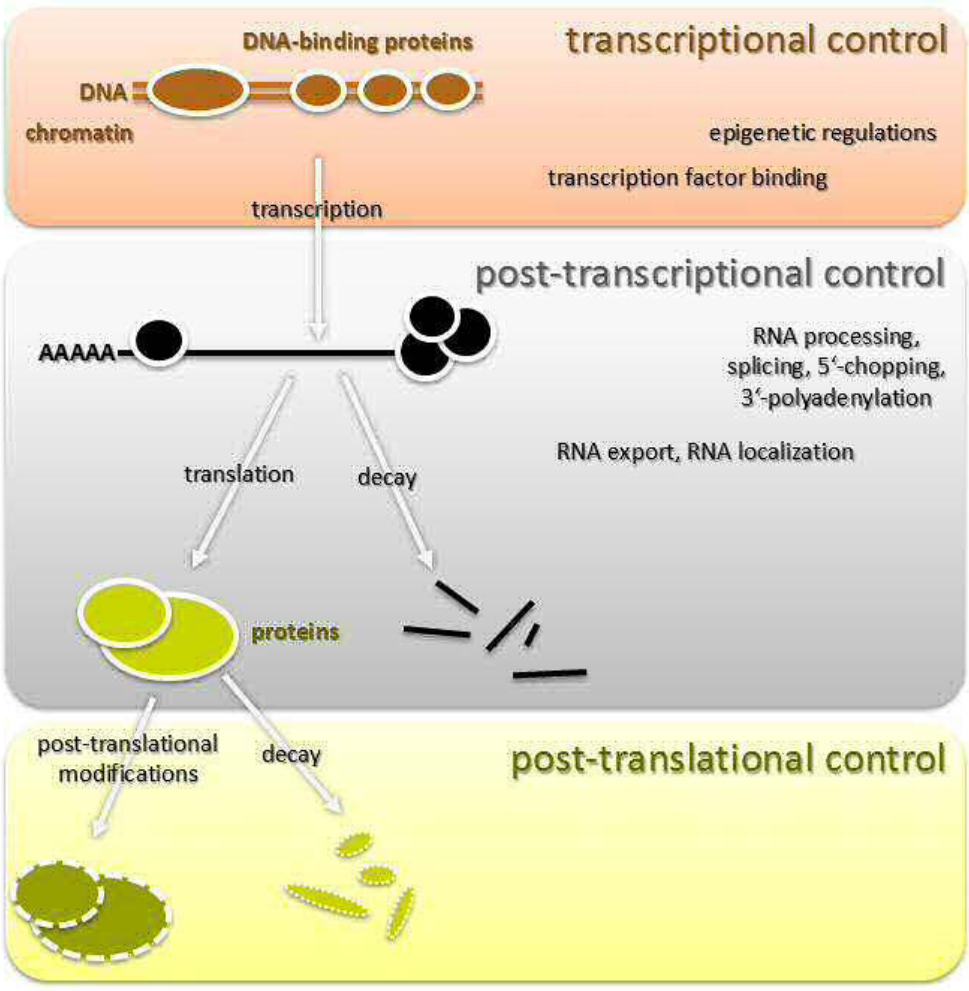


**Figure 2. Stages of regulation of protein abundance.**

Protein-to-mRNA ratios are indicators of post-transcriptional regulation. In case of purely transcriptional regulation of all proteins, these ratios would be the same throughout the proteome. Beyer et al (2004) report a median protein-to-mRNA ratio of 2,500 protein molecules per mRNA molecule in yeast. Dependent on compartment, these median values vary cell-wide by a factor of two. Extracellular and cell wall proteins need the most mRNA per produced protein molecule, lipid particles the least. The data also showed that proteins related to energy metabolism

had the largest protein-mRNA ratios. An explanation for that was given to be their constant need throughout the cell cycle and under all environmental conditions. In Figure 3, mRNA and protein levels are plotted against each other for proteins grouped into different compartments or functions. Put simply, relatively large protein-to-mRNA ratios can be found above the dashed line, smaller ones below that line (Beyer et al. 2004).

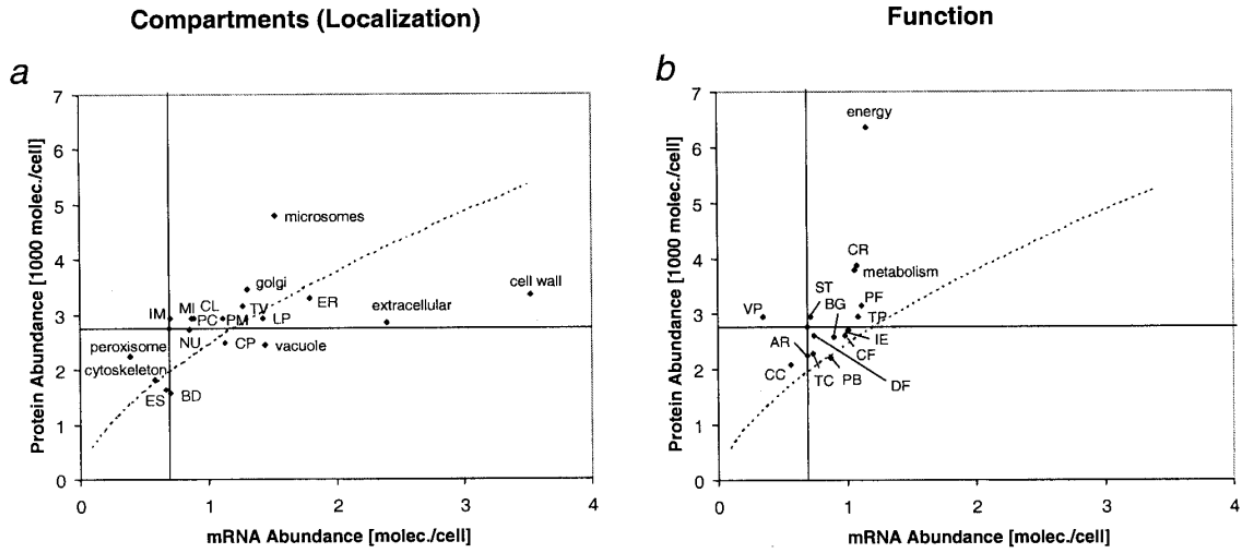


**Figure 3 (Beyer et al. 2004). Median values protein levels versus mRNA abundance of protein properties grouped by (a) localization (cell compartments) and (b) function.** Solid lines indicate median values for the whole cell (i.e. all available ORFs); dashed lines in a and b are power-law regressions of protein versus mRNA levels for all ORFs (exponent = 0.6). Two-letter code for protein groups are as follows. Compartments: BD, bud; CL, cytoplasm; CP, cell periphery; ER, endoplasmic reticulum; ES, endosome; GG, Golgi; IM, integral membrane/endomembrane; LP, lipid particles; MI, mitochondria; NU, nucleus; PC, punctate composite; PM, plasma membrane; PX, peroxisome; TV, transport vesicle. Functional modules: AR, protein activity regulation; BG, biogenesis; CC, cell cycle/DNA processing; CF, cell fate; CR, cell rescue/defense/virulence; DF, differentiation; IE, interaction with cellular environment; MB, metabolism; PB, protein with binding function; PF, protein fate; ST, cellular communication/signal transduction; TC, transcription; TP, transport; VP, transposable elements/viral and plasmid proteins.

In yeast, a large protein-per-mRNA ratio tends to be correlated to low molecular weight of the protein but is also positively correlated with codon usage. Elongation-related features of translation have the most pronounced influence on the protein-per-mRNA ratio rather than initiation-related features. A positive correlation between

transcript half-life and ribosome occupancy (which is more pronounced for short-lived transcripts) suggests competition between translation and mRNA degradation. Eukaryotic genes with low protein-per-mRNA ratios may undergo “translation on demand”: translation of the mRNA is held back until the protein is needed (de Sousa Abreu et al. 2009).

The cell can achieve high protein concentrations via frequent transcription, high mRNA stability, frequent translation and high protein stability. Many ribosomal proteins maximize all these processes. Some metabolic proteins with a high protein-to-mRNA ratio have high translation rates, are stable in regard to degradation, while their transcription rate is relatively low. Less stable molecules are costly for the cell, but allow flexible responses to environmental stimuli. This may be appropriate for genes whose expression needs to change rapidly, for example genes of the TCA cycle, glycolysis and gluconeogenesis in yeast. Sometimes, genes that are upregulated in their transcription under different conditions become even more efficiently translated, resulting in a potentiation effect. Gene expression regulation has been optimized at multiple levels during evolution, and there are different strategies of gene expression regulation (Vogel et al. 2010).

In eukaryotes, mRNA processing and modification, such as poly-adenylation, impacts mRNA stability and translation. A generally positive correlation was found between the length of the poly(A) tail and translation efficiency (Vogel et al. 2010). RNA-binding proteins can determine the fate of the transcript by influencing splicing events, initiation of translation or subcellular mRNA localization (Glisovic et al. 2008). Proteasomal protein degradation also has a huge influence on protein abundance. Here, poly-ubiquitinylation is an important regulatory mechanism: E3 ubiquitin ligases detect degradation signals and promote the attachment of a poly-ubiquitin chain (Vogel et al. 2010).

Analysis of proteomics data contributes to new insights in translation: (Beyer et al. 2004) analyzed the relation of transcription, translation, and protein turnover on a genome-wide scale based on data for the yeast *Saccharomyces cerevisiae* (protein and mRNA abundance, translational status, transcript length). Transcripts often face suppressed translation under normal conditions. “Translation on demand” results in an increase of ribosome density on a transcript, elevating translational activity and

hence, observed protein levels. Another group also showed in yeast that the shorter the coding sequence is, the higher the protein concentrations in general was. Subsequent ribosome foot-printing revealed a much higher ribosome density for the first 30-40 codons compared to the mRNA sections more downstream (Coghlan and Wolfe 2000).

Differences between species in terms of the proportionality of transcript and protein levels can be explained by different mRNA and protein synthesis and degradation rates, resulting in different turnover rates. For example, in mammalian cells, mRNA copies are produced relatively slowly in comparison to copies of the corresponding protein. The half-lives of mRNA are also much longer in mammalian cells than in bacteria. The overall long half-lives of proteins in most examined systems also make it necessary to take into account, that a substantial amount of proteins is decreased simply by dilution, i.e. lost due to cell division. It is not surprising that strong difference in protein turn-over rates depending on the biological function were observed: Structural or housekeeping proteins tend to be much longer-lived than regulatory proteins. Studies in different species showed that transcription and transcript degradation are not the strongest influence on protein abundance but share their impact with regulation of post-transcription, translation and protein degradation (Plotkin 2010; Vogel and Marcotte 2012). Another interesting yet not surprising finding was that – throughout even evolutionary distant species – functionally related proteins (orthologues) share similar protein abundance. One can easily think of a strong correlation of a protein´s half-life and its function, i.e. the stoichiometry with other interacting proteins or within its physical compound (Vogel and Marcotte 2012). Thus, the fact that evolution conserves rather protein functions than corresponding gene sequences offers a sound explanation for the observation that protein levels are more conserved than the corresponding mRNA levels. This was shown across many species (Laurent et al. 2010).

### 2.3.4 Metabolic flux analysis

Knowledge about intracellular metabolite levels and their fluxes through various pathways is the key of understanding of both, the effects of transcript and protein levels and the actual cause for such changes. Metabolomics and derived fluxomics are powerful tools for the investigation of cellular reactions to genetic or

environmental modifications. The choice of analytical tools is highly dependent on the nature (i.e. size, chemical properties) of the metabolites to be measured and the complexity of the sample. Early work in this field was based on enzymatic assays but has evolved into more sensitive and precise instruments using liquid (LC) or gas chromatography (GC) in combination with mass spectroscopy. The chromatographic method determines range and resolution of the measurement, hence LC-MS and GC-MS analysis is often combined (Kell 2004).

Metabolic fingerprinting aims at the quantification of a large number of intracellular metabolites, while targeted analysis focuses on a smaller subset of metabolites that need to be quantified with higher accuracy. Therefore, internal and external reference components are used (Roberts et al. 2012).

It clearly is not enough to analyse intracellular concentrations of certain metabolites because additional information about metabolite turnover and distribution to and flux through different pathways provides a more useful picture of the cell´s phenotype, its physiological state. Metabolic fluxes are both the ultimate result and origin of all regulatory events within the cell.

Metabolic flux analysis combines experimental data with mathematical models. The mathematical model can be simplified by using data from steady-state experiments (most notably chemostat cultivation). In this so-called Flux Balance Analysis (FBA), constant intracellular metabolite levels are assumed so that flux distributions can be fitted to stoichiometric constraints by solving a set of linear algebraic equations. This method requires information on the stoichiometric properties of the network as well as the kinetics of metabolite uptake (e.g. substrates) and output (e.g. secretion, diffusion) of the investigated biological system (Orth et al. 2010). While FBA can be used for resolving less complex networks, $^{13}$C-Metabolic Flux Analysis adds to a more detailed picture: Isotopically labeled carbon substrate is fed to the culture and leads to steady-state labeling patterns of products and intermediates (most notably amino acids) of the central carbon metabolism. The intracellular flux distributions are calculated in an iterative manner to fit the fluxes to experimental data (Sauer 2006; Wiechert 2002).

Figure 4 shows different levels of cellular organization with their corresponding analytical tools used in the work for this thesis.

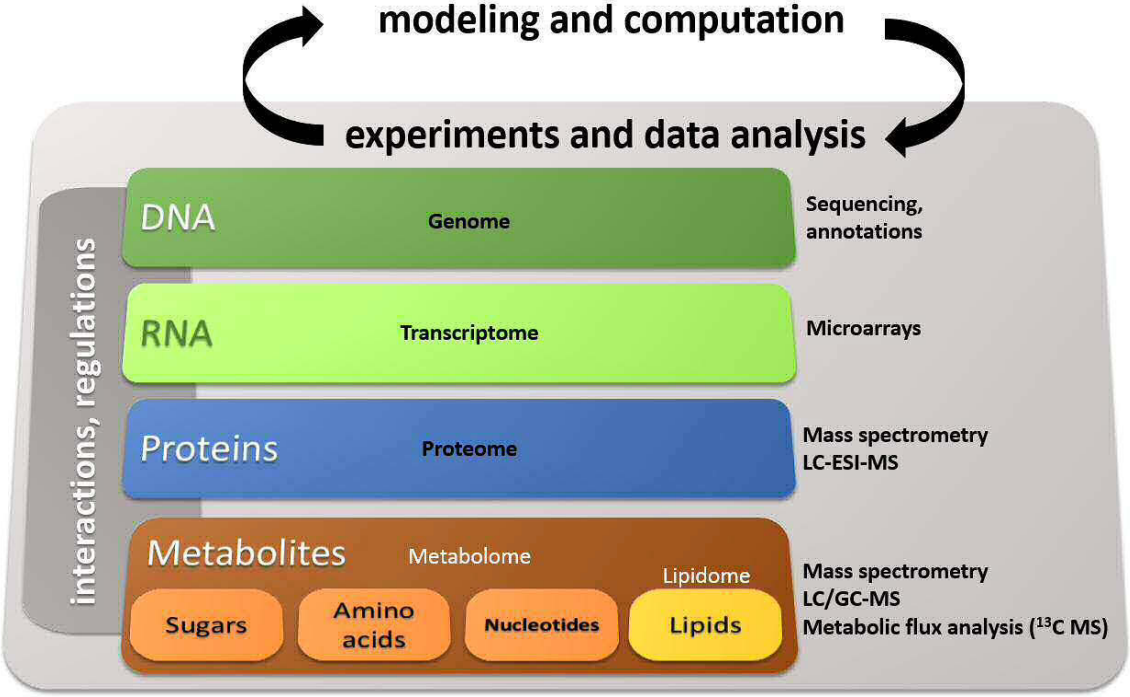


**Figure 4. Tools for Systems Biology.** Levels of cellular organization are shown and their corresponding analytical tools used in the work for this thesis.

# 3 Aim of the work

In recent years, the *P. pastoris* genome was published, automated and manual genome annotation has improved, technical advances were made in the fields of high-throughput protein detection and quantification and protocols were developed for the sampling for metabolome analysis. These were the pre-requisites for systems-level studies of the cell under conditions which we assumed to be relevant for a better understanding of its physiology but also its behavior during recombinant protein production. For the work of this thesis, the organism *P. pastoris* was approached from different angles:

Working with a microbial host for protein production usually comes with the need to cultivate a high number of clones or groups of clones in order to screen them for certain characteristics like productivity. The improvement of small-scale cultivation protocols is the first task as the selection of clones for further work or the conclusions drawn from comparative clone screenings should be based on reproducible, but also scalable results. *P. pastoris* is often grown on methanol, which not only imposes several technical drawbacks on large-scale bioprocesses but also on cultivation in small volumes, when non-uniform induction of methanol utilization genes or unwanted methanol accumulation can cause erratic deviations in growth or productivity among a group of clones. For the establishing of new small-scale cultivation procedures, we wanted to take all knowledge about this yeast´s methylotrophic nature into account while also respecting its actual use in fed batch bioprocesses.

Secretion of recombinant proteins is the primary and ultimate application for *P. pastoris*. Hence, strain engineering was also in our focus: Transcription factors are able to simultaneously regulate different proteins involved in folding and secretion. A few transcription factors have been reported to assist protein folding and secretion in *P. pastoris* (Damasceno et al. 2007; Gasser et al. 2006; Gasser et al. 2007; Guerfal et al. 2010; Stadlmayr et al. 2010a). So rather than co-overexpressing single genes, bottlenecks in folding and secretion could be overcome by regulating a whole set of evolutionary co-regulated genes. After several genes have been reported over the years to assist secretion when overexpressed, we decided to take a closer look at the promoter regions of such genes and browse them for putative binding sites in

order to find novel transcription factors for secretion engineering to overcome cellular bottleneck of recombinant protein production.

Of course, we also wanted to investigate possible bottlenecks in the cell´s metabolism. The main effort of this thesis was the systems-level analysis of *P. pastoris* during production of different recombinant proteins and/or during growth on methanol. The integration of genome, transcriptome, proteome and metabolome analyses offers unprecedented data quality for the detailed exploration of the host´s physiology. Based on different observations for growth on methanol vs. glucose in terms of productivities (irrespective of promoter strength), we assumed that the presence of the C1 substrate methanol has huge implications on the metabolism of *P. pastoris* and is highly relevant for its capability to secrete recombinant proteins.

# 4 Results

## 4.1 Manuscript 1: High-throughput cultivation techniques for *Pichia pastoris*

(Manuscript accepted for publication)

**Establishing a reliable and transferable small scale screening method for a large number of clones**

Strain engineering and performance analysis requires cultivation techniques that fit the purpose of your work in terms of reliability, reproducibility and – equally important – convenience in order to achieve reasonable throughput, hence increase the quantitative scale of the work. Although there are more or less well-established cultivation protocols available for *Pichia pastoris*, they often lack overall transferability for several reasons like for example using a different strain background, different lab facilities and devices or cultivation media. The "right" choice of a cultivations strategy is more like a compromise that takes into account the actual purpose of the work: the creation of production strains which ultimately have to prove their benefits in actual production processes or at least scalable bioreactor experiments, where oxygen supply or pH control is less an issue.

It is known that the rates of recombinant protein production are coupled to biomass and its growth in various fashions (Maurer et al. 2006), depending on the use of promoter(s) or recombinant gene product(s) as well as feed strategies for both glucose and methanol fed *Pichia pastoris* cultures.

Two major feed/induction strategies for the recombinant strains (secreting e.g. enzymes or antibody fragments) were pursued, using either glucose or methanol as main carbon source for $P_{GAP}$ or $P_{AOX1}$ driven protein expression, respectively. In both cases, the goal was to keep the productive phase as reproducible as possible. Avoiding oxygen limitation or methanol accumulation in small scale pays tribute to the tasks and limitations of the bioreactor scale. Hence, it was of importance to only use specific growth rates resembling the fed batch in bioreactor cultivation by limiting the respective carbon source supply rate. Therefore, methanol is added repeatedly in microliter volumes and glucose is administered via slowly-releasing glucose tablets.

The presented protocols allow for screening for high-level production strains of *P. pastoris* wildtype as well as of mut$^S$ (methanol utilization slow) type using either promoter/C-source strategy. For methanol-induced screenings, we describe two protocols: Cultivation in 96 deep well plates is suitable for high throughput production clone screening, while higher volumes in the 24 deep well plate protocol allow for more subsequent analytics procedures and is therefore designed for comparative strain engineering purposes. The following cultivation protocols were developed by the author together with the industry partners in a series of experiments using the production strains generated in the course of this work.

# High-throughput cultivation techniques for *Pichia pastoris*

Markus Buchetics[1], Brigitte Gasser[2], Roland Weis[3]

[1] Austrian Centre of Industrial Biotechnology (ACIB) c/o Department of Biotechnology

[2] BOKU - University of Natural Resources and Life Sciences Vienna, Austria

[3] VTU Technology GmbH, Grambach, Austria

Manuscript accepted for publication in Pichia protocols, Third Edition; edited by Harald Pichler, TU Graz; published by Springer Science+Business Media, LLC)

## Abstract

The provided protocols describe fast and reproducible methods to screen for *Pichia pastoris* production clones in micro-scale in order to select strains for bioreactor cultivations, or to assess outcomes of host strain engineering with a larger number of clones.

Culture volumes of 0.5 mL (in square-bottomed 96-well plates) or 2 mL (in round-bottomed 24-deep well plates) allow for high throughput screening and provide enough sample material for analyses. The screenings were tested with several recombinant protein products, such as antibody fragments and enzymes. Cultivation procedures for the two most commonly used expression strategies are presented, i.e. methanol-induced and glucose-induced recombinant protein production.

**Key words**: small-scale cultivation, deep-well plates, high-throughput, protein production, methanol, glucose tablet, screening

## 1 Introduction

Some key aspects of methanol utilization have to be considered for a reproducible cultivation in a small scale where precise control of methanol feed rate control is not possible. The availability of glucose-releasing feed tablets enables equivalent cultivation principles as applied in most glucose fed batches. Alternatively, also a repeated (spike) feeding protocol for 96-well cultivation is described.

## Methanol-induced recombinant protein production in *Pichia pastoris*

Heterologous gene expression under the control of the *AOX1* promoter is strong, and methanol-driven fed batch processes are often the first choice for protein production in *P. pastoris*. Nevertheless, a variety of aspects have to be considered for a reproducible cultivation of this methylotrophic yeast in a smaller scale where feed rate control is not feasible. It is preferable to perform already the initial clone screenings to identify the best production strains in conditions that resemble the later bioprocess as much as possible.

The first enzyme of the methanol utilization pathway, alcohol oxidase, oxidizes methanol to formaldehyde and hydrogen peroxide (1). Although this takes place in peroxisomes, where toxic hydrogen peroxide and formaldehyde are separated from the rest of the cell (2), high concentrations of methanol in the medium are toxic to *P. pastoris* with increasing formaldehyde and hydrogen peroxide accumulation (3, 4).

Throttling the methanol utilization therefore increases culture viability and process robustness. There are several strategies to reduce the formation rate of these toxic metabolites, however, some of those can only be applied in bioreactor cultivations.

As alcohol oxidase has a low affinity for oxygen, *P. pastoris* compensates for this by generating large amounts of the enzyme (2). A reduction of the intracellular amount of alcohol oxidase is achieved by deletion of the *AOX1* gene. The functional homolog *AOX2* enzyme has a weaker native promoter, thus transcript and protein levels of alcohol oxidase are much lower (about 15%). On methanol, so-called mut$^S$ (methanol utilization slow) strains have a drastically reduced maximum specific growth rate of 0.04 $h^{-1}$ compared to 0.14 $h^{-1}$ of the mut$^+$ (methanol utilization plus) wild type variant with two genes for alcohol oxidase protein present (1, 5). As a result, mut$^S$ cultures are prone to methanol accumulation in the medium in case the methanol feed rates exceed the uptake/metabolization rates over a longer process time. In fed batch cultivations, this is avoided by $C_1$-limiting feed strategies with slowly increasing feed rates or by coupling the feed rate to a constant methanol concentration in the culture with the help of on-line measurement of methanol in the culture broth or in the off-gas stream, respectively (6). Another process control strategy is oxygen limitation, leading to increased cell fitness (7) as formaldehyde and hydrogen peroxide formation rates are reduced – at the expense of growth and *AOX1* promoter activity, which drives foreign gene expression (7, 8). This could be compensated for by even higher concentrations of methanol (9).

In small scale screenings, limitations in oxygen transfer are less severe in 24-well plates due to increased well-size compared to 96-well plates. It has to be mentioned, however, that oxygen transfer rates in deep-well plates (both 24- and 96-well, respectively) are comparable to those obtained in baffled shake flasks (10). In addition to that, the use of mut$^S$ strains (for *AOX1*-promoter driven production) results in less oxygen demand due to slower specific growth rates. This can also be achieved by glucose-limited tablet feed for e.g. *GAP*-promoter driven production ((11), see next section).

In order to cultivate *P. pastoris* at higher specific growth rates, mixed-substrate feed media are often used, especially for mut$^S$ phenotype strains (5): Pure methanol might overburden the cells when feed rates are applied that exceed the maximum uptake rate. However, when methanol is combined with e.g. glycerol-limited feed, it serves as inducer of recombinant protein production while process times can be reduced due to faster growth with the help of the co-substrate. Process optimization is more time-consuming compared to pure methanol feed strategies due to the necessity of staying below the repressive concentration of glycerol. Improved volumetric productivities were reported in individual cases for mixed-feed processes (5). However, a reduction of the total process time most times comes at the expense of total product titers, because there is no simple positive correlation of specific product formation rates with specific growth rates (12, 13) for protein production under the *AOX1* promoter (14). A mixed-feed strategy is not applicable for smaller scale screenings because of severe repression (1) of the *AOX1* (and *AOX2*) promoter and insufficient peroxisome proliferation (7) when co-substrates such as glycerol or sorbitol can only be administered batch-wise, i.e. repeatedly in excess (15). In order to reduce the time for main culture of the clone screening down to two days (compared to three days in the described protocols), there are commercial solutions to co-feed a methanol-pulsed culture with glucose by using continuous enzymatic hydrolysis of starch (16).

Use of methanol for *AOX1* promoter driven protein production has required the development of multi-step bioreactor cultivation strategies (5). In the batch phase, the recombinant yeast grows on a salt medium of a usually non-fermentable carbon source, e.g. glycerol. In a first short fed batch phase – often referred to as transition phase (5) – glycerol is added in a fed-batch manner to further increase the biomass and to de-repress the *AOX1* promoter prior to the addition of methanol. Additionally, by-products generated in the batch phase (e.g. ethanol when using glucose as C-source, *per se* a repressor of the *AOX1* promoter (17)) are consumed. During the transition phase, methanol can already be fed with gently increasing rates to further support the cellular production of alcohol oxidase. A starvation phase (e.g. one hour) is reported to remove residual glycerol or accumulated metabolites before the methanol induction phase is initiated (5). A methanol pulse (0.5%) prior to continuous feeding is sometimes used to support adaptation to the subsequent growth on methanol. Interpretation of pulse consumption characteristics, i.e. methanol uptake rates, can help finding the maximum feasible starting feed rates to avoid methanol accumulation. Methanol uptake was described to vary significantly between different production strains (18). Empirical data indicate massive variations in productivity of strains secreting different recombinant proteins, especially when changing specific growth rates (5).

Therefore, one must conclude that there will not be one general small scale screening strategy (where methanol is administered batch-wise) that allows for comparison of different production hosts. Just as for

a later cultivation in controlled bioreactors, the optimal process for each strain must be determined individually.

The described protocols are reliable methods for a first screening for high-level production strains that significantly outperform other co-cultivated transformants. Depending on the nature of the recombinant target protein(s), such small scale cultivations may or may not give valid information about achievable titers in controlled fed batch bioreactor cultivations. In general, one can expect transferability of lab bench to bioreactor performances in terms of relative productivities of compared transformants (19).

Two different protocols are presented for methanol-induced screenings. While cultivation in 96 deep well plates allows high throughput production clone screening, the 24 deep well plate protocol has higher volumes for subsequent analytics procedures and is better suited for comparative strain engineering purposes.

*Type 1 Basic protocol in 24-well plates*

The 24-deep well plate screening protocol (24-deep-well plate) - depicted in Fig. 1 - has a 4-day (~96 hours) time frame and reflects some key principles of the bioreactor process, i.e. the sequential use of carbon sources necessary to de-repress and induce the *AOX1* promoter. Apart from the predetermined carbon sources, it can be performed in any cultivation medium that supports production of the desired recombinant protein (synthetic, minimal or complex media at different pH conditions).

It requires some additional effort but has been shown to deliver more reproducible performances, in particular for strains with reduced growth rates (e.g. due to production of hard-to-express recombinant proteins). It compensates for the potential influence of different starvation times prior to first *AOX1* promoter induction as they can occur when one clone finishes batch phase much earlier than another. The key feature of this method is the combined glucose batch and fed batch phase for fast biomass accumulation with automatic transition into a glucose-limited growth phase. Here, a glucose batch phase is directly followed by glucose-limited, slow growth phase. Then, the inoculum for the methanol induced culture has to be washed in order to remove residual glucose (and ethanol, if present), which would repress the *AOX1* promoter. Unlike for glycerol, there are options for a slow continuous release of glucose into the growth medium. This method relies on glucose feed tablets (11) that work in any aqueous medium irrespective of pH or the addition of enzymes for e.g. starch hydrolyzation (20).

An additional positive effect on reproducibility comes from the removal of still-present pre-culture media components that can influence productivities due to varying initial inoculum volume.

The dense inoculation of this culture to $OD_{600}$=2 ensures that within one day all cultures finish the batch phase – especially when using minimal salt media – and continue to grow slowly for several hours. The subsequent obligatory washing step ensures total removal of glucose, ethanol and old media components

(from the pre-culture) before 2 mL of production medium are inoculated to $OD_{600}$=4 and all cultures are induced by the addition of 0.5% (10 µL) pure methanol. During the following 2-day induction phase, 0.5% (10 µL) methanol is added once more, followed by two times 1% (20 µL). The cultures are then harvested after ~48h.

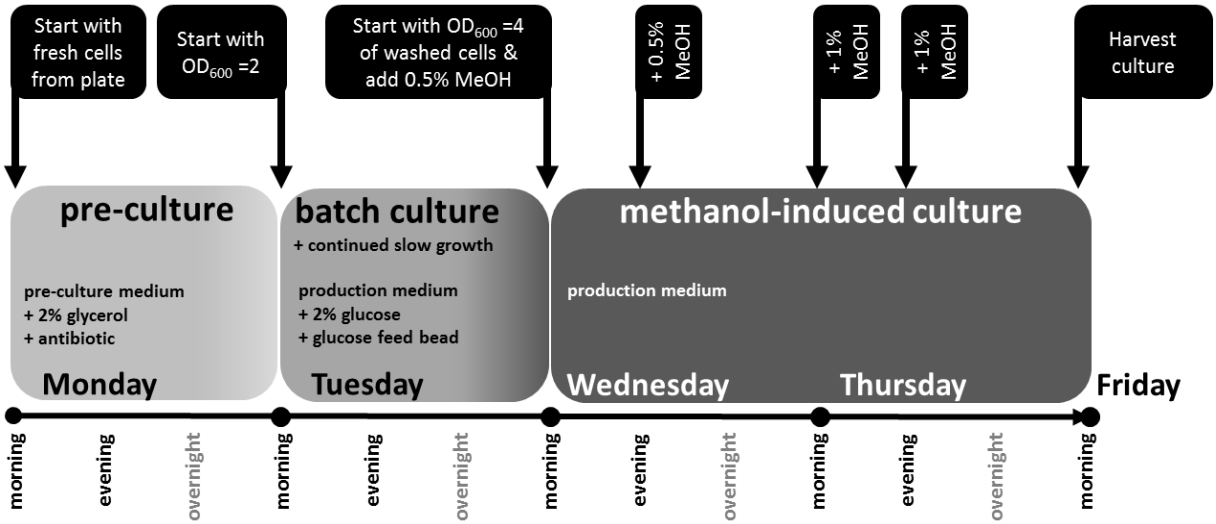


**Figure 1. Clone screening protocol for *AOX1* promoter/methanol driven recombinant production in *P. pastoris***

*Type 2 Basic protocol in 96-well plates*

The small-scale screening protocol in 96-deep-well plates depicted in Fig. 2 has a 5 ½ day (~130 - 136 hours) time frame.

The long initial phase of 60-70 hours in batch medium serves as time buffer to ensure that all colonies consumed the supplied carbon source (glucose). The limiting glucose amount in this special medium formulation secures physiological competence for heterologous protein expression also for those cultures that utilize glucose faster than others and consequently rest longer in the stationary phase before being induced by methanol. Additionally, this procedure and the media formulation lead to reduced fractions of necrotic and apoptotic cells as compared to other media compositions (19).

Methanol induction is carried out for a total of 72 hours, thereby maximising the time-frame of the productive phase of cultivation. The inducer is supplied 4 times in a buffer formulation thereby being compatible to standard multi-channel liquid handling tools.

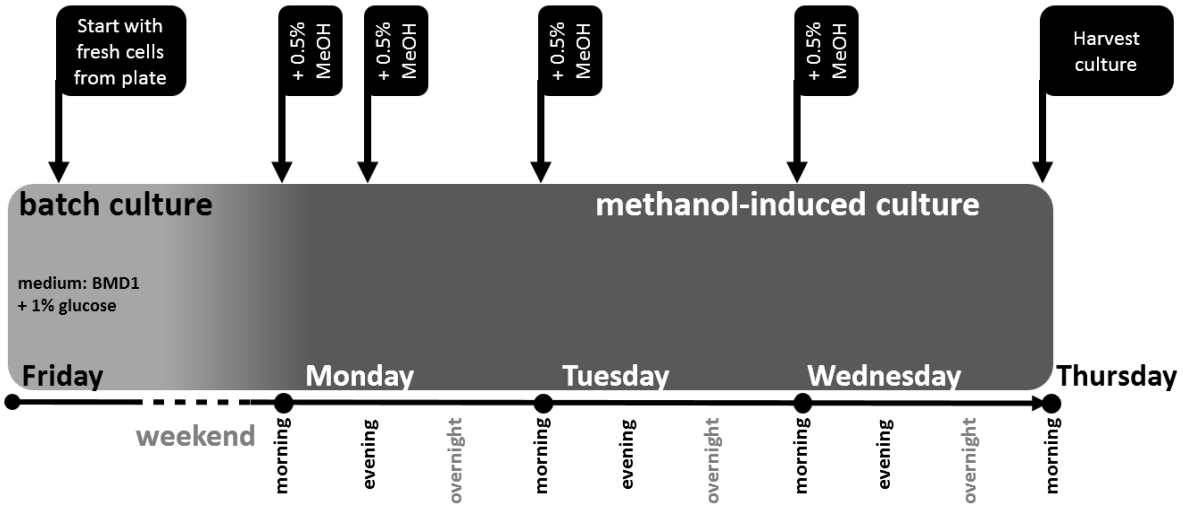


**Figure 2. Clone screening protocol for methanol-driven recombinant production in 96-well format**

## Glucose-driven recombinant protein production in *Pichia pastoris*

Glucose-limited feeding is the cultivation method of choice for fed batches of strains utilizing the strong glycolytic GAP promoter for heterologous gene expression. The GAP promoter is constitutively active and not repressed by glucose in the culture medium (21). Glucose-based production eliminates hazard and costs associated with large quantities of methanol in bioreactor cultivation (16, 22).

The most common cultivation strategy consists of a glycerol batch followed by a glucose-limited fed batch phase. At C-limiting conditions, the culture remains free of by-products or aggregates (23), and specific growth rates ($< \mu_{max}$) can be modulated to optimize specific or volumetric productivities (12). In order to transfer this principle of C-limitation by controlled substrate release into small-scale for clone screening applications, commercial cultivation media systems for slow enzymatic release of glucose from a starch-polymer have been successfully tested for the production of a variety of products (24, 25). However, these protocols require a specific commercial media composition, which is prone to become acidic at long term cultivations. Furthermore, the enzyme has to be added repeatedly. An alternative is to use slow glucose releasing polymer particles. The 12 mm feed beads (Kuhner, CH) described in the protocol below are liberating glucose at a non-linear rate of $1.63 \cdot t^{0.74}$ mg per disc (t = time [h]), which equals to 28.6 mg per disc after 48 hours (11). Apart from cultivation of clones utilizing the GAP promoter, both methods are also suitable to test glucose-methanol co-feeding strategies for AOX-based expression (16, 26) as well as for promoters that need glucose-derepression such as modified AOX variants (27, 28) or G-promoters (29, 30).

*Type 1 Basic protocol for 24-well plates*

The described protocol using glucose liberating feed beads allows for free choice of cultivation medium and pH and offers a relatively cheap and easy method to simulate glucose-limited fed batch conditions in small scale cultivation plates (Fig. 3). The three day time frame and the relatively simple procedure allow for high throughput in screenings.

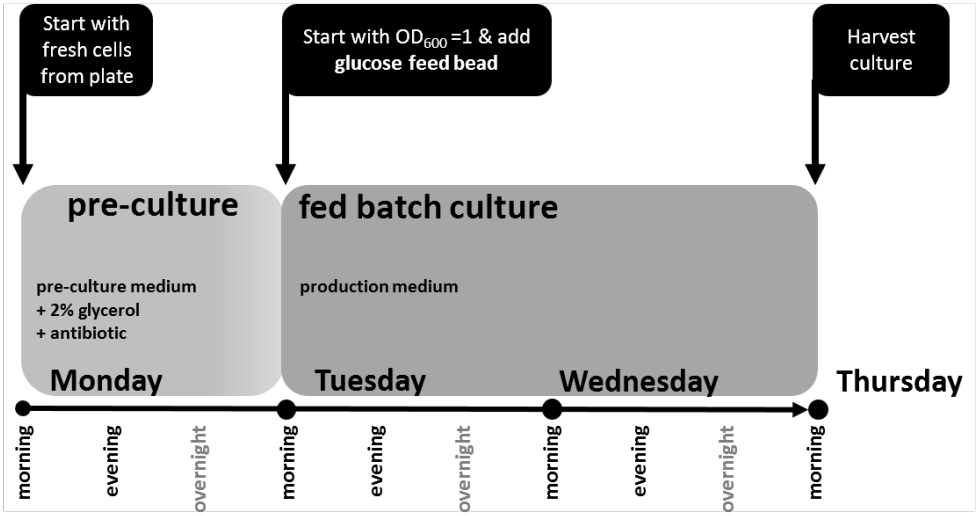


**Figure 3. Clone screening protocol for glucose driven recombinant production in *P. pastoris***

Figure 4 charts the theoretical course of the specific growth rate μ as a result of a roughly constant release of substrate from a feed bead, in this case a 12 mm diameter tablet (11). After dense inoculation of the main culture to $OD_{600}$ = 1, the cells grow for ~12 hours in a non-limited fashion when substrate release exceeds maximum specific growth rate. Then, μ steadily and slowly decreases. Until the culture is harvested, it has run through a range of μ that is similar in principle to bioreactor fed batch cultivations: from a phase of maximum growth rates and substrate surplus to slower C-limited growth while the bulk of biomass, hence recombinant product, is generated in the latter phase. An over-proportionate batch phase is avoided by using solely a glucose tablet from the beginning instead of adding one after a day of batch phase (on e.g. 2% glucose).

*Type 1 Basic protocol for 24-well plates*

The described protocol using glucose liberating feed beads allows for free choice of cultivation medium and pH and offers a relatively cheap and easy method to simulate glucose-limited fed batch conditions in small scale cultivation plates (Fig. 3). The three day time frame and the relatively simple procedure allow for high throughput in screenings.

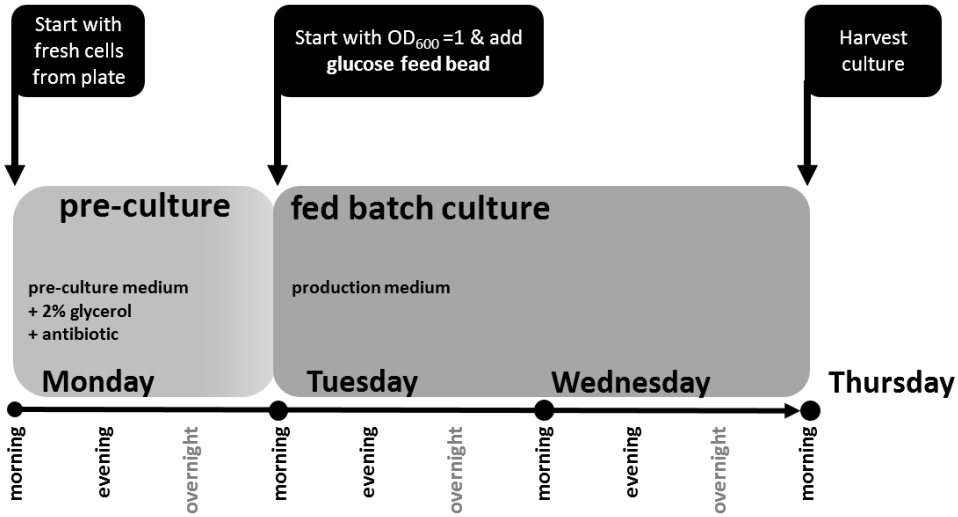


**Figure 3. Clone screening protocol for glucose driven recombinant production in *P. pastoris***

Figure 4 charts the theoretical course of the specific growth rate μ as a result of a roughly constant release of substrate from a feed bead, in this case a 12 mm diameter tablet (11). After dense inoculation of the main culture to $OD_{600}$ = 1, the cells grow for ~12 hours in a non-limited fashion when substrate release exceeds maximum specific growth rate. Then, μ steadily and slowly decreases. Until the culture is harvested, it has run through a range of μ that is similar in principle to bioreactor fed batch cultivations: from a phase of maximum growth rates and substrate surplus to slower C-limited growth while the bulk of biomass, hence recombinant product, is generated in the latter phase. An over-proportionate batch phase is avoided by using solely a glucose tablet from the beginning instead of adding one after a day of batch phase (on e.g. 2% glucose).

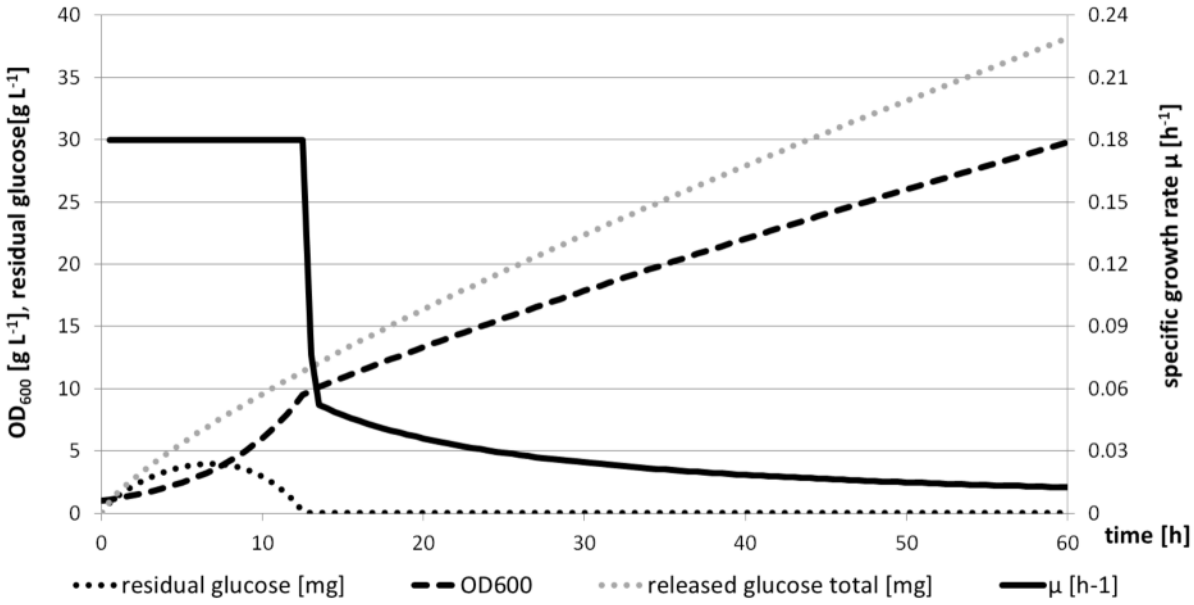


**Figure 4. Specific growth rate model** of tablet-fed (Ø 12mm) cultures inoculated to $OD_{600}$ = 4.0

*Type 2* *Basic protocol in 96-well plates*

The small-scale screening protocol in 96-deep-well plates depicted in Fig. 5 has a 4 ½ day (~108 - 114 hours) time frame.

As for the methanol-induction cultivation regime, the long initial phase of 60-70 hours in batch medium serves as time buffer to ensure that all colonies consumed the supplied carbon source (glucose). For glucose supply, addition of glucose in medium formulation is carried 4 times in total, and culture harvest is done after a total of approximately 110 hours of cultivation.

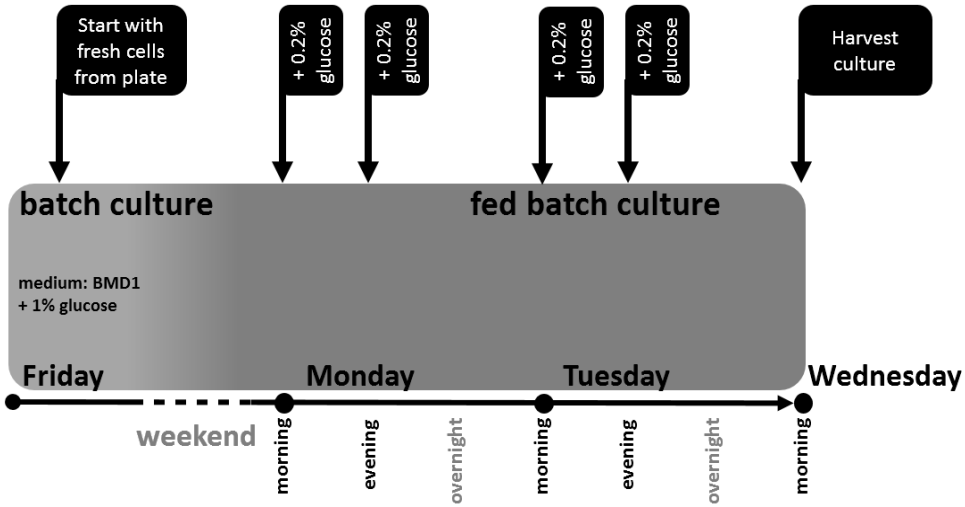


**Figure 5. Clone screening protocol for glucose-driven recombinant production in 96-well format**

# 2 Materials

## 2.1 24-deep well plate scale

- pre-culture medium (selective YPG) (*see* **Notes 1 and 2**)
- main culture medium (e.g. M2 medium)
- 24 deep well cultivation plate(s) with square-bottom; obligatory for main culture and also recommended for pre-culture (50 mL falcon tubes are also possible)
- breathable sealing films (see **Note 3**)
- Shaking device (*see* **Note 4**)
- Centrifuge

For pre-cultures, we recommend selective YP-medium containing 2% glycerol (*see* **Note 2**). The main cultures can be performed in any soluble growth/production medium background.

**Media compositions:**

YPG medium (see **Note 1**):

- 20 g $L^{-1}$ soy peptone
- 10 g $L^{-1}$ yeast extract
- 20 g $L^{-1}$ glycerol

pH is set to 7.4 with NaOH.

Synthetic citrate buffered minimal medium "M2" without carbon source (*see* **Note 5** and **Note 6**):

- 3.15 g $L^{-1}$ $(NH_4)_2PO_4$
- 0.49 g $L^{-1}$ $MgSO_4*7H_2O$
- 0.80 g $L^{-1}$ KCl
- 0.0268 g $L^{-1}$ $CaCl_2*2H_2O$
- 22.0 g $L^{-1}$ Citric acid monohydrate
- 1.47 mL $L^{-1}$ PTM1 trace metals
- 2 mL $L^{-1}$ Biotin (0.2 g $L^{-1}$)
- set selected pH (possible range: 3.0-6.0) with KOH (solid)
- sterilize by filtration (0.22 μm)

PTM1 trace metal solution (*see* **Note 6**):

- 5 mL $L^{-1}$ $H_2SO_4$ (95-98%)
- 65 g $L^{-1}$ $FeSO_4*7H_2O$
- 20 g $L^{-1}$ $ZnCl_2$
- 6 g $L^{-1}$ $CuSO_4*5H_2O$
- 3.36 g $L^{-1}$ $MnSO_4*H_2O$
- 0.82 g $L^{-1}$ $CoCl_2*7H_2O$
- 0.2 g $L^{-1}$ $Na_2MoO_4*2H_2O$
- 0.08 g $L^{-1}$ NaI
- 0.02 g $L^{-1}$ $H_3BO_3$

Carbon source supplements

- For all 24-deep well plate protocols:
- for AOX1 clone screening: glucose, glucose feed beads (e.g. 12 mm diameter by Kuhner, Switzerland; one per cultivated clone), methanol
- for glucose-driven clone screening: glucose, glucose feed beads (e.g. 12 mm diameter by Kuhner, Switzerland; one per cultivated clone)

## 2.2 96-deep well plate scale

**Media compositions:**

Buffered minimal medium (BM):

- 1.34% YNB
- 4x $10^{-5}$ % biotine
- 100 – 200 mM potassium phosphate buffer pH 6.0 (*see* **Note 7**)

BMD1:

- Buffered minimal medium supplemented with 1% glucose

BMM2:

- Buffered minimal medium supplemented with 1% methanol

BMM10:

- Buffered minimal medium supplemented with 5% methanol

Further BMD-media described (BMDx):

- Buffered minimal medium supplemented with between 0.1 and 0.9% glucose

Valid for all media: in case the selected strain is auxotroph for one or more amino acids, the appropriate amount of the necessary amino acid must be supplied accordingly (*see* **Note 8**).

96 deep well plate(s) with square-bottom and cover (*see* **Note 9**)

# 3 Methods

## 3.1 24-deep well plate scale cultivation

A relatively high screening throughput can be achieved with 24-deep-well plates. Operational effort can be reduced by the use of multi-channel pipettes and tube/cuvette-racks that allow for a 6x4 (24) format handling during culture dilutions (optical density measurements), potential washing steps or sampling/harvesting.

When using one´s laboratory shaker the first time for 24-deep well plates containing 2 mL culture, it is strongly recommended to perform a shaking test at the given range to check for proper homogenisation of the culture without wetting the sealing membrane.

It is disadvised to use cultivation temperatures of 30°C or above for 2 mL cultures as this usually results in massive evaporation effects, even when using incubators with humidity control. At 25°C, culture volume reduction can also be observed after three days of 2 mL main culture. However, this does not necessarily affect reproducibility as long as plate position effects for evaporation as well as potential insolubility of medium components can be excluded. A beneficial side-effect of otherwise unwanted evaporation is the improvement of oxygen transfer, especially during later growth phases, when biomass levels are higher.

The protocol descriptions are fitted to a schedule of Monday - Friday for methanol-induced screenings or Monday – Thursday for glucose-based screenings. Note that the time span between “morning” and “evening” should be at least 6 hours. A maximum number of eight 24-deep well plates is considered to be manageable per operator and week.

Pay attention to aseptic handling during all steps, especially when working with rich growth media that are prone to bacterial contamination (*see* **Note 10**).

As product yield reproducibility of biological replicates deviates more among different 24-deep well plates than within one cultivation plate, it is recommended to use "normalizer clones" on each plate for optimal comparison of screening results of different plates or screening rounds (see Fig. 6 for suggested plate layouts) (*see* **Note 11**).

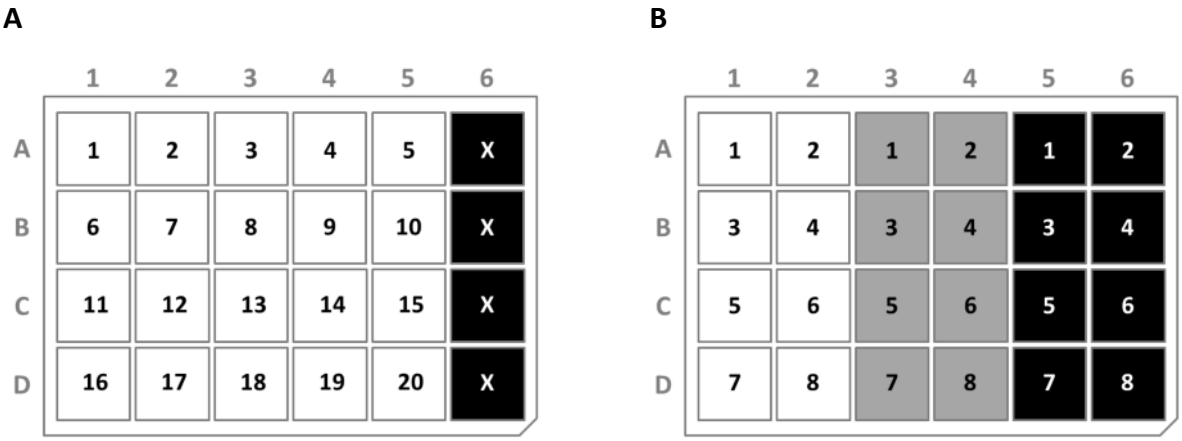


**Figure 6. Suggested plate layouts for 24-deep well plate cultivations using normalizer clones or clone families**:

(A) Recommended layout for regular production clone screening. Up to 20 novel clones per plate are cultivated and analysed in comparison to four biological replicates (black wells) of a normalizer clone X. Biomass and product yields as well as standard deviations of the four replicates can be used as parameters of quality control in terms of screening robustness and reproducibility.

(B) Layout for the analysis of production clone families with e.g. different strain engineering background (*see* **Note 12**). Up to two groups of transformants (white and light grey wells; e.g. versions of a production strain) are cultivated in parallel to an equally-sized group of control clones (black wells; e.g. the parental strain transformed with an empty vector control). Alternatively, one might also compare different media compositions with respect to basic media supplements, pH or buffer strength using this layout.

## *Basic clone screening protocol for AOX1 promoter/methanol driven recombinant production in P. pastoris*

Traces of complex (pre)culture components can have a strong influence on growth and productivity, hence the reliability of the clone screening. As described below, a washing step prior to methanol induction can help to improve performance reproducibility. This protocol is recommended in particular for strains where production of the recombinant proteins leads to reduced growth performance due to cell stress.

*1. Pre-culture, Day 1, morning*

- Inoculate 2 mL selective YPG medium (*see* **Notes 1 and 2**) with cells from a master plate (*see* **Notes 13 and 14**) or transformation plate (*see* **Note 15 and 16**)
- Incubate at 25°C at 280-320 rpm

*2. Glucose batch culture with continued slow growth, Day 2, morning*

- Pipet 2 mL M2 medium (containing 2% glucose) into each well of a sterile 24-deep-well-plate
- Measure $OD_{600}$ of the pre-culture and calculate the inoculation volume for a starting $OD_{600}$ of 2.0: $V_{inoculum}$ = 2 mL ($V_{main\ culture}$)* 2.0 ($OD_{600,main\ culture}$) / $OD_{600,pre\text{-}culture}$
- Inoculate the main cultures (2 mL M2 + 2%glucose) (see **Note 17**)
- Add one glucose tablet (12 mm diameter) to each well
- Incubate at 25°C at 280-320 rpm

*3. Methanol induction, Day 3, morning*

- Pipet 2 mL M2 medium (without carbon source!) into each well of a fresh 24-deep-well-plate
- Wash the batch cultures: transfer ~1 mL culture into a sterile eppi (1.5 mL) and centrifuge at 13.000 rpm for 1 min at room temperature. Remove the supernatant completely, add 1 mL M2 medium (without carbon source) and resuspend by vortexing
- Measure $OD_{600}$ of the washed culture and calculate the inoculation volume for a starting $OD_{600}$ of 4.0: $V_{inoculum}$ = 2 mL ($V_{main\ culture}$)* 4.0 ($OD_{600,main\ culture}$) / $OD_{600,washed\ culture}$
- Inoculate the main cultures (2 mL M2) (see **Note 17**)
- Add 10 µL (0.5 %) pure methanol to each well (*see* **Note 18**)
- Incubate at 25°C at 280-320 rpm

*4. Methanol induction, Day 3, evening*

- Add 10 µL (0.5 %) pure methanol to each well (*see* **Note 18**).
- Incubate at 25°C at 280-320 rpm

*5. Methanol induction, Day 4, morning*

- Add 20 µL (1.0 %) pure methanol to each well (*see* **Note 18**).
- Incubate at 25°C at 280-320 rpm

*6. Methanol induction, Day 4, evening*

- Add 20 µL (1.0 %) pure methanol to each well (*see* **Note 18**).
- Incubate at 25°C at 280-320 rpm

*7. Harvest of culture, Day 5, morning*

- After ~48 hours of induced culture, harvest the culture (see next section: Harvest).

***Clone screening protocol for glucose driven recombinant production in P. pastoris***

*1. Pre-culture, Day 1, morning*

- Inoculate 2 mL selective YPG medium (*see* **Notes 1 and 2**) with cells from a master plate (*see* **Note 13 and 14**) or transformation plate (*see* **Notes 13, 15 and 16**)
- Incubate at 25°C at 280-320 rpm

*2. Glucose fed batch culture, Day 2, morning*

- Pipet 2 mL M2 medium (containing 2% glucose) into each well of a sterile 24-deep-well-plate
- Measure $OD_{600}$ of the pre-culture and calculate the inoculation volume for a starting $OD_{600}$ of 1.0: $V_{inoculum}$ = 2 mL ($V_{main\ culture}$)* 1.0 ($OD_{600,main\ culture}$) / $OD_{600,pre\text{-}culture}$
- Inoculate the main cultures (2 mL M2) (see **Note 17**)
- Add one glucose tablet (12 mm diameter) to each well
- Incubate at 25°C at 280-320 rpm

*3. Harvest of culture, Day 4, morning*

- After ~48 hours of induced culture, harvest the culture (see next section: Harvest).

## 3.2 96-deep well plate cultivations

Screening in 96-deep-well plates confers a high screening throughput, but comes with an almost mandatory use of multi-channel pipettes and/or other (automated) liquid handling devices. The applied shaking device needs to have temperature control and the ability to humidify the chamber. Due to the relatively high shaking speed, it is also advised to select the correct shelf board for fixing the 96-deep well plates: boards with a sticky tape will not be able to keep all plates fixed (especially if the panel is filled at all possible positions), but there are solutions for microtiter- and deep-well plates.

When using one´s laboratory shaker the first time for 96-deep well plates containing 0.6 mL culture, it is strongly recommended to perform a shaking test at the given range to check for proper homogenisation of the culture without wetting the cover.

Cultivation temperature should not exceed 28°C at most due to uncontrollable evaporation effects. At 28°C, significant culture volume reduction can also be observed in the perimeter wells which leads to the advice to either fill rows A and H and columns 1 and 12 with mock strain or other strains not being inflicted in the screening (reduction of throughput), or decrease the cultivation temperature to 25°C or below (especially recommended for the production phase).

The composition of both, the batch medium as well as the induction/carbon source supply medium (see below) should not be changed to e.g. a complex medium because the cells could not bear the predetermined long period of stationary phase prior to induction without severe physiological implications.

The methanol-induced protocol description spans the range of Friday afternoon until Thursday morning in the next week with only 6 manipulation steps (see below). Glucose-driven expression requires the same amount of manipulation steps from Friday afternoon until Wednesday morning.

Note that the time span between “morning” and “evening” should be at least 8 hours. A person without experience in this particular handling can easily handle 24 plates per week, given that the necessary instruments are at hand (liquid handling device, shaker).

One always needs to carry out handling steps very carefully and aseptically, even though only minimal media are applied in this protocol.

Deviations between biological replicates occur mainly among different plates, hence “normalizer clones” are advised on each separate plate (Fig. 7A). For clone screening applications, i.e. the initial and first liquid cultivation of freshly prepared transformants, it is not necessarily required to run these “normalize clones” on each plate because a rescreening of selected clones from this initial round is mandatory in order to validate results obtained from this first liquid cultivation, anyhow (Fig. 7B)

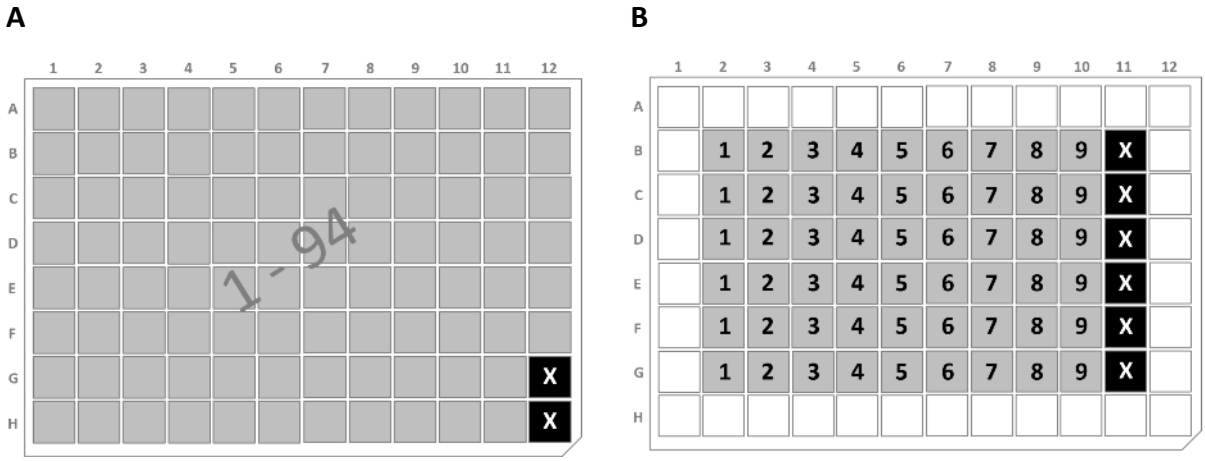


**Figure 7. Plate layouts for 96-deep well plate cultivations for initial clone screening and rescreening**:

(A) Recommended layout for regular production clone screening. Up to 96 novel clones per plate are cultivated and analyzed in comparison to biological replicates of a normalizer clone X (wells G and H 12, not necessarily performed for each plate). In case of massive evaporation of the perimeter wells (e.g. through insufficient humidity control), inoculate these wells with a mock strain, a normalizer strain or any other *Pichia* strain.

(B) Layout for rescreening of selected strains from the initial screening round, or for parameter check with respect to media formulation, pH, buffer strength and the like. In order to anticipate a potential evaporation in perimeter wells, the inoculation of rows A and H as well as columns 1 and 12 with mock strain, a normalizer strain or any other *Pichia* strain may be performed.

## *Basic protocol in 96-well plates: methanol induced expression*

*1. Inoculation of culture with transformants from plates, Day 1 (Friday), afternoon*

Inoculate 0.25 mL BMD1 medium with cells from a master plate (*see* **Notes 13, 14 and 18**) or a transformation plate (*see* **Notes 15 and 19**)

incubate at max. 28°C at 260-340 rpm

*2. Methanol induced culture, Day 4 (Monday), morning*

- Add 0.25 mL of BMM2 into each well (*see* **Note 20**)
- Incubate at max. 28°C at 260-340 rpm

*3. Methanol induction, Day 4 (Monday), evening*

- Add 0.05 mL of BMM10 into each well (*see* **Note 20**)
- Incubate at max. 28°C at 260-340 rpm

*4. Methanol induction, Day 5 (Tuesday), morning*

- Add 0.05 mL of BMM10 into each well (*see* **Note 20**)
- Incubate at max. 28°C at 260-340 rpm

*5. Methanol induction, Day 6 (Wednesday), morning*

- Add 0.05 mL of BMM10 into each well (*see* **Note 20**)
- Incubate at max. 28°C at 260-340 rpm

*6. Harvest of culture, Day 7 (Thursday), morning*

- After ~72 hours of methanol induction, harvest the culture (see next section: Harvest).

### *Basic protocol in 96-well plates: glucose driven expression*

*1. Inoculation of culture with transformants from plates, Day 1 (Friday), afternoon*

- Inoculate 0.25 mL BMD1 medium with cells from a master plate (*see* **Notes 13 and 14**) or a transformation plate (*see* **Notes 13, 15 and 19**)
- Incubate at max. 28°C at 260-340 rpm

*2. Carbon source supply, Day 4 (Monday), morning*

- Add 0.25 mL of BMDx medium to each well (*see* **Notes 20 and 22**)
- Incubate at max. 28°C at 260-340 rpm

*3. Carbon source supply, Day 4 (Monday), evening*

- Add 0.05 mL of BMDx to each well (*see* **Notes 20 and 22**)
- Incubate at max. 28°C at 260-340 rpm

*4. Carbon source supply, Day 4 (Tuesday), morning*

- Add 0.05 mL of BMM10 to each well (*see* **Notes 20 and 22**)
- Incubate at max. 28°C at 260-340 rpm

*5. Carbon source supply, Day 4 (Tuesday), evening*

- Add 0.05 mL of BMM10 to each well (*see* **Notes 20 and 22**)
- Incubate at max. 28°C at 260-340 rpm

*6. Harvest of culture, Day 7 (Wednesday), morning*

- After ~110 hours of methanol induction, harvest the culture (see next section).

## Harvest

<u>Variants for biomass determination (*see* **Note 23**)</u>

Wet cell weight analysis:

- Mix the culture by pipetting up and down
- transfer 1 mL culture into a pre-weighed 1.5 mL eppendorf tube (*see* **Note 24**)
- centrifuge at 13.000 rpm for 5 min
- pour the supernatant into a fresh tube for analysis or storage
- remove residual supernatant by pipetting or suction
- determine the wet cell weight from the pellet by subtracting the tared weight from the obtained weight: the obtained mg-value gives the wet cell weight in mg $L^{-1}$

Optical density measurement at 600nm

- Mix the culture by pipetting up and down
- Transfer culture into dilution tubes (e.g. Eppendorf 1.5mL)
- For appropriate dilution, start with a dilution of 60-fold and adjust accordingly
- Measure OD600 in spectrophotometer

Harvest of culture supernatant (*see* **Note 25**)

In case of available centrifuge adaptor for deep well plates (often compatible with standard microtiter format):

- Centrifuge the plate(s) at 2,500x g for 10 minutes
- collect the supernatants for analysis or storage

If no equivalent centrifuge or adapters are available, perform supernatant harvesting in individual vessels (e.g. Eppendorf 1.5mL)

Harvest of cell mass (pellet) (*see* **Note 25**)

In case of available centrifuge adaptor for deep well plates (often compatible with standard microtiter format):

- pipet the intended amount of cell broth to e.g. 96-well PCR plates (*see* **Note 26**)
- centrifuge at 2,500x g for 10 minutes
- remove the supernatant by pipeting and subsequent banging of the plate(s) on absorbent paper (*see* **Note 27**)
- use the pellet fraction for direct lysis (see **Note 28**) or storage (*see* **Note 29**)

If no equivalent centrifuge or adapters are available, perform harvesting of cell mass (pellet) in individual vessels (e.g. Eppendorf 1.5 mL)

# 4 Notes

1. The peptone described in early publications and in the "Invitrogen Pichia Protocols" (www.lifetechnologies.com) is not available with a certificate that no animal-derived material (as e.g. proteases) was used to produce the peptone. In order to comply with possible rules to use only certified animal-free material already in early stage development, refer to other peptones as e.g. phytone from Beckton-Dickinson or plant peptones (e.g. soy peptone).
2. Selective medium means applying the selection pressure initially used for the transformation, i.e. the method of antibiotic resistance or auxotrophy complementation. In case the selected strains are still auxotroph, use minimal medium supplemented with the required amino acid for the pre-culture as well.
3. Any commercially available sterile breathable sealing film is appropriate. EXCEL Scientific´s AeraSeal™ (CAT# B525) was especially convenient due to the films´ relative robustness, which made it possible to remove/attach the film when inoculating or feeding the cultures without destroying the film.
4. A variety of shakers are available. It is required that the device features temperature control (ideally cooling to 20°C, but definitely to 25°C), humidity control (80% or more) and a shaking speed (with a shelf board suitable for deep-well plates) of up to 340 rpm. Suitable shelf boards for deep-well plates must be available.
5. It has been tested for both methanol- and glucose-driven screenings, and can be used within a pH range from 3.0 to 6.0. The citric acid used as buffer was shown not to be metabolized by *P. pastoris*.
6. Due to the low solubility of biotin, it is not a component of the PTM1 trace metal solution but dissolved and added separately.
7. Any other buffer composition can be applied, with the prerequisite that it suitable for pH ranges of 4 to above 6 (most frequently used range is 5 to 6). As word of caution, it must be said that the applicability of the protocol must be newly assessed when changing the buffer system. Also, the absence of cross-reactivity of any buffer system used (also the described potassium phosphate buffer) with the subsequent assay employed to validate expression of the target protein must be checked individually.
8. Appropriate concentrations for amino acid supplementations can be found in the "Invitrogen Pichia Protocols" via www.lifetechnologies.com.

9. Best results were obtained with plates from Bel-Art Biosciences (US-based company, www.belart.com, available in Europe via VWR).
10. After sterilization of the plates, it is advised to open the plates for inoculation and any other manipulation step only in the clean bench in order to exclude contamination.
11. Such “normalizer clones” can be e.g. the best performing clone from previous screenings or a randomly chosen clone from 1 transformation round that has not yet been tested in liquid culture, or any other *Pichia* strain producing a target protein (also different from the one tested in this cultivation) which can be easily assayed. Including such a “normalizer clone” facilitates the assessment of potentially occurring systemic differences between individual deep-well plates, be it through unequal treatment during cultivation, problematic differences of oxygen supply or humidity between certain positions on the shaker or any other reason.
12. Due to clonal variations, a statistical comparison (by Student´s t-test) of groups of samples (n=8) is helpful to assess the significance of changes caused by a host strain engineering strategy.
13. Transfer of cellular material from agar-plates is best done with sterile double-sharp toothpicks. The density of transformants on an agar-plate should not be too high, i.e. it is advisable to plate several times little volume of the transformation liquid onto several plates. As rule of thumb, a transformant count of 200 per agar-plate is the upper limit in order to guarantee a fast work-flow.
14. A master plate represents an agar-plate with cell mass originating from a re-streak from a colony on a transformation plate (not older than ~2 weeks).
15. A transformation plate means to be the selection agar-plate from the initial transformation (not older than ~2 weeks)
16. Use either 50 mL falcon tubes or 24-deep well plates (recommended due to operational convenience). Use cells from a plate not older than ~2 weeks. Inoculate with more cells rather than less because it is more convenient to transfer uniformly grown individual pre-cultures into main cultures. As the subsequent main culture is given enough time for batch phase, it doesn´t matter if the pre-culture is in stationary phase when transferred to the main culture. High variations in transfer volume due to different optical densities of the pre-cultures has a higher influence on product yields, which is avoided by early enough and dense inoculation of the pre-cultures.
17. Do not use more than 10% (200 µL) of main culture volume for inoculation (preferably, main culture inoculation should be performed with smaller volumes of 20-100 µL). Otherwise the final volumes of main culture have to be adjusted accordingly.
18. Accurate pipetting of methanol is essential. “Stepper” pipettes reduce time effort for feeding. It is also possible to dispense a diluted methanol solution to increase pipetting volumes.

19. Use cells from a plate not older than ~2 weeks. Inoculate with the approximate equal cell mass per colony on transformation plates of master plates: as it is impossible to estimate the exact cell mass to apply to the well, one should strive for avoiding the transfer of the entire cell mass of the largest colonies on the plate while taking only a fraction of small colonies on the plate. Ideally, a well-balanced cell mass per transferable colony will do.
20. Without proper liquid handling tools this step is extremely time-consuming when cultivating more than 2-3 plates (true for all methanol induction steps).
    As growth on methanol (especially for mut$^S$ strains) is limited anyhow, a reduction of the incubation temperature to 25°C or even as low as 20°C is not detrimental to the cell´s fitness but may well positively influence your screening results (valid for the whole production phase).
21. The optimal concentration of glucose to be supplied in the applied BMDx medium needs to be evaluated individually. BMDx might reach from BMD0.1 to BMD1 (meaning 0.1% to 1% glucose in the buffered minimal medium). As a rule of thumb, application of BMD0.5 (i.e. 0.5% glucose in the buffered minimal medium) for this first supply delivered good results for a variety of assayed strains producing different target proteins.
22. The optimal concentration of glucose to be supplied in the applied BMDx medium needs to be evaluated individually. BMDx might reach from BMD0.1 to BMD1 (meaning 0.1% to 1% glucose in the buffered minimal medium). As a rule of thumb, application of BMD0.3 (i.e. 0.3% glucose in the buffered minimal medium) at this stage delivered good results for a variety of assayed strains producing different target proteins.
23. This procedure saves tubes and cuvettes for separate $OD_{600}$ dilution and measuring, is accurate and relatively time-efficient.
24. Use a balance capable of weighing to 0.1mg
25. In case a wet cell weight determination and/or an $OD_{600}$ measurement is to be performed, make sure the appropriate volume is transferred to a different vessel before centrifuging the plates.
26. The withdrawal of the culture supernatant works more efficiently in PCR-plates. Also, it is much more laborious to resuspend cell mass pelleted on the bottom of square shaped deep-well plates as compared to the conical bottom of a PCR-plate.
27. The cell mass on the bottom of the PCR-plate is bound quite tightly, hence a vigorous banging of the plates on absorbent paper will efficiently remove remaining liquid while it does not loosen the pellet from the bottom. Essentially, no washing of the pellet is necessary for follow-up work using the cell mass.
28. Treatment of the cell mass for lysis (e.g. glassbeads or Y-PER) can be performed directly in the PCR-plate. For the subsequent centrifugation to obtain the soluble cytosolic fraction cleared from

glassbeads and/or debris, the PCR-plate provides a much better means as compared to the deep-well plate.

29. For storage in a freezer, it is advised to seal the PCR-plate (or the deep-well plate in case harvest of the pellet has been performed directly with the cultivation plate) with an aluminium seal. Make sure the seal adheres properly.

# 5 References


1. **Hartner FS, Glieder A.** 2006. Regulation of methanol utilisation pathway genes in yeasts. Microb Cell Fact **5:**39.
2. **Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA.** 1989. Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris. Mol Cell Biol **9:**1316-1323.
3. **Couderc R, Baratti J.** 1980. Oxidation of Methanol by the Yeast, *Pichia pastoris.* Purification and Properties of the Alcohol Oxidase. Agricultural and Biological Chemistry **44:**2279-2289.
4. **Murray WD, Duff SJB, Lanthier PH.** 1989. Induction and stability of alcohol oxidase in the methylotrophic yeast *Pichia pastoris*. Applied Microbiology and Biotechnology **32:**95-100.
5. **Cos O, Ramon R, Montesinos JL, Valero F.** 2006. Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast Pichia pastoris under different promoters: a review. Microb Cell Fact **5:**17.
6. **Gurramkonda C, Adnan A, Gabel T, Lunsdorf H, Ross A, Nemani SK, Swaminathan S, Khanna N, Rinas U.** 2009. Simple high-cell density fed-batch technique for high-level recombinant protein production with Pichia pastoris: Application to intracellular production of Hepatitis B surface antigen. Microb Cell Fact **8:**13.
7. **Kim S, Warburton S, Boldogh I, Svensson C, Pon L, d'Anjou M, Stadheim TA, Choi BK.** 2013. Regulation of alcohol oxidase 1 (AOX1) promoter and peroxisome biogenesis in different fermentation processes in Pichia pastoris. J Biotechnol **166:**174-181.
8. **Cereghino JL, Cregg JM.** 2000. Heterologous protein expression in the methylotrophic yeast Pichia pastoris. FEMS Microbiol Rev **24:**45-66.
9. **Khatri NK, Hoffmann F.** 2006. Impact of methanol concentration on secreted protein production in oxygen-limited cultures of recombinant Pichia pastoris. Biotechnol Bioeng **93:**871-879.
10. **Duetz WA, Witholt B.** 2001. Effectiveness of orbital shaking for the aeration of suspended bacterial cultures in square-deepwell microtiter plates. Biochem Eng J **7:**113-115.
11. **Jeude M, Dittrich B, Niederschulte H, Anderlei T, Knocke C, Klee D, Buchs J.** 2006. Fed-batch mode in shake flasks by slow-release technique. Biotechnol Bioeng **95:**433-445.

12. **Maurer M, Kuhleitner 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.
13. **Buchetics M, Dragosits M, Maurer M, Rebnegger C, Porro D, Sauer M, Gasser B, Mattanovich D.** 2011. Reverse engineering of protein secretion by uncoupling of cell cycle phases from growth. Biotechnol Bioeng.
14. **Kupcsulik B, Sevella B.** 2004. Effect of methanol concentration on the recombinant Pichia pastoris mut$^S$ fermentation. *Periodica Polytechnica Chemical Engineering* **48:**73-87.
15. **Inan M, Meagher MM.** 2001. Non-repressing carbon sources for alcohol oxidase (AOX1) promoter of Pichia pastoris. J Biosci Bioeng **92:**585-589.
16. **Panula-Perala J, Vasala A, Karhunen J, Ojamo H, Neubauer P, Mursula A.** 2014. Small-scale slow glucose feed cultivation of Pichia pastoris without repression of AOX1 promoter: towards high throughput cultivations. Bioprocess Biosyst Eng **37:**1261-1269.
17. **Lin-Cereghino GP, Godfrey L, de la Cruz BJ, Johnson S, Khuongsathiene S, Tolstorukov I, Yan M, Lin-Cereghino J, Veenhuis M, Subramani S, Cregg JM.** 2006. Mxr1p, a key regulator of the methanol utilization pathway and peroxisomal genes in Pichia pastoris. Mol Cell Biol **26:**883-897.
18. **Dietzsch C, Spadiut O, Herwig C.** 2011. A fast approach to determine a fed batch feeding profile for recombinant Pichia pastoris strains. Microb Cell Fact **10:**85.
19. **Weis R, Luiten R, Skranc W, Schwab H, Wubbolts M, Glieder A.** 2004. Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena. FEMS Yeast Res **5:**179-189.
20. **Krause M, Ukkonen K, Haataja T, Ruottinen M, Glumoff T, Neubauer A, Neubauer P, Vasala A.** 2010. A novel fed-batch based cultivation method provides high cell-density and improves yield of soluble recombinant proteins in shaken cultures. Microb Cell Fact **9:**11.
21. **Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM.** 1997. Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. Gene **186:**37-44.
22. **Mattanovich D, Jungo C, Wenger J, Dabros M, Maurer M.** 2014. Yeast Suspension Culture, p. 95-129. *In* Meyer H-P, Schmidhalter D (ed.), Industrial Scale Suspension Culture of Living Cells. Wiley-Blackwell.
23. **Garcia-Ortega X, Ferrer P, Montesinos JL, Valero F.** 2013. Fed-batch operational strategies for recombinant Fab production with Pichia pastoris using the constitutive GAP promoter. Biochemical Engineering Journal **79:**172-181.

24. **Glazyrina J, Krause M, Junne S, Glauche F, Storm D, Neubauer P.** 2012. Glucose-limited high cell density cultivations from small to pilot plant scale using an enzyme-controlled glucose delivery system. N Biotechnol **29:**235-242.
25. **Hemmerich J, Adelantado N, Barrigon JM, Ponte X, Hormann A, Ferrer P, Kensy F, Valero F.** 2014. Comprehensive clone screening and evaluation of fed-batch strategies in a microbioreactor and lab scale stirred tank bioreactor system: application on Pichia pastoris producing Rhizopus oryzae lipase. Microb Cell Fact **13:**36.
26. **Scheidle M, Jeude M, Dittrich B, Denter S, Kensy F, Suckow M, Klee D, Buchs J.** 2010. High-throughput screening of Hansenula polymorpha clones in the batch compared with the controlled-release fed-batch mode on a small scale. FEMS Yeast Res **10:**83-92.
27. **Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, Lin-Cereghino J, Kovar K, Cregg JM, Glieder A.** 2008. Promoter library designed for fine-tuned gene expression in Pichia pastoris. Nucleic Acids Res **36:**e76.
28. **Hartner F, Glieder A.** 2005. Mutant AOX promoters.
29. **Prielhofer R, Maurer M, Klein J, Wenger J, Kiziak C, Gasser B, Mattanovich D.** 2013. Induction without methanol: novel regulated promoters enable high-level expression in Pichia pastoris. Microb Cell Fact **12:**5.
30. **Mattanovich D, Gasser B, Maurer M, Prielhofer R, Klein J, Wenger J.** 2011. Regulatable promoter.

# 6 Acknowledgements

MB was funded by Biomin Research Center, Boehringer-Ingelheim RCV, Lonza AG, Biocrates Life Sciences AG, VTU Technology GmbH, Sandoz GmbH, the Austrian Federal Ministry of Science, Research and Economy (BMWFW), the Federal Ministry of Traffic, Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol and ZIT – Technology Agency of the City of Vienna through the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG.

## 4.2 Manuscript 2: *Pichia pastoris* Aft1 - a novel transcription factor, enhancing recombinant protein secretion

Genome wide analysis can highlight candidate genes for secretion engineering. In this study, a transcription factor, namely Aft1, was found to act as regulator for many secretory genes. It was then co-overexpressed in a *P. pastoris* strain producing a carboxylesterase and secretion was shown to increase up to 2.5-fold in bioreactor experiments. Additionally, the role of Aft1 was investigated by transcriptomic analysis, indicating a role in regulating carbon-responsive genes in *P. pastoris* rather than iron homeostasis genes. Aft1 was originally annotated as an *a*ctivator of *f*errous *t*ransport due to sequence homology to *S. cerevisiae´*s Aft1/2, extensively studied regulators of iron uptake and homeostasis. As in most fungi (except *S. cerevisiae*), the iron-regulatory pathway of *P. pastoris* is under control of a zinc finger GATA-type repressor, in this case Fep1.

By computational genome analysis tools, we investigated the upstream (promoter) regions of a set of genes, previously identified to improve secretion (when co-overexpressed) or to be differentially regulated when a human sialic acid transporter was secreted. The goal was to find transcription factor binding sites and ultimately a novel transcription factor for improving protein secretion. *P. pastoris* Aft1 binding sites were predicted for the regulatory regions of the genes *PDI1, KAR2, SSA4, KIN2* and *NCP1*. Regulatory Sequence Analysis Tools (RSAT) was used to identify additional putative binding sites (for the binding motifs known for *S. cerevisiae* Aft1/2) in *P. pastoris* promoters. Such binding motifs were found in 561 annotated genes, only 6 thereof annotated to be involved in iron regulation, but many other genes were found to play a role in the secretory pathway: 23 gene hits were found for protein targeting, 19 for Golgi to vesicle transport, 12 for response to oxidative stress, 11 for folding, 10 for glycosylation and vesicle organization, 9 for vacuole organization and regulation of transport, and 6 for exocytosis.

The amino acid sequence of *P. pastoris* Aft1 was also compared to its *S. cerevisiae* homolog(s). Although still highly homologous (especially in the DNA binding region), the *P. pastoris* Aft1 N-terminal sequence is shorter and lacks the conserved characteristic amino acid residues conferring iron sensitivity.

As a first experiment, *AFT1* overexpression (*AFT1*-OE) and *Δaft1* disruption mutants were generated and tested for their growth in low and high iron containing media to examine whether *P. pastoris* Aft1 is involved in iron regulation at all. No differences were observed. Then, we analysed transcriptional patterns of the knock out strain in comparison to the wildtype using DNA microarrays. It was shown that *AFT1* disruption led to both, down-regulation (34 genes) but also up-regulation (54 genes). Among the regulated ones were 9 putative transcriptional regulators and genes containing at least one putative Aft1 binding site were clearly overrepresented compared to the total *P. pastoris* genes. Regarding iron regulation, no changes in transcript levels were observed for the GATA-type repressor *FEP1* or other genes involved in iron uptake and homeostasis. Moreover, carbon-source responsive transporters of the multifacilitator superfamily were among the regulated genes in the *Δaft1* mutant, implicating that Aft1 is involved in the regulation of glucose-repressed genes.

The *AFT1* promoter was cloned into a plasmid in front of a gene for the intracellular expression of green fluorescent protein (GFP) and *P. pastoris* was transformed with this vector. Clones expressing GFP from the *AFT1* promoter were studied for intracellular fluorescence levels in comparison to clones expressing GFP from the constitutive *GAP* promoter. $P_{AFT1}$ turned out to be a fairly strong promoter as $P_{AFT1}$ clones reached about 60% (50% in terms of yield) of the fluorescent levels of the $P_{GAP}$ clones. Bioreactor cultivations with different glucose feed strategies showed that $P_{AFT1}$ performs best under glucose limiting conditions. A glucose surplus significantly decreased its strength.

*AFT1* was then co-overexpressed in two carboxylesterase secreting strains: The better secreting strain was shown to have 6 expression cassettes integrated in its genome while an average strain had only one copy. Parallel bioreactor cultivations confirmed: In both strains, *AFT1* overexpression enhanced product secretion, with the effects being more pronounced in the single-copy production strain (2-fold increase).

MICROBIAL CELL FACTORIES

RESEARCH Open Access

# *Pichia pastoris* Aft1 - a novel transcription factor, enhancing recombinant protein secretion

Claudia Ruth[1†], Markus Buchetics[2†], Viktorija Vidimce[1], Daniela Kotz[3], Stefan Naschberger[3], Diethard Mattanovich[2,4], Harald Pichler[1,5*] and Brigitte Gasser[2,4*]

## Abstract

**Background:** The methylotrophic yeast *Pichia pastoris* is frequently used for the production of recombinant proteins. However, expression levels can vary depending on the target protein. Allowing for simultaneous regulation of many genes, which may elicit a desired phenotype like increased protein production, overexpression of transcription factors can be used to overcome expression bottlenecks. Here, we present a novel *P. pastoris* transcription factor currently annotated as Aft1, activator of ferrous transport.

**Results:** The promoter regions of key secretory *P. pastoris* genes were screened for fungal transcription factor binding sites, revealing Aft1 as an interesting candidate for improving secretion. Genome wide analysis of transcription factor binding sites suggested Aft1 to be involved in the regulation of many secretory genes, but also indicated possible novel functions in carbohydrate metabolism. No Aft binding sites were found in promoters of characteristic iron homeostasis genes in *P. pastoris*. Microarrays were used to study the Aft1 regulon in detail, confirming Aft1 involvement in the regulation of carbon-responsive genes, and showing that iron regulation is dependent on *FEP1*, but not *AFT1* expression levels. The positive effect of *AFT1* overexpression on recombinant protein secretion was demonstrated for a carboxylesterase from *Sphingopyxis* sp. MTA144, for which secretion was improved 2.5-fold in fed batch bioreactor cultivations.

**Conclusion:** This study demonstrates that the transcription factor Aft1 can be used to improve recombinant protein secretion in *P. pastoris*. Furthermore, we discovered possible novel functions of Aft1 in carbohydrate metabolism and provide evidence arguing against a direct role of Aft1 in *P. pastoris* iron regulation.

**Keywords:** *Pichia pastoris*, Aft1, Transcription factor, Novel functions, Enhanced secretion

## Background

The methylotrophic yeast *Pichia pastoris* (syn. *Komagataella phaffii*) is among today's most frequently used yeast systems for the production of recombinant proteins [1]. Benefits of this yeast are the capability of high cell density cultivations, eukaryotic posttranslational modifications and good secretion capacity. A low level of endogenously secreted proteins allows for the production of relatively pure, recombinant secretory proteins. The recent availability of the genomic sequence boosted the generation of a versatile *P. pastoris* toolbox, including various strains, plasmids and promoters of different strength. To overcome individual bottlenecks during protein folding and secretion, a variety of helper factors such as the ER foldases Pdi1 or BiP (Kar2) have been studied in recent years [2,3]. The capability of transcription factors (TFs) as expression helpers was demonstrated by Guerfal *et al.* and Gasser *et al.* [4,5], who improved the secretion of the mIL-10 protein and antibody fragments by overexpression of the UPR (unfolded protein response) transcription factor *HAC1*. Also overexpression of the gene encoding the TF Nrg1 was shown to positively influence the secretion of recombinant porcine and human trypsinogen as well as the antibody Fab fragment 2 F5 [6].

Allowing for the simultaneous regulation of different proteins involved in e.g. folding and secretion, TFs have

* Correspondence: harald.pichler@tugraz.at; brigitte.gasser@boku.ac.at
†Equal contributors
[1]Austrian Centre of Industrial Biotechnology (ACIB GmbH), Petersgasse 14, 8010 Graz, Austria
[2]Austrian Centre of Industrial Biotechnology (ACIB GmbH), Muthgasse 11, 1190 Wien, Austria
Full list of author information is available at the end of the article

huge potential to overcome bottlenecks in the cellular protein production machinery.

Here, we present a novel *P. pastoris* TF which was annotated as Aft1 (Activator of ferrous transport) by sequence homology to *Saccharomyces cerevisiae* Aft1/2. While no information is currently available on the function of *P. pastoris* Aft1, its two *S. cerevisiae* homologs have been studied extensively. In *S. cerevisiae* the transcriptional activators Aft1/2 are responsible for iron uptake and homeostasis fulfilling overlapping but non-redundant roles [7,8]. It was shown that iron homeostasis is primarily maintained by Aft1, while Aft2 is the weaker transcriptional activator [8,9]. Still, both TFs can interact with the same iron-responsive element (FeRE) found within promoters of the genes of the iron regulon such as the iron reductases *FRE1–6* or the multicopper oxidases *FET3/FET5* [8]. Under iron limiting conditions, Aft1 binds to promoters of the iron regulon genes and increases their expression [10,11]. Under iron repletion, the monothiol glutaredoxins (Grx3/4) attached to an iron-sulfur cluster bind Aft1 and initiate its dissociation from target promoters [12]. Export of Aft1 from the nucleus is mediated by the nuclear export receptor Msn5 [13]. Several amino acid (aa) residues have been shown to be important for the iron-responsive regulation of Aft1. While phosphorylation of Ser210/Ser224 and an intermolecular interaction are essential for recognition by Msn5 [13], the residues Leu99, Leu102, Cys291, Cys293 are involved in the interaction with Grx3/4 and iron dependent regulation [12].

In contrast to *S. cerevisiae*, in most fungi such as *Candida albicans*, *Pichia stipitis* or *Schizosaccharomyces pombe*, the iron-regulatory pathway is controlled by a conserved zinc finger GATA-type repressor [14,15]. Also in *P. pastoris* iron uptake was shown to be under control of a GATA-factor, named Fep1. Fep1 was shown to bind to DNA only under iron repletion and disruption of *FEP1* led to constitutively high expression of *FET3*, independent of the availability of iron [16]. Interestingly, several species such as *C. albicans, P. stipitis* and also *P. pastoris* have been found to possess an Aft-type regulator in addition to the GATA-type repressor, leading to the question on the function of Aft in these species [16,17].

In this study, we investigated the functions of the *P. pastoris* Aft1 regulator based on the prediction of putative Aft binding sites in promoters, focusing on the secretion of recombinant proteins. We provide evidence that Aft1 is not directly involved in *P. pastoris* iron regulation, but rather in carbon-responsive regulation. Furthermore, we show that *AFT1*, when overexpressed under its natural promoter, increased the secretion of a model protein up to 2.5-fold in fed batch bioreactor cultivations.

## Results and discussion

### Analysis of transcription factor binding sites in selected genes shown to enhance protein secretion

To identify novel TFs useful for improving protein secretion in *P. pastoris*, the promoter regions, i.e. 1000 bp upstream of the ATG start codon, of key secretory genes involved in folding, transport and exocytosis were studied for putative fungal TF binding sites using the program MatInspector (Genomatix, [18]). The following genes, which were previously identified to improve secretion if modified [4,5,19,20], and/or differentially regulated upon expression of a human sialic acid transporter [21], were studied: *ERO1, HAC1, KAR2, PDI1*, *YDJ1*, *CNE1*, *SSE1*, *SSA4*, *SSB1*, *IRE1*, *UBI4*, *KIN2*, *SSO2*, *CUP5*, *CDS1*, *PGS1*, *ERG1*, *ERG3*, *ERG25*, *NCP1* and *INO1.*

Obtained TF hits were analysed according to frequency, function (link to the secretory machinery or stress response), matrix similarity (similarity of the input sequence to TF matrixes stored in the database, cut-off 0.9) and distance to the start codon (proximal more weighty than distal).

*S. cerevisiae* Aft1/2 binding sites were found enriched in the *P. pastoris* genes *PDI1*, *KAR2*, *SSA4*, *KIN2* and *NCP1* (Table 1). By sequence homology, the gene product of PAS_chr1-4_0361/PP7435_Chr1-1146 [NCBI] was identified as the single *P. pastoris* Aft protein, therefore called Aft1.

**Table 1 Putative *P. pastoris* Aft1 binding sites found in promoters of selected genes known to enhance protein secretion**

| Gene | Functions[a] | Aft1 UBS (bp) | Sequence |
|---|---|---|---|
| *PDI1* | Protein disulfide isomerase, ER chaperone, formation of disulfide bonds | -423[b] to -409 | gatcacaCACCctct |
| *KAR2* | ER chaperone, mediating protein folding, UPR regulation | -870[b] to -856 | tcgtataCACCctca |
| | | -194[b] to -180 | tgtaataCACCcttg |
| *SSA4* | HSP70 protein, co-translational protein-membrane targeting and translocation of nascent proteins into the ER | -608[c] to -594 | actcatgCACCctta |
| | | -342[b] to -328 | catggaaCACCccat |
| *KIN2* | Serine/threonine kinase involved in exocytosis | -177[c] to -163 | ataactgCACCcaga |
| *NCP1* | P450 reductase, involved in lipid metabolism (ergosterol biosynthesis) | -539[c] to -525 | ggttttgCACCcagg |

[a]Functions are derived for the *S. cerevisiae* homologs of the *P. pastoris* genes in the Saccharomyces Genome Database.
[b]Aft2 matrix.
[c]Aft1 matrix.
Upstream binding sites (UBS): 5′→ 3′, -1000 to -1 A(0)TG. The core binding sequences, representing the highest conserved, consecutive positions are highlighted. TFBS analysis: MatInspector [18] using the search groups fungi and general core promoter elements.

*P. pastoris* Aft1 binding sites were predicted for the regulatory regions of Pdi1 and BiP, two important ER chaperones, indicating that Aft1 is involved in oxidative protein folding. Consistently, Blaiseau *et al.* [7] showed *S. cerevisiae* Aft2 involved in oxidative stress resistance. Also, the chaperone Ssa4 and the serine/threonine kinase Kin2, both recently identified as secretion enhancing factors [3], were found to have an upstream Aft1 binding site, confirming Aft1 as an interesting candidate for improving recombinant protein secretion in *P. pastoris*. In addition, Aft1 binding was predicted for the promoter region of *NCP1*, suggesting that Aft1 is also involved in lipid metabolism [22].

### Prediction of Aft1 binding sites in the *P. pastoris* genome

To elucidate the function of Aft1 in *P. pastoris*, Regulatory Sequence Analysis Tools [23] was used to search for putative Aft1 binding sites in *P. pastoris* promoters. Using the binding motifs known for *S. cerevisiae* Aft1/2, ANTGCACCC and BRCACCCB, resulted in 972 genes with a putative Aft binding site. Thereof, 561 were found annotated [24] and mapped to broader parent terms, GO slim terms, using AmiGO GO Slimmer [25]. Aft1 was found involved in the regulation of approximately 100 different biological processes (Additional file 1), including nucleobase-containing small molecule metabolic processes (57 hits), carbohydrate metabolic processes (50 hits), transcription from RNA polymerase II promoter (47 hits) or amino acids metabolic processes (47 hits). Interestingly, within the categories ion transport (20 hits) and cellular ion homeostasis (11 hits) only 6 proteins with a possible connection to iron regulation were found: Ccc1, mediating transfer of iron from cytosol to vacuole [26]; Gef1, a chloride channel localized to the Golgi or endomembrane system, which has also been reported to be governing iron-limited growth [27]; Hmx1, a heme oxygenase, required for the reutilization of iron from heme, also involved in oxidative stress resistance [28,29] and Nfu1, a protein involved in mitochondrial iron sulfur cluster assembly [30]. Notably, none of these are characteristic iron regulatory genes such as the iron reductases *FRE1-6* or the multicopper oxidases *FET3/5*. However, a putative Aft1 binding site was found within the upstream region of *FEP1*, the GATA type repressor, which has been shown to negatively control iron uptake/homeostasis in *P. pastoris* [16]. Thus, though no Aft1 binding site was found within the upstream region of characteristic iron regulators, the presence of such within the upstream region of *FEP1* raises the possibility of a latent, indirect involvement of Aft1 in *P. pastoris* iron regulation.

Analysing possible other functions of *P. pastoris* Aft1, a high number of biological processes related to the cellular secretory pathway machinery was found (Table 2). 23 gene hits were found for protein targeting, 19 for Golgi to vesicle transport, 12 for response to oxidative stress, 11 for folding, 10 for glycosylation and vesicle organization, 9 for vacuole organization and regulation of transport, and 6 for exocytosis. Additionally to the already discussed ER chaperones Pdi1 and BiP, several important secretory regulators were identified. Sec12 and Sec23, both involved in COPII vesicle formation and ER to Golgi transport [31], Sec61, forming a channel for protein translocation into and out of the ER [32], Gos1, a v-SNARE protein found to be involved in ER to Golgi and/or intra-Golgi transport [33], Yap1, a transcription factor involved in oxidative stress response [34], Och1, a Golgi resident mannosyltransferase initiating the hypermannosylation of glycoproteins [35] and Kin1, a protein kinase involved in exocytosis [36], were found to possess a putative upstream Aft1 binding site. In accordance with the above-mentioned data, an Aft1 binding site was found in the upstream region of several genes conferring resistance to oxidative stress, such as *YAP1* or *HMX1*, supporting the hypothesis of a role of Aft1 in oxidative protein folding and/or stress response.

**Table 2 *P. pastoris* genes with putative Aft1 binding site (s) related to the cellular secretory pathway machinery**

| GO slim term - Biological process | Number (Genes) |
|---|---|
| GO:0006605 protein targeting | 23 (*ATG1 ATG12 ATG18 GET3 GET4 IMP2 KAR2 LHS1 NPL6 NUP188 NUP84 PAM17 PEX14 PEX7 PEX8 SEC61 SPC1 SPC3 TAM41 VPS21 VPS64 VPS68 VPS8*) |
| GO:0048193 Golgi vesicle transport | 19 (*AGE2 APL4 CHS5 DRS2 EMP24 GET3 GOS1 HRR25 KES1 SEC12 SEC23 SFB3 SNC2 SVP26 TED1 TLG1 TRS31 VTI1 YPT32*) |
| GO:0006979 response to oxidative stress | 12 (*AFT1 CTA1 EOS1 GCY1 GSH1 HMX1 HSP104 MXR1 POS5 TSA1 YAP1 YBL055C*) |
| GO:0006457 protein folding | 11 (*CAJ1 CCT2 CCT6 CPR6 EGD2 ERV2 HLJ1 HSP104 PDI1 SIS1 TSA1*) |
| GO:0006486 protein glycosylation | 10 (*ALG7 EOS1 KRE5 MNN11 OCH1 OST6 PMT1 PMT5 STT3 SVP26*) |
| GO:0016050 vesicle organization | 10 (*EMP24 GOS1 SEC12 SEC23 SFB3 SNC2 TLG1 UBC1 VPS4 VTI1*) |
| GO:0007033 vacuole organization | 9 (*ATG1 ATG12 ENO1 GYP7 RDI1 TPM2 VPS21 VTI1 YHC3*) |
| GO:0051049 regulation of transport | 9 (*AKL1 ARG81 CUP9 FPK1 GEF1 SEC12 SEC23 SLG1 TUP1*) |
| GO:0006887 exocytosis | 6 (*ARG81 KES1 KIN1 SNC2 TPM2 YPT32*) |

Search field: -1000 bp upstream of the *P. pastoris* GS115 coding sequences. Binding motifs: ScAft1 (ANTGCACCC) and ScAft2 (BRCACCCB). Gene hits were categorized into biological function GO terms using AmiGO GO Slimmer (Yeast GO slim set, [25]).

### *P. pastoris* Aft1 has a conserved Aft-domain, but does not contain the iron-responsive motif

The amino acid sequence of *P. pastoris* (Pp) Aft1 was compared to *S. cerevisiae* (Sc) Aft1/2 and *Kluyveromyces*

*lactis* (Kl) Aft using ClustalW2 [37]; all of these proteins display the positive mode of regulation and the characteristic residues Leu99, Leu102, Cys291 and Cys293 conferring iron sensitivity (numbering based on ScAft1, [14]). Interestingly, the *P. pastoris* N-terminal sequence is 41-98 amino acids shorter, lacking the conserved residues Leu99 and Leu102 (PpAft1: Met1, Ile4; Figure 1). Also the residues Cys291 and Cys293 were not found conserved in the *P. pastoris* sequence (PpAft1: Ser180, Ser182), suggesting a role different from iron regulation for PpAft1. Still, also regions of high homology were found in the N-terminal part of the *P. pastoris* protein, particularly the Aft-domain (supposedly the DNA binding domain [38]) between the amino acids 10-130. Especially high conservation was found for the residues 10-45 (identity 53%) and 102-130 (identity 45%), including the two conserved cysteines and histidines, which have been suggested to be part of a WRKY-motif involved in zinc binding [14,39]. Regarding nuclear export, no conservation was found for Ser210 in *P. pastoris* (PpAft1: Arg97) and *K. lactis* (KlAft: Gly224), but Ser224 is conserved in all 4 sequences (PpAft1: Ser111, KlAft: Ser238). However, as Ueta *et al.* [13] showed that individual serine mutations did not affect ScAft1 localization, export via Msn5p seems still possible for PpAft1 and KlAft. In contrast to ScAft1 and KlAft, a glutamine-rich region was found not at the C-terminal end, but between the residues 164-177. This glutamine-rich region is followed by serine (aa 180-213) and asparagine repeats (aa 216-269), maybe as part of a protein-protein interaction domain [40]. No homology was found for the C-terminal part of the PpAft1 protein (identity <1% ), in particular the residues 162 to 363. Interestingly, with only 363 amino acids PpAft1 is considerably shorter than ScAft1 (690 aa), ScAft2 (416 aa) and KlAft (823 aa).

PpAft1 was further compared to *C. albicans* Aft [NCBI protein: XP_714862], which also lacks the characteristic residues Leu99, Leu102, Cys291 and Cys293. Similar to *P. pastoris*, in *C. albicans* iron regulation was shown to be under control of a GATA-type repressor [17]. Sequence comparison of these two proteins revealed only one short N-terminal region of high homology between the residues 3-47 (identity 67%, data not shown). This region was also found to be highly similar in ScAft1/2 and KlAft. Additionally, we searched the genome of the genetically and biochemically close methylotrophic yeast *Hansenula polymorpha* for a homolog of PpAft1. The protein encoded by HPODL_04658 in *H. polymorpha* (NCBI protein: ESX01890) has high N-terminal sequence homology including the two conserved cysteines und histidines (aa 3-132: identity 55%, Additional file 2). A region of high similarity was also found between the residues 244-332 (identity 31%), suggesting that HPODL_04658 functions similar to PpAft1. As PpAft1 and CaAft, also HPODL_04658 lacks the characteristic iron sensitivity residues. Accordingly, we also found a homolog of the GATA-type repressor in the *H. polymorpha* genome (HPODL_03720; 48% sequence identity to PpFep1). In summary, Aft homologs from species that are known to contain also a Fep1-like GATA type repressor (*C. albicans*, *Pichia stipitis*, *Debaryomyces hansenii* according to [14] and *H. polymorpha*) share high sequence homology to PpAft1 in the N-terminal DNA binding region, but lack the characteristic residues that are conferring iron sensitvity in ScAft1/2 (data not shown). Interestingly, all these species possess another protein of unknown function having a domain with sequence similarity to parts of the Aft DNA-binding domain, but low similarity to ScAft1/2 (i.e. less than 15% overall identity).

### The *P. pastoris* Aft1 regulon

To investigate if *P. pastoris* Aft1 is involved in iron regulation, we generated an *AFT1* overexpression (AFT1-OE) and an Δ*aft1* disruption mutant, and tested their growth in low and high iron containing media. AFT1-OE was achieved by expressing an additional copy of *AFT1* under control of its native promoter (see below), while the Δ*aft1* strain was generated by exchanging parts of the gene for the KanMX marker cassette using the split marker approach as described by Heiss *et al.* [41]. Low iron media was prepared by addition of the iron chelator BPS (bathophenanthroline disulphonate) to YPD or YNB-Glucose agar plates as described by Miele *et al.* [16]. In order to exclude the possibility to be outside the sensitivity range we tried several concentrations of BPS in the iron sensitivity assay. Miele et al. [16] reported the use of YPD containing 80–160 μM BPS for *P. pastoris*. We tested concentrations ranging from 80-200 μM BPS in YP and minimal medium (YNB or M2) using either glucose or methanol as carbon source. We did not observe growth impairment of *P. pastoris* CBS7435 wild type using 80 μM BPS in YPD or YPM (data not shown). When using higher BPS concentrations, growth of all strains was significantly delayed. However, we did not observe differences between the Δ*aft1* strain, the AFT1-OE and the wild type control on any of the media tested (Figure 2). Thus, disruption of *AFT1* does not render *P. pastoris* sensitive to iron-limited conditions. This behaviour is contrary to the phenotype observed for Δ*aft1/2* in *S. cerevisiae* and Δ*aft1* in *K. lactis*, which grow only poorly or are unable to grow in the same low iron conditions [7,14], suggesting that *P. pastoris* Aft1 is not involved in iron regulation. Additionally, the *P. pastoris* Δ*aft1* mutant did not show impaired growth on the cell wall disturbing agent Calcofluor White, which is again in contrast to what is reported for *S. cerevisiae* Δ*aft1* (not shown).


**Figure 1** (See legend on next page.)

(See figure on previous page.)
**Figure 1 Aft1 protein sequence comparison.** Amino acid sequences: *S. cerevisiae* Aft1 [NCBI protein: NP_011444] and Aft2 [NCBI protein: NP_015122], *K. lactis* Aft [NCBI protein:CAH00307], and *P. pastoris* Aft1 [NCBI protein:CCA37276] were aligned using ClustalW2 [37]. asterisk: "indicates positions which have a single, fully conserved residue", colon: "indicates conservation between groups of strongly similar properties"; period: "indicates conservation between groups of weakly similar properties"; shaded grey: conserved residues (Leu99*, Leu102, Cys291 and Cys293 conferring iron sensitivity. Ser210 and Ser224 involved in Msn5 recognition and nuclear export. Cys143, Cys215, His239 and His241, which are suggested to be part of a WRKY-motif involved in zinc binding [39]). Differences are boxed and/or highlighted red; *numbering is based on ScAft1.

In order to identify the regulon of Aft1 in *P. pastoris*, we analysed the transcription patterns of the Δ*aft1* strain and control using DNA microarrays. All strains were cultivated in synthetic minimal medium with glucose feed beads for 5 h in three biological replicates. Interestingly, although Aft-like proteins are described as transcriptional activators, an even higher number of genes was up-regulated (54 genes, 33 thereof more than 1.5-fold) than down-regulated (34 genes, 13 thereof more than 1.5-fold) in the Δ*aft1* mutant compared to the wild type (an adjusted p-value of < 0.05 was applied as cut-off to identify significantly regulated genes, Additional file 3). 42% of the up-regulated and 38% of the down-regulated genes contain at least one putative Aft1 binding site in their promoters (compared to 18% of total *P. pastoris* genes), suggesting that both, up- and down-regulation, are a direct consequence of *AFT1* disruption. Table 3 summarizes differentially regulated genes in the Δ*aft1* mutant compared to the wild type strain according to their GO term category.

In the Δ*aft1* strain, 9 putative transcriptional regulators are among the regulated genes, seven of them carrying putative Aft1 binding sites in their promoters. PAS_chr4_0324, a fungal specific transcription factor of unknown function with a Zn2/Cys6 DNA-binding domain is the highest up-regulated gene, while *MIG1-1*, encoding a transcription factor connected to glucose repression, has lower expression levels in the Δ*aft1* mutant. Most other regulated transcription factors are also of the fungal-specific Zn2/Cys6 type, but lack annotated function and defined target genes. Furthermore, 23 of the 54 up-regulated and 9 of the 34 down-regulated genes encode hypothetical proteins of unknown function. The second largest group of regulated genes comprises transport proteins, mainly transmembrane transporters, however, there is no clear preference for transported substrates, which range from ions to polyamines and sugars.

Regarding iron regulation, no changes in transcript levels were observed for the GATA-type repressor *FEP1*, in line with the unaltered expression levels of genes involved in iron uptake and homeostasis (i.e. *FRE1–6* or *FET3/FET5*). Re-analysis of our previous microarray data (obtained in different environmental conditions, [42,43]) for expression changes of iron regulatory genes indeed

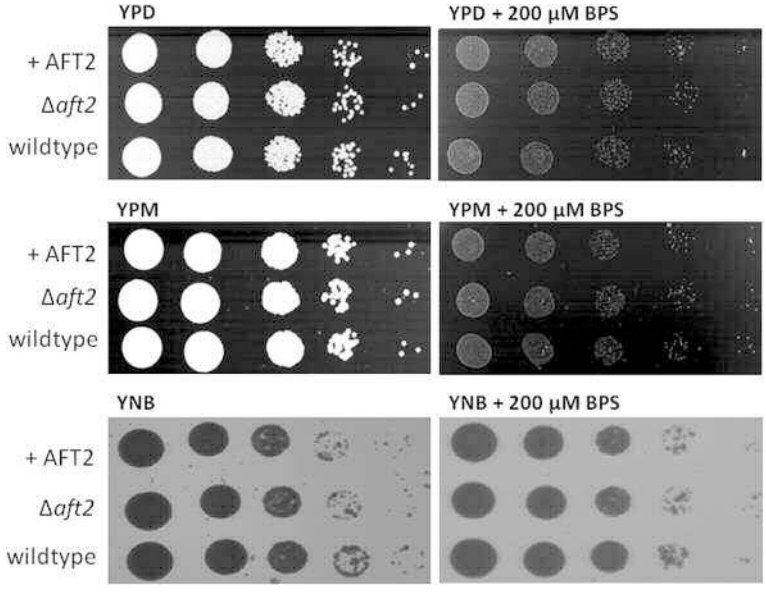


**Figure 2 Spotting assay on iron deficient media.** Serial 1:10 dilutions of *P. pastoris* CBS7435 wild type, AFT1-OE and Δ*aft1* cell suspensions were spotted on YP plates containing either 2% glucose or 1% methanol as carbon source and on YNB agar plates containing 2% glucose as carbon source. Iron-deficient conditions were generated by addition of the iron chelator BPS (200 µM). Plates were incubated at 28°C for 70 h. While addition of BPS resulted in generally delayed growth on both substrates, no difference between the strains was seen.

**Table 3 Differentially expressed and annotated genes in the Δ*aft1* vs. control (wild type) strain**

| Up-regulated genes (Δ*aft1* vs. control) | |
|---|---|
| **GO slim term - Biological process** | **Number (Genes)** |
| GO:0006811 ion transport | 7 ( *AQR1 ATX1 GEF1 MEP2 TPO3 YHL008C ZRT1* ) |
| GO:0008150 biological_process* | 6 ( *ECM13 GPM3 YEL023C YLR156C-A YLR278C YOR292C* ) |
| GO:0055085 transmembrane transport | 6 ( *AQR1 GEF1 MEP2 STL1 TPO3 ZRT1* ) |
| GO:0006366 transcription from RNA polymerase II promoter | 5 ( *GCR1 PHD1 RPM2 UGA3 YLR278C* ) |
| GO:0042221 response to chemical | 5 ( *AQR1 ATX1 GCR1 SLI1 UGA3* ) |
| GO:0006520 cellular amino acid metabolic process | 4 ( *ADH2 ECM4 SFA1 UGA3* ) |
| GO:0007124 pseudohyphal growth | 4 ( *HMS1 MEP2 PHD1 PTP1* ) |
| GO:0051186 cofactor metabolic process | 4 ( *ADH2 ALD4 BIO2 THI21* ) |
| GO:0005975 carbohydrate metabolic process | 3 ( *DOG1 GCR1 GPM3* ) |
| GO:0006091 generation of precursor metabolites and energy | 2 ( *GCR1 GPM3* ) |
| GO:0006766 vitamin metabolic process | 2 ( *BIO2 THI21* ) |
| GO:0006873 cellular ion homeostasis | 2 ( *ATX1 GEF1* ) |
| GO:0055086 nucleobase-containing small molecule metabolic process | 2 ( *ADH2 ALD4* ) |
| GO:0001403 invasive growth in response to glucose limitation | 1 ( *PTP1* ) |
| GO:0002181 cytoplasmic translation | 1 ( *RPM2* ) |
| GO:0006325 chromatin organization | 1 ( *GCR1* ) |
| GO:0006397 mRNA processing | 1 ( *RPM2* ) |
| GO:0006470 protein dephosphorylation | 1 ( *PTP1* ) |
| GO:0006629 lipid metabolic process | 1 ( *YJU3* ) |
| GO:0006865 amino acid transport | 1 ( *AQR1* ) |
| GO:0006979 response to oxidative stress | 1 ( *ATX1* ) |
| GO:0007005 mitochondrion organization | 1 ( *RPM2* ) |
| GO:0008033 tRNA processing | 1 ( *RPM2* ) |
| GO:0008643 carbohydrate transport | 1 ( *STL1* ) |
| GO:0070647 protein modification by small protein conjugation or removal | 1 ( *PCI8* ) |

**Table 3 Differentially expressed and annotated genes in the Δ*aft1* vs. control (wild type) strain** *(Continued)*

| Down-regulated genes (Δ*aft1* vs. control) | |
|---|---|
| **GO slim term - Biological process** | **Number (Genes)** |
| GO:0006811 ion transport | 4 ( *ATO2 JEN1 TAT2 VMA8* ) |
| GO:0055085 transmembrane transport | 4 ( *ITR1 JEN1 PEX13 TAT2* ) |
| GO:0006281 DNA repair | 3 ( *PCD1 RVB1 YRA1* ) |
| GO:0006325 chromatin organization | 3 ( *ACS1 RLF2 RVB1* ) |
| GO:0006366 transcription from RNA polymerase II promoter | 3 ( *MIG1 RVB1 YHP1* ) |
| GO:0006974 cellular response to DNA damage stimulus | 3 ( *PCD1 RVB1 YRA1* ) |
| GO:0008150 biological_process* | 3 ( *NBA1 TOS8 YHR177W* ) |
| GO:0005975 carbohydrate metabolic process | 2 ( *CTS1 MIG1* ) |
| GO:0006766 vitamin metabolic process | 2 ( *SNO1 SNZ3* ) |
| GO:0051186 cofactor metabolic process | 2 ( *ACS1 SNZ3* ) |
| GO:0055086 nucleobase-containing small molecule metabolic process | 2 ( *PCD1 YRA1* ) |
| GO:0070271 protein complex biogenesis | 2 ( *PEX13 RLF2* ) |
| GO:0000278 mitotic cell cycle | 1 ( *YHP1* ) |
| GO:0000910 cytokinesis | 1 ( *CTS1* ) |
| GO:0002181 cytoplasmic translation | 1 ( *RPL4A* ) |
| GO:0006091 generation of precursor metabolites and energy | 1 ( *ACS1* ) |
| GO:0006520 cellular amino acid metabolic process | 1 ( *SNO1* ) |
| GO:0006605 protein targeting | 1 ( *PEX13* ) |
| GO:0006865 amino acid transport | 1 ( *TAT2* ) |
| GO:0006873 cellular ion homeostasis | 1 ( *VMA8* ) |
| GO:0007031 peroxisome organization | 1 ( *PEX13* ) |
| GO:0008643 carbohydrate transport | 1 ( JEN1 ) |
| GO:0015931 nucleobase-containing compound transport | 1 ( *YRA1* ) |
| GO:0016570 histone modification | 1 ( *ACS1* ) |
| GO:0018193 peptidyl-amino acid modification | 1 ( *ACS1* ) |
| GO:0042221 response to chemical | 1 ( *MIG1* ) |
| GO:0042594 response to starvation | 1 ( *MIG1* ) |
| GO:0043543 protein acylation | 1 ( *ACS1* ) |
| GO:0048193 Golgi vesicle transport | 1 ( *BRE5* ) |
| GO:0051049 regulation of transport | 1 ( *BRE5* ) |
| GO:0051169 nuclear transport | 1 ( *YRA1* ) |
| GO:0070647 protein modification by small protein conjugation or removal | 1 ( *BRE5* ) |

*Biological process is unkown.
Genes were categorized into biological function GO terms using AmiGO GO Slimmer (Yeast GO slim set, [25]).

revealed that induction of the iron regulon is dependent on the levels of *FEP1*, but not *AFT1* in *P. pastoris*.

Notably, Atx1 and Gef1, which were described to be required for the correct assembly of the high affinity iron transporter Fet3 in *S. cerevisiae*, were found among the up-regulated genes in the Δ*aft1* mutant, pointing towards a different regulation of iron acquisition genes in *P. pastoris*.

Among the up-regulated genes several genes which are described to be repressed by glucose in *S. cerevisiae* are found, including alcohol dehydrogenase *ADH2* and the two putative mitochondrial aldehyde dehydrogenases *ALD4-1* and *ALD5*. All of these contain putative Aft1 binding sites in their promoter regions. Interestingly, deletion of these gene functions has recently been predicted by the genome scale metabolic model of *P. pastoris* to enhance recombinant protein production [44]. Moreover, carbon-source responsive transporters of the multifacilitator superfamily are among the regulated genes in the Δ*aft1* mutant. While the genes encoding the high affinity glucose transporter Ght1 and the putative glycerol transporter PAS_c034_0021 are up-regulated in the Δ*aft1* mutant, the second high affinity glucose transporter of *P. pastoris*, encoded by *GHT2* as well as a homolog of the *S. cerevisiae* myo-inositol transporter *ITR1*, are repressed in this strain. This implicates that Aft1 is involved in the regulation of glucose-repressed genes, at least in an indirect manner by regulating *MIG1-1* expression, but maybe also directly as some of these genes also contain Aft1 binding sites in their promoters. In this respect, it should be noted that *AFT1* expression levels are significantly higher not only in glucose-limited conditions, but also in cells grown on methanol as compared to glucose or glycerol surplus (see below and own unpublished data). A correlation of *AFT1/2* expression levels and the carbon source has also been observed in *S. cerevisiae*, where differences between the fermentable carbon source glucose and the non-fermentable carbon source glycerol were reported [38].

### Regulation of the *P. pastoris* AFT1 promoter

To study the expression strength of the natural *AFT1* promoter and to test if it is applicable for overexpression studies, intracellular expression of green fluorescent protein was analysed (cycle-3-GFP, [45]). 96-well deep-well plate screening, which has been applied for *P. pastoris* promoter studies before [46], was used for cultivation. However, the protocol was adjusted for *GAP* promoter driven expression, which was used for the expression of our model protein carboxylesterase.

*P. pastoris* was transformed with plasmid pPpKan-S-GFP. Clones expressing GFP from the *AFT1* promoter were studied for intracellular fluorescence levels in comparison to clones expressing GFP from the constitutive *GAP* promoter. On average, the tested $P_{AFT1}$ clones reached 724 relative fluorescence units (RFU) after 69 h of cultivation, which was 60% in comparison to the respective $P_{GAP}$ clones. In terms of average yields, the $P_{AFT1}$ clones reached 343 RFU $OD^{-1}$, corresponding to 50% of the yield of the $P_{GAP}$ clones, characterising $P_{AFT1}$ as fairly strong promoter under the conditions tested. Analysing expression at 24, 48 and 69 h of cultivation, differences in regulation were observed. While the $P_{GAP}$ clones showed the expected constitutive expression profile with stable titers of 1201 ± 64 RFU at the different time points, $P_{AFT1}$ controlled expression significantly increased with longer cultivation times, reaching a maximum at 69 h of cultivation (Figure 3A). Assuming that the observed effect was related to a decrease in glucose concentration, a batch series starting with 1, 2 and 4% glucose was studied for about 48 h (data not shown). The results confirmed enhanced $P_{AFT1}$ activity at lower

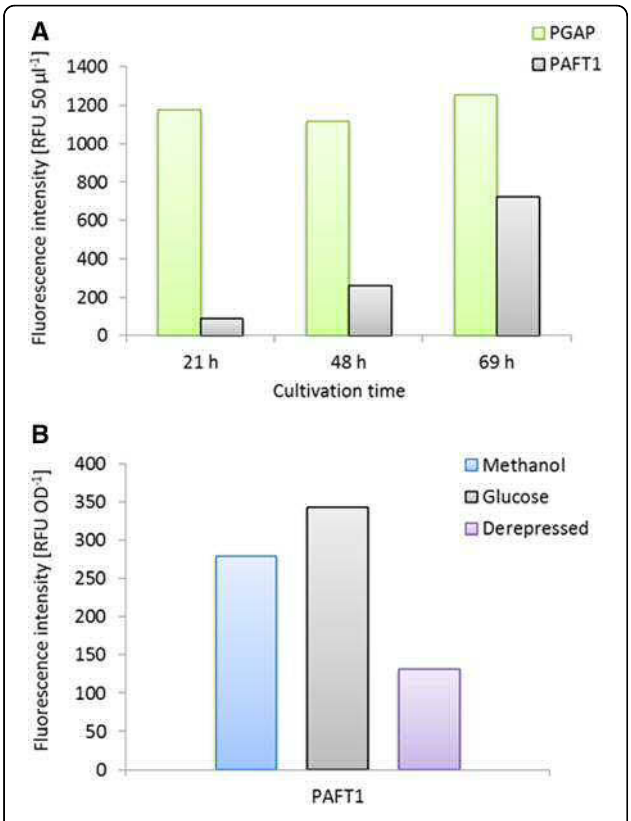


**Figure 3 Characterization of the *AFT1* promoter. (A)** GFP expression levels under control of $P_{GAP}$ and $P_{AFT1}$. Cultivation in 96-well plate. RFU: relative fluorescence units, average of 87 *P. pastoris* clones. **(B)** GFP expression levels under control of $P_{AFT1}$ using different cultivation conditions. Glucose: cultivation time 69 h, 2% glucose batch, 3 times addition of glucose to 0.5%. Methanol: cultivation time 92 h, 48 h glucose batch (1%), 3 times addition of methanol to 0.5%. Derepressed: cultivation time 69 h, 2% glucose batch, no additional supplementation.

glucose concentrations, showing average yields of 104, 81 and only 16 RFU $OD^{-1}$, for the 1, 2 and 4% glucose batches, respectively.

Interestingly and matching above data, the *AFT1* promoter was also found inducible by methanol, reaching 80% (280 RFU $OD^{-1}$) of the yield of glucose cultivated cells (Figure 3B). In comparison, using a 2% glucose batch without supplementation (derepressed protocol), a yield of only 40% was reached. While $P_{AFT1}$ is obviously a rather strong and also methanol-inducible promoter, unlimited access of glucose negatively affects $P_{AFT1}$ activity and presumably decreases overexpression effects.

### The effect of *AFT1* overexpression on the secretion of recombinant proteins

Although none of the secretion-related genes was differentially expressed in the Δ*aft1* mutant under the analysed conditions, our next step was to analyse whether *AFT1* overexpression influences the secretion of recombinant proteins. Our model protein of choice was a carboxylesterase from *Sphingopyxis* sp. MTA144. This carboxylesterase is under development for use as a feed additive enzyme, because it hydrolyses an antinutritive substance that may be naturally contained in animal feed [47]. *P. pastoris* has recently been shown to secrete active carboxylesterase [48]. However, for a technological application of this carboxylesterase as feed enzyme for gastrointestinal detoxification in animals, a high yield recombinant production process is required.

### Generation of carboxylesterase secreting strains

Carboxylesterase secreting strains were generated by transformation of *P. pastoris* CBS7435 with plasmid pPM2dZ30-PGAPα-CE expressing carboxylesterase under control of the *GAP* promoter and using the *S. cerevisiae* alpha mating factor signal sequence as secretion leader. Carboxylesterase represented the major fraction of total secreted protein of each transformant, yielding a strong band at the calculated size of 52 kDa. However, due to varying numbers of integrated expression cassettes, strains secreting lower and higher levels of carboxylesterase were observed (Additional file 4). The best-secreting strain CE#18 contained 6 copies of the expression cassette, while the average strain CE#12 had only one expression cassette integrated in its genome, confirming a positive correlation between secretion level and gene copy number. Using ELISA, an expression level of 80 μg $mL^{-1}$ was determined for CE#18, whereas a titer of approximately 20 μg $mL^{-1}$ was determined for strain CE#12.

### *AFT1* co-expression studies

The strains CE#12 and CE#18 were chosen to study whether overexpression of *AFT1* can positively influence recombinant protein secretion. Both strains were transformed with linearized pPM2aK21-AFT1, comprising the *AFT1* gene under control of its natural promoter and terminator. Supernatants of 8 transformants of each strain were studied for carboxylesterase secretion after 48 h of cultivation in shake flask. ELISA was used to reveal small, but significant differences in carboxylesterase secretion levels, showing improved secreting clones for both strain backgrounds. Figure 4 shows the results of the best two clones of each strain. Interestingly, while the best CE#12-AFT1 clones yielded a strong improvement of 62 ± 5%, the best two CE#18-AFT1 strains showed considerably less improvement, reaching an increase of only 19 ± 6%. However, it has to be considered that strain CE#18 produced four times more carboxylesterase (80 μg $mL^{-1}$) than strain CE#12 (20 μg $mL^{-1}$). Assuming that Aft1 is involved in oxidative protein folding, high overproduction of carboxylesterase could have led to an overload of recombinant protein in the ER and subsequent induction of ERAD. Newly elicited upstream or downstream bottlenecks could also be the reason for the reduced influence of *AFT1* overexpression and the only modest improvement for strain CE#18.

The growth behaviour of *AFT1* overexpressing strains was studied in addition. In contrast to secretion levels, growth under these conditions was seemingly unaffected by overexpression of *AFT1*, showing neither significant changes in final OD levels (<15% , 48 h) nor any detectable alteration in the growth curve (data not shown).

### Bioreactor cultivations

To confirm the positive effect of *AFT1* overexpression on carboxylesterase secretion, the best strains from screening were studied in bioreactor cultivations. First, the performance of the *AFT1* overexpressing strain CE#12-AFT1-A was compared to the starting strain (CE#12). Cultivations were performed using aerobic and

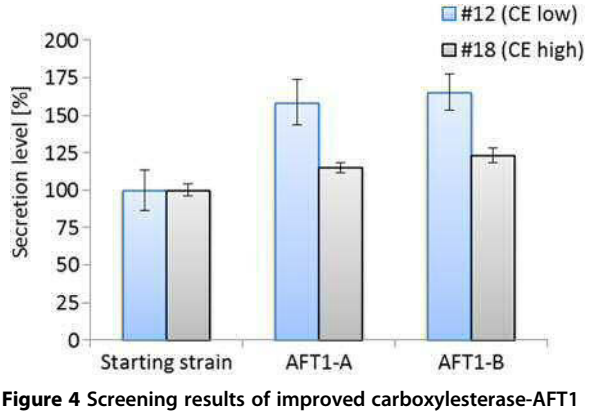


**Figure 4 Screening results of improved carboxylesterase-AFT1 *P. pastoris* strains.** Secretion levels: determined by ELISA and normalized to the respective starting strain. Starting strains: CE#12 (blue, low level expressing strain), CE#18 (grey, high level expressing strain). AFT1-A and AFT1-B: *AFT1* overexpressing clones.

hypoxic production conditions as preceding bioreactor experiments had indicated that a low oxygen supply is beneficial for carboxylesterase secretion (data not shown), an observation that had also been made for the production of antibody Fab fragments and trypsinogen [49]. While under aerobic production conditions a dissolved oxygen (DO) concentration of 20% was maintained throughout the whole process, a DO level of 5% was applied during the feed phase for hypoxic conditions. Bioreactor cultivations were monitored for 100 h. In addition to the quantification of secreted recombinant carboxylesterase by SDS-PAGE, also the functional quality of the enzyme was assessed by an activity assay.

The positive effect of *AFT1* overexpression on the secretion rate of strain CE#12 was indeed confirmed in bioreactor. A maximum of 141 U $L^{-1}$ was reached for strain CE#12-AFT1-A after 100 h of cultivation under hypoxic conditions, while the starting strain reached only a level of 76 U $L^{-1}$ at this time point (Figure 5A). Confirming the beneficial effect of hypoxic production conditions, carboxylesterase activity levels were 3-4 fold lower under aerobic conditions. Applying 20% DO, a maximum of only 37 U $L^{-1}$ was reached for strain CE#12-AFT1-A after 100 h of cultivation. However, comparing to the starting strain under the same conditions a 2.5-fold higher activity level was achieved for the *AFT1* overexpressing strain. It seems that the amount of dissolved oxygen does not abrogate the beneficial impact of *AFT1* overexpression on carboxylesterase secretion. As observed previously [43], lower amounts of biomass were reached in hypoxic compared to normoxic conditions due to the production of ethanol at the low DO setpoint. Consistently, about 2-fold higher carboxylesterase yields (U $g^{-1}$ biomass) were obtained for the *AFT1* overexpressing strain under both conditions. The results of the activity assays were confirmed by SDS-PAGE (Figure 5B). Comparing to a bovine serum albumin (BSA) standard, a carboxylesterase protein level of about 0.75 g $L^{-1}$ was estimated for strain CE#12-AFT1-A after 100 h of cultivation under hypoxic conditions, while the starting strain produced below 0.5 g $L^{-1}$ until this time point.

Also the performance of the multi-copy carboxylesterase *AFT1* overexpressing strain CE#18-AFT1-A was investigated in bioreactor under hypoxic conditions (Figure 5C). Again the positive effect of *AFT1* overexpression was confirmed. Strain CE#18-AFT1-A reached 169 U $L^{-1}$ after 74 h of cultivation, while the starting strain (CE#18) only reached 133 U $L^{-1}$ at the same time point. As before, results were confirmed by SDS-PAGE, showing increased protein levels for strain CE#18-AFT1-A after 74 and 98 h of cultivation (Figure 5D). Comparing yields, secretion per biomass was increased 24% for strain CE#18-AFT1-A. The *AFT1* overexpressing strain also yielded 25% higher ethanol levels, reconfirming the involvement of *P. pastoris* Aft1 in carbon-responsive regulation.

The *AFT1* copy number of the overexpressing strains was also investigated. RT-PCR revealed one *AFT1* overexpression cassette for strain CE#12-AFT1-A and three for strain CE#18-AFT1-A. Notably, strain CE#18-AFT1-

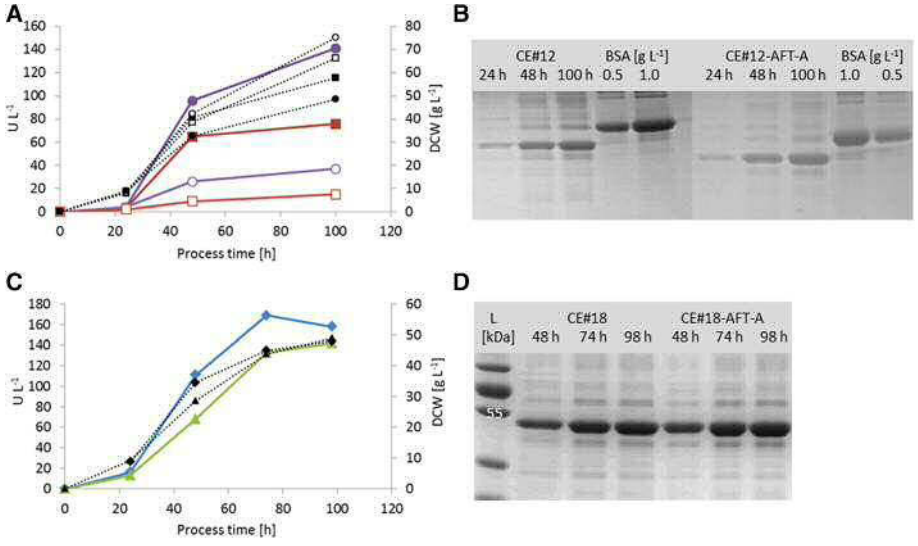


**Figure 5 Bioreactor cultivations of improved carboxylesterase-AFT1 *P. pastoris* strains. (A)** Carboxylesterase activity levels (U $L^{-1}$) of strain CE#12 (red squares) and strain CE#12-AFT1-A (purple circles). Closed symbols: hypoxic conditions; open symbols: aerobic conditions. **(C)** Carboxylesterase activity levels (U $L^{-1}$) of strain CE#18 (green triangles) and strain CE#18-AFT1-A (blue rhomboids). Dashed lines: biomass values (DCW, g $L^{-1}$). **(B)** and **(D)** Reducing SDS-PAGE of supernatant samples from different cultivation time points (hypoxic conditions). BSA: bovine serum albumin standard (66.5 kDa).

B, which had only one *AFT1* overexpression cassette integrated, did not show superior behaviour in bioreactor (data not shown). Seemingly, several *AFT1* expression cassettes are necessary to positively influence the secretion of the high-level carboxylesterase producer #18, which might allow for further improvement by targeted *AFT1* copy number amplification or the use of stronger promoters for *AFT1* overexpression.

Both carboxylesterase-strains, CE#12 and CE#18, were also transformed with an empty vector control plasmid (pPM2aK21-empty). However, carboxylesterase secretion levels were not influenced, neither in shake flask (CE#12-empty) nor in bioreactor cultivations (CE#18-empty).

## Conclusions

By analysing the promoter regions of secretion enhancing *P. pastoris* genes, the ortholog of the *S. cerevisiae* iron regulators Aft1/2 was selected as a novel factor to improve recombinant protein secretion. A genome-wide analysis of putative Aft1 binding sites in *P. pastoris* showed Aft1 to be involved in the regulation of many secretory genes, in addition to genes involved in carbohydrate metabolism. The absence of Aft1 binding sites in iron regulatory genes, i.e. *FRE1–6* or *FET3/FET5*, led us to assume iron-independent functions of Aft1 in *P. pastoris*. These findings were supported by primary amino acid sequence analysis, showing that the DNA binding domain, but not the iron-responsive motif is conserved in the *P. pastoris* protein. Using DNA microarrays we unveiled further evidence that Aft1 might not directly be required for iron regulation, but is rather involved in regulatory mechanisms in response to carbon source availability, showing e.g. higher transcriptional activation at low glucose concentrations. We also discovered an involvement of *P. pastoris* Aft1 in the expression of glucose-repressible genes, which needs to be analysed in more detail in future studies. Finally, though microarray data did not reveal differential regulation of any secretion related genes, the secretion enhancing effect of *AFT1* was confirmed in overexpressing strains, yielding up to 2.5-fold more secreted carboxylesterase.

## Materials and methods

### Strains and plasmids

*P. pastoris* wild type CBS7435 (Centraalbureau voor Schimmelcultures, NL) was used as host strain [50]. The plasmid pPpKan-S-GFP for GFP expression was described by Näätsaari *et al.* ([GenBank:JQ519694], [51]). Plasmid pPM2dZ30-PGAPα, a derivative of pPUZZLE [6], was used for the expression of a carboxylesterase from *Sphingopyxis macrogoltabida* (aa residues 48 to 540, [GenBank: ACS27056]) under control of the *P. pastoris GAP* promoter, with the *S. cerevisiae* α-MF leader sequence for secretion and a Zeocin resistance marker cassette. Prior to transformation, the expression vector was linearized within the *GAP* promoter using the restriction enzyme *Bln*I for homologous integration into the native *GAP* promoter locus of the *P. pastoris* genome.

For overexpression of *P. pastoris AFT1* the pPUZZLE derived plasmid pPM2aK21 was used, which contains the KanMX4 cassette conferring resistance to Kanamycin/Geneticin (G418), and an *AOX1* terminator sequence, which, when linearized with *Asc*I, provides the homologous stretches for integration into the native *P. pastoris AOX1* terminator locus. The *AFT1* expression cassette, including the *AFT1* gene [*P. pastoris* gene identifier: PAS_chr1-4_0361] and 1000 bp up- and 400 bp downstream sequences, was amplified from *P. pastoris* genomic DNA using following primers: ApaI-AFT-fw (AAAGGGC CCCCAGGTGAATGTACGTAATGGAG) and AgeI-AFT-rv (TTTACCGGTGGGGAGAAGCCGAATTGGAAG). After *Apa*I/*Age*I digestion, the PCR product was cloned into pPM2aK21, creating pPM2aK21-AFT.

### *AFT1* gene knock out

A split marker cassette approach was used as described by Heiss *et al.* [41] to generate transformants with a disrupted *AFT1* gene locus. The *AFT* flanking regions (A upstream, D downstream) were fused to the G418 resistance cassette fragments B and C, respectively, by overlap PCR. Then, equal amounts of both split marker fragments (AB and CD) were pooled and simultaneously transformed into *P. pastoris*. The G418 resistance gene is reconstituted when both split marker fragments integrate at the correct locus. Verification of positive Δ*aft1* transformant strains was done by PCR using a primer pair designed to bind on the native *AFT1* locus up- and downstream of the split-marker cassette on genomic DNA of Geneticin-resistant transformants. Primers and split marker fragment sizes are shown in Additional file 5.

### Transformation

Electrocompetent *P. pastoris* cells were transformed using the following parameters: 1.5-2.0 kV, 25 μF and 200 Ω. After two hours of regeneration on YP medium, containing 20 g $L^{-1}$ glucose, selection of positive transformants was done by incubation for 48 h and 28°C on YPD agar plates (20 g $L^{-1}$ yeast extract, 10 g $L^{-1}$ peptone, 20 g $L^{-1}$ glucose, 20 g $L^{-1}$ agar-agar) supplemented with antibiotic, 50 μg $mL^{-1}$ Zeocin and/or 450 μg $mL^{-1}$ Geneticin, respectively.

### Media and cultivation

If not stated otherwise chemicals were purchased from BD, Carl Roth, Merck and Sigma Aldrich. YP medium contained 20 g peptone and 10 g yeast extract per liter. Buffered minimal (BM)-medium contained 10 g yeast extract, 10 g peptone, 100 mM potassium phosphate

buffer (pH 6.0), 13.4 g yeast nitrogen base without amino acids and 0.4 mg biotin per liter.

Microscale cultivations for GFP expression were performed in 96-well deep-well plates. 300 μL of BM-medium, containing 2% glucose, were inoculated using a toothpick and incubated at 25°C and 360 rpm. The culture was supplemented with glucose to 0.5% at 24 and 36 h of cultivation. Cells were finally harvested after 69 h of cultivation. Methanol induced cultivations were carried out based on the protocol of Weis *et al.* [52]. Cells were grown for 48 h in 300 μL of BM-medium, containing 1% glucose. Cells were induced by addition of methanol (0.5%) after 48, 56 and 72 h of cultivation. Finally, cells were harvested after 92 h of cultivation.

Shake flask cultivations were performed in 100 mL shake flasks without baffles. As pre-culture, 2.5 mL YP medium, containing 2% glycerol and the respective antibiotic(s), were inoculated in a 50 mL falcon tube and incubated for a minimum of 24 h at 25°C and 180 rpm. The main cultures, containing 10 mL BM-medium with 2% glucose, were inoculated to an $OD_{600}$ of 0.1. Cultures were then incubated at 25°C and 180 rpm, and supplemented three times with glucose to 0.5% in 12 h intervals. Cells were harvested after 48 h by centrifugation at 4000 rpm. Subsequently, supernatants were analyzed for carboxylesterase production and cell pellets were used for cell weight analysis. Alternatively, $OD_{600}$ was measured.

Pre-cultures for bioreactor cultivations were performed in 500 mL baffled shake flasks and 50 mL YP medium, containing 2% glucose and 0.01 g $L^{-1}$ Glanapon DG160 (Bussetti). The main medium used for fed batch cultivations was used as described by Zhao *et al.* [53], supplemented with biotin and Glanapon DG160. The batch medium consisted of 4 g $L^{-1}$ $KH_2PO_4$, 4 g $L^{-1}$ $(NH_4)_2SO_4$, 0.38 g $L^{-1}$ $CaCl_2$, 18.2 g $L^{-1}$ $K_2SO_4$, 9.4 g $L^{-1}$ $MgSO_47H_2O$, 40 g $L^{-1}$ glucose monohydrate (Tereos Syral), 1 g $L^{-1}$ Glanapon DG160 and 1 mL $L^{-1}$ trace element solution. The trace element solution contained 2.50 g $L^{-1}$ $MnSO_4H_2O$ (Riedel-de-Haën), 54.17 g $L^{-1}$ $FeSO_47H_2O$ (Merck), 16.67 g$L^{-1}$ $ZnCl_22H_2O$ (Riedel-de-Haën) and 0.17 g$L^{-1}$ $Na_2MoO_42H_2O$. The batch medium was supplemented with 2 mL of 0.2 g $L^{-1}$ biotin stock solution per liter medium. The feed medium consisted of 600 g $L^{-1}$ glucose monohydrate, 2 g $L^{-1}$ $(NH_4)_2PO_4$ and 1 g $L^{-1}$ Glanapon DG160 and was supplemented with 2 mL of biotin stock solution per liter medium.

High cell-density fermentations were carried out in 1 L bioreactors (DASGIP). A starting volume of 500 mL batch medium was inoculated to an $OD_{600}$ of 0.3. The pH was measured using a glass electrode (Mettler Toledo) and maintained at $5.0 \pm 0.1$ by automatic addition of 25% ammonium hydroxide (AppliChem). Dissolved oxygen (DO) was monitored using an optical, dissolved oxygen electrode (Hamilton) and maintained at 20% of saturation during the batch phase by a DO cascade of agitation and aeration. Hypoxic conditions were applied during the feed phase. Maximum agitation was 1200 rpm (corresponding to a maximum tip speed of 2.89 m $s^{-1}$); maximum aeration was set to 1.3 vvm. Temperature was maintained at 25°C.

### SDS-PAGE and western blot

10 μL of culture supernatant (containing secreted carboxylesterase enzyme) were run on a reducing sodium dodecyl sulfate (SDS) NuPAGE® 12% Bis-Tris polyacrylamide gel (Life technologies™) with NuPAGE® morpholinepropanesulfonic acid (MOPS) buffer at 180 V for 60 min. Protein bands were visualized using Coomassie staining solution. For Western blotting, SDS-PAGE separated proteins were transferred to a nitrocellulose membrane using the XCell II™ Blot Module for wet (tank) transfer (Life technologies™) according to the manufacturer's instructions. Carboxylesterase was detected using anti-carboxylesterase antiserum as described by Heinl *et al.* [48].

### Gene copy number determination using real time PCR

Copy numbers were determined by Real time (RT)-PCR as described by Abad *et al.* [54]. The ABI PRISM 7300 Real Time PCR System and Power SYBR® Green PCR Master Mix were used (Life technologies™). Normalization of the data was achieved using the *P. pastoris ARG4* gene as reference. The number of copies per μL was calculated using Avogadro's number. Following primers were used for amplification: ARG4-RTfw (TCCTCCGGTGGCAG TTCTT), ARG4-RTrv (TCCATTGACTCCCGTTTGAG), AFT-RTfw (GGGCAATATCCAATAGGGCTAA), AFT-RTrv (GGGTGCGCCAAGACTAACA), ZEO-RTfw (CG GCCTGGACGAGCTGTA), ZEO-RTrv (GGCTGCTCG CCGATCTC). Genomic DNA was prepared according to Hoffman and Winston [55].

### Fluorescence measurements

GFP fluorescence levels were determined per 50 μL of cell culture. Measurements were performed in microplate using the SynergyMX plate reader (Biotek) applying the following settings: excitation 395 nm and emission 507 nm.

### Carboxylesterase activity assay

Enzymatic activity was determined photometrically at 405 nm by hydrolysis of p-Nitrophenyl 2-(trimethylsilyl) ethyl carbonate (pNSi) to p-Nitrophenol + 2-(trimethylsilyl) ethanol + carbon dioxide. The reaction was performed at 37°C in microtiter plates and followed over time using a Tecan Sunrise™ plate reader (XFLUOR4 version: V 4.51). Clarified fermentation supernatants were used undiluted and in several dilutions ($10^{-1}$ - $10^{-5}$) using 1 × FCE buffer.

10 × FCE buffer consisted of 200 mM Tris-Cl, pH 8.0 and 1 mg $mL^{-1}$ BSA. 100 mM pNSi stock solution was prepared by dissolving 141.68 mg pNSi in 5 mL of 96% ethanol. 1 mM pNSi reaction solution was made by mixing 100 mM pNSi stock solution with 1× FCE buffer. 20 μL of the test samples were provided in a microtiter plate and the reaction was started by addition of 180 μL reaction solution. Absorbance was measured every 30 s for 2 h and activity was determined according to Lambert-Beer's law.

### Iron-dependent growth analysis: spotting assay

The ferrous iron chelator Bathophenanthrolinedisulfonate disodium salt (BPS; Sigma) reduces the amount of free iron in the medium. YP agar plates containing 80 - 200 μM BPS were prepared by addition of BPS to the melted YP or YPD (YP + 2% glucose) agar right before pouring the plates. Methanol (1%) was applied directly in 250 μL volumes onto YP agar plates and spread using a sterile spatula. YNB agar plates containing 200 μM BPS were prepared by addition of BPS to the melted YNB agar right before pouring the plates. YNB agar contained 3.4 g $L^{-1}$ Yeast Nitrogen Base (Becton Dickinson, NJ), 10 g $L^{-1}$ ammonium sulfate, 20 g $L^{-1}$ glucose and 100 mM potassium phosphate buffer (pH 6.0). Cells from a YPD agar plate were resuspended in 1 mL sterile PBS, the optical density ($OD_{600}$) was determined and set to $OD_{600}$ = 0.3. Five μL of serial 1:10 dilutions (in sterile PBS) were applied on each agar plate. The plates were then incubated at 30°C for 72 h.

### Regulatory sequence analysis tools (RSAT)

RSAT subcategory pattern matching and string genome-scale dna-pattern was used to search for the *S. cerevisiae* Aft1 and Aft2 binding sites within -1000 bp upstream of the *P. pastoris* GS115 coding sequences. Default settings were applied.

### AmiGO Go Slimmer

AmiGO Go Slimmer (version 1.8) was used to map genes into GO slim terms according to their biological process. SGD was used as database filter (Evidence Code: all) and Yeast GO slim as pre-existing GO slim set (GO database release: 27.10.2012 (analysis of *P. pastoris* Aft1 binding sites) and 15.07.2013 (analysis of microarray data: *Δaft1* vs. control). Advanced results option was used to display gene products and counts for each slim term.

### Microarray hybridization and data analysis

For the generation of samples for microarray analysis the Δ*aft1* strain and the wild type control were cultivated in three biological replicates. Pre-cultures were cultivated as described above. The main cultures, containing 2 mL M2 medium, were inoculated to an optical density $OD_{600}$ of 4.0 and a 12 mm glucose FeedBead (Kuhner Shaker) was added. The synthetic medium M2 contained per liter: 22.0 g Citric acid monohydrate, 3.15 g $(NH_4)_2PO_4$, 0.49 g $MgSO_4$*7H2O, 0.80 g KCl, 0.0268 g $CaCl_2$*2$H_2O$, 1.47 mL of PTM1 trace metals and 4 mg Biotin; pH was set to 5 with KOH (solid). Cultures were shaken at 180 rpm and 25°C. Slow release of glucose ensured glucose limited growth. Samples were taken after five h of main culture (estimated specific growth rate: 0.08 $h^{-1}$), fixed in phenol/ethanol (5% phenol (v/v) in pure ethanol, ice-cold), and stored at -80°C until total RNA extraction.

Total RNA extraction was performed using Trizol as described in Graf *et al.* [42]. cDNA synthesis and labelling as well as the microarray hybridizations (in-house designed *P. pastoris* specific oligonucleotide arrays, AMAD-ID 034821, 8x15K custom arrays, Agilent) were carried out according to the Agilent protocols Quick Amp Labelling Kit (Cat. No. 5190-0444) and Gene Expression Hybridisation Kit (Cat. No. 5188-5242) using a reference design. Therefore, each sample was labelled in a dye-swap manner and hybridized against a reference cDNA, which was generated from a pool of cells grown under different culture conditions. Normalization steps and statistical analysis of microarray data included removal of color bias using locally weighted MA-scatterplot smoothing (LOESS) followed by between array normalization using the "Aquantile" method. For identifying differentially expressed genes and calculating p-values a linear model fit with an eBayes correction was used. P-values were adjusted for multiple testing with the false discovery method (FDR) by Benjamini & Yekutieli. For identifying differentially expressed genes, a fold change cut-off of at least 1.5 > FC >1/1.5 was applied. All steps were done using the R software (http://www.rproject.org) and the limma package. The expression changes of some genes selected based on their regulation pattern was confirmed using quantitative real time PCR (Additional file 6).

## Additional files

**Additional file 1: *P. pastoris* genes with putative Aft1 binding site(s).** Search field: -1000 bp upstream of the *P. pastoris* GS115 coding sequences. Binding motifs: ScAft1 (ANTGCACCC) and ScAft2 (BRCACCCB). Genes were categorized into biological function GO-Terms using AmiGO GO Slimmer (Yeast GO slim set, [25]). Genes not found in the database and removed from the calculation: *FLO104*, *ATG30*, *SHL23* and *FEP1*.

**Additional file 2: Aft1 protein sequence comparison (ClustalW2, [37]).** Amino acid sequences: *P. pastoris* Aft1 [NCBI protein:CCA37276] and *H. polymorpha* HPODL_04658 [NCBI protein: ESX01890]; asterisk: "indicates positions which have a single, fully conserved residue", colon: "indicates conservation between groups of strongly similar properties"; period: "indicates conservation between groups of weakly similar properties"; red boxed: conserved residues (Cys45*, Cys102, His126 and His128), shaded grey: Aft domain (pfam_Is:AFT, Motif Scan-My Hits-SIB), *numbering is based on PpAft1.

**Additional file 3 List of significantly regulated genes in Δ*aft1* compared to wild type.** GO term enrichment was analysed using

Amigo GO Slimmer [25]. Significantly regulated GO terms were determined by GO term Finder. List of homologs of the *S. cerevisiae* iron regulon which are not regulated in Δ*aft1* in *P. pastoris*.

**Additional file 4: SDS-PAGE of carboxylesterase secreting *P. pastoris* strains.** Clones #8-20: CBS7435 transformed with plasmid pPM2dZ30-PGAP-CE; negative control: CBS7435.

**Additional file 5: Primers used for the construction of the split marker cassette for *AFT1* disruption in *P. pastoris*.**

**Additional file 6: Primers used for the verification of gene regulation in Δ*aft1* and wild type.** Comparison of gene regulation patterns from microarrays and qPCR for genes selected based on their regulatory behaviour.

#### Competing interests
The authors declare that they have no competing interests.

#### Authors' contributions
CR, SN, DM, HP and BG participated in the design of the study. CR and VV carried out the sequence analysis, promoter and overexpression studies. Bioreactor cultivations were performed by DK. MB and BG carried out the DNA microarrays and data analysis. CR and MB drafted the manuscript. DK, SN, DM, HP and BG revised the manuscript. All authors read and approved the final manuscript.

#### Acknowledgements
This work has been supported by the Federal Ministry of Science, Research and Economy (BMWFW), the Federal Ministry of Traffic, Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol and ZIT - Technology Agency of the City of Vienna through the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG. Further support by Biomin Research Center, Boehringer-Ingelheim RCV, Lonza AG, Biocrates Life Sciences AG, VTU Technology GmbH, and Sandoz GmbH is acknowledged. The carboxylesterase activity assay with pNSi as substrate originated at evocatal GmbH, Monheim am Rhein, Germany. The authors thank Dr. Alexandra B. Graf (University of Applied Sciences - FH Campus Vienna) for the processing of the microarray data, Dr. Minoska Valli (ACIB) for help with *P. pastoris* gene annotation, and Gabriele Wilt (ACIB) and Elena Pietrich (ACIB) for their excellent technical support.

#### Author details
[1]Austrian Centre of Industrial Biotechnology (ACIB GmbH), Petersgasse 14, 8010 Graz, Austria. [2]Austrian Centre of Industrial Biotechnology (ACIB GmbH), Muthgasse 11, 1190 Wien, Austria. [3]BIOMIN Research Center, Technologiezentrum Tulln, Technopark 1, 3430 Tulln, Austria. [4]Department of Biotechnology, BOKU University of Natural Resources and Life Sciences Vienna, Muthgasse 18, 1190 Vienna, Austria. [5]Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria.

Received: 8 May 2014 Accepted: 10 August 2014
Published: 3 September 2014

#### References
1. Gasser B, Prielhofer R, Marx H, Maurer M, Nocon J, Steiger M, Puxbaum V, Sauer M, Mattanovich D: **Pichia pastoris: protein production host and model organism for biomedical research.** *Future Microbiol* 2013, **8:**191–208.
2. Damasceno LM, Anderson KA, Ritter G, Cregg JM, Old LJ, Batt CA: **Cooverexpression of chaperones for enhanced secretion of a single-chain antibody fragment in Pichia pastoris.** *Appl Microbiol Biotechnol* 2007, **74:**381–389.
3. 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:**6499–6507.
4. Guerfal M, Ryckaert S, Jacobs PP, Ameloot P, Van Craenenbroeck K, Derycke R, Callewaert N: **The HAC1 gene from Pichia pastoris: characterization and effect of its overexpression on the production of secreted, surface displayed and membrane proteins.** *Microb Cell Fact* 2010, **9:**49.
5. Gasser B, Maurer M, Gach J, Kunert R, Mattanovich D: **Engineering of Pichia pastoris for improved production of antibody fragments.** *Biotechnol Bioeng* 2006, **94:**353–361.
6. Stadlmayr G, Benakovitsch K, Gasser B, Mattanovich D, Sauer M: **Genome-scale analysis of library sorting (GALibSo): isolation of secretion enhancing factors for recombinant protein production in Pichia pastoris.** *Biotechnol Bioeng* 2010, **105:**543–555.
7. Blaiseau PL, Lesuisse E, Camadro JM: **Aft2p, a novel iron-regulated transcription activator that modulates, with Aft1p, intracellular iron use and resistance to oxidative stress in yeast.** *J Biol Chem* 2001, **276:**34221–34226.
8. Rutherford JC, Jaron S, Winge DR: **Aft1p and Aft2p mediate iron-responsive gene expression in yeast through related promoter elements.** *J Biol Chem* 2003, **278:**27636–27643.
9. Rutherford JC, Ojeda L, Balk J, Mühlenhoff U, Lill R, Winge DR: **Activation of the iron regulon by the yeast Aft1/Aft2 transcription factors depends on mitochondrial but not cytosolic iron-sulfur protein biogenesis.** *J Biol Chem* 2005, **280:**10135–10140.
10. Yamaguchi-Iwai Y, Dancis A, Klausner RD: **AFT1: a mediator of iron regulated transcriptional control in Saccharomyces cerevisiae.** *EMBO J* 1995, **14:**1231–1239.
11. Yamaguchi-Iwai Y, Stearman R, Dancis A, Klausner RD: **Iron-regulated DNA binding by the AFT1 protein controls the iron regulon in yeast.** *EMBO J* 1996, **15:**3377–3384.
12. Ueta R, Fujiwara N, Iwai K, Yamaguchi-Iwai Y: **Iron-induced dissociation of the Aft1p transcriptional regulator from target gene promoters is an initial event in iron-dependent gene suppression.** *Mol Cell Biol* 2012, **32:**4998–5008.
13. Ueta R, Fujiwara N, Iwai K, Yamaguchi-iwai Y: **Mechanism underlying the iron-dependent nuclear export of the iron-responsive transcription factor Aft1p in Saccharomyces cerevisiae.** *Mol Cell Biol* 2007, **18:**2980–2990.
14. Conde e Silva N, Gonçalves IR, Lemaire M, Lesuisse E, Camadro JM, Blaiseau PL: **KlAft, the Kluyveromyces lactis ortholog of Aft1 and Aft2, mediates activation of iron-responsive transcription through the PuCACCC Aft-type sequence.** *Genetics* 2009, **183:**93–106.
15. Haas H, Eisendle M, Turgeon BG: **Siderophores in fungal physiology and virulence.** *Annu Rev Phytopathol* 2008, **46:**149–187.
16. Miele R, Barra D, di Patti MC B: **A GATA-type transcription factor regulates expression of the high-affinity iron uptake system in the methylotrophic yeast Pichia pastoris.** *Arch Biochem Biophys* 2007, **465:**172–179.
17. Lan C-Y, Rodarte G, Murillo LA, Jones T, Davis RW, Dungan J, Newport G, Agabian N: **Regulatory networks affected by iron availability in Candida albicans.** *Mol Microbiol* 2004, **53:**1451–1469.
18. Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T: **MatInspector and beyond: promoter analysis based on transcription factor binding sites.** *Bioinformatics* 2005, **21:**2933–2942.
19. 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, **7:**11.
20. Idiris A, Tohda H, Kumagai H, Takegawa K: **Engineering of protein secretion in yeast: strategies and impact on protein production.** *Appl Microbiol Biotechnol* 2010, **86:**403–417.
21. Vogl T, Thallinger GG, Zellnig G, Drew D, Cregg JM, Glieder A, Freigassner M: **Towards improved membrane protein production in Pichia pastoris: general and specific transcriptional response to membrane protein overexpression.** *N Biotechnol* 2014, S1871-6784(14)00022-3. doi: 10.1016/j.nbt.2014.02.009. [Epub ahead of print].
22. Daum G, Tuller G, Nemec T, Hrastnik C, Balliano G, Cattel L, Milla P, Rocco F, Conzelmann A, Vionnet C, Kelly DE, Kelly S, Schweizer E, Schüller HJ, Hojad U, Greiner E, Finger K: **Systematic analysis of yeast strains with possible defects in lipid metabolism.** *Yeast* 1999, **15:**601–614.
23. Thomas-Chollier M, Defrance M, Medina-Rivera A, Sand O, Herrmann C, Thieffry D, van Helden J: **RSAT 2011: regulatory sequence analysis tools.** *Nucleic Acids Res* 2011, **39**(Web Server issue):W86–W91.
24. Mattanovich D, Callewaert N, Rouzé P, Lin Y-C, Graf A, Redl A, Tiels P, Gasser B, De Schutter K: **Open access to sequence: browsing the Pichia pastoris genome.** *Microb Cell Fact* 2009, **8:**53.
25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A,

Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. the Gene Ontology Consortium.** *Nat Genet* 2000, **25:**25–29.
26. Li L, Chen OS, McVey Ward D, Kaplan J: **CCC1 is a transporter that mediates vacuolar iron storage in yeast.** *J Biol Chem* 2001, **276:**29515–29519.
27. Greene JR, Brown NH, DiDomenico BJ, Kaplan J, Eide DJ: **The GEF1 gene of Saccharomyces cerevisiae encodes an integral membrane protein; mutations in which have effects on respiration and iron-limited growth.** *Mol Gen Genet* 1993, **241:**542–553.
28. Kim D, Yukl ET, Moënne-Loccoz P, Montellano PR: **Fungal heme oxygenases: functional expression and characterization of Hmx1 from Saccharomyces cerevisiae and CaHmx1 from Candida albicans.** *Biochemistry* 2006, **45:**14772–14780.
29. Collinson EJ, Wimmer-Kleikamp S, Gerega SK, Yang YH, Parish CR, Dawes IW, Stocker R: **The yeast homolog of heme oxygenase-1 affords cellular antioxidant protection via the transcriptional regulation of known antioxidant genes.** *J Biol Chem* 2011, **286:**2205–2214.
30. Schilke B, Voisine C, Beinert H, Craig E: **Evidence for a conserved system for iron metabolism in the mitochondria of Saccharomyces cerevisiae.** *Proc Natl Acad Sci U S A* 1999, **96:**10206–10211.
31. Lee MCS, Miller EA: **Molecular mechanisms of COPII vesicle formation.** *Semin Cell Dev Biol* 2007, **18:**424–434.
32. Römisch K: **Surfing the Sec61 channel: bidirectional protein translocation across the ER membrane.** *J Cell Sci* 1999, **112**(Pt 2):4185–4191.
33. McNew JA, Coe JG, Søgaard M, Zemelman BV, Wimmer C, Hong W, Söllner TH: **Gos1p, a Saccharomyces cerevisiae SNARE protein involved in Golgi transport.** *FEBS Lett* 1998, **435:**89–95.
34. Kuge S, Jones N: **YAP1 dependent activation of TRX2 is essential for the response of Saccharomyces cerevisiae to oxidative stress by hydroperoxides.** *EMBO J* 1994, **13:**655–664.
35. Vervecken W, Kaigorodov V, Callewaert N, Geysens S, De Vusser K, Contreras R: **In vivo synthesis of mammalian-like, hybrid-type N-glycans in pichia pastoris.** 2004, **70:**2639–2646.
36. Elbert M, Rossi G, Brennwald P: **The yeast par-1 homologs kin1 and kin2 show genetic and physical interactions with components of the exocytic machinery.** *Mol Biol Cell* 2005, **16:**532–549.
37. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23:**2947–2948.
38. Rutherford JC, Jaron S, Ray E, Brown PO, Winge DR: **A second iron-regulatory system in yeast independent of Aft1p.** *Proc Natl Acad Sci U S A* 2001, **98:**14322–14327.
39. Babu MM, Iyer LM, Balaji S, Aravind L: **The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons.** *Nucleic Acids Res* 2006, **34:**6505–6520.
40. Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD: *Molecular Biology of the Cell, 3rd Addition.* New York: Garland Science; 1994.
41. Heiss S, Maurer M, Hahn R, Mattanovich D, Gasser B: **Identification and deletion of the major secreted protein of Pichia pastoris.** *Appl Microbiol Biotechnol* 2013, **97:**1241–1249.
42. 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, **9:**390.
43. Baumann K, Carnicer M, Dragosits M, Graf AB, Stadlmann J, Jouhten P, Maaheimo H, Gasser B, Albiol J, Mattanovich D, Ferrer P: **A multi-level study of recombinant Pichia pastoris in different oxygen conditions.** *BMC Syst Biol* 2010, **4:**141.
44. Nocon J, Steiger MG, Pfeffer M, Sohn SB, Kim TY, Maurer M, Rußmayer H, Pflügl S, Ask M, Haberhauer-Troyer C, Ortmayr K, Hann S, Koellensperger G, Gasser B, Lee SY, Mattanovich D: **Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production.** *Metab Eng* 2014, **24:**129–138. doi:10.1016/j.ymben.2014.05.011.
45. Crameri A, Whitehorn EA, Tate E, Stemmer WP: **Improved green fluorescent protein by molecular evolution using DNA shuffling.** *Nat Biotechnol* 1996, **14:**315–319.
46. Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, Lin-Cereghino J, Kovar K, Cregg JM, Glieder A: **Promoter library designed for fine-tuned gene expression in Pichia pastoris.** *Nucleic Acids Res* 2008, **36:**e76.
47. Hartinger D, Moll W-D: **Fumonisin elimination and prospects for detoxification by enzymatic transformation.** *World Mycotoxin J* 2011, **4:**271–283.
48. Heinl S, Hartinger D, Thamhesl M, Vekiru E, Krska R, Schatzmayr G, Moll W-D, Grabherr R: **Degradation of fumonisin B1 by the consecutive action of two bacterial enzymes.** *J Biotechnol* 2010, **145:**120–129.
49. Baumann K, Maurer M, Dragosits M, Cos O, Ferrer P, Mattanovich D: **Hypoxic fed-batch cultivation ofPichia pastoris increases specific and volumetric productivity of recombinant proteins.** *Biotechnol Bioeng* 2008, **100:**177–183.
50. Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, Jaenicke S, Brinkrolf K, Goesmann A, Szczepanowski R, Pühler A, Schwab H, Glieder A, Pichler H: **High-quality genome sequence of Pichia pastoris CBS7435.** *J Biotechnol* 2011, **154:**312–320.
51. Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A: **Deletion of the Pichia pastoris KU70 homologue facilitates platform strain generation for gene expression and synthetic biology.** *PLoS One* 2012, **7:**e39720.
52. Weis R, Luiten R, Skranc W, Schwab H, Wubbolts M, Glieder A: **Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena.** *FEMS Yeast Res* 2004, **5:**179–189.
53. Zhao W, Wang J, Deng R, Wang X: **Scale-up fermentation of recombinant Candida rugosa lipase expressed in Pichia pastoris using the GAP promoter.** *J Ind Microbiol Biotechnol* 2008, **35:**189–195.
54. Abad S, Kitz K, Hörmann A, Schreiner U, Hartner FS, Glieder A: **Real-time PCR-based determination of gene copy numbers in Pichia pastoris.** *Biotechnol J* 2010, **5:**413–420.
55. Hoffman CS, Winston F: **A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli.** *Gene* 1987, **57:**267–272.

doi:10.1186/s12934-014-0120-5
**Cite this article as:** Ruth *et al.*: *Pichia pastoris* Aft1 - a novel transcription factor, enhancing recombinant protein secretion. *Microbial Cell Factories* 2014 **13**:120.

## 4.3 Manuscript 3: Systems-level organization of yeast methylotrophic lifestyle

### The cellular reaction to methanol induction of the methylotrophic yeast *Pichia pastoris*

Although other strong promoters have been investigated recently for recombinant protein production in *P. pastoris* (Prielhofer et al. 2013), the two most commonly used expression strategies either rely on the strong constitutive $P_{GAP}$ or the strong inducible $P_{AOX1}$. The first promoter regulates transcription of a gene encoding a key glycolytic enzyme found in all eukaryotes. The latter one regulates expression of the *AOX1* gene in *P. pastoris*. Its gene product, alcohol oxidase, converts methanol into formaldehyde in the very first step of the methanol utilization pathway. The strong promoter with low affinity to its substrate methanol induces heterologous gene expression when switching the carbon source to methanol.

Transcriptional studies showed that $P_{GAP}$ can compete with $P_{AOX1}$ in terms of promoter strength. $P_{GAP}$ is constitutively active and recombinant gene expression therefore active during the whole process time. This feature allows for constant protein expression, shorter process times and simpler cultivation procedures but is less ideal when the protein product has toxic properties, which could lead to cell lyse or reduced growth. Apart from technical disadvantages in large scale production processes like heat formation or security risks linked to high flammability or toxicity, methanol-based protein production often suffers from host cell protein contaminations in the supernatant of the fermentation broth due to increased cell lysis.

Promoter strength cannot be regarded as the sole cause for beneficial effects in heterologous protein secretion. The carbon source undoubtedly has an impact on cellular processes which again might affect protein secretion. Thus, a systems-level comparison of *P. pastoris* cultures grown on either glucose or methanol/glycerol can deliver insights in if or how differences in productivities can be traced back to the use of carbon source. In order to eliminate as much unwanted experimental influences as possible – especially growth rate related – steady-state chemostat cultivations were performed in triplicates with each carbon source, followed by transcriptomics, proteomics, metabolomics and fluxomics analysis.

Thereby, regulation patterns of 5354 genes, 575 proteins, 141 metabolites, and fluxes through 39 reactions of P. pastoris were compared between growth on glucose and a methanol/glycerol mixed medium. In direct comparison, 406 genes were found to be significantly differentially expressed. Quantifications of the gene products was mainly achieved for proteins with already higher transcript levels. A positive co-regulation of proteins and their transcripts was expected according to the literature (Lee et al. 2011; Vogel and Marcotte 2012). Similar to the published data for *S. cerevisiae*, this experiment showed a 61% regulatory match of transcriptome and proteome abundance ($r = 0.78$, $r^2 = 0.61$).

When fed with methanol, the most obvious reaction of *P. pastoris* is a strong up-regulation of peroxisome proliferation. A novel discovery of this work was that in *P. pastoris* – and likely other methylotrophic yeasts as well – the entire methanol assimilation is localized to peroxisomes. The assimilation of three molecules of methanol into one molecule of glyeraldehyde-3-phosphate (GAP) net product requires a constant regeneration of xylulose-5-phosphate (Xyl5P). After methanol has been converted to the (toxic) intermediate formaldehyde, the mentioned pentose phosphate is needed for further conversion into GAP and dihydroxyacetone (DHA). Instead of using the cytosolic pentose phosphate pathway, *P. pastoris* regenerates Xyl5P entirely with a duplicated methanol inducible enzyme set localized in its peroxisomes. Similar to the Calvin cycle (for $CO_2$ assimilation in chloroplasts of plants), sedoheptulose-1,7-bisphosphate is used as intermediate in this so-called xylose-monophosphate cycle.

With methanol/glycerol as carbon source, central carbon metabolism is entered at the lower stage of glycolysis by glyceraldehyde-3-phosphate, a C-3 molecule. While key regulatory enzymes of this part of glycolysis are equally abundant as with glucose as carbon source, the upper part, which shares several enzymes with gluconeogenesis is up-regulated. This comes not surprising as on methanol, there is a higher demand for hexose and pentose synthesis than when directly fed with glucose.

Looking at the fluxes, we get a more differentiated picture: Fluxes through the lower part of glycolysis as well as the TCA cycle (mainly for glutamate production) are reduced when methanol is the carbon source and its dissimilation is a major source for energy and NADH (needed for ATP production). During growth on methanol, the

peroxisomes also harbor an active glyoxylate cycle, corresponding to lower carbon fluxes through the TCA cycle. The mitochondrial import of NADH is mediated by the also up-regulated malate-aspartate shuttle, i.e. the glutamate-aspartate transporter *AGC1* and the malate-α-ketoglutarate transporter *ODC1*.

Unlike its peroxisomal isoforms (for methanol assimilation), cytosolic pentose phosphate genes were not differentially regulated on methanol/glycerol. However, higher fluxes through the pentose phosphate pathway (PPP) were observed, fitting to a high synthesis rate of NADPH and to a higher demand of PPP intermediates which serve as precursors of e.g. riboflavin biosynthesis. Higher levels of riboflavin are needed for flavin adenine dinucleotide (FAD), the prosthetic group of alcohol oxidase, the highly abundant enzyme of methanol assimilation. High levels of methanol utilization enzymes require overproduction of other vitamins and cofactors as well: Thiamine synthesis was shown to be strongly up-regulated due to demand of peroxisomal transketolases. High nicotinamide levels are a result of increased formaldehyde detoxification. Heme synthesis is up-regulated for its incorporation in catalase needed for peroxide detoxification.

A highly interesting observation on *P. pastoris* grown on methanol/glycerol is that there is a higher protein content due to the strong induction of the C-1 metabolism enzymes and their cofactors (both seen on the transcript and protein level). This corresponds to increased ribosome biogenesis and translation and amino acid biosynthesis. Again, the higher NADPH synthesis rate fits to these findings because NADPH is the main electron donor in amino acid biosynthesis. There are less free amino acids as they are constantly needed to meet the increased demand for protein production. Pathways linked to protein folding, secretion and degradation were not affected by methanol as substrate. The higher translational capacity of methanol-grown cells however could indicate higher recombinant protein production capacities. So the change in carbon source leads to significant changes in the amino acid metabolism. The more efficient supply of intracellular amino acids may be responsible for higher productivities.

Rußmayer et al. BMC Biology (2015) 13:80
DOI 10.1186/s12915-015-0186-5

BMC Biology

RESEARCH ARTICLE Open Access

# Systems-level organization of yeast methylotrophic lifestyle

Hannes Rußmayer[1,2†], Markus Buchetics[1,2†], Clemens Gruber[2,3], Minoska Valli[1,2], Karlheinz Grillitsch[4,5], Gerda Modarres[2,6], Raffaele Guerrasio[2,3,10], Kristaps Klavins[2,3,11], Stefan Neubauer[2,3,12], Hedda Drexler[2,3], Matthias Steiger[1,2], Christina Troyer[2,3], Ali Al Chalabi[7], Guido Krebiehl[7], Denise Sonntag[7], Günther Zellnig[8], Günther Daum[4,5], Alexandra B. Graf[2,5], Friedrich Altmann[2,3], Gunda Koellensperger[9], Stephan Hann[2,3], Michael Sauer[1,2], Diethard Mattanovich[1,2*] and Brigitte Gasser[1,2]

## Abstract

**Background:** Some yeasts have evolved a methylotrophic lifestyle enabling them to utilize the single carbon compound methanol as a carbon and energy source. Among them, *Pichia pastoris* (syn. *Komagataella sp.*) is frequently used for the production of heterologous proteins and also serves as a model organism for organelle research. Our current knowledge of methylotrophic lifestyle mainly derives from sophisticated biochemical studies which identified many key methanol utilization enzymes such as alcohol oxidase and dihydroxyacetone synthase and their localization to the peroxisomes. C1 assimilation is supposed to involve the pentose phosphate pathway, but details of these reactions are not known to date.

**Results:** In this work we analyzed the regulation patterns of 5,354 genes, 575 proteins, 141 metabolites, and fluxes through 39 reactions of *P. pastoris* comparing growth on glucose and on a methanol/glycerol mixed medium, respectively. Contrary to previous assumptions, we found that the entire methanol assimilation pathway is localized to peroxisomes rather than employing part of the cytosolic pentose phosphate pathway for xylulose-5-phosphate regeneration. For this purpose, *P. pastoris* (and presumably also other methylotrophic yeasts) have evolved a duplicated methanol inducible enzyme set targeted to peroxisomes. This compartmentalized cyclic C1 assimilation process termed xylose-monophosphate cycle resembles the principle of the Calvin cycle and uses sedoheptulose-1,7-bisphosphate as intermediate. The strong induction of alcohol oxidase, dihydroxyacetone synthase, formaldehyde and formate dehydrogenase, and catalase leads to high demand of their cofactors riboflavin, thiamine, nicotinamide, and heme, respectively, which is reflected in strong up-regulation of the respective synthesis pathways on methanol. Methanol-grown cells have a higher protein but lower free amino acid content, which can be attributed to the high drain towards methanol metabolic enzymes and their cofactors. In context with up-regulation of many amino acid biosynthesis genes or proteins, this visualizes an increased flux towards amino acid and protein synthesis which is reflected also in increased levels of transcripts and/or proteins related to ribosome biogenesis and translation.

**Conclusions:** Taken together, our work illustrates how concerted interpretation of multiple levels of systems biology data can contribute to elucidation of yet unknown cellular pathways and revolutionize our understanding of cellular biology.

**Keywords:** Metabolome, Methanol, Peroxisome, *Pichia pastoris*, Proteome, Transcriptome, Xylulose-monophosphate cycle

* Correspondence: diethard.mattanovich@boku.ac.at
†Equal contributors
[1]Department of Biotechnology, BOKU - University of Natural Resources and Life Sciences Vienna, Muthgasse 18, 1190 Vienna, Austria
[2]Austrian Centre of Industrial Biotechnology, A-1190 Vienna, Austria
Full list of author information is available at the end of the article

## Background

Methylotrophic yeasts accept a broad range of carbon sources. Multicarbon sources, such as sugars and sugar alcohols like glucose, glycerol, or mannitol, are utilized at similar efficiency as reduced C1-compounds like methanol [1]. Besides the proper equipment of the cells with enzymes necessary for substrate metabolism, their coordinated expression is a prerequisite for the successful utilization of different carbon and energy sources. The methylotrophic yeast *Pichia pastoris* (syn. *Komagataella sp.*) is widely used for recombinant protein production with several biopharmaceuticals on the market [2] and an expanding portfolio of industrial enzymes produced [3]. Recently, the application of *P. pastoris* as a model system for peroxisome and secretory organelle proliferation has also expanded [4, 5]. The methylotrophic lifestyle has been the main driving force for this development, as it involves strong and regulated promoters used for expression of recombinant genes [6], as well as specialized organelles, the peroxisomes. Peroxisomes are defined as intracellular compartments accommodating hydrogen peroxide ($H_2O_2$) forming oxidases together with the $H_2O_2$ detoxifying enzyme catalase. Also the fatty acid beta-oxidation pathway of *P. pastoris* is located in these organelles [7]. Yeast peroxisomal oxidases are predominantly involved in the metabolism of various unusual carbon and nitrogen sources (e.g. alcohols, fatty acids, D-amino acids, or primary amines) [8]. In methylotrophic yeasts, peroxisomes, which harbor the initial steps of the methanol utilization pathway, are highly abundant in methanol-grown cells but become heavily decreased in both number and volume upon catabolite repression [9]. When grown on glucose, *Hansenula polymorpha*, another methylotrophic yeast, harbors only a single, small peroxisome which can serve as a source for proliferation by fission when induction is triggered by shifting the cells to methanol [10, 11]. In addition to genes encoding structural peroxisomal proteins, the expression of methanol utilization related genes is strongly induced on methanol. The first steps of methanol assimilation involve an alcohol oxidase (AOX) to convert methanol to formaldehyde, and a special transketolase named dihydroxyacetone synthase (DAS) to form a C-C bond with the C1 molecule formaldehyde. The reactions of these two enzymes and their localization to peroxisomes are well described [12, 13]. The further reaction cycle of methanol assimilation is supposed to involve pentose phosphate reactions, but the details are not fully clarified to date.

While there are several studies analyzing cellular reactions of *P. pastoris* to methanol induction in context of recombinant protein production [14–18], the response of non-recombinant strains to the different carbon sources is largely unknown. Thus, we decided to investigate the cellular responses of *P. pastoris* cells not producing a recombinant protein to methanol and glucose, respectively, which are the two most widely used substrates for cultivation. To enable the same chemostat-controlled constant specific growth rates for direct comparability the methanol cultures were co-fed with glycerol. Availability of whole genome sequences made a number of transcriptome regulation studies of *P. pastoris*, analyzing the implications of growth rate [19], unfolded protein response (UPR) induction [20], oxygen availability [21], osmotic stress [22], or heterologous protein production [16, 23], become feasible. Analyses of the host proteome gave further insights into characteristics of *P. pastoris* grown at different temperatures [24], osmolarity [22], UPR induction [25], and oxygen supply [21]. More recently, *P. pastoris* strains producing an insulin precursor were analyzed for changes in the cellular proteome as adaptation response to methanol induction during fed batch cultivation using 2D-DIGE and subsequent mass spectrometry identification of differentially abundant proteins. High abundance of enzymes from the dissimilatory methanol metabolism and induction of the UPR were observed [14]. Regulation of cellular enzyme concentrations will cause changes in metabolic fluxes, eventually also leading to changes in free metabolite concentrations. Quantitative determination of intracellular fluxes is the key to a better understanding of metabolic networks. First genome-scale metabolic network models of *P. pastoris* [26, 27] and flux distributions of central carbon metabolism [28–30] indicate growth rate-related methanol (co-)assimilation with proposed implications for the pentose phosphate pathway [31].

The work at hand incorporates transcriptomics, proteomics, metabolomics, and fluxomics analyses of non-producing *P. pastoris* in steady-state cultures at a uniform specific growth rate comprising the carbon source as the investigated variable. This integrated systems level analysis allowed to reveal cellular processes that are co-regulated with methanol metabolism, such as vitamin biosynthesis and amino acid metabolism. Furthermore, these co-regulation patterns were the pre-requisite to elucidate the thus far unidentified steps of sugar phosphate rearrangements recycling xylulose-5-phosphate for methanol fixation. We propose, herein, a new model for the assimilation of methanol as a separate strictly regulated pathway, originating from duplication of the involved genes.

## Results and discussion

### Growth parameters of *P. pastoris* differ significantly on different substrates

*P. pastoris* CBS7435 was cultivated in chemostat cultivations at a fixed specific growth rate of 0.1 $h^{-1}$, corresponding to approximately 60 % of $\mu_{max}$ on glucose [19].

Constant growth is a prerequisite to avoid growth rate-dependent effects during genome-scale analyses. As the maximum specific growth rate on pure methanol as a carbon source would be significantly lower, and intracellular carbon fluxes could not be analyzed on methanol alone, a mixed feed strategy applying glycerol-methanol co-feeding was employed. A methanol-glycerol mix of 8.5 g/L methanol and 49.0 g/L glycerol was employed based on experiments with *P. pastoris* proving that total methanol utilization and full induction of the methylotrophic pathway were realized under these conditions. Chemostats were run in three biological replicates per condition and samples for transcriptomics, proteomics, and metabolite analyses were taken in steady state after seven residence times as described in the Methods section. For metabolic flux analysis, separate chemostat cultivations employing $^{13}C$-labelled substrates were performed. Substrate limitation of all cultures, i.e. no residual glucose or methanol/glycerol, respectively, was confirmed by HPLC. The growth parameters derived from these cultures are summarized in Table 1. The $CO_2$ exchange rate of cells grown on methanol/glycerol was 13 % lower compared to those grown on glucose while their oxygen uptake rate was 30 % higher. The higher oxygen uptake rate of methanol/glycerol-grown *P. pastoris* can be explained by the higher degree of reduction of methanol and glycerol compared to glucose. As methanol oxidation to formaldehyde by AOX is an exothermic oxygen consuming reaction, an equimolar amount of oxygen is needed only to pass methanol into cellular metabolism. The biomass yield was slightly higher for cells grown on methanol/glycerol compared to glucose, which is in good agreement with data from the literature [28, 30]. Transcriptional regulation was analyzed using *P. pastoris*-specific DNA microarrays [20, 32], liquid chromatography-tandem mass spectrometry (LC/MS-MS) was used for differential proteomics and quantification of metabolites. Additionally, distribution of specific lipid classes were analyzed. Flux ratios were calculated from $^{13}C$ labelling patterns in proteinogenic amino acids. The numerical results of these genome scale analyses can be found in Additional file 1.

### Transcriptome and proteome are significantly co-regulated

At the transcriptional level, 406 of 5,354 genes were significantly differentially expressed on methanol/glycerol and glucose. As protein abundance, however, does not necessarily directly correlate with transcription [33], we also measured differential proteome regulation using 2D-LC-MS of Tandem Mass Tag labelled total protein samples and obtained quantitative data for 575 cellular proteins. In agreement with the literature [33–35], where a positive correlation between protein concentration and the abundance of the transcript has been described, we could mainly quantify proteins with higher transcript levels. Figure 1 shows 575 data pairs with mean $\log_2$ fold changes in transcript and protein levels of the methanol/glycerol experiments compared to the glucose experiments. Proteins and their transcripts were significantly co-regulated ($r = 0.78$, $r^2 = 0.61$). Lu et al. [34] have shown that, in *Saccharomyces cerevisiae*, protein levels are determined to 73 % by transcription. Similarly, we observed that transcriptional control determined the regulation of protein abundance by 61 %.

Of the 575 genes with available transcriptomics and proteomics data, 130 (23 %) were differentially regulated at the transcript and/or protein level, the largest group being up-regulated at both levels. Based on the differential changes at both protein and transcript levels, data have been allocated to seven groups and analyzed for overrepresentation of functional groups (Fig. 1 and Additional file 1). As expected, during growth on methanol/glycerol there are strongly increased levels of transcripts and proteins involved in methanol metabolism and peroxisome formation, while two proteins needed mainly on glucose, hexokinase and high affinity glucose transporter, had lower abundance on both transcript and protein level. Higher levels of proteins of the translation machinery and cytoskeleton organization in cells grown on methanol/glycerol were not met by higher transcript levels, thus indicating a post-transcriptional regulation while the significant down-regulation of transcripts for lower glycolysis and fatty acid beta-oxidation was not reflected in protein levels (Fig. 1). These processes will be described in more detail below.

To confirm that the gene regulations attributed to methanol cultivation in this work are truly due to methanol, and not to glycerol as a co-substrate, we compared these regulation patterns to transcript regulation obtained in fed batch cultivations using methanol, glycerol, or glucose (Additional file 2). Thereby, we could confirm that all genes discussed to be regulated

**Table 1** Growth parameters of *P. pastoris* grown on methanol/glycerol and glucose in chemostats at $\mu = 0.1\ h^{-1}$

| | Glucose | Glycerol | Methanol | CER | OUR | Biomass | $Y_{X/S}$ |
|---|---|---|---|---|---|---|---|
| | [mmol/(gCDW*h)] | [mmol/(gCDW*h)] | [mmol/(gCDW*h)] | [mmol/(gCDW*h)] | [mmol/(gCDW*h)] | [g/L] | [gCDW/gSubstrate] |
| Glucose | 1.02 ± 0.03 | – | – | 2.11 ± 0.07 | 2.39 ± 0.07 | 28.1 ± 0.3 | 0.54 ± 0.01 |
| Glycerol/Methanol | – | 1.64 ± 0.06 | 0.81 ± 0.04 | 1.86 ± 0.05 | 3.09 ± 0.08 | 31.6 ± 0.3 | 0.57 ± 0.02 |

*CDW* Cell dry weight, *CER* $CO_2$ exchange rate, *OUR* Oxygen uptake rate, $Y_{XS}$ Biomass yield

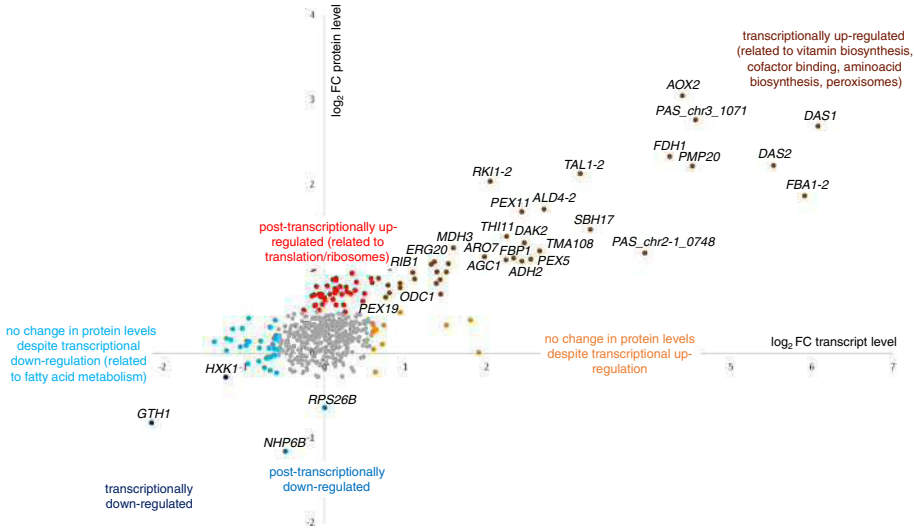


**Fig. 1** Co-regulation of transcripts and proteins of *P. pastoris* grown on methanol/glycerol compared to glucose in chemostats at $\mu = 0.1\ h^{-1}$. $Log_2$ fold changes of transcripts versus $log_2$ fold changes of proteins for the 575 data pairs are shown which have been grouped based on their regulation patterns. Significantly enriched GO terms in these groups are indicated. Criteria for up/down-regulation of transcript and protein levels are described in the Methods section

by methanol utilization in the present study are actually induced by methanol.

### Peroxisome proliferation is strongly up-regulated on methanol

Methanol-induced cells show an up-regulation in genes encoding proteins essential for peroxisome biogenesis and proliferation. Fourteen *PEX*-genes were markedly up-regulated on the transcriptome level, three of them also on the proteome level (Additional file 1). Genes of the peroxisomal import machinery encoding the docking complex (*PEX13*, *PEX14*, *PEX17*) and RING-finger complex (*PEX2*, *PEX10*, *PEX12*) are up-regulated at almost equal levels. Among peroxins required for the import of peroxisomal matrix proteins [36], *PEX11* and its isoform *PEX11C* were amongst the highest up-regulated genes (3- to 6-fold higher levels), while members of the Pex23-family were not induced. This distinction seems to be specific for *P. pastoris*, as both gene groups were up-regulated upon methanol induction in *H. polymorpha* [10]. A similar regulation of *P. pastoris PEX* genes has also been described recently by Prielhofer et al. [37], who observed that genes encoding the Pex7/Pex20-mediated import machinery and the Pex23-family were only up-regulated in conditions inducing expression of beta-oxidation genes but not upon methanol induction. Most key players of the peroxisomal methanol utilization pathway, such as AOX, catalase, and DAS, rely on the Pex5-mediated PTS1 import pathway [9]; thus, up-regulation of receptors, which recognize a peroxisomal targeting signal sequence (PTS), was restricted to PTS1-specific *PEX5* [38]. Conversely, *PEX7*, encoding a signal receptor for the signal sequence PTS2, was slightly down regulated. This finding is in agreement with previous studies [39, 40] demonstrating that *H. polymorpha* and *P. pastoris* do not require the PTS2 import pathway for growth on methanol. Unchanged expression levels of Pex7 and Pex20 encoding its accessory protein also correlated with unaltered protein levels of beta-oxidation enzymes. Genes encoding auxiliary functions in matrix protein import and quality control were induced: the putative peroxisomal Lon-protease Pim1-2 and the peroxisomal ATP importer Pmp47 (PAS_chr3_0099) required for import of DAS were strongly up-regulated ($log_2$FC +3.78). Further, up-regulation at both transcriptome and protein level was detected for the glutathione peroxidase Pmp20, a peroxisomal protein which might also be involved in the detoxification of $H_2O_2$ in the peroxisome of methanol growing cells.

### The xylulose-monophosphate cycle of methanol assimilation utilizes a duplicated methanol inducible enzyme set and is entirely localized to peroxisomes

The key players of the methanol utilization pathway have been identified during the last 30 years [12, 13]; however, major steps of the assimilation pathway still remain to be resolved. Briefly, methanol is oxidized to formaldehyde by AOX (Aox1 and Aox2 in *P. pastoris*) within the peroxisomes, thereby generating stoichiometric amounts of $H_2O_2$. Formaldehyde is further converted in two possible routes, either dissimilatory by glutathione-dependent formaldehyde dehydrogenase, S-formyl glutathione hydrolase, and formate dehydrogenase yielding NADH and $CO_2$, or assimilatory by

the action of DAS (Das1 and Das2 in *P. pastoris*). DAS catalyzes the fusion of formaldehyde to xylulose-5-phosphate (XYL5P), thereby generating dihydroxyacetone and glyceraldehyde-3-phosphate (GAP). These intermediates are further converted by dihydroxyacetone kinase (DAK), fructose-1,6-bisphosphate aldolase, and fructose-1,6-bisphosphatase, to finally yield one molecule of GAP per three molecules of methanol, which is then used for the generation of biomass and energy. It is generally assumed that XYL5P gets recycled through rearrangements in the pentose phosphate pathway (PPP), although the detailed mechanism of these rearrangements as well as the interplay of PPP and peroxisomes is still unknown.

All known enzymes of the methanol utilization pathway have significantly higher transcript and protein levels when methanol is present (Fig. 1, upper right quadrant). Interestingly, our analysis revealed that *P. pastoris* does not only have a second isoform of fructose-1,6-bisphosphate aldolase (designated as Fba1-2) as reported by Küberl et al. [41], but also isoforms of the PPP enzymes transaldolase (Tal1-2), ribose-5-phosphate ketol-isomerase (Rki1-2), and ribulose-5-phosphate 3-epimerase (Rpe1-2). All these isoforms were found among the group of up-regulated gene-protein pairs, except for Rpe1-2 which was not identified at the proteomic level. Sequence analysis predicted that Fba1-2, Tal1-2, Rki1-2, and Rpe1-2 each contain a PTS1 peroxisomal targeting signal [42, 43], indicating their potential involvement in a separate peroxisomal methanol assimilation pathway. On the contrary, their respective cytosolic or mitochondrial isoforms (Fba1-1, Tal1-1, Rki1-1, and Rpe1-1) were not differentially regulated and do not contain a peroxisomal targeting sequence. The same regulation pattern was observed comparing cultures grown on methanol alone to those grown on glycerol or glucose (Additional file 2).

Subsequently, cellular fractions enriched of highly pure peroxisomes were isolated from methanol- or glucose-grown *P. pastoris* according to the protocol established by Wriessnegger et al. [44], and subjected to proteomics analyses. In this way, we demonstrated that all relevant enzymes for methanol assimilation were present only in methanol-derived peroxisomal fractions, but not in glucose-derived peroxisome fractions (Table 2). The relative enrichment of proteins in the peroxisomal fractions compared to total cell homogenates was quantified as average weighted ratios of the peak areas of respective peptides. This was consequently only possible with methanol-derived samples where peptides of the proteins of interest had been identified. Table 3 shows MASCOT scores as indicators of identification, and the average ratios of protein abundance in peroxisomal vs. homogenate samples, normalized to Aox1. The methanol assimilation pathway enzymes discussed above were enriched at the same level or higher than Aox1 in the peroxisomal fractions, just as several selected peroxisomal proteins, while cytosolic proteins of glycolysis, PPP, and methanol dissimilation were either not identified at all or markedly depleted in peroxisomal preparations. Only DAK of the methanol assimilation pathway was rather depleted compared to Aox1. Luers et al. [45] described that DAK localizes to the cytosol despite having a PTS1 signal. We could, however, quantify DAK also in peroxisomal fractions of methanol-grown cells, indicating that this enzyme can localize in more than one compartment.

Additionally, one of the unidentified ORFs present among the up-regulated gene/protein pairs was identified to be the homolog of *S. cerevisiae*, YKR043C, which was recently reported to encode sedoheptulose-1,7-bisphosphatase (Shb17) [46]. Further, this protein was found to be enriched in the peroxisomes in methanol-grown *P. pastoris* (Table 2). Shb17 was shown to hydrolyze sedoheptulose-1,7-bisphosphate (S1,7BP) to sedoheptulose-7-phosphate in a thermodynamically driven pathway for the synthesis of pentose-5-phosphates alternative to PPP [46]. S1,7BP was not among the quantified metabolites in our initial metabolomics analyses due to the lack of a commercially available standard. After receiving purified S1,7BP from Amy Caudy (University of Toronto, CA), a previously unidentified substance with differential abundance could be unambiguously assigned as S1,7BP. The signal-to-noise ratios of all samples with glucose-treatment were below 5. In the methanol-grown samples, the signal-to-noise ratios were 84, 111, and 116, clearly indicating the presence of S1,7BP in methanol-grown *P. pastoris* cells, contrary to glucose-grown cells (Fig. 2). This result prompted us to reconsider the pentose phosphate rearrangements leading to the formation of XY L5P for methanol assimilation.

### Methanol assimilation employs an alternative xylulose-5-phosphate forming pathway via sedoheptulose-1,7-bisphosphate

While it would be stoichiometrically possible that XYL5P is regenerated through the canonical non-oxidative branch of the pentose phosphate pathway, our genomic, transcriptomic, and proteomic data point to another direction of C1 assimilation. It appears most likely that *P. pastoris* and other methylotrophic yeasts evolved a specialized set of enzymes for sugar phosphate rearrangements which is specifically induced by growth on methanol and localizes to peroxisomes. Figure 3 shows our proposed pathway for the rearrangement reactions, with 1 GAP molecule per 3 molecules of methanol as the net result. Thereby, F6P (generated from GAP and dihydroxyacetone phosphate (DHAP) by the action of Fba1 and fructose-1,6-bisphosphatase) and another GAP are interconverted to erythrose-4-phopsphate and XYL5P in a transketolase

**Table 2** Transcriptional and post-transcriptional regulation of genes related to the methanol metabolism, the pentose phosphate pathway, and the glyoxylate cycle. Presence of the corresponding protein in the peroxisomal fraction (of methanol- or glucose-grown *P. pastoris*) is indicated as well as prediction of peroxisomal targeting based on the C-terminal amino residues of the proteins using the PTS1 predictor [43]

| Pathway | Short name [a] | ORF name [b] | Description | Transcript (methanol/ glycerol vs glucose) [c] | Protein (methanol/ glycerol vs glucose) [d] | Presence in the peroxisome fraction (methanol) [e] | Presence in the peroxisome fraction (glucose) [e] | Prediction of peroxisomal targeting [f] | Last 12 C-terminal amino acid residues |
|---|---|---|---|---|---|---|---|---|---|
| Methanol assimilation | AOX1 | PP7435_Chr4-0130/ PAS_chr4_0821 | Alcohol oxidase | up | n.i. | yes | no | yes | LGTYEKTGLARF |
| | AOX2 | PP7435_Chr4-0863/ PAS_chr4_0152 | Alcohol oxidase | up | up | n.i. | n.i. | yes | LGTYEKTGLARF |
| | DAS1 | PP7435_Chr3-0352/ PAS_chr3_0832 | Dihydroxyacetone synthase variant 1 | up | up | yes | no | no | HDLKGKPKHDKL |
| | DAS2 | PP7435_Chr3-0350/ PAS_chr3_0834 | Dihydroxyacetone synthase variant 2 | up | up | yes | no | no | TDLKGKPKHDKL |
| | DAK2 | PP7435_Chr3-0343/ PAS_chr3_0841 | Dihydroxyacetone kinase | up | up | yes | no | Twilight zone | ITDAYFKSETKL |
| | FBA1-2 | PP7435_Chr1-0639/ PAS_chr1-1_0319 | Fructose-1,6-bisphosphate aldolase | up | up | yes | no | yes | HAAGTFKSESKL |
| | FBP1 | PP7435_Chr3-0309/ PAS_chr3_0868 | Fructose-1,6-bisphosphatase | up | up | yes | no | no | LTKKIKIQSVNL |
| | SHB17 | PP7435_Chr2-0185/ PAS_chr2-2_0177 | Sedoheptulose-1,7-bisphosphatase | up | up | yes | no | no | VVPVEEAEADRA |
| | RKI1-2 | PAS_chr4_0212 | Ribose-5-phosphate ketol-isomerase | up | up | yes | no | yes | ITSLSVSVPARL |
| | TAL1-2 | PAS_chr2-2_0338 | Transaldolase | up | up | yes | no | yes | VPSLFRRVLSKL |
| | RPE1-2 | PP7435_Chr3-0772 | D-ribulose-5-phosphate 3-epimerase | up | n.i. | n.i. | n.i. | Twilight zone | QKKAKAKPKPNL |
| Peroxisomal protein | CTA1 | PP7435_Chr2-0137/ PAS_chr2-2_0131 | Catalase A | up | n.q. | yes | no | yes | QLSPRGDSAARL |
| | PMP20 | PP7435_Chr1-1351/ PAS_chr1-4_0547 | Peroxiredoxin | up | up | yes | no | yes | KHSSADRVLAKL |
| Methanol dissimilation | FLD | PP7435_Chr3-0140/ PAS_chr3_1028 | Bifunctional alcohol dehydrogenase and formaldehyde dehydrogenase | up | n.q. | no | no | no | AGNCIRAVITMH |
| | FGH1 | PP7435_Chr3-0312/ PAS_chr3_0867 | Esterase that can function as an S-formylglutathione hydrolase | up | n.q. | no | no | no | HAAHHAKYLGLN |
| | FDH1 | PP7435_Chr3-0238/ PAS_chr3_0932 | NAD(+)-dependent formate dehydrogenase | up | up | yes | no | no | KTKAYGNDKKVA |
| Pentose phosphate pathway oxidative branch | ZWF1 | PP7435_Chr2-0993/ PAS_chr2-1_0308 | Glucose-6-phosphate dehydrogenase | not changed | not changed | no | no | no | WPVTRPDVLHKM |
| | SOL3 | PP7435_Chr3-0037/ PAS_chr3_1126 | 6-phosphogluconolactonase | not changed | n.q. | no | no | Twilight zone | ALSGVSVSTSKY |

**Table 2** Transcriptional and post-transcriptional regulation of genes related to the methanol metabolism, the pentose phosphate pathway, and the glyoxylate cycle. Presence of the corresponding protein in the peroxisomal fraction (of methanol- or glucose-grown *P. pastoris*) is indicated as well as prediction of peroxisomal targeting based on the C-terminal amino residues of the proteins using the PTS1 predictor [43] *(Continued)*

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | GND2 | PP7435_Chr3-0944/ PAS_chr3_0277 | 6-phosphogluconate dehydrogenase | not changed | not changed | yes | yes | no | KGGNVSASTYDA |
| Pentose phosphate pathway non-oxidative branch | RPE1-1 | PP7435_Chr3-0771 | D-ribulose-5-phosphate 3-epimerase | n.a. | up | n.i. | n.i. | no | QDSLKKKGLLDE |
| | RKI1-1 | PAS_chr4_0213 | Ribose-5-phosphate ketol-isomerase | up | n.q. | n.i. | n.i. | no | GNEDGSVATLTL |
| | TKL1 | PP7435_Chr1-0919/ PAS_chr1-4_0150 | Transketolase | not changed | not changed | no | no | no | SPLNKAFESVHA |
| | TAL1-1 | PP7435_Chr2-0357/ PAS_chr2-2_0337 | Transaldolase | not changed | not changed | yes | yes | no | TLLNLLKEKVQA |
| Glyoxylate cycle | CIT1 | PP7435_Chr1-0426/ PAS_chr1-1_0475 | Citrate synthase | not changed | n.q. | yes | no | no | EKYIELVKGLGK |
| | ACO1 | PP7435_Chr1-0105/ PAS_chr1-3_0104 | Aconitase | not changed | not changed | yes | no | no | ALNNMAAVKASK |
| | ACO2 | PP7435_Chr3-0541/ PAS_chr3_0659 | Aconitase | not changed | n.q. | no | no | no | INYIGRLKREQQ |
| | ICL1 | PP7435_Chr1-1123/ PAS_chr1-4_0338 | Isocitrate lyase | not changed | n.q. | no | no | no | GAGVTEDQFKDH |
| | MLS1 | PP7435_Chr4-0820/ PAS_chr4_0191 | Malate synthase | not changed | up | n.i. | n.i. | no | LESSPVDLDSLK |
| | MDH3 | PP7435_Chr4-0136/ PAS_chr4_0815 | Peroxisomal malate dehydrogenase | up | up | yes | no | no | NIAKGTAFIAGN |
| | MLS2 | PP7435_Chr1-1255/ PAS_chr1-4_0459 | Malate synthase | up | n.i. | n.i. | n.i. | Twilight zone | STIPINIHQQKL |
| | AAT1 | PP7435_Chr1-0511/ PAS_chr1-1_0200 | Aspartate aminotransferase | up | n.q. | yes | no | no | YLANAIHEVTTN |
| | AAT2 | PAS_chr4_0974 | Aspartate aminotransferase | not changed | n.q. | yes | no | no | RVAAAIDQVVRV |
| | ODC1 | PP7435_Chr3-1205/ PAS_chr3_0040 | Oxoglutarate-malate shuttle | up | up | yes | no | no | FTTCMDFFRTLQ |
| | OSM1 | PP7435_Chr3-1001/ PAS_chr3_0225 | Fumarate reductase | up | n.i. | n.i. | n.i. | Twilight zone | YLLKSLSNYHKL |

[a]In some cases, *P. pastoris* has two homologs of the same *S. cerevisiae* gene (i.e. TAL1-1 and TAL1-2)
[b]ORF names of two *P. pastoris* strains: *P. pastoris* CBS7435/*P. pastoris* GS115 (the sequences are identical in the two strains; however, in a few cases only the ORF name of one strain is reported because the sequence of the other strain is not or wrongly annotated.)
[c]*n.a.* not available on microarray
[d]*n.i.* not identified; *n.q.* identified but could not be quantified
[e]*n.i.* not identified in the peroxisome fraction
[f] Prediction of peroxisomal targeting with PTS1 predictor [43] (classification according to [42]: yes: predicted; twilight zone: questionable but with reasonable estimated false-positive rate; no: not predicted)

**Table 3** Identification and quantification of methanol metabolic enzymes and control proteins in peroxisomal preparations (Pex) and homogenates (Hom) of *P. pastoris* grown on methanol. MASCOT scores indicate identification of the respective proteins in the samples while peak areas of the identified peptides were used for quantification. To normalize the dataset, average ratios of the summarized peak areas of Aox1 peptides of peroxisomal samples vs homogenates were set to 1, and all ratios were calculated in relation to this. Peroxisomal proteins serve as positive control, while methanol dissimilation, pentose phosphate pathway (PPP), and glycolysis-related enzymes are negative controls localized to the cytosol

| Short name | Function/ localization | Description | MASCOT Score Pex1 | MASCOT Score Pex2 | MASCOT Score Hom1 | MASCOT Score Hom2 | ratio peak area Pex/Hom |
|---|---|---|---|---|---|---|---|
| AOX1 | Methanol assimilation | Alcohol oxidase 1 | 1542.8 | 1157.2 | 1021.4 | 1061 | 1.00 |
| DAS1 | Methanol assimilation | Dihydroxyacetone synthase 1 | 1918.9 | 1503 | 971.2 | 970.4 | 14.79 |
| DAS2 | Methanol assimilation | Dihydroxyacetone synthase 2 | 1797.9 | 1473.2 | 986.3 | 961.1 | 7.10 |
| DAK2 | Methanol assimilation | Dihydroxyacetone kinase | 226.9 | 0 | 666.2 | 550.4 | 0.26 |
| FBA1-2 | Methanol assimilation | Fructose-bisphosphate aldolase | 464.8 | 185.7 | 287 | 291.5 | 0.96 |
| FBP1 | Methanol assimilation | Fructose-1,6-bisphosphatase | 623.8 | 419.8 | 585 | 576.8 | 2.42 |
| SHB17 | Methanol assimilation | Sedoheptulose-1,7-bisphosphatase | 357.4 | 314 | 191 | 145.2 | 3.05 |
| RKI1-2 | Methanol assimilation | Ribose-5-phosphate ketol-isomerase | 217.1 | 0 | 0 | 139.8 | 2.97 |
| TAL1-2 | Methanol assimilation | Transaldolase | 374.3 | 279.8 | 0 | 0 | > > 1 |
| CTA1 | Peroxisomal protein | Catalase | 907.3 | 434.4 | 414.3 | 304.3 | 1.26 |
| PEX3 | Peroxisomal protein | Peroxisomal biogenesis factor | 119.2 | 109.9 | 0 | 0 | > > 1 |
| PEX5 | Peroxisomal protein | Peroxisomal targeting signal 1 receptor | 86.1 | 62.5 | 127.8 | 36.3 | 1.09 |
| PEX11 | Peroxisomal protein | Peroxisomal membrane protein | 523.4 | 262.1 | 221.5 | 151.8 | 4.26 |
| PEX14 | Peroxisomal protein | Peroxisomal membrane protein | 145.2 | 103.7 | 0 | 0 | > > 1 |
| PMP20 | Peroxisomal protein | Peroxiredoxin | 536 | 437.3 | 260.7 | 274.2 | 1.49 |
| PMP47 | Peroxisomal protein | Peroxisomal membrane protein | 539.5 | 345.7 | 165.9 | 162.5 | 8.02 |
| FLD | Methanol dissimilation | Formaldehyde dehydrogenase | 0 | 0 | 577.9 | 424.4 | 0.00 |
| FGH1 | Methanol dissimilation | S-formylglutathione hydrolase | 0 | 0 | 455.4 | 432.4 | 0.00 |
| FDH1 | Methanol dissimilation | Formate dehydrogenase | 491.1 | 303.5 | 910.5 | 856.7 | 0.12 |
| TAL1-1 | PPP | Transaldolase | 220.2 | 0 | 376.7 | 206.5 | 0.38 |
| TKL1 | PPP | Transketolase | 0 | 0 | 85 | 166 | 0.00 |
| ZWF1 | PPP | Glucose-6-phosphate 1-dehydrogenase | 0 | 0 | 0 | 46.7 | 0.00 |
| GND2 | PPP | 6-phosphogluconate dehydrogenase | 247.1 | 66.6 | 626.8 | 522.9 | 0.26 |
| FBA1-1 | Glycolysis | Fructose-bisphosphate aldolase | 0 | 0 | 347.3 | 379.1 | 0.00 |
| HXK1 | Glycolysis | Hexokinase | 0 | 0 | 100.6 | 195.6 | 0.00 |
| TDH3 | Glycolysis | | 0 | 0 | 629.6 | 705.5 | 0.00 |

**Table 3** Identification and quantification of methanol metabolic enzymes and control proteins in peroxisomal preparations (Pex) and homogenates (Hom) of *P. pastoris* grown on methanol. MASCOT scores indicate identification of the respective proteins in the samples while peak areas of the identified peptides were used for quantification. To normalize the dataset, average ratios of the summarized peak areas of Aox1 peptides of peroxisomal samples vs homogenates were set to 1, and all ratios were calculated in relation to this. Peroxisomal proteins serve as positive control, while methanol dissimilation, pentose phosphate pathway (PPP), and glycolysis-related enzymes are negative controls localized to the cytosol *(Continued)*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Glyceraldehyde-3-phosphate dehydrogenase | | | | | |
| PGK1 | Glycolysis | Phosphoglycerate kinase | 0 | 0 | 372.1 | 224.3 | 0.00 |
| GPM1 | Glycolysis | Phosphoglycerate mutase | 0 | 0 | 121.7 | 108.4 | 0.00 |

reaction. Erythrose-4-phopsphate is then condensed with DHAP to form S1,7BP, a reaction shown to be catalyzed by the aldolase Fba1 in yeast and plants [46]. We propose that peroxisomal Fba1-2 or Tal1-2 might be the responsible enzyme for this reaction in *P. pastoris*. Shb17 catalyzes the dephosphorylation of S1,7BP to sedoheptulose-7-phosphate, which is finally converted to two XYL5P by transketolase, Rki1-2, and Rpe1-2. As *P. pastoris* Tkl1 is cytosolic and not induced in the presence of methanol, we propose that Das1 and/or Das2, both homologs of Tkl1, catalyze this reaction. Overall, in a process driven by the net loss of one high-energy phosphate bond, Sbh17, together with transaldolase and transketolase, convert five moles of triose-phosphate into the three moles of XYL5P required for fixation of three moles of formaldehyde by DAS. Localization of this entire pathway in the same compartment makes import of XYL5P into peroxisomes obsolete, which was proposed to be necessary by Douma et al. [47] according to the classical model of methanol assimilation. Thus, the net peroxisomal flux of carbon would be one mole DHAP or GAP out of peroxisomes per three moles of methanol.

Based on our data, we propose here a novel carbon assimilation pathway (Fig. 3, left) that shares the concept of compartmentalization with plants [48] and cyanobacteria [49]. According to this model, DAS is responsible for C-C bond formation similar to the mechanism of RuBisCO, followed by a cyclic pathway (the equivalent to the Calvin cycle; Fig. 3, right) for regeneration of the pentose phosphate substrate of the carboxylation reaction. Shb17 has been shown to drive the flux from erythrose-4-phosphate and DHAP toward ribose-5-phosphate during riboneogenesis in *S. cerevisiae* in a reaction similar to the Calvin cycle [46]. A similar mechanism driving the flux towards XYL5P is proposed here in methanol-induced *P. pastoris* and probably also in other methylotrophic yeasts. Supporting this hypothesis we also found PTS1 containing isoforms of Fba1 and Tal1 in the *H. polymorpha* genome sequence by BLAST analysis.

Tandem gene duplication occurs with high frequency and has been reported to be a major contributor of new genetic material [50]. Several models for the occurrence of gene duplications have been proposed (reviewed in [50, 51]). Among them unequal crossing over can lead to tandem duplication, as it is observed here. Duplicated genes have a high probability of being lost again unless they acquire a new function [50]. Byun-McKay and Geeta [52] have proposed that subcellular relocalization of duplicate gene products may play an important role in stabilizing duplications and acquiring new functions. While they extend their idea only to N-terminal mutations modifying targeting sequences to the endoplasmic reticulum, mitochondria, or chloroplasts, it may well be that C-terminal mutations may have enabled peroxisomal relocation of duplicate gene products in an ancestor of methylotrophic yeasts. One may envisage that compartmentalized xylulose-5-monophosphate pathway enzymes would constitute novel functions which underwent positive selection, leading to a highly regulated peroxisomal pathway as observed herein, while leaving the PPP unaffected.

### The central carbon metabolism is reverted to gluconeogenesis

Growth on non-carbohydrate carbon sources necessitates the synthesis of hexoses and pentoses for the biosynthesis of macromolecules, which is accomplished by reverting the carbon flux to gluconeogenesis. Glycolysis and gluconeogenesis share several enzymes, while the irreversible, highly exergonic steps of glycolysis are bypassed. Therefore, exactly these reactions are the control steps of flux direction, and their regulation indicates the activity of gluconeogenesis. Both methanol and glycerol enter the central carbon metabolism at the level of C3-molecules (DHAP and GAP). One of the key regulatory enzymes of the upper part of glycolysis/gluconeogenesis is fructose-1,6-bisphosphatase, which we found to be up-regulated in methanol/glycerol-grown cells at the transcriptomic and proteomic level (Fig. 4). The other key regulatory enzymes of the lower part, pyruvate carboxylase and phosphoenolpyruvate carboxykinase, showed no differential regulation comparing both conditions, which is consistent with the fact that carbon flux from methanol and glycerol enters the central carbon metabolism at the point of glyceraldehyde-3-phosphate. Correspondingly, there was only a minor

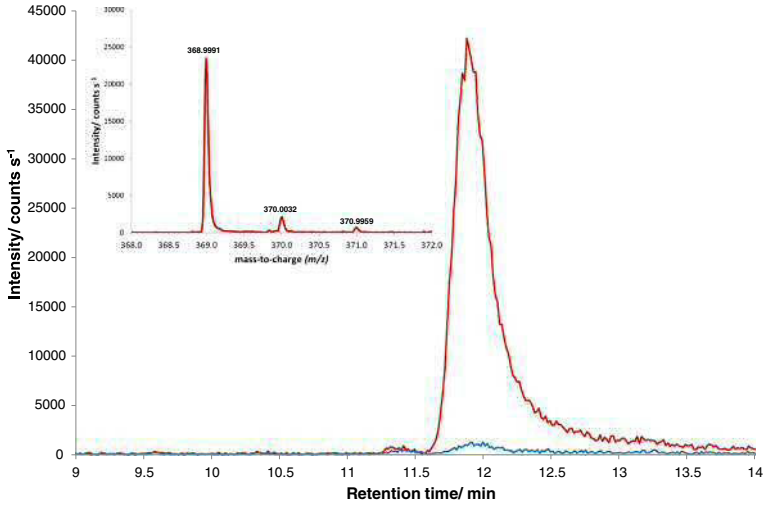


**Fig. 2** Extracted ion chromatogram ($m/z = 368.9993 \pm 10$ ppm) of a sample grown on methanol (red) or glucose (blue) showing the sedoheptulose-1,7-bisphosphate peak at a retention time of 12 min. In the upper left corner the mass spectrum ($m/z$) of the peak of the methanol-grown sample after background subtraction is shown. The difference between the measured accurate mass and the calculated exact mass is 0.53 ppm (0.2 mDa)

difference in the calculated fluxes from glyceraldehyde-3-phosphate towards pyruvate while the upper glycolytic flux was reverted on methanol/glycerol towards glucose-6-phosphate (Fig. 5). Additionally, the absence of extracellular glucose rendered low and high affinity glucose transporters *HXT1* and *GTH1* as well as hexokinase *HXK1* obsolete which all had significantly lower transcript (and protein) levels on methanol/glycerol (Fig. 1, lower left).

Flux through the lower branch of glycolysis was lower on methanol/glycerol (Fig. 5) which fits to the observed lower TCA-cycle flux and the accumulation of three glycolytic intermediates (2-phosphoglycerate, 3-phosphoglycerate, and

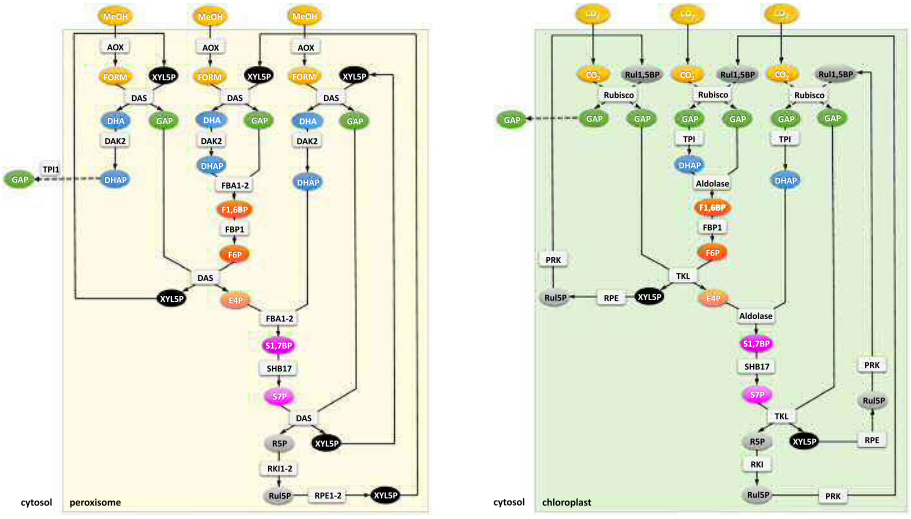


**Fig. 3** Regeneration of pentose phosphates. Left: Methanol assimilation through the xylulose-monophosphate cycle: proposed rearrangements employing an alternative XYL5P forming pathway via S1,7BP. The net reaction of methanol assimilation is the formation of one GAP molecule from three methanol molecules. Right: Rearrangement reactions of the Calvin cycle. Regeneration of ribulose-1,5-bisphosphate (Rul-1,5-BP) needed for $CO_2$ fixation in chloroplasts of plants via S1,7BP. For simplicity, the initial reaction steps after carbon fixation are condensed. The enzyme RuBisCO catalyzes the fixation of $CO_2$ to Rul-1,5-BP, which yields two 3-phosphoglycerate molecules, which are phosphorylated to 1,3-bisphosphoglycerate by phosphoglycerate kinase, and then reduced to GAP by glyceraldehyde 3-phosphate dehydrogenase. Involved metabolites are in oval signs, genes/proteins are shown in rectangular signs. The colors of the individual metabolites serve for better readability of the figure, that is, chemically related compounds share the same color. The regulation pattern and the cellular localization of the proteins is given in Table 2

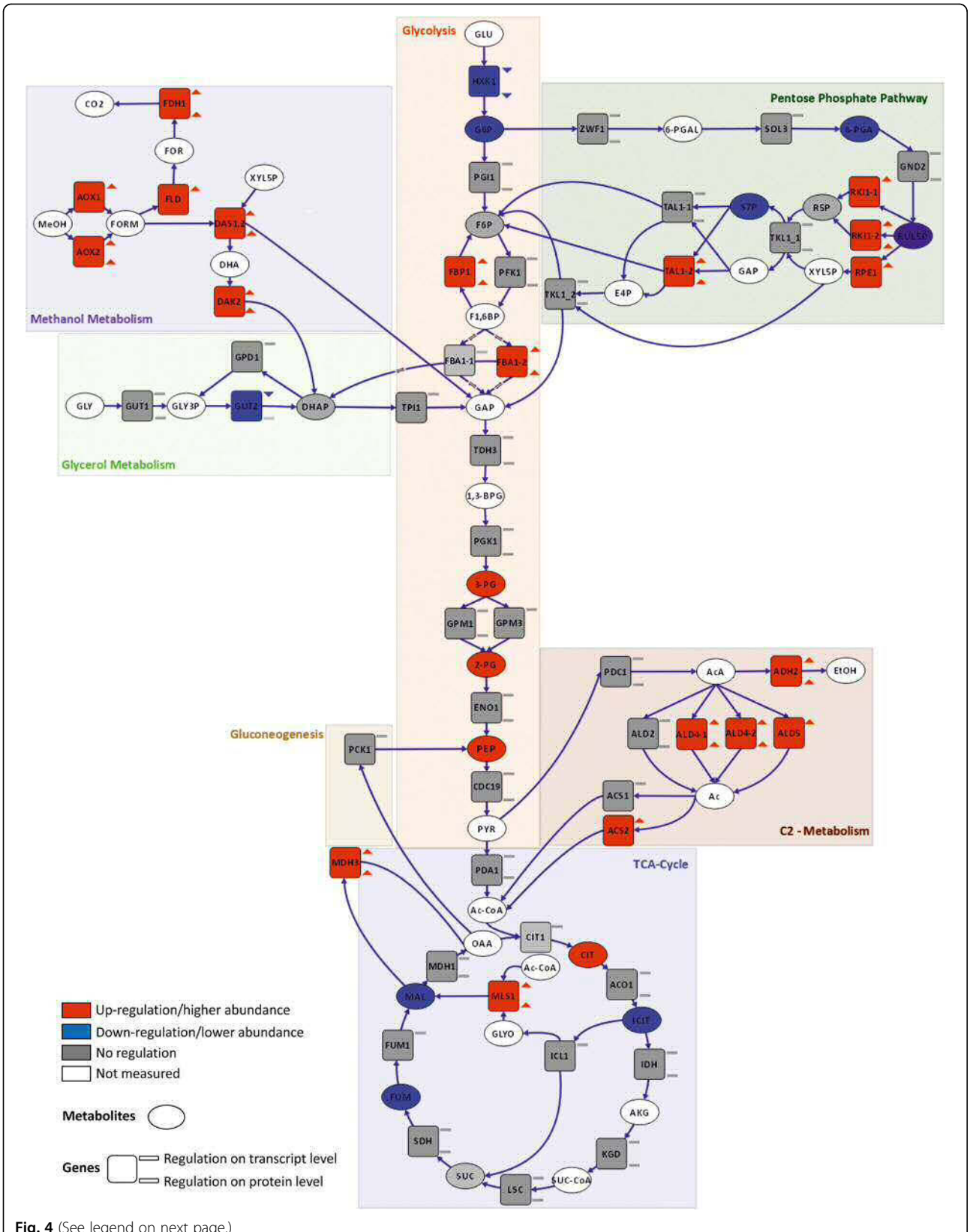


**Fig. 4** (See legend on next page.)

(See figure on previous page.)
**Fig. 4** Differential regulation of central carbon metabolism comparing methanol/glycerol- and glucose-grown cells. Visualization of changes in transcript (square, upper symbol), protein (square, lower symbol), and metabolite (oval) levels. Red: up-regulation on methanol/glycerol; blue: down-regulation on methanol/glycerol; gray: not differentially regulated; white/no symbol: not measured. Criteria for up-/down-regulation of transcript, protein, and metabolite levels are described in the Methods section

phosphoenolpyruvate). Transcript and protein levels of most glycolytic enzymes did not change (Fig. 4). Similarly, in *S. cerevisiae*, a poor correlation between fluxes and transcript levels of genes of this pathway was observed [53]. The low TCA-cycle flux indicates that methanol dissimilation is a major source for NADH and energy production in cells grown on methanol/glycerol.

### Pentose phosphate pathway flux is increased on methanol/glycerol

The PPP serves for the generation of NADPH for reductive assimilatory processes and for the generation of ribose-5-phosphate as a precursor for nucleic acids. Other PPP intermediates are used as precursors for other metabolic pathways like synthesis of histidine, nucleotides, and riboflavin. Methanol assimilation, using sugar phosphate intermediates in a cyclic fashion, has to be regarded separately from PPP, as outlined above.

The high gluconeogenic flux on methanol/glycerol was accompanied by a high flux through the PPP (Fig. 5), enabling a high specific production rate of reduced NADPH of appr. 3 mmol $g^{-1}$ $h^{-1}$. About 10 % of this higher NADPH production is needed for amino acid synthesis for the higher protein content of cells grown on methanol/glycerol. A higher PPP flux also provides for more ribose for nucleotide and riboflavin synthesis (see below). Cytosolic pentose phosphate pathway genes, however, were not differentially regulated at the different media (Table 2), contrary to the methanol-induced peroxisomal isoforms suggested to be employed in methanol assimilation in this compartment.

### Up-regulation of the malate-aspartate shuttle serves for mitochondrial import of NADH generated by methanol dissimilation and correlates with decreased TCA-cycle flux in methanol/glycerol-grown *P. pastoris*

The TCA-cycle as a central hub for cellular metabolism is dedicated to energy production and supply of precursors for several other metabolic pathways. Again, the most striking C-source-dependent differences were observed on the flux level, being 3.2 times lower on methanol/glycerol, mainly being controlled by the lower influx of acetyl-CoA into the TCA-cycle (Fig. 5). The low TCA-cycle flux on methanol is mainly diverted to glutamate, thus contributing only marginally to energy production. No significant changes in transcript or protein levels of genes connected to the TCA-cycle were observed. Nevertheless, we found marked differences in TCA-cycle metabolites. Citrate levels in methanol/glycerol-grown cells were higher than in glucose-grown cells, whereas levels for isocitrate, fumarate, and malate were lower (Fig. 4). Taken together, these data indicate that, on methanol/glycerol, the TCA-cycle reactions are mainly employed for production of metabolic precursors for biomass formation rather than producing energy through the respiratory chain. Methanol utilization has a major impact on the energy state of the cells, as two moles of NADH are produced via dissimilation of one mole of methanol to $CO_2$. Intracellular flux calculation showed that about half of the methanol was dissimilated to $CO_2$ and therefore additional NADH was produced which may consequently lead to down-regulation of TCA-cycle flux.

Dissimilatory oxidation of formaldehyde takes place in the cytosol. Therefore, the produced NADH has to be transported via the inner mitochondrial membrane to drive the generation of ATP. For the transport of electrons via the mitochondrial membrane several shuttle systems exist, most importantly the malate-aspartate shuttle. The homologs of the malate-α-ketoglutarate transporter Odc1 [54] and the glutamate-aspartate transporter Agc1 [55] were both highly up-regulated at transcript and protein levels, indicating the relevance of this NADH shuttle for methylotrophic ATP generation in mitochondria.

### The glyoxylate cycle is active in methanol/glycerol-grown *P. pastoris* and mainly localizes to the peroxisomes

The glyoxylate cycle is necessary for the utilization of non-fermentable carbon sources because of its ability to convert acetyl-CoA into C4 compounds that can be used for ATP generation in the mitochondria [56]. Isocitrate lyase converts isocitrate to glyoxylate and succinate, the former intermediate is then condensed with acetyl-CoA to form malate by malate synthase. Additionally, malate dehydrogenase, citrate synthase, and aconitase are required. This process is assumed to take place in the peroxisomes in non-*Saccharomyces* yeasts [8]. Indeed, we found most of the enzymes to be present in the peroxisomal fraction in methanol-grown cells (Table 2), only isocitrate lyase was found solely on methanol but predominantly in the cytosolic fraction. Furthermore, methanol-grown cells had increased transcript and/or protein levels of both putative peroxisomal malate synthase (PAS_chr1-4_0459, which we named Mls2) as well as of the cytosolic malate synthase and malate dehydrogenase. In agreement with this data, we also found increased glyoxylate cycle fluxes in cells grown

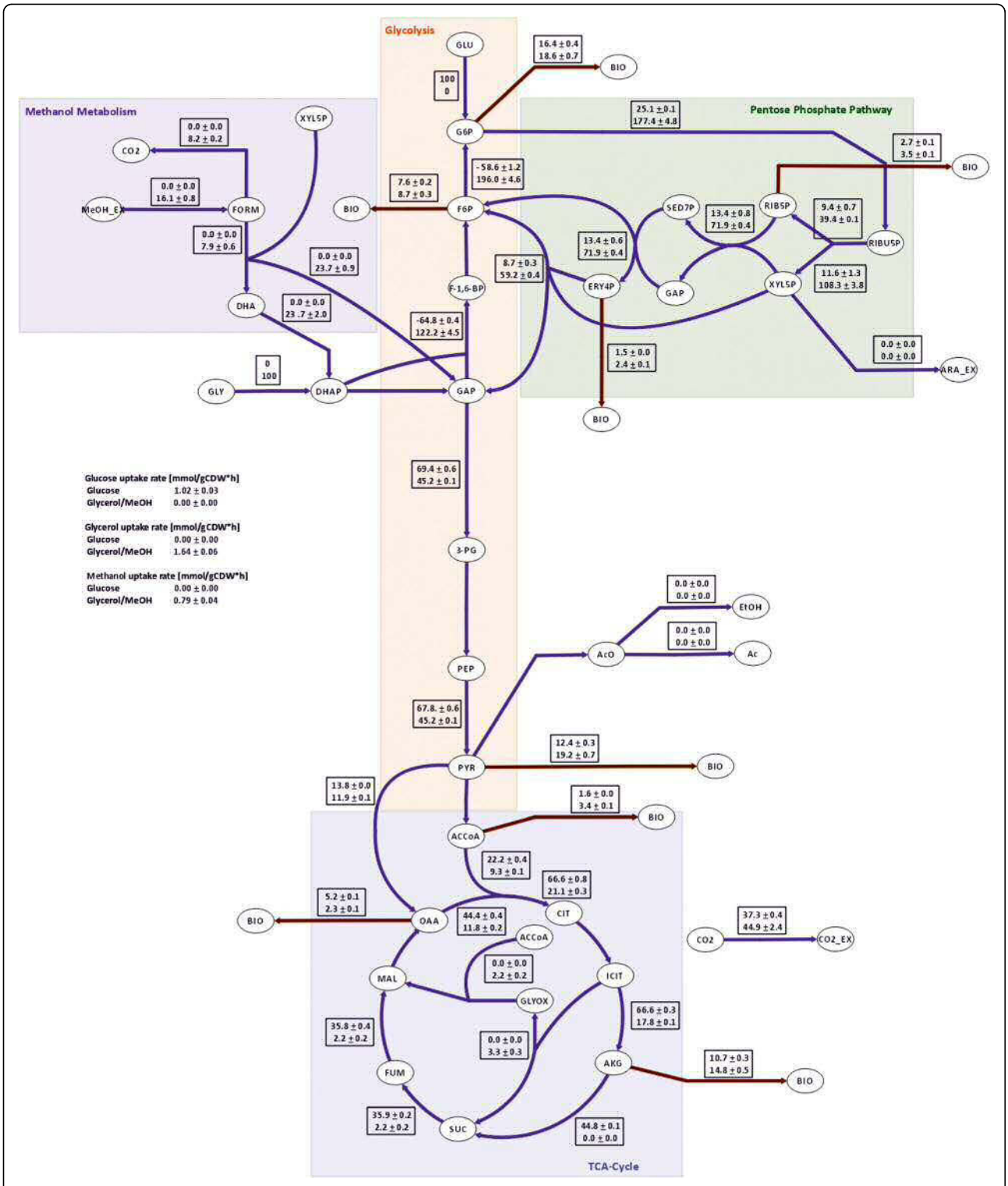


**Fig. 5** Intracellular metabolic flux distributions of methanol/glycerol- and glucose-grown cells. The flux values are normalized to glycerol or glucose uptake, respectively, and presented in [% Cmol]. The upper value in the rectangular boxes represents the flux distribution on glucose and the lower value the flux distribution on methanol/glycerol. For reversible reactions only the net fluxes are presented

on methanol/glycerol. We propose that the generated C4 compounds are mainly used as precursors for the biosynthesis of TCA-cycle-derived amino acids when using methanol/glycerol as substrate, rather than being shuttled to gluconeogenesis. In this line, no up-regulation of phosphoenolpyruvate carboxykinase, a key enzyme in lower gluconeogenesis, could be seen.

### Methanol-grown *P. pastoris* cells have a higher protein but lower free amino acid content

Protein is the largest macromolecular component of cells, creating the highest demand for energy, reduction equivalents, and carbon flux. A change in substrate forces the cell to adapt, e.g. by varying the total protein content. Methanol/glycerol-grown cells had a 35 % higher protein content, with 0.54 g(protein) g(cell dry weight (CDW))$^{-1}$ compared to glucose-grown cells with 0.40 g(protein) g(CDW)$^{-1}$ (Table 4). Consequently, the levels of protein-bound amino acids were generally higher in cells grown on methanol/glycerol, while the level of free intracellular amino acids was approximately 20 % lower in methanol/glycerol-grown cells. The higher specific protein synthesis rate on methanol/glycerol (0.054 g(protein) g(CDW)$^{-1}$ h$^{-1}$ vs 0.040 g(protein) g(CDW)$^{-1}$ h$^{-1}$ on glucose) creates a higher drain of amino acids towards protein synthesis, which explains the generally low levels of intracellular free amino acids. The higher demand for amino acids creates a metabolic pull for the respective synthesis pathways. The higher demand for amino acids was well supported by the transcriptional and/or post-transcriptional up-regulation of genes involved in the biosynthesis of all twenty amino acids (Fig. 6). For histidine which derives from one intermediate of the pentose phosphate pathway, we saw regulation of *HIS1*, while for the amino acids which derive from glycolytic intermediates, regulation of Ser2 (serine), *AGX1* and *GLY1* (glycine), *TRP5-2* (tryptophan), Aro7 and Aro8 (tyrosine and phenylalanine), Ilv2 and Ilv5 (leucine, valine, isoleucine), and Alt1 (alanine) was observed. If we consider amino acids which derive from intermediates of the TCA-cycle, we saw regulation of *LYS20*, *LYS21* and Lys2 (lysine), Gdh2 and Gdh3 (glutamate), *AAT1* (aspartate), *ASP1* (asparagine), Hom2 (precursor of threonine, methionine, and cysteine), *THR1* (threonine), Met17 (precursor of methionine and cysteine), and Cys3 (cysteine).

While methanol utilization enzymes were up-regulated at the transcriptional level and thus increased at the protein level, the second major class of more abundant proteins (ribosomal proteins) were not transcriptionally regulated.

**Table 4** Composition of protein bound and free intracellular amino acids of *P. pastoris* grown on glucose or methanol/glycerol in chemostats at $\mu = 0.1$ h$^{-1}$

| | Protein bound amino acids | | | Free intracellular amino acids | | | |
|---|---|---|---|---|---|---|---|
| | Methanol/glycerol | Glucose | $Log_2FC$ | Methanol/glycerol | Glucose | $Log_2FC$ | *P* value |
| | Average [mg/gCDW] | Average [mg/gCDW] | | Average [mg/gCDW] | Average [mg/gCDW] | | |
| Asx | 48.8 | 30.2 | 0.69 | 5.42 | 2.91 | 0.89 | 0.00 |
| Ala | 32.1 | 20.2 | 0.67 | 0.90 | 1.02 | −0.18 | 0.28 |
| Arg | 29.6 | 25.9 | 0.19 | 11.1 | 12.1 | −0.12 | 0.47 |
| Cys | 5.59 | 4.56 | 0.30 | – | – | – | – |
| Glx | 87.0 | 68.0 | 0.36 | 22.3 | 29.4 | −0.40 | 0.00 |
| Gly | 18.7 | 11.5 | 0.70 | – | – | – | – |
| His | 11.9 | 7.33 | 0.70 | 0.76 | 0.74 | 0.03 | 0.78 |
| Ile | 17.8 | 10.6 | 0.76 | 0.05 | 0.04 | 0.28 | 0.00 |
| Leu | 36.0 | 22.6 | 0.67 | 0.10 | 0.08 | 0.35 | 0.00 |
| Lys | 37.5 | 24.5 | 0.62 | 0.76 | 0.92 | −0.28 | 0.18 |
| Met | 5.34 | 4.33 | 0.30 | 0.04 | 1.08 | −4.63 | 0.00 |
| Phe | 20.0 | 12.9 | 0.63 | 0.03 | 0.03 | 0.03 | 0.56 |
| Pro | – | – | – | 0.92 | 2.76 | −1.58 | 0.00 |
| Ser | 25.9 | 19.0 | 0.44 | 0.25 | 0.52 | −1.04 | 0.00 |
| Thr | 28.2 | 19.3 | 0.55 | 0.24 | 0.27 | −0.14 | 0.42 |
| Tyr | 17.3 | 8.89 | 0.96 | 0.06 | 0.05 | 0.18 | 0.02 |
| Val | 26.4 | 15.7 | 0.75 | 0.25 | 0.14 | 0.83 | 0.00 |
| Total amino acids [mg/g Cell dry weight (CDW)] | 448 | 306 | 0.55 | 43.2 | 52.1 | −0.27 | 0.03 |
| Total protein content [mg/gCDW] | 540 | 390 | 0.47 | – | – | – | – |

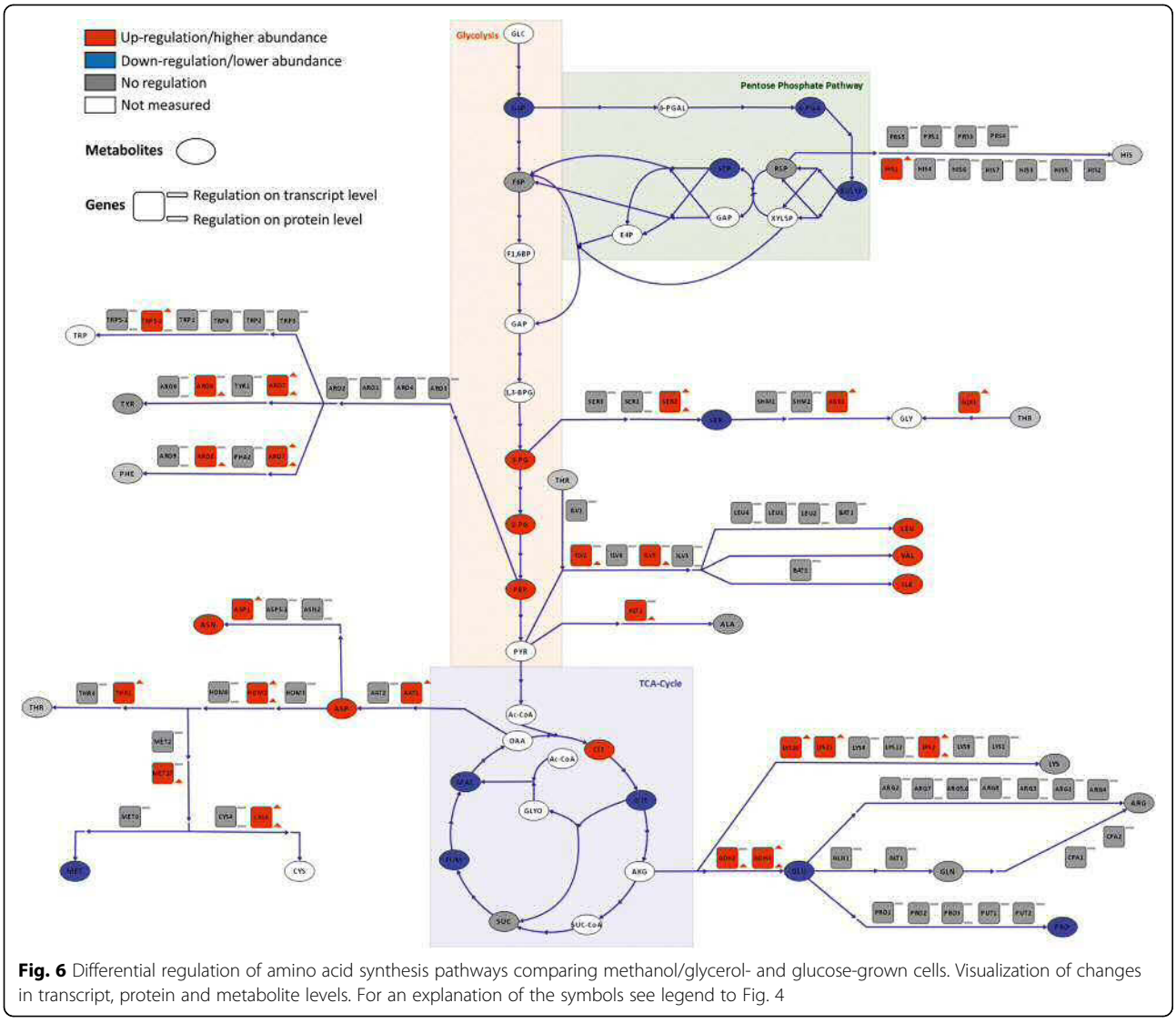


**Fig. 6** Differential regulation of amino acid synthesis pathways comparing methanol/glycerol- and glucose-grown cells. Visualization of changes in transcript, protein and metabolite levels. For an explanation of the symbols see legend to Fig. 4

This finding indicates an efficient post-transcriptional regulation mechanism, as described for *S. cerevisiae* [57], and a higher steady state demand of translational capacity on methanol. This last observation is supported also by a higher total protein content of biomass grown on methanol/glycerol and by the up-regulation of amino acid synthesis pathways. It remains to be elucidated whether this higher translational capacity on methanol is related to the observed higher recombinant protein production capacity of methanol-based expression strains. The general upregulation of protein synthesis in methanol-induced cultures did not coincide with a higher abundance of enzymes in the protein folding machinery. An accumulation of misfolded proteins as a result of heterologous gene expression has been observed many times in recombinant *P. pastoris*, leading to UPR activation (reviewed by Puxbaum et al. [58]) but appears to be absent in non-recombinant *P. pastoris* cultivated on methanol.

### Protein folding, secretion, and degradation pathways are not affected by methanol as substrate

While we and others have observed a transient up-regulation of the UPR immediately following methanol induction [59, 60], no such regulation pattern was noticed in the methanol-adapted cells in steady state in the present study, thus ruling out the possibility that a permanently induced UPR positively influences recombinant protein production in methanol-grown cells. Contrary to Liang et al. [16], who detected up-regulation of endoplasmic reticulum protein processing, N-glycan biosynthesis, and protein export pathways when comparing recombinant protein secreting *P. pastoris* in chemostats with methanol as substrate, we did not see any changes in protein folding, secretory pathway, N-glycosylation, or proteasome both at the proteome and transcriptome level in the non-expressing strains in our study (Additional file 1). Protein synthesis,

however, was obviously up-regulated on methanol, as described above.

## High levels of methanol utilization enzymes require overproduction of vitamins and cofactors

### *Alcohol oxidase requires high riboflavin synthesis*

AOX, which catalyzes the first reaction of methanol utilization, is a homooctamer with flavin adenine dinucleotide (FAD) as non-covalently bound prosthetic group. When methylotrophic yeast cells grow on methanol, AOX can account for up to 30 % of total cellular protein [61], and the FAD content of AOX alone amounts to 1.7 mg/g biomass. AOX predominantly oligomerizes in the peroxisomal matrix of methylotrophic yeasts [62, 63]. Experiments with *H. polymorpha* and *P. pastoris* revealed that insertion of FAD is an essential step prior to the assembly of AOX [63].

Almost the entire pathway leading to FAD is transcriptionally up-regulated when methanol is present (Fig. 7 and Additional files 1 and 2): *RIB1* and *RIB3* encode the first steps of the riboflavin biosynthesis pathway (with GTP and ribulose-5-phosphate as precursors, respectively), while *RIB4* and *RIB5* code for the last enzymes in the pathway. The induction of the riboflavin pathway during growth on methanol has been previously observed [16], but was not linked to AOX biosynthesis. Via the up-regulated *FMN1*, riboflavin is converted to flavin mononucleotide (FMN), a strong oxidizing cofactor of mitochondrial NADH-dehydrogenases (which, however, are not regulated). The generation of FAD from FMN is catalyzed by Fad1, which is strongly transcriptionally up-regulated (~8-fold). The bulk of FAD apparently goes into AOX as other cellular flavoproteins [64] are rather unaffected during growth on methanol with the FAD-requiring Gut2, Hem14 (both up-regulated), Pox1, Fmo1-1, and Fre2 (all three down-regulated) as exceptions. We observed no changes in free riboflavin, indicating that flux to riboflavin is up-regulated upon its high demand while synthesis is tightly regulated by its free intracellular concentration as described by Marx et al. [65].

## Thiamine synthesis is strongly up-regulated due to demand of peroxisomal transketolases

There is also significant up-regulation of genes involved in thiamine (vitamin B1) and thiamine pyrophosphate (TPP) biosynthesis (*THI20*, *THI6*, *THI80*, *THI4*, *THI13*, *THI73*, *THI21*; Fig. 7 and Additional files 1 and 2) in cells grown on methanol. We have shown before that severe thiamine limitation is required for the induction of *THI13* [66], indicating that induction of the methanol utilization pathway leads to intracellular thiamine deficiency.

TPP, the active derivative of thiamine, is the co-factor of decarboxylases, transketolases, and phosphoketolases. The homodimeric enzymes bind one $Mg^{2+}$ ion and one TPP per subunit. In *P. pastoris* cultures grown on methanol, upregulation of TPP biosynthesis coincided with the high abundance of the TPP-containing enzymes Das1 and Das2, catalyzing the fixation of formaldehyde to XYL5P in the peroxisome. In this study, Das1 and Das2 had very high changes in transcript levels (~68-fold and 46-fold up-regulation, respectively) and one of the highest changes in protein level (~6.5-fold and 4.7-fold, respectively). On the contrary, the level of cytosolic transketolase (Tkl1) was unaffected. Thus, we conclude that the strong induction of DAS1/2 led to a limitation of thiamine availability which was compensated by induction of the thiamine synthesis pathway.

## High nicotinamide levels are required for formaldehyde detoxification

In the methanol dissimilation pathway, formaldehyde is oxidized to carbon dioxide by two consecutive reactions catalyzed by formaldehyde dehydrogenase and formate dehydrogenase. On methanol, both enzymes are strongly increased both on transcript (5-fold and 19-fold, respectively) and protein (formaldehyde dehydrogenase not quantified, formate dehydrogenase 5-fold) levels. The two enzymes, which are mainly located to the cytosol, are required for detoxification of formaldehyde and formate and both use nicotinamide adenine dinucleotide as cofactor. The generated NADH provides energy for growth on methanol. In this respect, the total amount of nicotinamide in cells grown in the presence of methanol was nearly 10-fold higher than in the glucose-grown cells, and total NAD content is approx. 50 % higher. Expression of *NMA1*, encoding nicotinic acid mononucleotide adenylyltransferase, which is involved in the *de novo* biosynthesis of NAD as well as in the NAD salvage pathway [67], was up-regulated 2.4-fold.

## Heme synthesis is up-regulated upon catalase demand for peroxide detoxification

Toxic $H_2O_2$ and formaldehyde are generated in the first step of methanol metabolism. The peroxisomal enzyme catalase, which is involved in the detoxification of $H_2O_2$, is transcriptionally up-regulated when methanol is present. Properly folded catalase incorporates a heme cofactor with an iron ion in the center, and needs to tetramerize to become active [63]. *CTA1* expression is up-regulated on methanol and, consequently, expression of almost all heme biosynthesis genes was up-regulated, including the rate-limiting steps *HEM2* and *HEM3* ($log_2$FC +0.43 and +0.54). Pet18, a heme oxygenase-like protein, was also up-regulated at both the transcript and protein levels. Heme oxygenases

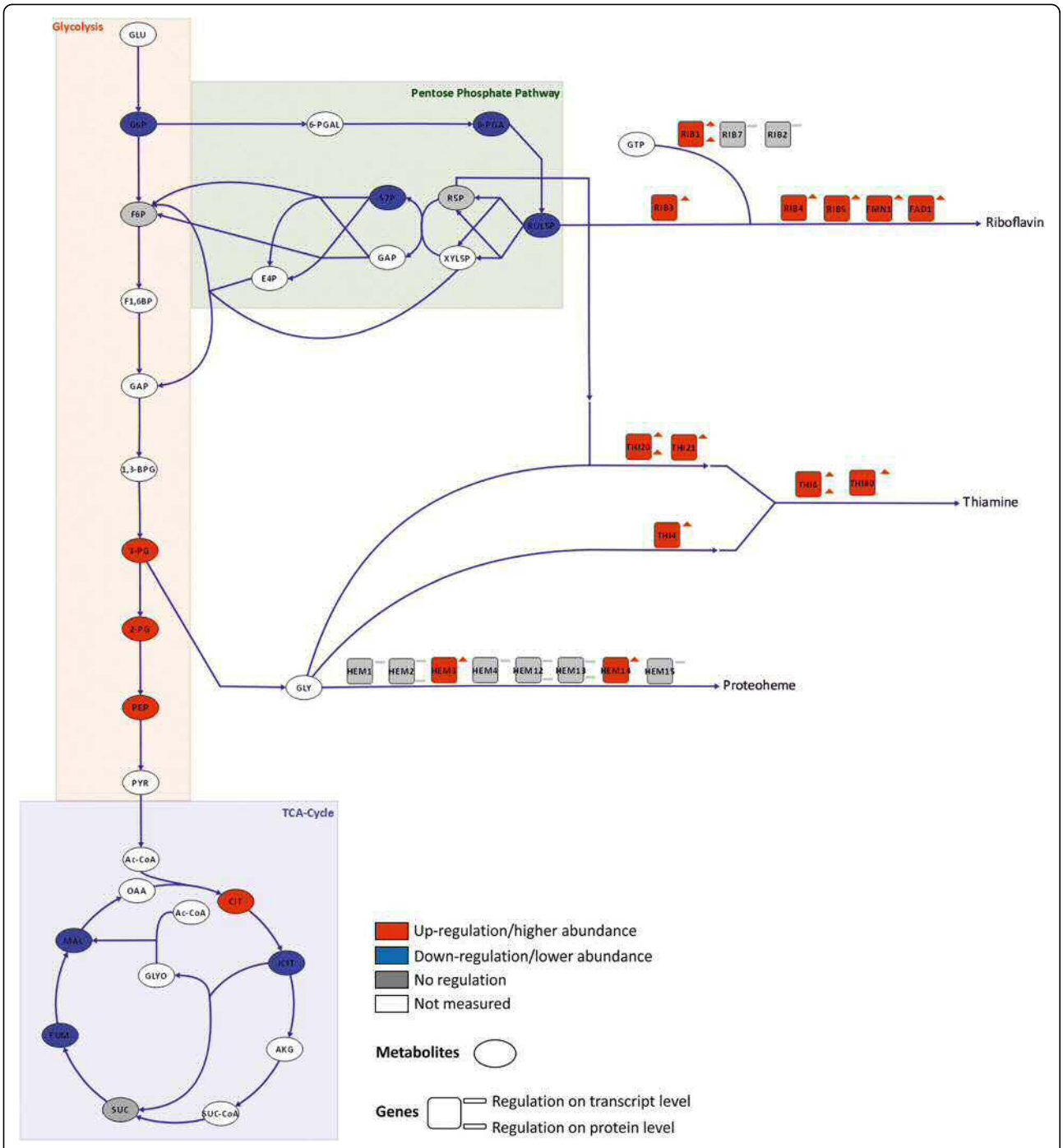


**Fig. 7** Differential regulation of riboflavin, thiamine, and heme synthesis pathways comparing methanol/glycerol- and glucose-grown cells. Visualization of changes in transcript, protein, and metabolite levels. For an explanation of the symbols see legend to Fig. 4

catalyze the degradation of heme and produce iron. Down-regulation (0.6-fold) of a low-affinity Fe(II) transporter (*FET4-2*) and up-regulation (1.65-fold) of *FTH1-1*, a putative high affinity iron transporter involved in intravacuolar iron storage, points towards low iron levels in the presence of methanol.

### On methanol, the general lipid metabolism is altered to allow peroxisome formation at the expense of lipid droplets

Environmental conditions and nutritional modifications often have dramatic effects on the composition of cellular membranes, which becomes apparent in lipid composition

and regulation of lipid metabolism. Major key enzymes of lipid-related pathways were apparently not affected when culture conditions varied between the supply of glucose or methanol/glycerol. Important components of biological membranes, sterols and phospholipids, were elevated only slightly in methanol/glycerol-grown cells (Table 5). The observed increase of building blocks for membranes can most likely be explained by the enhanced occurrence of internal membranes. The total amount of peroxisomes was strongly increased in methanol/glycerol-grown cells which caused a weak effect on the total amounts of phospholipids and sterols of internal membranes. Wriessnegger et al. [44, 68] already showed in previous work that utilization of glucose or methanol as the sole carbon source does not lead to major differences in the distribution of phospholipids, although the culture conditions and sampling points were not the same as in the present study. The slight increase in the total amount of phospholipids observed here was not matched by any significant regulation of lipid biosynthetic genes involved in the complex pathways of phospholipid formation, except for *INO1* and *OPI3*, which were both down-regulated ($log_2$FC of −1.38 and −0.89). The pattern of fatty acids from methanol-grown cells as well mostly resembled glucose-grown cells, although some minor changes were detected. A decrease in oleic acid (C18:1) by roughly 20 % was accompanied by an increase in palmitic acid (C16:0), palmitoleic acid (C16:1), and linolenic acid (C18:3). Again, the influence of the intracellularly predominant peroxisomal membranes most likely was the reason for the observed changes of bulk membrane fatty acid composition.

**Table 5** Glycerophospholipid, non-polar lipid (TG, triacylglycerol; SE, steryl esters), unesterified ergosterol, and free and total fatty acid content in total cell extracts of *Pichia pastoris* grown on glucose (GAP) or methanol (AOX) as the sole carbon source. Data are listed as μg lipid/mg Wet Cell Weight which have been calculated from at least two independent experiments with standard deviation (±). Significance was estimated by Student's *t*-test (two tailed, unpaired)

| | Glucose | Methanol/glycerol | *P* value |
|---|---|---|---|
| Glycerophospholipids | 8.07 ± 0.14 | 8.92 ± 0.30 | 0.01 |
| Non-polar lipids | | | |
| TG | 3.16 ± 0.71 | 1.47 ± 0.28 | 0.06 |
| SE | 0.22 ± 0.03 | 0.30 ± 0.03 | 0.07 |
| Free ergosterol | 1.67 ± 0.06 | 1.83 ± 0.18 | 0.20 |
| Free fatty acids | 2.11 ± 0.49 | 3.38 ± 0.72 | 0.09 |
| Total fatty acids | | | |
| C16:0 | 1.22 ± 0.03 | 1.43 ± 0.16 | 0.01 |
| C16:1 | 0.75 ± 0.01 | 0.85 ± 0.06 | 0.003 |
| C18:0 | 0.33 ± 0.02 | 0.34 ± 0.09 | 0.67 |
| C18:1 | 4.28 ± 0.18 | 3.30 ± 0.31 | 0.0002 |
| C18:2 | 3.26 ± 0.14 | 3.45 ± 0.30 | 0.21 |
| C18:3 | 0.96 ± 0.04 | 1.30 ± 0.11 | 0.0001 |
| ∑ of fatty acids | 10.79 ± 0.40 | 10.67 ± 1.01 | 0.79 |

The strongest effect on lipid classes resulting from cultivation on different carbon sources was on triacylglycerols (TAG), the major non-polar lipid of *P. pastoris*. Both TAG synthases, *DGA1* and *LRO1*, were transcriptionally down-regulated on methanol/glycerol ($log_2$FC −1.07 and −0.43). As a direct result, TAG were reduced in methanol/glycerol-grown cells by more than 50 % (Table 5). The significant decrease of TAG was accompanied by a severe reduction of lipid droplets in *P. pastoris* cultivated on methanol, which was observed by electron microscopy (Fig. 8). While the amount of TAG was severely reduced, precursors of TAG (diacylglycerols and free fatty acids) were increased by approximately 40 %. Upon mobilization of TAG by TAG lipases, activated fatty acids could serve as substrates either for β-oxidation or as building blocks for membrane formation. In comparison to glucose we observed on methanol/glycerol a down-regulation of transcripts encoding β-oxidation relevant genes as well as TAG forming enzymes, which was not followed at the protein level. Notably, it has been previously shown that genes involved in fatty acid utilization are differentially regulated upon using glycerol or glucose as the carbon source, and depend on substrate availability [37] (Additional file 2). The utilization of methanol enables *P. pastoris* cells for proper growth based on energy supply by alcohol oxidation, but apparently does not provide excess carbons to be incorporated in storage material. Therefore, non-polar lipid synthesizing enzymes are down-regulated. As a direct consequence, no alternative supply of fatty acids may be available and β-oxidation relevant enzymes are shut down as well because of the limited substrate available.

*ERG20*, encoding farnesyl pyrophosphate synthetase, is the only lipid biosynthetic gene which was found to be up-regulated when comparing glucose to methanol-grown cells. Erg20 is part of the sterol biosynthetic pathway, which is composed of more than 20 enzymes. However, all other sterol biosynthetic genes remained transcriptionally unaffected. Erg20 is located at an important branching point of this biosynthetic pathway. The product of the Erg20 catalyzed reaction, farnesyl pyrophosphate, cannot only serve as a substrate for the formation of structural lipid compounds which is one of the major routes, but can be directed towards several other pathways, among them heme biosynthesis. As the formation of heme was found to be transcriptionally up-regulated to serve as a prosthetic group of catalase, we anticipated that *ERG20* was up-regulated predominantly to provide sufficient substrate for the *de-novo* formation of heme.

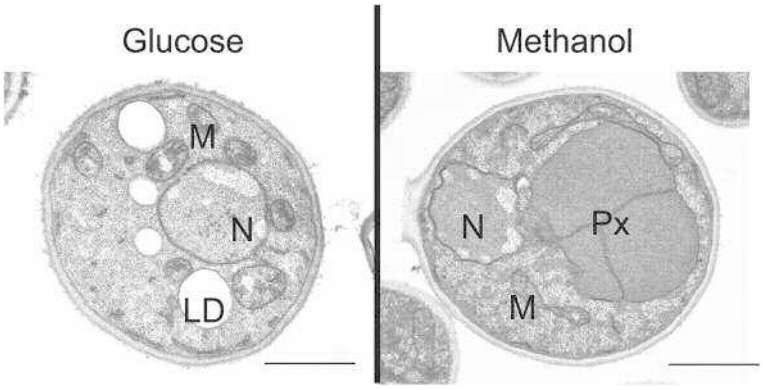


**Fig. 8** Electron microscopy of *P. pastoris* grown on glucose or methanol. Cells were cultivated in complex media containing either glucose or methanol as the sole carbon source until they reached the late exponential growth phase. N, Nucleus; M, Mitochondria; LD, Lipid droplet; Px, Peroxisome. Scale bar: 1 μm

## Conclusions

Methylotrophy is a unique ability of microorganisms to live on C1 molecules that requires efficient pathways to form C-C bonds and to oxidize C1 compounds via toxic intermediates. This systems level investigation provides comprehensive insight into regulatory and metabolic specificities of the methylotrophic yeast *P. pastoris*. Co-regulation of enzymes with AOX and DAS at the transcript and protein level allowed us to identify in detail the putative pathway for XYL5P regeneration during methanol assimilation. We revealed that the xylulose-monophosphate cycle is employing a specialized set of methanol-induced enzymes located in the peroxisome, rather than the PPP proteins, which are essentially not transcriptionally or translationally regulated in this study. For this purpose, *P. pastoris* has acquired a second copy of the relevant genes, each adjacent to the canonical PPP gene. During growth on methanol, the peroxisomes also harbor an active glyoxylate cycle, while the TCA cycle flux is reduced, indicating that methanol dissimilation is a major source for NADH and energy production on this substrate. Furthermore, growth on methanol/glycerol leads to a higher amino acid synthesis rate and a higher translational capacity which is reflected by a higher total protein content and may indicate a higher capacity for production of heterologous proteins as well. The observed changes in lipid metabolism can be explained by the high abundance of peroxisomes and the absence of lipid droplets in methanol-grown *P. pastoris*. The methylotrophic lifestyle reflects a low energy status, thus impeding lipid storage. During growth on methanol, the methanol utilization enzymes are produced in high amounts. Consequently, the biosynthetic pathways for the corresponding prosthetic groups and co-enzymes are also strongly up-regulated. Up-regulation of the pathways to riboflavin, thiamine, nicotinamide, and heme clearly indicates their high steady state demand in methanol-grown cells.

This work provides a unique data set on the methylotrophic metabolism of *P. pastoris*, and enables the redefinition of the methanol assimilation pathway. These findings will also have major impact on the understanding and evolution of methylotrophy in other yeasts.

## Methods

### Strains & chemostat cultivation

The chemostat cultivations were performed in a 1.4-L bioreactor (DASGIP Parallel Bioreactor System, Germany) with a working volume of 400 mL.

Briefly, 100 mL pre-culture medium (per liter: 10 g yeast extract, 20 g peptone, 10 g glycerol) were inoculated with 750 μL cryostock of *P. pastoris* CBS7435 and grown at 28 °C and 150 rpm overnight. This culture was used for inoculation of the bioreactor at an optical density ($OD_{600}$) of 1.0. After a batch phase of approximately 24 h, the cells were grown in carbon-limited chemostats with a dilution rate of 0.1 $h^{-1}$ for at least seven residence times before taking the samples. For each condition, three independent chemostat cultivations were performed. Temperature, pH, and dissolved oxygen were maintained at 25 °C, 5.0 (with 8 M KOH) and 20 % (by controlling the stirrer speed and inlet air), respectively.

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 32 % (w/w) HCl.

Chemostat medium (Glucose) contained per liter: 55 g glucose·$H_2O$, 2.3 g citric acid, 21.75 g $(NH_4)_2HPO_4$, 0.04 g $CaCl_2{\cdot}2H_2O$, 2.5 g KCl, 1.0 g $MgSO_4{\cdot}7H_2O$, 2 g biotin (0.2 g $L^{-1}$), and 2.43 g trace salts stock solution. The pH was set to 5.0 with 32 % (w/w) HCl.

Chemostat medium (methanol/glycerol) contained per liter: 57 g glycerol (86 %), 8.5 g methanol (100 %), 2.3 g citric acid, 21.75 g $(NH_4)_2HPO_4$, 0.04 g $CaCl_2{\cdot}2H_2O$,

2.5 g KCl, 1.0 g $MgSO_4{\cdot}7H_2O$, 2 g biotin (0.2 g $L^{-1}$), and 2.43 g trace salts stock solution. The pH was set to 5.0 with 32 % (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).

### Sampling and quenching

For transcriptomics, 9 mL of culture were added to 4.5 mL of freshly prepared pre-chilled (−20 °C) fixing solution (5 % v/v phenol in ethanol abs.), mixed, and 1.5 mL were aliquoted into ribolyzer tubes and centrifuged at 13,000 rpm for 1 min at 4 °C. The supernatant was discarded and the tubes containing the fixed cell pellets were immediately stored at −80 °C. For protein analysis, 2 mL of culture were centrifuged and the cell pellet was stored at −80 °C. The supernatant was also stored at −80 °C for analysis of extracellular metabolites.

Samples for analysis of intracellular metabolites were taken immediately by using a pump. Approximately 50 mL fermentation broth were quenched in 200 mL of 60 % (v/v) methanol at −27 °C. After quenching, 2 mL of quenched cells (corresponding to approximately 10 mg biomass per filter) were filtered with cellulose acetate filter (0.45 μm, Satorius Biolab Products) using a vacuum pump.

The cells were washed once with cold 60 % (v/v) methanol and then the filter was kept on dry ice. Using two filtration units (Polycarbonat Filter Holders, Satorius Lab Technologies Product), 6 samples per chemostat cultivation were taken.

Biomass was determined by drying duplicates of 2 mL chemostat culture to constant weight at 105 °C in preweight beakers.

### Total protein determination

Cell pellets from 2 mL chemostat culture were washed with 0.9 % NaCl and resuspended in 1 mL of PBS (pH 7.0). The protein extraction was done accordingly to Verduyn et al. [69], by addition of NaOH and incubation at 95 °C. After incubation, 0.8 M HCl were added and cell debris were collected via centrifugation. The supernatant was used for the determination of the total protein content using Bradford. The total protein content was related to the yeast dry mass (%).

### Lipid analysis

Lipid extraction from *P. pastoris* chemostat samples was performed as described by Folch et al. [70]. For quantitative determination of non-polar lipids (TG, SE), free fatty acids and free ergosterol, lipid extracts were loaded to Silica Gel 60 plates (Macherey-Nagel, Düren, Germany) and chromatograms were developed in an ascending manner by using the solvent system light petroleum/diethyl ether/acetic acid (70:30:2; per vol.) for approximately the first third of the distance. Subsequently, plates were briefly dried and further developed using the solvent system light petroleum/diethyl ether (49:1; per vol.) until the solvent front reached the top of the plate. Unesterified sterols and steryl esters were quantified densitrometrically using a TLC scanner (Camag TLC Scanner 3) at 275 nm using ergosterol as standard. Other lipids were irreversibly stained by dipping the TLC plates into a charring solution (0.63 g $MnCl_2{\cdot}4H_2O$, 60 mL water, 60 mL methanol, and 4 mL concentrated sulfuric acid) and heated at 100 °C for 30 min. Densitrometric scanning was performed at a wavelength of 400 nm, and lipids were quantified with ergosterol, oleic acid, or triolein as standard.

For estimation of total amounts of glycerophospholipids separate bands from non-polar lipid analysis (see above) were used. Glycerophospholipids were visualized on plates by reversible staining with iodine vapor, scraped off, and subjected to quantification by the method of Broekhyuse [71].

Analysis of total fatty acids was achieved by conversion to methyl esters by methanolysis using 2.5 % sulfuric acid in methanol and heating at 85 °C for 90 min. FAMEs (fatty acid methyl ester) were extracted twice in a mixture of light petroleum and water (3:1; v/v.) and subjected to gas liquid chromatography (Hewlett-Packard 6890 Gas-chromatograph) using an HP-INNOWax capillary column (15 m × 0.25 mm i.d. × 0.50 μm film thickness) with helium as carrier gas. Fatty acids were identified by comparison to commercially available fatty acid methyl ester standard mix GLC-68B (NuCheck, Inc., Elysian, MN, USA) and quantified by using pentadecanoic acid (Sigma) as an internal standard.

### Microarrays and data analysis

The RNA was isolated from chemostat sample cells using TRI reagent according to the supplier's instructions (Ambion, USA). RNA integrity was analyzed using RNA nano chips (Agilent). In-house designed *P. pastoris*-specific oligonucleotide arrays (AMAD-ID: 034821, 8x15K custom arrays, Agilent) were used [20, 32]. cRNA synthesis, hybridization, and scanning were performed according to the Agilent protocol for two-color expression arrays. Each sample was hybridized against an RNA reference pool sample in dye swap. The microarray data was not background normalized. Normalization steps and statistical analysis of microarray data included removal of color bias using locally weighted MA-scatterplot smoothing (LOESS) followed by between array normalization using the "Aquantile" method. The *P* values associated with the differential expression values were calculated using a linear model fit (limma R package), subsequently they were adjusted for multiple testing using the method of Benjamini

and Yekutieli [72] using the BY method of limma R package. To identify differentially expressed genes, the following criteria were applied: fold change cut-off of at least 1.5 > FC >1/1.5 and adjusted *P* value <0.05. All steps were performed using the R software package [73], and the limma package. Transcriptomics data were deposited at Gene Expression Omnibus with the accession number GSE67690. Data can be accessed with following link http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=stopswyszunfkf&acc=GSE67690.

## Proteomics

### *Cell lysis and sample preparation*

Cells were lysed with glass beads as described by Dragosits et al. [24] in 100 mM triethylammoniumbicarbonate (TEAB) buffer, containing 30 mM tris(2-carboxyethyl)phosphine hydrochloride and 2 % SDS. After incubation for 45 min at 56 °C (to reduce cysteine bridges) cellular proteins were extracted with chloroform/methanol, dried, dissolved in TEAB buffer, and digested with trypsin. Tandem Mass Tag (Thermo Scientific) labelling was performed as described by Pichler et al. [74] following the manufacturer's protocol.

### *2D-LC and MS analysis*

Samples were separated by high pH C18 HPLC applying an elution gradient of 12.5–80 % acetonitrile at pH 10 (200 mM ammonium formiate). Eighteen fractions were collected, partially pooled and applied to a C18 nanocolumn on a Bruker maxis 4G ETD QTOF LC-MS instrument, and separated with a 5–32 % acetonitrile gradient with 0.1 % formic acid (followed by a 32–80 % gradient to elute large peptides). The mass spectrometer was equipped with the captive spray source (1350 V capillary voltage, 3 L/min dry gas). Mass spectrometry scans were recorded in DDA mode (range: 150–2200 Da) and the 10 highest peaks were selected for fragmentation. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD002036.

## Peptide/protein identification

The software Mascot was used for the identification of peptides and proteins by matching the observed spectra with a database containing unique *P. pastoris* protein sequences. Mascot uses the MOWSE (MOlecular Weight SEarch) score: the more matches, the higher the peptide score. Protein scores are the sum of the peptide scores. Protein identification requires the match of at least two independent peptides with a score of >25.

## Data processing

For quantitative analysis of the proteomics data, the software Isobar Version 1.7.5 was used [75]. Mascot identification and quantification data were normalized using Isobar's default normalization method, which corrects for differences in reporter channel median intensities. Intensity measurement noise was corrected with a noise model comparing identical samples in multiple channels. For obtaining the protein ratios, Isobar calculates a weighted average of the peptide spectra after eliminating outliers. Comparing different distributions showed that a t-distribution fitted the random protein ratio distribution of our data best, and was selected for *P* value calculation of differentially expressed proteins.

Three biological replicates, with two technical replicates each, had been analyzed leading to six replicate data sets of both growth conditions. For every identified peptide, Isobar calculated the $\log_2$ of the ratio between the methanol/glycerol samples and the glucose samples ($\log_2$ FC) and the *P* value. Peptides with ion intensity values smaller then 300 and protein ratios deriving from single peptide spectra were excluded from the analysis, as well as proteins that were identified only in one or two replicates; 1,066 proteins fulfilled those criteria.

Proteins meeting the following criteria were defined as significantly changed between growth conditions: |mean FC| >1.5 and *P* values <0.1 or 1.3 < |mean FC| ≤1.5 and *P* value <0.05. Proteins with |mean FC| ≤1.3 and *P* value >0.05 were defined as not changed between the two growth conditions. To further increase stringency of evaluation we defined that >50 % of the replicates in which a given protein could be identified must have the same regulatory characteristics. Proteins that did not fulfill these criteria were not further considered. From the 1,066 proteins identified, 575 could be quantitatively evaluated.

## Metabolomics

### *Extraction and measurement of intracellular metabolites*

For the measurement of intracellular concentrations of free metabolites quenched cells on cellulose acetate filters were used. Prior to the extraction, uniformly labelled $^{13}C$ internal standard was added to the samples. Free intracellular metabolites were extracted by addition of 4 mL boiling HPLC grade ethanol (82 %; v/v; tempered at 85 °C). After addition of the boiling ethanol the quenched cells were immediately suspended by vortexing for 30 s. Suspended cells were heated for 3 min in total at 85 °C using a water bath. After 1.5 min of extraction samples were vortexed for 10 s and put back to the water bath at 85 °C. After 3 min of extractions extracted cells were immediately cooled down on dry ice. The cooled sample was then centrifuged to remove cell debris (10 min, −20 °C, 4000 *g*). The ethanolic extract was decanted into a fresh cooled

15 mL tube and kept on dry ice until sample preparation for LC-MS/MS and GC-MS/MS analysis. LC-MS/MS analysis of free intracellular metabolites was performed according to Klavins et al. [76], whereas GC-MS/MS analysis of sugar phosphates was performed after automated derivatization via ethoximation followed by trimethylsilylation. Both methods employed quantification by external calibration utilizing a uniformly $^{13}$C-labeled ethanolic extract of *P. pastoris* for internal standardization [77].

#### *Detection of sedoheptulose-1,7-bisphosphate in cell extracts of P. pastoris*

Acetonitrile, water, and formic acid (all LC-MS grade) were purchased at Sigma-Aldrich. *P. pastoris* cells were grown in glucose- or methanol-limited conditions in bioreactors. A set of three samples from glucose-fed cellular extracts was compared to a set of three methanol-grown samples. Each sample was derived from a separate biological replicate. After extraction (see above) the samples were stored at −80 °C until analysis. 500 μL of the sample were evaporated to dryness using a Savant SPD 121P SpeedVac Concentrator (Thermo Scientific). The residues were reconstituted in 100 μL water and directly analyzed.

Liquid chromatography separation was performed on a Hypercarb 150 × 2.1 mm, 3 μm particle size column (Thermo Scientific) with a Hypercarb guard cartridge (10 × 2.1 mm, 3 μm) using a 1260 BinPumpSL (Agilent Technologies) combined with a CTC Pal autosampler (CTC Analytics AG). The flow rate was 250 μL min$^{-1}$ and the column oven was set to 40 °C. Sample injection volume was 5 μL. Mobile phase A was 100 % water, whereas mobile phase B contained 80 % acetonitrile, 10 % water, and 10 % formic acid. A gradient was applied as follows: starting conditions of 1 % B were held for 2.5 min and then increased to 40 % within 14 min. This composition was held for 1 min, before returning to 1 % B in 0.1 min for re-equilibration. The total analysis time was 20 min.

An Agilent 6220 LC-TOFMS system equipped with a dual-ESI-Source was used for the LC-MS analysis. Source parameters for negative mode were set as follows: 350 °C gas temperature, 10 L min$^{-1}$ drying gas flow, 25 psig nebulizer gas pressure, 3500 V capillary voltage, 140 V fragmentor voltage, and 60 V skimmer voltage. The mass spectrometer was operated in the 2 GHz mode (extended dynamic range) recording the mass range from 50 to 1000 m/z with an acquisition rate of 1.03 spectra s-1 (9644 transients per spectrum). Data evaluation was performed using the Agilent MassHunter Qualitative Analysis B.07.00.

The identification of sedoheptulose-1,7-bisphosphate was confirmed by comparing the signals obtained in the samples to a standard which was provided by Amy A. Caudy (University of Toronto, Canada). The difference between the measured accurate mass and the calculated exact mass was below 2 ppm for all samples where sedoheptulose-1,7-bisphosphate was detected.

### $^{13}$C-Metabolic flux analysis

$^{13}$C-labelling experiments were performed as described in Baumann et al. [21]. The cells grew in a chemostat at a constant growth rate of 0.1 h$^{-1}$ on a mixture of 20 % fully $^{13}$C-labelled substrate and 80 % naturally labelled substrate, either glucose or methanol/glycerol. The labelling pattern of protein-bound amino acids was determined via GC-MS. The GC-MS spectra were used for the calculation of mass distribution vectors of the protein bound amino acids [78]. The Matlab-based software package Openflux was used for $^{13}$C-Metabolic flux analysis. For the calculation standard settings were applied [79]. The flux calculation was performed with a stoichiometric model of *P. pastoris* central carbon metabolism. The model is analogous to the model already published by Jorda et al. [28]. As a constraint, the labelling pattern of protein-bound amino acids and the calculated uptake and segregation rates of extracellular metabolites were used.

### Isolation and proteome characterization of the peroxisomes

*P. pastoris* cells were cultivated on glucose (YPD) or methanol (YPM) until they reached the late logarithmic growth phase. Cellular fractions enriched of highly pure peroxisomes were isolated following the procedure which had previously been established for *P. pastoris* by Wriessnegger et al. [44]. Isolated peroxisomes from methanol- and glucose-grown cells and the respective homogenates were evaluated for specific marker protein enrichment by Western blots (Additional file 3) and subjected to proteomics identification.

Samples were analyzed with a nano LC system as described above in 2D-LC and MS analysis. A standard 180 min gradient, using 0.1 % formic acid and 80 % acetonitrile as solvents, was applied. Data interpretation was performed manually (quantification) using DataAnalysis 4.0 and the files were converted to XML files for protein identification. XML files are suitable for performing a MS/MS ion search with ProteinScape (Bruker software, MASCOT embedded). At least two peptides and a MASCOT score of 30 were minimum thresholds for a positive hit. For quantification, the extracted ion chromatograms of the most intense peptides of each protein were integrated and peak areas were calculated. The sum of peak areas of each protein quantified of peroxisomal preparations was set in relation to the sum of peak areas of homogenate samples. Thus, a value higher than 1 reflects a relative enrichment in comparison to Aox1 in the peroxisomal preparation and a value lower than 1 shows a lower abundance in the peroxisomal

fraction in comparison to Aox1. Peptide and protein hits, and the peak areas of peptides used for quantification are provided in Additional file 4. The mass spectrometry proteomics data were deposited in the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD002831.

### Electron microscopy

Cells were cultivated at 25 °C with shaking at 150 rpm in baffled flasks using YPD until reaching the late exponential phase. Washed cells were fixed for 5 min in a 1 % aqueous solution of $KMnO_4$ at room temperature, washed with double distilled water, and fixed again in a 1 % aqueous solution of $KMnO_4$ for 20 min. Fixed cells were washed four times in distilled water and incubated in 0.5 % aqueous uranyl acetate overnight at 4 °C. Samples were then dehydrated for 20 min, in a graded series of 50 %, 70 %, 90 %, and 100 % ethanol, each. Pure ethanol was then changed to propylene oxide and specimens were gradually infiltrated with increasing concentrations (30 %, 50 %, 70 %s and 100 %) of Agar 100 epoxy resin mixed with propylene oxide for a minimum of 3 h per step. Samples were embedded in pure, fresh Agar 100 epoxy resin and polymerized at 60 °C for 48 h. Ultrathin 80-nm sections were stained for 3 min with lead citrate and viewed with a Philips CM 10 transmission electron microscope.

## Additional files

**Additional file 1: Transcriptomic, proteomic, and metabolomic regulation of *P. pastoris* during methylotrophic growth.** Containing the following eight sheets: Summary Omics Data: number of significantly regulated genes, proteins or metabolites (e.g. "up" refers to up-regulation in methanol/glycerol compared to glucose). Transcriptomics and proteomics: Average fold changes and *P* values of transcriptomics and proteomics comparing *P. pastoris* cultivated with methanol/glycerol or glucose as carbon source in chemostat. Average values derive from three biological replicates per condition. Metabolomics: Average fold changes and *P* values of metabolomics measurements comparing *P. pastoris* cultivated with methanol/glycerol or glucose as carbon source in chemostat cultivations. Average values derive from three biological replicates per condition. Co-regulation (related to Fig. 1 in the text): Regulation of the 575 gene-protein pairs with transcriptomics and proteomics data available and assignment to regulatory groups. Central carbon metabolism (related to Fig. 4 in the text): Average fold changes and *P* values of transcriptomics, proteomics, and metabolomics measurement depicted in Fig. 4. Amino acid metabolism (related to Fig. 6 in the text): Average fold changes and *P* values of transcriptomics, proteomics, and metabolomics measurement depicted in Fig. 6. Vitamin biosynthesis (related to Fig. 7 in the text): Average fold changes and *P* values of transcriptomics, proteomics, and metabolomics measurement depicted in Fig. 7. Peroxisomal gene regulation: Average fold changes and *P* values of transcriptomics and proteomics for all mentioned peroxisomal genes. Average values derive from three biological replicates per condition. (XLSX 2348 kb)

**Additional file 2: Comparison of gene regulation in *P. pastoris* cultivated with methanol/glycerol or glucose as carbon source in chemostat to transcriptomics data obtained in fed batch cultivation with methanol glycerol or glucose as carbon source.** Average fold changes and *P* values of regulation patterns for methanol vs glycerol and methanol vs glucose 1 h our after starting the methanol feed are shown. Average values derive from three biological replicates per condition. Containing the following four sheets: Description: Experimental setup of the fed batch cultivations, description, and discussion of observed similarities and differences in gene regulations. Upregulated_methanol: Comparison of all genes upregulated in *P. pastoris* cultivated with methanol/glycerol or glucose as carbon source in chemostat to transcriptomics data obtained by analyzing methanol vs glycerol and methanol vs glucose fed batches. Downregulated_methanol: Comparison of all genes downregulated in *P. pastoris* cultivated with methanol/glycerol or glucose as carbon source in chemostat to transcriptomics data obtained by analyzing methanol vs glycerol and methanol vs glucose fed batches. All_data: Transcriptomics data of Additional file 1 (*P. pastoris* cultivated with methanol/glycerol or glucose as carbon source in chemostat) compared to transcriptomics data obtained by analyzing methanol vs glycerol and methanol vs glucose fed batches. (XLSX 1178 kb)

**Additional file 3: Enrichment of the peroxisomal marker protein Pex3p in the peroxisomal fraction.** (PDF 271 kb)

**Additional file 4: Proteomic identification and quantification of methanol metabolic enzymes and control proteins in peroxisomal fractions and homogenates of *P. pastoris* cells grown on methanol.** Containing the following three sheets: Protein hits: contains all identified proteins that met the threshold in at least one sample, with their respective MASCOT scores, number of peptides, and percent sequence coverage. Peptide hits: list of all identified peptides, their MASCOT scores, mass and charge values, and intensities. Peptides used for quant + areas: lists all peptides of the proteins in Table 3 that were used for quantification, and their respective peak areas in the different samples. (XLSX 879 kb)

#### Abbreviations

AOX: Alcohol oxidase; CDW: Cell dry weight; DAK: Dihydroxyacetone kinase; DAS: Dihydroxyacetone synthase; DHAP: Dihydroxyacetone phosphate; FAD: Flavin adenine dinucleotide; Fba1-2: Fructose-1,6-bisphosphate aldolase; FMN: Flavin mononucleotide; GAP: Glyceraldehyde-3-phosphate; $H_2O_2$: Hydrogen peroxide; PPP: Pentose phosphate pathway; PTS: Peroxisomal targeting signal sequence; Rki1-2: Ribose-5-phosphate ketol-isomerase; Rpe1-2: Ribulose-5-phosphate 3-epimerase; S1,7BP: Sedoheptulose-1,7-bisphosphate; Shb17: Sedoheptulose-1,7-bisphosphatase; Tal1-2: Transaldolase; TAG: Triacylglycerols; TPP: Thiamine pyrophosphate; UPR: Unfolded protein response; XYL5P: Xylulose-5-phosphate.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

HR and MB performed all cultivations, sampling, and transcriptomics analyses. CG measured quantitative proteomes, supervised by FA. MV coordinated data management and evaluation, and analyzed proteomics data together with GM and CG. KHG performed lipid analysis and purification of peroxisomes. RG, KK, SN, and AC measured the metabolomes under the guidance of GK, SH, and DS. CT measured $^{13}C$ labelling patterns for metabolic flux analysis which was calculated by HR. HD analyzed S1,7BP with support by SH. GZ provided electron microscopy. ABG supervised and participated in evaluation of transcriptomics and proteomics data. DM and BG designed this study with support of MS, MSt, and GD. Systems level data interpretation was performed by HR, MB, MV, KHG, GD, MSt, MS, DM, and BG. HR, MB, MV, and KHG drafted the manuscript, which was revised by BG and DM. All authors read and approved the final manuscript.

#### Acknowledgements

This work has been supported by the Federal Ministry of Science, Research and Economy (BMWFW), the Federal Ministry of Traffic, Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol, the Government of Lower Austria and ZIT – Technology Agency of the City of Vienna through the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG. Further support by Biomin Research Center, Boehringer-Ingelheim RCV, Lonza AG, Biocrates Life Sciences AG, VTU Technology GmbH, and Sandoz GmbH is

acknowledged. EQ BOKU VIBT GmbH is acknowledged for providing mass spectrometry and bioreactor instrumentation. The authors are grateful to Amy A. Caudy (University of Toronto, Canada) for kindly providing a standard of seduheptulose bisphosphate and thank Jonas Hohlweg for excellent support in transcriptomics analyses.

**Author details**
[1]Department of Biotechnology, BOKU - University of Natural Resources and Life Sciences Vienna, Muthgasse 18, 1190 Vienna, Austria. [2]Austrian Centre of Industrial Biotechnology, A-1190 Vienna, Austria. [3]Department of Chemistry, BOKU - University of Natural Resources and Life Sciences Vienna, A-1190 Vienna, Austria. [4]Institute of Biochemistry, Graz University of Technology, A-8010 Graz, Austria. [5]Austrian Centre of Industrial Biotechnology, A-8010 Graz, Austria. [6]School of Bioengineering, University of Applied Sciences FH Campus, A-1190 Vienna, Austria. [7]BIOCRATES Life Sciences AG, A-6020 Innsbruck, Austria. [8]Institute of Plant Sciences, NAWI Graz, University of Graz, A-8010 Graz, Austria. [9]Institute of Analytical Chemistry, University of Vienna, A-1090 Vienna, Austria. [10]Present addresses: Sandoz GmbH, A-6250 Kundl, Austria. [11]Present addresses: BIOCRATES Life Sciences AG, A-6020 Innsbruck, Austria. [12]University of Tübingen, D-72076 Tübingen, Germany.

Received: 2 September 2015 Accepted: 3 September 2015
Published online: 23 September 2015

**References**

1. Anthony C. The biochemistry of methylotrophs. London: Academic; 1982.
2. Meehl MA, Stadheim TA. Biopharmaceutical discovery and production in yeast. Curr Opin Biotechnol. 2014;30:120–7.
3. Liu L, Yang H, Shin HD, Chen RR, Li J, Du G, et al. How to achieve high-level expression of microbial enzymes: strategies and perspectives. Bioengineered. 2013;4:212–23.
4. Ma C, Agrawal G, Subramani S. Peroxisome assembly: matrix and membrane protein biogenesis. J Cell Biol. 2011;193:7–16.
5. Losev E, Reinke CA, Jellen J, Strongin DE, Bevis BJ, Glick BS. Golgi maturation visualized in living yeast. Nature. 2006;441:1002–6.
6. Vogl T, Glieder A. Regulation of *Pichia pastoris* promoters and its consequences for protein production. N Biotechnol. 2013;30:385–404.
7. Kohlwein SD, Veenhuis M, van der Klei IJ. Lipid droplets and peroxisomes: key players in cellular lipid homeostasis or a matter of fat–store 'em up or burn'em down. Genetics. 2013;193:1–50.
8. van der Klei IJ, Veenhuis M. Yeast peroxisomes: function and biogenesis of a versatile cell organelle. Trends Microbiol. 1997;5:502–9.
9. van der Klei IJ, Yurimoto H, Sakai Y, Veenhuis M. The significance of peroxisomes in methanol metabolism in methylotrophic yeast. Biochim Biophys Acta. 1763;2006:1453–62.
10. van Zutphen T, Baerends RJ, Susanna KA, de Jong A, Kuipers OP, Veenhuis M, et al. Adaptation of *Hansenula polymorpha* to methanol: a transcriptome analysis. BMC Genomics. 2010;11:1.
11. Nagotu S, Krikken AM, Otzen M, Kiel JA, Veenhuis M, van der Klei IJ. Peroxisome fission in *Hansenula polymorpha* requires Mdv1 and Fis1, two proteins also involved in mitochondrial fission. Traffic. 2008;9:1471–84.
12. Hartner FS, Glieder A. Regulation of methanol utilisation pathway genes in yeasts. Microb Cell Fact. 2006;5:39.
13. Yurimoto H, Oku M, Sakai Y. Yeast methylotrophy: metabolism, gene regulation and peroxisome homeostasis. Int J Microbiol. 2011;2011:101298.
14. Vanz AL, Lunsdorf H, Adnan A, Nimtz M, Gurramkonda C, Khanna N, et al. Physiological response of *Pichia pastoris* GS115 to methanol-induced high level production of the Hepatitis B surface antigen: catabolic adaptation, stress responses, and autophagic processes. Microb Cell Fact. 2012;11:103.
15. Vanz AL, Nimtz M, Rinas U. Decrease of UPR- and ERAD-related proteins in *Pichia pastoris* during methanol-induced secretory insulin precursor production in controlled fed-batch cultures. Microb Cell Fact. 2014;13:23.
16. Liang S, Wang B, Pan L, Ye Y, He M, Han S, et al. Comprehensive structural annotation of *Pichia pastoris* transcriptome and the response to various carbon sources using deep paired-end RNA sequencing. BMC Genomics. 2012;13:738.
17. Sauer M, Branduardi P, Gasser B, Valli M, Maurer M, Porro D, et al. Differential gene expression in recombinant *Pichia pastoris* analysed by heterologous DNA microarray hybridisation. Microb Cell Fact. 2004;3:17.
18. Jorda J, Rojas HC, Carnicer M, Wahl A, Ferrer P, Albiol J. Quantitative metabolomics and instationary $^{13}$C-metabolic flux analysis reveals impact of recombinant protein production on Trehalose and energy metabolism in *Pichia pastoris*. Metabolites. 2014;4:281–99.
19. Rebnegger C, Graf AB, Valli M, Steiger MG, Gasser B, Maurer M, et al. In *Pichia pastoris*, growth rate regulates protein synthesis and secretion, mating and stress response. Biotechnol J. 2014;9:511–25.
20. Graf A, Gasser B, Dragosits M, Sauer M, Leparc G, Tuechler T, et al. Novel insights into the unfolded protein response using *Pichia pastoris* specific DNA microarrays. BMC Genomics. 2008;9:390.
21. Baumann K, Carnicer M, Dragosits M, Graf AB, Stadlmann J, Jouhten P, et al. A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions. BMC Syst Biol. 2010;4:141.
22. Dragosits M, Stadlmann J, Graf A, Gasser B, Maurer M, Sauer M, et al. The response to unfolded protein is involved in osmotolerance of *Pichia pastoris*. BMC Genomics. 2010;11:207.
23. Hesketh AR, Castrillo JI, Sawyer T, Archer DB, Oliver SG. Investigating the physiological response of *Pichia* (*Komagataella*) *pastoris* GS115 to the heterologous expression of misfolded proteins using chemostat cultures. Appl Microbiol Biotechnol. 2013;97:9747–62.
24. Dragosits M, Stadlmann J, Albiol J, Baumann K, Maurer M, Gasser B, et al. The effect of temperature on the proteome of recombinant *Pichia pastoris*. J Proteome Res. 2009;8:1380–92.
25. Lin XQ, Liang SL, Han SY, Zheng SP, Ye YR, Lin Y. Quantitative iTRAQ LC-MS/MS proteomics reveals the cellular response to heterologous protein overexpression and the regulation of *HAC1* in *Pichia pastoris*. J Proteomics. 2013;91:58–72.
26. Chung BK, Selvarasu S, Andrea C, Ryu J, Lee H, Ahn J, et al. Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. Microb Cell Fact. 2010;9:50.
27. Sohn SB, Graf AB, Kim TY, Gasser B, Maurer M, Ferrer P, et al. Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production. Biotechnol J. 2010;5:705–15.
28. Jorda J, Jouhten P, Camara E, Maaheimo H, Albiol J, Ferrer P. Metabolic flux profiling of recombinant protein secreting *Pichia pastoris* growing on glucose:methanol mixtures. Microb Cell Fact. 2012;11:57.
29. Jorda J, Suarez C, Carnicer M, ten Pierick A, Heijnen JJ, van Gulik W, et al. Glucose-methanol co-utilization in *Pichia pastoris* studied by metabolomics and in stationary $^{13}$C flux analysis. BMC Syst Biol. 2013;7:17.
30. Jorda J, de Jesus SS, Peltier S, Ferrer P, Albiol J. Metabolic flux analysis of recombinant *Pichia pastoris* growing on different glycerol/methanol mixtures by iterative fitting of NMR-derived $^{13}$C-labelling data from proteinogenic amino acids. N Biotechnol. 2014;31:120–32.
31. 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;153:281–90.
32. Prielhofer R, Maurer M, Klein J, Wenger J, Kiziak C, Gasser B, et al. Induction without methanol: novel regulated promoters enable high-level expression in *Pichia pastoris*. Microb Cell Fact. 2013;12(1):5.
33. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2012;13:227–32.
34. Lu P, Vogel C, Wang R, Yao X, Marcotte EM. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol. 2007;25:117–24.
35. Lee MV, Topper SE, Hubler SL, Hose J, Wenger CD, Coon JJ, et al. A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. Mol Syst Biol. 2011;7:514.
36. Kiel JA, Veenhuis M, van der Klei IJ. PEX genes in fungal genomes: common, rare or redundant. Traffic. 2006;7:1291–303.
37. Prielhofer R, Cartwright SP, Graf AB, Valli M, Bill RM, Mattanovich D, et al. *Pichia pastoris* regulates its gene-specific response to different carbon sources at the transcriptional, rather than the translational, level. BMC Genomics. 2015;16:167.
38. Terlecky SR, Nuttley WM, McCollum D, Sock E, Subramani S. The *Pichia pastoris* peroxisomal protein PAS8p is the receptor for the C-terminal tripeptide peroxisomal targeting signal. EMBO J. 1995;14:3627–34.
39. Faber KN, Haima P, Gietl C, Harder W, Ab G, Veenhuis M. The methylotrophic yeast *Hansenula polymorpha* contains an inducible import pathway for peroxisomal matrix proteins with an N-terminal targeting signal (PTS2 proteins). Proc Natl Acad Sci U S A. 1994;91:12985–9.

40. Elgersma Y, Elgersma-Hooisma M, Wenzel T, McCaffery JM, Farquhar MG, Subramani S. A mobile PTS2 receptor for peroxisomal protein import in *Pichia pastoris*. J Cell Biol. 1998;140:807–20.
41. Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, et al. High-quality genome sequence of *Pichia pastoris* CBS7435. J Biotechnol. 2011;154:312–20.
42. Neuberger G, Maurer-Stroh S, Eisenhaber B, Hartig A, Eisenhaber F. Prediction of peroxisomal targeting signal 1 containing proteins from amino acid sequence. J Mol Biol. 2003;328:581–92.
43. PTS1 Predictor. http://mendel.imp.ac.at/mendeljsp/sat/pts1/PTS1predictor.jsp. Last access April 28, 2015.
44. Wriessnegger T, Gübitz G, Leitner E, Ingolic E, Cregg J, de la Cruz B, et al. Lipid composition of peroxisomes from the yeast *Pichia pastoris* grown on different carbon sources. Biochim Biophys Acta. 1771;2007:455–61.
45. Luers GH, Advani R, Wenzel T, Subramani S. The *Pichia pastoris* dihydroxyacetone kinase is a PTS1-containing, but cytosolic, protein that is essential for growth on methanol. Yeast. 1998;14:759–71.
46. Clasquin MF, Melamud E, Singer A, Gooding JR, Xu X, Dong A, et al. Riboneogenesis in yeast. Cell. 2011;145:969–80.
47. Douma AC, Veenhuis M, de Koning W, Evers M, Harder W. Dihydroxyacetone synthase is localized in the peroxisomal matrix of methanol-grown *Hansenula polymorpha*. Arch Microbiol. 1985;143:237–43.
48. Anderson LE, Carol AA. Enzyme co-localization with rubisco in pea leaf chloroplasts. Photosynth Res. 2004;82:49–58.
49. Rae BD, Long BM, Whitehead LF, Forster B, Badger MR, Price GD. Cyanobacterial carboxysomes: microcompartments that facilitate $CO_2$ fixation. J Mol Microbiol Biotechnol. 2013;23:300–7.
50. Hahn MW. Distinguishing among evolutionary models for the maintenance of gene duplicates. J Hered. 2009;100:605–17.
51. Zhang J. Evolution by gene duplication: an update. Trends Ecol Evol. 2003;18:292–8.
52. Byun-McKay SA, Geeta R. Protein subcellular relocalization: a new perspective on the origin of novel genes. Trends Ecol Evol. 2007;22:338–44.
53. Daran-Lapujade P, Rossell S, van Gulik WM, Luttik MA, de Groot MJ, Slijper M, et al. The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. Proc Natl Acad Sci U S A. 2007;104:15753–8.
54. Tibbetts AS, Sun Y, Lyon NA, Ghrist AC, Trotter PJ. Yeast mitochondrial oxodicarboxylate transporters are important for growth on oleic acid. Arch Biochem Biophys. 2002;406:96–104.
55. Cavero S, Vozza A, del Arco A, Palmieri L, Villa A, Blanco E, et al. Identification and metabolic role of the mitochondrial aspartate-glutamate transporter in *Saccharomyces cerevisiae*. Mol Microbiol. 2003;50:1257–69.
56. Kunze M, Pracharoenwattana I, Smith SM, Hartig A. A central role for the peroxisomal membrane in glyoxylate cycle function. Biochim Biophys Acta. 1763;2006:1441–52.
57. Warner JR, Mitra G, Schwindinger WF, Studeny M, Fried HM. *Saccharomyces cerevisiae* coordinates accumulation of yeast ribosomal proteins by modulating mRNA splicing, translational initiation, and protein turnover. Mol Cell Biol. 1985;5:1512–21.
58. Puxbaum V, Mattanovich D, Gasser B. Quo vadis? The challenges of recombinant protein folding and secretion in *Pichia pastoris*. Appl Microbiol Biotechnol. 2015;99:2925–38.
59. 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;85:367–75.
60. 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.
61. Ozimek P, van Dijk R, Latchev K, Gancedo C, Wang DY, van der Klei IJ, et al. Pyruvate carboxylase is an essential protein in the assembly of yeast peroxisomal oligomeric alcohol oxidase. Mol Biol Cell. 2003;14:786–97.
62. Stewart MQ, Esposito RD, Gowani J, Goodman JM. Alcohol oxidase and dihydroxyacetone synthase, the abundant peroxisomal proteins of methylotrophic yeasts, assemble in different cellular compartments. J Cell Sci. 2001;114:2863–8.
63. Waterham HR, Russell KA, Vries Y, Cregg JM. Peroxisomal targeting, import, and assembly of alcohol oxidase in *Pichia pastoris*. J Cell Biol. 1997;139:1419–31.
64. Gudipati V, Koch K, Lienhart WD, Macheroux P. The flavoproteome of the yeast *Saccharomyces cerevisiae*. Biochim Biophys Acta. 1844;2014:535–44.
65. Marx H, Mattanovich D, Sauer M. Overexpression of the riboflavin biosynthetic pathway in *Pichia pastoris*. Microb Cell Fact. 2008;7:23.
66. Stadlmayr G, Mecklenbrauker A, Rothmuller M, Maurer M, Sauer M, Mattanovich D, et al. Identification and characterisation of novel *Pichia pastoris* promoters for heterologous protein production. J Biotechnol. 2010;150:519–29.
67. Sporty J, Lin SJ, Kato M, Ognibene T, Stewart B, Turteltaub K, et al. Quantitation of $NAD^+$ biosynthesis from the salvage pathway in *Saccharomyces cerevisiae*. Yeast. 2009;26:363–9.
68. Wriessnegger T, Sunga AJ, Cregg JM, Daum G. Identification of phosphatidylserine decarboxylases 1 and 2 from *Pichia pastoris*. FEMS Yeast Res. 2009;9:911–22.
69. Verduyn C, Postma E, Scheffers WA, van Dijken JP. Physiology of *Saccharomyces cerevisiae* in anaerobic glucose-limited chemostat cultures. J Gen Microbiol. 1990;136:395–403.
70. Folch J, Lees M, Sloane Stanley GH. A simple method for the isolation and purification of total lipides from animal tissues. J Biol Chem. 1957;226:497–509.
71. Broekhuyse RM. Phospholipids in tissues of the eye. I. Isolation, characterization and quantitative analysis by two-dimensional thin-layer chromatography of diacyl and vinyl-ether phospholipids. Biochim Biophys Acta. 1968;152:307–15.
72. Benjamini Y, Heller R, Yekutieli D. Selective inference in complex research. Philos Trans A Math Phys Eng Sci. 2009;367:4255–71.
73. R project. http://www.r-project.org. Last access January 31, 2012.
74. Pichler P, Kocher T, Holzmann J, Mazanek M, Taus T, Ammerer G, et al. Peptide labeling with isobaric tags yields higher identification rates using iTRAQ 4-plex compared to TMT 6-plex and iTRAQ 8-plex on LTQ Orbitrap. Anal Chem. 2010;82:6549–58.
75. Breitwieser FP, Muller A, Dayon L, Kocher T, Hainard A, Pichler P, et al. General statistical modeling of data from protein relative expression isobaric tags. J Proteome Res. 2011;10:2758–66.
76. Klavins K, Neubauer S, Al Chalabi A, Sonntag D, Haberhauer-Troyer C, Russmayer H, et al. Interlaboratory comparison for quantitative primary metabolite profiling in *Pichia pastoris*. Anal Bioanal Chem. 2013;405:5159–69.
77. Neubauer S, Haberhauer-Troyer C, Klavins K, Russmayer H, Steiger MG, Gasser B, et al. $U^{13}C$ cell extract of *Pichia pastoris*–a powerful tool for evaluation of sample preparation in metabolomics. J Sep Sci. 2012;35:3091–105.
78. Zamboni N, Fischer E, Sauer U. FiatFlux–a software for metabolic flux analysis from $^{13}C$-glucose experiments. BMC Bioinformatics. 2005;6:209.
79. Quek LE, Wittmann C, Nielsen LK, Kromer JO. OpenFLUX: efficient modelling software for $^{13}C$-based metabolic flux analysis. Microb Cell Fact. 2009;8:25.

## 4.4 Comparison of gene and protein regulations in various *P. pastoris* production strains with different carbon sources

Product titers in recombinant protein production with *P. pastoris* were often observed to vary tremendously, depending on the heterologous gene product (and its copy number), the promoter used for expression thereof, the strain/clone background, or the carbon source used for cultivation and/or induction of protein production.

In order to elucidate the implications of protein production conditions on the proteome of the cell, an experimental setup was created to compare *P. pastoris* production strains with non-producing control strains. In case of the recombinant product HyHEL Fab, this was done under control of the GAP promoter with glucose as carbon source but also under control of the AOX promoter with a glycerol/methanol medium for growth and induction of protein production. Figure 5 shows the different comparisons that were achieved by cultivating five different strains (three production strains, two control strains) in chemostats in three biological replicates in order to obtain samples for transcriptomic and proteomic analysis.

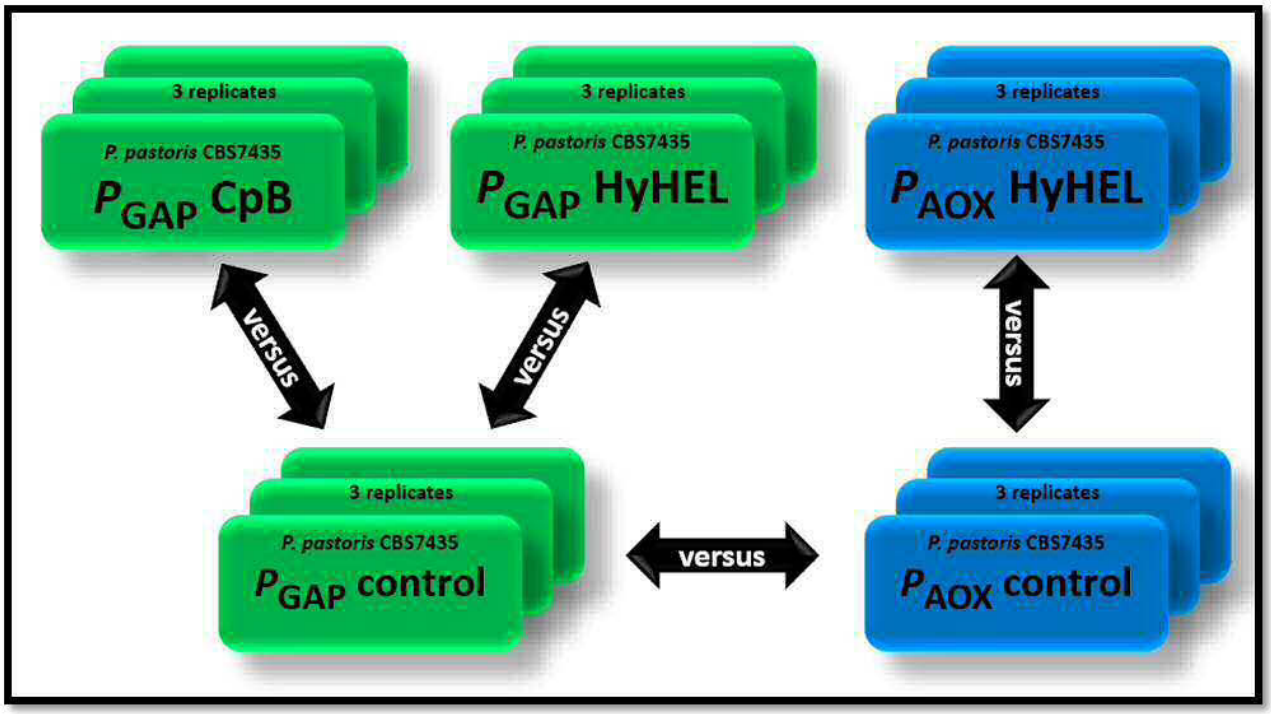


**Figure 5. Strain comparison experiments for the comparative analysis of the proteome and transcriptome of *P. pastoris* production strains.** The strains were cultivated in chemostats in three biological replicates each.

### 4.4.1 Methods and Materials

#### 4.4.1.1 Generation of production strains

Three different production strains were generated in *P. pastoris* CBS7435, for the secretion of three different protein products using two different promoter/induction strategies: AOX_HyHEL, GAP_HyHEL and GAP_CpB. Two strains, namely AOX and GAP, both without any gene of interest inserted in the respective expression vector, were used as controls for this study. HyHEL is a heterodimeric antibody Fab fragment, CpB is Carboxypeptidase B*. Pichia pastoris (Komagataella phaffii)* CBS7435 is the parental strain of commonly used *P. pastoris* recombinant protein production hosts. Its genome has been fully sequenced (Küberl et al. 2011; Valli et al. 2016).

Cloning and transformation was done using the in-house vector pPuzzle (Stadlmayr et al. 2010b), which contains a Zeocin resistance cassette for selection in both *E. coli* and yeast, an expression cassette for the gene of interest (GOI) consisting of a multiple cloning site and the *S. cerevisiae* CYC1 transcription terminator, and a locus for integration into the *P. pastoris* genome (3´ *AOX1* region). Vectors carrying the respective model protein (or a blank multiple cloning site for the control strains) under control of $P_{GAP}$ as well as $P_{AOX}$ were used throughout the study. For all model proteins (CpB and HyHEL Fab) the *S. cerevisiae* alpha mating factor signal sequence was used as secretion leader. Prior to transformation, the expression vector was linearized within the GAP promoter using the restriction enzyme *Bln*I for homologous integration into the native GAP promoter locus of the *P. pastoris* genome.

For the expression of heterodimeric HyHEL antibody Fab fragment (HyHEL Fab), the expression cassettes of light chain and Fab heavy chain (each under control of $P_{GAP}$ or $P_{AOX}$) were combined into one vector: The LC fragments were ligated into a variant of the vector backbone, where one restriction enzyme site in the promoter region was exchanged for another to allow subsequent linearization. The HC fragments were ligated into the unmodified versions of the vectors. Then, the expression cassettes for both chains were combined onto one vector. The previously exchanged restriction site now allowed for linearization of the vector without cutting in both AOX or GAP

promoter regions prior to electroporation.

Transformation into *P. pastoris* was done using a standard electroporation protocol as described in (Gasser et al. 2006). Selection of positive transformants was performed on YPD plates (per liter: 10 g yeast extract, 20 g peptone, 20 g glucose, 20 g agar-agar) containing 50 µg/mL of Zeocin.

#### 4.4.1.2 Chemostat Cultivation for the analysis of differential gene expression and protein abundance

The three *P. pastoris* CBS7435 AOX or GAP production strains as well as their control strains were cultivated in chemostats at a fixed specific growth rate of 0.1 $h^{-1}$. Constant and carbon-limited growth was applied to avoid growth rate dependent effects during genome-scale analyses. Heterologous gene expression under the control of the *AOX1* promoter is strong, and methanol-driven fed batch processes are often the first choice for protein production in *P. pastoris*. Pure methanol would overburden the cells at feed rates exceeding the maximum uptake rate but combined with e.g. glycerol, it serves as inducer of recombinant protein production while faster growth and therefore higher metabolic rates are achieved with the help of the co-substrate, in this case, glycerol. In order to cultivate *P. pastoris* at higher specific growth rates, mixed-substrate feed media are often used, especially for $mut^S$ phenotype strains (Cos et al. 2006). A methanol:glycerol mix of 8.5 g/L methanol and 49.0 g/L glycerol was employed based on pre-experiments with *P. pastoris* production strains proving that methanol is utilized entirely and the methylotrophic pathway is fully induced under these conditions.

Chemostats were run in three biological replicates per condition and samples for transcriptomic analysis were taken during steady state after 7 residence times. Substrate limitation of all cultures, i.e. no residual glucose or methanol/glycerol, respectively, was confirmed by HPLC. Transcriptional regulation was analyzed using *P. pastoris* specific DNA microarrays (Graf et al. 2008; Prielhofer et al. 2013).

The chemostat cultivations were performed in a 1.4 L bioreactor (DASGIP Parallel Bioreactor System, Germany) with a working volume of 400 mL. 100 mL pre-culture medium (per liter: 10 g yeast extract, 20 g peptone, 10 g glycerol) were inoculated with 750 µL cryostock of *P. pastoris* CBS7435 and grown at 28°C and 150 rpm

overnight. This culture was used for inoculation of the bioreactor at an optical density ($OD_{600}$) of 1.0. After a batch phase of approximately 24 hours, the cells were grown in carbon limited chemostats with a dilution rate of 0.1 $h^{-1}$ for at least 7 residence times before taking the samples. For each strain, three independent chemostat cultivations were performed. Temperature, pH and dissolved oxygen were maintained at 25°C, 5.0 (with 8M KOH) and 20% (by controlling the stirrer speed and inlet air), respectively.

Batch medium contained per liter: 39.9 g glycerol, 1.8 g citric acid, 12.6 g $(NH_4)_2HPO_4$, 0.022 g $CaCl_2.2H_2O$, 0.9 g KCl, 0.5 g $MgSO_4.7H2O$, 2 mL Biotin (0.2 g $L^{-1}$), 4.6 mL trace salts stock solution. The pH was set to 5.0 with 32% (w/w) HCl.

Chemostat medium contained per liter: 57 g glycerol (86%), 8.5 g methanol (100%) 2.3 g citric acid, 21.75 g $(NH_4)_2HPO_4$, 0.04 g $CaCl_2.2H_2O$, 2.5 g KCl, 1.0 g $MgSO_4.7H_2O$, 2 g Biotin (0.2 g $L^{-1}$), and 2.43 g trace salts stock solution. The pH was set to 5.0 with 32% (w/w) HCl.

Trace salts stock solution contained per liter: 6.0 g $CuSO_4.5H_2O$, 0.08 g NaI, 3.0 g $MnSO_4.H_2O$, 0.2 g $Na_2MoO_4.2H_2O$, 0.02 g $H_3BO_3$, 0.5 g $CoCl_2$, 20.0 g $ZnCl_2$, 5.0 g $FeSO_4.7H_2O$, and 5.0 mL $H_2SO_4$ (95–98% w/w).

#### 4.4.1.3 Transcriptomics workflow

For transcriptomics with microarray analysis, 9 mL of culture were added to 4.5 mL of freshly prepared pre-chilled (−20 °C) fixing solution (5 % v/v phenol in ethanol abs.), mixed, and 1.5 mL were aliquoted into ribolyzer tubes and centrifuged at 13,000 rpm for 1 min at 4 °C. The supernatant was discarded and the tubes containing the fixed cell pellets were immediately stored at −80 °C. For protein analysis, 2 mL of culture were centrifuged and the cell pellet was stored at −80 °C. The supernatant was also stored at −80 °C for analysis of extracellular metabolites.The RNA was isolated from chemostat sample cells using TRI reagent according to the supplier's instructions (Ambion, USA). RNA integrity was analyzed using RNA nano chips (Agilent). In-house designed *P. pastoris*-specific oligonucleotide arrays (AMAD-ID: 034821, 8x15K custom arrays, Agilent) were used (Graf et al. 2008; Prielhofer et al. 2013). cRNA synthesis, hybridization, and scanning were performed according to the Agilent protocol for two-color expression arrays. Each sample was hybridized against an

RNA reference pool sample in dye swap. The microarray data was not background normalized. Normalization steps and statistical analysis of microarray data included removal of color bias using locally weighted MA-scatterplot smoothing (LOESS) followed by between array normalization using the “Aquantile” method. The P values associated with the differential expression values were calculated using a linear model fit (limma R package), subsequently they were adjusted for multiple testing using the method of Benjamini and Yekutieli (Benjamini et al. 2009) using the BY method of limma R package. To identify differentially expressed genes, the following criteria were applied: fold change cut-off of at least $1.5 > FC > 1/1.5$ and adjusted P value $<0.05$. All steps were performed using the R software package, and the limma package.

#### 4.4.1.4 Proteomics workflow

For peptide/protein identification and quantification with LC-ESI-MS/MS, cell pellets were stored at -80°C. After mechanical cell lysis, protein isolation and proteolytic digestion with trypsin, proteins were labeled with Tandem Mass Tag for relative quantitation (TMT labeling). Detection was performed with a Q-TOF mass spectrometer equipped with the captive spray source (Bruker maXis 4G) in the positive ion, DDA mode (= switching to MSMS mode for eluting peaks). The LC-ESI-MS/MS method was developed within the Strategic program of ACIB at the CF Proteomics at BOKU-DCh (by Clemens Gruber in the lab of Friedrich Altmann)

Accurate peptide to protein inference requires a database containing unique protein sequences. Such a list was created by downloading all available and annotated protein entries of *P. pastoris* from Uniprot. By comparing the protein sequences annotated for the different *P. pastoris* strains, we could identify over 800 sequences showing differences between the strains and delete proteins homolog to proteins already present in the list as there were sequences either wrongly annotated or having a wrong ORF or intron prediction.

The software Mascot has been used for the identification of peptides and proteins by matching the observed spectra with a database. Mascot uses the MOWSE (MOlecular Weight SEarch) score - the more matches, the higher the peptide score. Protein scores are the sum of the peptide scores. Protein identification requires the

match of at least 2 independent peptides with a score of >25. The software Isobar has been used to process and normalize Mascot data.

Three biological replicates, with two technical replicates each, have been analyzed but data elaboration has been performed considering the six technical replicates independently. For every identified peptide and for every comparison the software Isobar calculated the $\log_2$ fold change ratio (FC) and the p-value. We excluded from the analysis peptides with ion intensity values < 300 and protein ratios deriving from single peptide spectra. Furthermore, we excluded proteins that could be identified in only 1 or 2 replicates. Finally, for every quantified protein we have a $\log_2$FC ($\log_2$ fold change) value and the regulatory characteristics (up, down or not changed).

The following criteria were used to define changes in the protein level: It was considered to be higher ("up"), when the fold change was bigger than 1.5 and p-values smaller than 0.1 or when the fold change was bigger than 1.3 with a corresponding p- value smaller than 0.05. It was considered to be lower ("down"), when the fold change was smaller than 0.67 and p-values smaller than 0.1 or when the fold change was smaller than 0.77 with a corresponding p- value smaller than 0.05. It was considered to have not changed in all other cases (fold change not strong enough in either direction and/or p-values indicating no statistical significance).

Additionally, to state the final change in the level of a defined protein, more than half of the replicates in which the protein could be identified must have the same regulatory characteristics. Proteins that do not fulfill these criteria have not been considered as significantly changed.

### 4.4.2 Results and Discussion

Comparative quantitative proteomics have been performed to analyze changes induced by the expression of heterologous proteins (CpB and HyHEL) or by the different carbon source (glycerol/methanol vs. glucose). As summarized in Table 3 we could identify more than 1,000 proteins in each of the 4 comparisons and 44-82% of the identified proteins could also be quantified. The table also lists the number of proteins that have higher, lower or same (no change) levels in the specific comparison.

In agreement with published data (Maier et al. 2009), we also observed that protein and transcript abundances correlate mainly positively. In our data, the transcripts showing high mean fluorescence intensity values on the microarray chips belonged to proteins that could be quantified in our proteomics analysis as well.

Comparative quantitative proteomics and transcriptomics analyses were combined for all IDs that could be quantified on both, transcript and protein level. The combined data allowed for a grouping into eight sub-groups that are characterized by different regulation patterns (see Table 4). Given group names and colours correspond to all tables and figures in this section. The three producing strains were compared to their relative controls (GAP CpB vs. GAP; GAP HyHEL vs. GAP and AOX HyHEL vs. AOX) and a strain growing on methanol was compared to a strain growing on glucose (AOX vs. GAP). For every comparison, the gene/proteins belonging to all the regulated sub-groups were looked at in detail, and checked which cellular functions or processes are enriched in the different regulatory subgroups.

**Table 3. Number of proteins identified, quantified, with higher/lower/unchanged abundance in the four comparisons.**

| | GAP CpB vs GAP control | GAP HyHEL vs GAP control | AOX HyHEL vs AOX control | AOX control vs GAP control |
|---|---|---|---|---|
| identified proteins | 1060 | 1036 | 1067 | 1071 |
| quantified proteins | 620 (58%) | 455 (44%) | 875 (82%) | 585 (55%) |
| higher proteins levels | 4 | 173 | 4 | 80 |
| lower protein levels | 1 | 4 | 5 | 4 |
| no change | 615 | 278 | 866 | 501 |

Changes on the proteome versus the transcriptome (as $log_2$FC of quantified proteins against the $log_2$FC of the transcripts) are summarized in Table 5 and visualized in Figure 6. The differently coloured symbols represent various combinations of changes in transcript and protein levels (according to the thresholds for higher/lower abundance described in the chapters for Proteomics and Transcriptomics workflow).

**Table 4. Regulatory sub-groups derived from comparative proteome and transcriptome analysis.**

| protein level | transcript level | regulatory sub-group |
|---|---|---|
| ↑ | ↑ | **protein up, transcript up:** Higher abundance in both analyses: The abundance of a protein is mainly regulated on transcriptional level. |
| ↑ | ↔ | **only protein up:** Higher protein levels, unaltered transcript levels: The abundance of a protein is probably mainly regulated post transcriptionally. |
| ↑ | ↓ | **protein up, transcript down (opposite trend):** The cell reacts to the different condition by down-regulating transcription of the respective gene but the abundance of the corresponding protein is higher because of e.g. even higher translational activity or an even lower protein degradation rate (relative to the amount of transcripts) or a high mRNA turnover rate. |
| ↔ | ↑ | **only transcript up:** Higher transcript levels, unaltered protein levels: The cell reacts to the different condition by up-regulating transcription of the respective gene but the abundance of the corresponding protein is probably mainly regulated post transcriptionally. |
| ↔ | ↔ | **protein & transcript unchanged:** There is no observable significant cellular reaction to the different condition on either transcript or protein level. |
| ↔ | ↓ | **only transcript down:** Lower transcript levels, unaltered protein levels: The cell reacts to the different condition by down-regulating transcription of the respective gene but the abundance of the corresponding protein is probably mainly regulated post transcriptionally. |
| ↓ | ↑ | **protein down, transcript up (opposite trend):** The cell reacts to the different condition by up-regulating transcription of the respective gene but the abundance of the corresponding protein is lower because of e.g. even lower translational activity or an even higher protein turnover rate (relative to the amount of transcripts). A lower mRNA turnover rate is also possible. This trend was not observed in any experiment in this study. |
| ↓ | ↔ | **only protein down:** Lower protein levels, unaltered transcript levels: The abundance of a protein is probably mainly regulated post transcriptionally. |
| ↓ | ↓ | **protein down, transcript down:** Lower abundance in both analyses: The abundance of a protein is mainly regulated on the transcriptional level. |

**Table 5. Number of IDs with higher (↑), lower (↓) or unchanged (↔) abundance on the protein vs the transcript level**

| protein level | transcript level | sub-group | GAP CpB vs GAP | GAP HyHEL vs GAP | AOX HyHEL vs AOX | AOX vs GAP |
|---|---|---|---|---|---|---|
| ↑ | ↑ | protein up, transcript up | 1 | 54 | 4 | 38 |
| ↑ | ↔ | only protein up | 3 | 114 | - | 42 |
| ↑ | ↓ | protein up, transcript down | - | 5 | - | - |
| ↔ | ↑ | only transcript up | 1 | 21 | 14 | 6 |
| ↔ | ↔ | protein & transcript unchanged | 605 | 200 | 810 | 471 |
| ↔ | ↓ | only transcript down | - | 54 | 31 | 17 |
| ↓ | ↑ | protein down, transcript up | - | - | - | - |
| ↓ | ↔ | only protein down | 1 | 1 | 2 | 2 |
| ↓ | ↓ | protein down, transcript down | - | 3 | 3 | 2 |

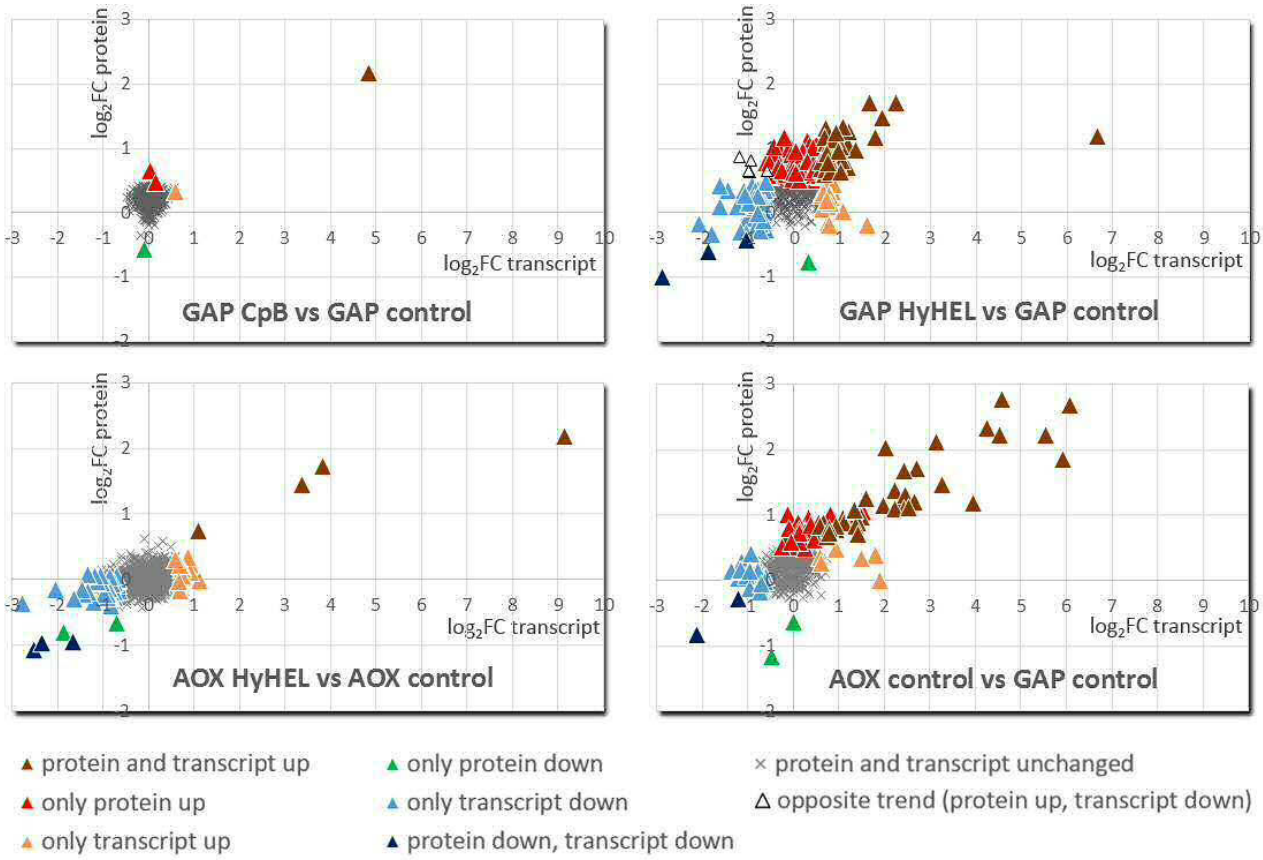


**Figure 6. Regulations on protein and/or transcript level of the four comparisons GAP CpB (vs control), GAP HyHEL (vs control), AOX HyHEL (vs control) and AOX control vs GAP control.** Correlations of $log_2$ fold changes of both, proteins and transcripts, are blotted against each other.

#### 4.4.2.1 GAP CpB vs. GAP control

When compared to the non-producing GAP control, 99% of all quantified proteins remained unchanged on the protein level, similar to the transcript level, with the heterologous gene transcript and product of CpB being the only exception as it is found in higher abundance on both levels, as expected. The comparison GAP CpB vs. GAP control shows, similarly to what is observed at transcript levels, that that more 99% of the quantified proteins keep the same level. A flocculin is the only protein with lower abundance but no signifant changes on the transcript level. Heterologous expression of CpB in the GAP strain does not affect proteome or transcriptome. All important cellular processes like central carbon metabolism, amino acid metabolism, secretion and translation remain unaltered according to this analysis. Table 6 lists the IDs assigned to the different subgroups according to their changes on the transcript and protein level. The individual IDs and their respective changes on the transcript and protein level are also plotted in Figure 7.

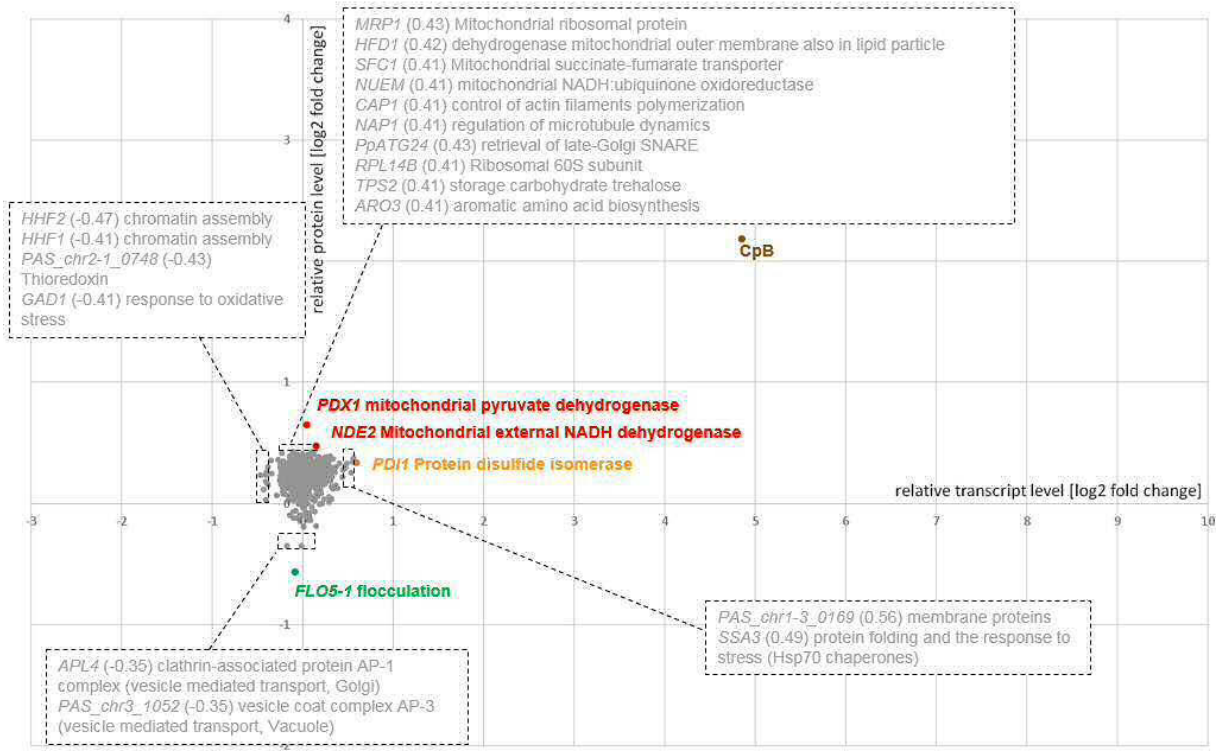


**Figure 7. Changes on the transcript and protein level when CpB ist produced under the GAP promoter.** $Log_2$ fold changes are plotted against each other and illustrate different regulatory sub-groups. The IDs marked in dashed lined boxes belong to the group “protein & transcript unchanged” but appear as outliers as they show a trend towards a group with a more pronounced regulatory pattern.

GO (gene ontology) term analyses of the sub-group “protein & transcript unchanged” (605 IDs) show that the metabolic and physiological state of the cell are not affected by the expression of CpB under the GAP promoter.

Only *PDI1* (and additionally *SSA3*), both involved in protein folding, showed higher transcript levels. On the protein level, only a few IDs related to mitochondrial activity and two proteins involved in cytoskeleton constitution (the filamentous proteins actin and tubulin) were detected in higher abundance.

Lower protein abundance (with no corresponding trend on the transcript level) was observed for one flocculin (*FLO5-1*) and two additional proteins involved in the clathrin vesicle mediated transport: one for sorting at the trans-Golgi network and one for the vacuolar protein sorting.

**Table 6. IDs with different regulation patterns in the GAP CpB production strain in comparison against GAP control.** IDs are listed for their respective sub-group according to differential regulation on the protein and/or transcript level. Additional IDs are marked in grey when they did not meet the chosen thresholds for significant regulation but appear closer to this sub-group than to the group of unregulated IDs.They are also shown in **Figure 7** in dashed lined boxes.

| sub-group | IDs | functional information about gene/protein |
|---|---|---|
| protein up, transcript up | 1 | only the recombinant product: CpB |
| only protein up | 2 | 2 mitochondrial: Mitochondrial external NADH dehydrogenase that provides cytosolic NADH to the mitochondrial respiratory chain (*NDE2*) and a protein which binds the E3 subunit of mitochondrial pyruvate dehydrogenase (*PDX1*)<br>10 additional genes/proteins:<br>4 mitochondrial: Mitochondrial ribosomal protein (*MRP1*), dehydrogenase mitochondrial outer membrane (*HFD1*), Mitochondrial succinate-fumarate transporter (*SFC1*) mitochondrial NADH:ubiquinone oxidoreductase (*NUEM*)<br>2 for cytoskeleton (*CAP1* for actin, *NAP1* for tubulin),<br>4 other: 1 late-Golgi SNARE (*PpATG24*), 1 ribosomal 60S subunit (*RPL14B*), 1 for storage of carbohydrate trehalose (*TPS2*) and 1 for aromatic amino acid metabolism (*ARO3*) |
| only transcript up | | 1 secretion: folding, protein disulfide isomerase (*PDI1*)<br>2 additional genes/proteins:<br>1 secretion related: Hsp70 chaperone for protein folding and the response to stress (*SSA3*)<br>1 membrane protein (*PAS_chr1-3_0169*) |
| only transcript down | - | only 4 additional gene/proteins:<br>2 for chromatin assembly (*HHF2*, *HHF1*)<br>2 related to oxidative stress/antioxidant (*GAD1*, *PAS_chr2-1_0748*) |
| only protein down | 1 | 1 flocculin protein (*FLO5-1*)<br>2 additional genes/proteins:<br>2 related to clathrin vesicle mediated transport: at trans-Golgi (*APL4*) and at vacuole (*PAS_chr3_1052*) |

#### 4.4.2.2 GAP HyHEL vs. GAP control

Unlike the CpB producing strain, expression of HyHEL apparently triggered more cellular regulatory reactions. The comparison of GAP HyHEL and its respective GAP control strain indicates many changes of the proteome and the transcriptome as well: only 61% of the quantified proteins remain at a constant level and 38% are more abundant in the HyHEL producing strain. Only 45% of the quantified IDs do not change at neither the transcript nor protein level.

The proteins with higher abundance are involved in many cellular processes like central carbon metabolism, beta-oxidation, glycerol metabolism, translation and cell wall. However, the analyzed clone (namely GAP HyHEL#3, a high producer according to clone screenings and bioreactor cultivations) turned out to possess two

nuclei. We suggest that the massive changes at both protein and transcript levels are not only a consequence of the expression of HyHEL alone, but can rather be explained by the presence of two nuclei. Only four proteins are present at lower level in the GAP HyHEL strain, three also at transcript level. Additionally - only in this comparison we observed a group of proteins where protein and transcript abundance behaved in an opposite manner: lower levels of transcripts but higher levels of the respective proteins.

**Fehler! Verweisquelle konnte nicht gefunden werden.**Table 7 lists the IDs assigned to the different subgroups according to their changes on the transcript and protein level. The individual IDs and their respective changes on the transcript and protein level are also plotted in Figure 8.

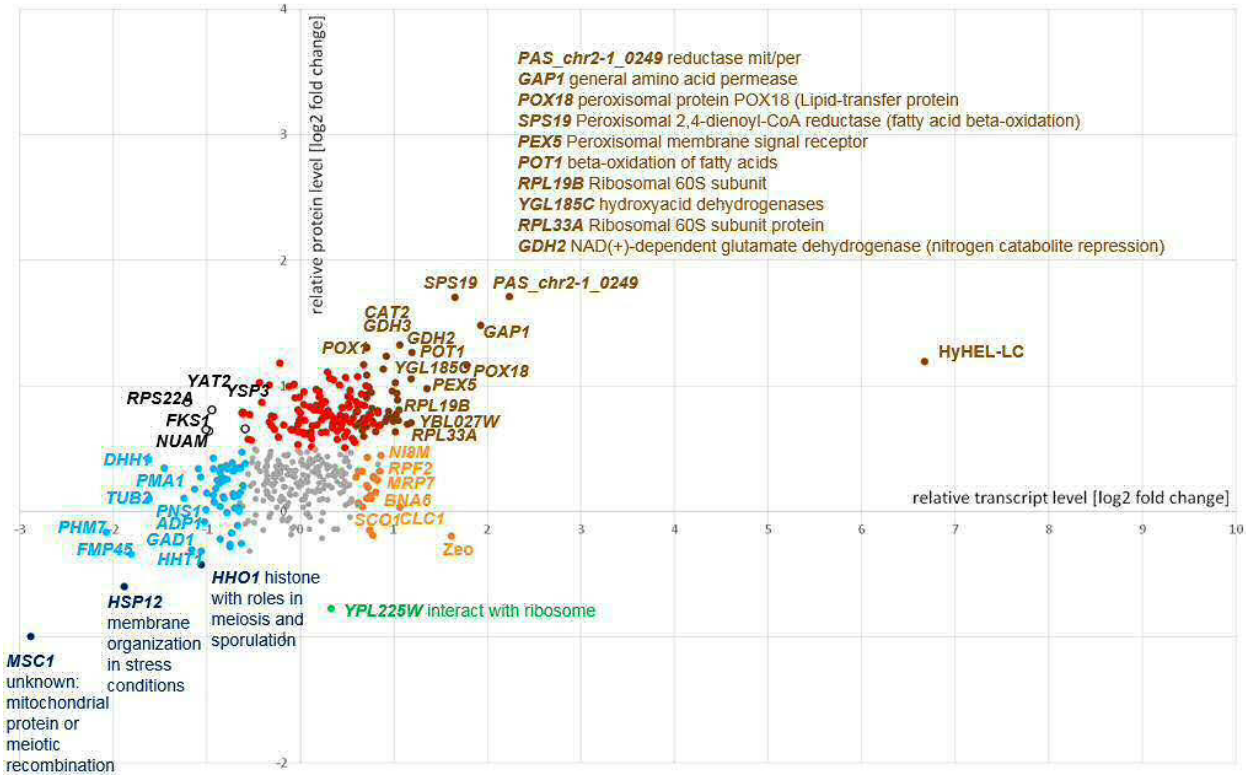


**Figure 8**. **Changes on the transcript and protein level when HyHEL Fab is produced under the GAP promoter.** $Log_2$ fold changes are plotted against each other and illustrate different regulatory sub-groups.

**Table 7. IDs with different regulation patterns in the GAP HyHEL production strain in comparison against GAP control.** IDs are listed for their respective sub-group according to differential regulation on the protein and/or transcript level.

| sub-group | IDs | functional information about gene/protein |
|---|---|---|
| protein up, transcript up | 54 | 22 ribosomal: 19 ribosomal protein subunites (11 60S and 8 40S); 2 proteins involved in the processing of pre-rRNA and 1 subunit of small nucleolar ribonucleoprotein particles (snoRNPs)<br>8 peroxisomal: proteins involved in fatty acid oxidation (*PAS_chr2-1_0249, POX1, POX18, SPS19, POT1, PEX11, SOU1*) and 1 Peroxisomal membrane PTS1 signal receptor (*PEX5*)<br>2 carnitine transporter (*CRC1, CAT2*)<br>10 mitochondria (*PET9, LSC1, GCV1, GGC1, PAS_chr3_0955, ISU1, CIR1, PUT2, NUIM, NUZM*)<br>7 amino acid metabolism: glutamate, histidine, threonine, serine, lysine (*YER152C, THR1, HIS3, GDH2, YMR226C, GDH3*) and 1 general amino acid transporter (*GAP1*) |
| only protein up | 113 | 19 amino acid metabolism (*SHM1, ARO9, ARG1, ARO3, AAT1, ADE3, SHM2, GLN1, GLT1, GLY1, ILV2, LEU1, LEU4, LYS20, DIP5, ARO4, ILV6, SAM2, AAT2*)<br>14 mitochondrial (*COR1, EHD3, NDE2, ATP4, COX4, PHB2, SSC1, POR1, MIR1, RMD9, NUEM, PDX1, COX13, TOM40*)<br>3 lipid: 2 proteins involved in fatty acid metabolism in the peroxisomes (*FOX2, FAA2*) and transcription factor involved in phospholipid synthesis (*YLR422W*)<br>26 ribosomal protein subunits (13 60S and 13 40S), 4 ribosomal and 16 involved in translation<br>3 protein post-translational modifications: O mannosylation (*DPM1*), mannose outer chain elongation of cell wall N-linked glycoproteins (*GUK1*) and addition of myristic acid (*NMT1*)<br>8 carbon metabolism: *HXK1, GUT1*, 4 in TCA (*CIT1, FUM1, KGD2, SDH1*), 1 in PPP (*ZWF1*) and glycerol degradation (*GUT2*)<br>5 purine and pyrimidine metabolism (*ADE4, ADE13, ADK1, URA2, PAS_chr2-2_0256*)<br>3 proteasomal (*PRE1, PRE3, RPT2*)<br>3 related to actin (*ACT1, ARP2, TWF1*)<br>2 ER related protein: ER morphology (*YOP1*) and ER redox buffer (*PAS_chr4_0192*)<br>10 other: ribosome-associated molecular chaperone (*SSB2*), mRNP complexes associated with polyribosomes (*BFR1*), snoRNP (*NOP56*), secretion (*SEC18*), signal recognition particle (SRP) complex: targeting to ER (*SRP68*), conversion of sulfite into sulfide (*MET10*), mating (*SCP160*), 2 unknown function (*PAS_chr3_0926, PAS_chr3_0527*), vitamin biosynthesis (*THI4*) |
| only transcript up | 21 | 3 mitochondrial (*MOS2, SCO1, NI8M*)<br>3 ribosomal (*MRP7, RPF2, TMA19*)<br>3 amino acid metabolism (*CIA1, ARG3, HOM2*)<br>2 clathrin (vesicle mediated transport) (*CLC1, PAS_chr3_1052*)<br>2 transport nucleus-cytoplasm (*MOG1, NTF2*)<br>8 other: NAD synthesis (*BNA6*), thiamin metabolism (*PET18-2*), proteasome (*PRE10*), secretion (*SEC4*), transcriptional activation (*TOA2*), DNA binding (*HNT1*), antibiotic resistance (Zeo) and unknown (*PAS_chr2-2_0136*) |

| | | |
|---|---|---|
| only transcript down | 54 | 4 carbon metabolism: glucose phosphorylation (*GLK1*), gluconeogenesis (*PGI1*), mobilization of glycogen (*GPH1*), monocarboxylic acid symporter (uptake of lactate, pyruvate, and acetate, *JEN1*)<br>4 chromatin structure (*HHT1, RSC9, RSC2, KAP114*) and 5 DNA related (*TOP2, RFA1, CRP1, PAS_chr4_0853, RNR1*)<br>4 oxidative stress (*TRX1, GSH1, ECM4, GAD1*)<br>2 mitochondrial (*DNM1, FMP45*)<br>2 protein folding (*TCP1, SSA3*)<br>2 protein modification (*PMT1, KTR1*) and 1 peptidase (*YOL057W*)<br>2 cell wall (*ROT2, SKG3*) and 1 cell separation during budding (*SDS23*)<br>3 transcription (*BDF1, PAS_chr3_0897, PAS_chr3_0261*)<br>3 vacuole (*YCF1, VTC2, PEP4*)<br>5 lipid (*ERG5, ERG10, OSH2, PNS1, PAS_chr4_0828*)<br>2 translation (*YHR020W, DHH1*)<br>13 other: amino acid metabolism (*TRP5-1*), Transmembrane nucleoporin (*POM33*), microtubule formation (*TUB2*), transmembrane transport (*ADP1*), clathrin vesicle mediated transport (trans-Golgi) (*APL4*), ER to Golgi transport (*SEC23*), Plasma membrane H+-ATPase (*PMA1*), RNA export from nucleus (*CRM1*), negative regulation of pseudohyphal differentiation (*SOK2*), NAD formation (*QNS1*), eisosomes (*EIS1*), 3 unknown (*PAS_chr4_0320, PAS_chr1-1_0257, PHM7*) |
| only protein down | 1 | Protein of unknown function; may interact with ribosomes (*YPL225W*) |
| protein down, transcript down | 3 | 1 mitochondrial protein may be involved in meiotic recombination with unknown function (*MSC1*)<br>1 Plasma membrane protein involved membrane organization in stress conditions (*HSP12*)<br>1 histone with roles in meiosis and sporulation (*HHO1*) |

22 IDs related to ribosomal activity and subunits were found to be transcriptionally regulated in the GAP HyHEL Fab production strain as they had both higher transcript and protein abundance. 7 genes for amino acid metabolism were also found in this group. This indicates a strong induction of the translation machinery in this strain with consequently a higher demand for amino acids.

The accumulation of proteins involved in post-translational modification processes, mainly glycosylation, could be explained considering that the alpha mating factor leader used to secrete the HyHEL Fab contains three glycosylation sites and is glycosylated before it leaves the ER. Furthermore, we might speculate that the cells need more energy to support all those induced intracellular processes. Indeed, there are increased levels of transcripts and/or proteins involved in the carbon metabolism, mitochondrial activities and fatty acids metabolism. Induction of purine and pyrimidine metabolism might also correlate with increased ribosome synthesis.

The high level of HyHEL Fab produced might be responsible of the increased level of transcript and/or proteins related to proteasome in the GAP HyHEL strain. Alternatively, higher abundance of proteasome subunits could correlate with a high turnover of cellular proteins due to stress. Similarly to the comparison GAP CpB vs.

GAP, there is accumulation of proteins related to actin, which plays an important role in polarized secretion by mediating the intracellular transport of secretory vesicles. , actin.

However, it has to be taken into account that many changes could also be mainly attributed to the peculiarity of the strain analyzed, which was afterwards demonstrated to have a binuclear phenotype. Higher overall levels of gene expression for constitutive proteins like actin or ribosomes are most likely a consequence of a second nucleus.

Among the down regulated IDs, there are 54 transcripts with lower abundance in the producing strain compared to the control. They belong to a large variety of cellular processes (e.g. carbon and amino acid metabolism, chromatin structure, oxidative stress, mitochondria, protein folding and modification, cell wall, transcription, vacuole, lipid, and translation), so we cannot identify specific cellular processes that are over-proportionally affected. Contrary to transcript levels, there are only few proteins with lower intracellular concentrations: One is related to ribosomes, one involved in plasma membrane organization and two related to meiosis.

#### 4.4.2.3 AOX HyHEL vs. AOX control

Only one percent of all quantified proteins undergo a change in intracellular concentration according to the the comparison of AOX HyHEL against the non-producing control strain. The few changes at proteomic level (4 gene products) are all also reflected on the transcript level. The four genes encoding the transcripts and proteins in higher abundance are: *CNE1* and - as expected - the heterologous gene products HyHEL-HC, HyHEL-LC and the Zeocin resistance. Among the proteins with reduced levels there are Aox2, Adh4-2, Idp2, Gdh2 and Mdh3. Interestingly, there is a group of proteins present at the same level but with increased or decreased transcript levels. Proteins involved in protein transport can be found in the up-regulated proteins, while among the down-regulated there are proteins involved in carbohydrate metabolism, generation of precursor metabolites and energy, cofactor metabolic process and lipid metabolism.

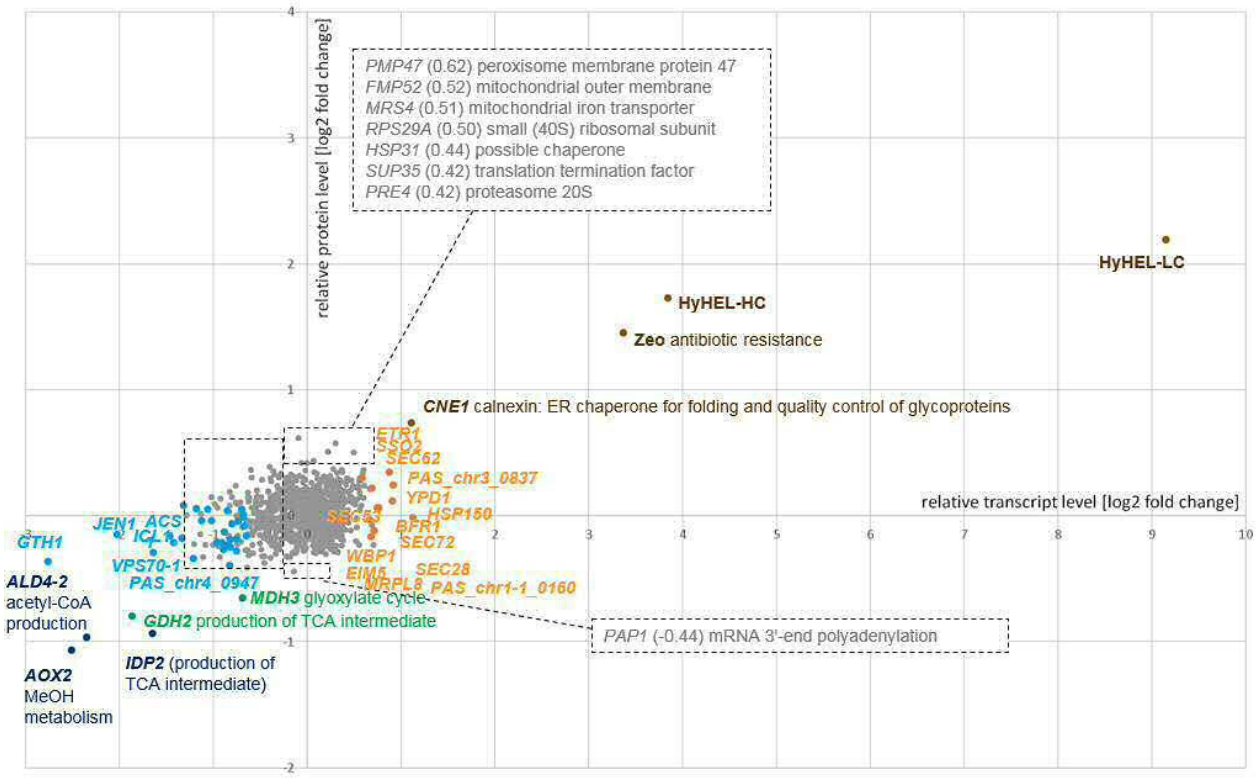


**Figure 9. Changes on the transcript and protein level when HyHEL Fab ist produced under the AOX promoter.** $Log_2$ fold changes are plotted against each other and illustrate different regulatory sub-groups. The IDs marked in dashed lined boxes belong to the group "protein & transcript unchanged" but appear as outliers as they show a trend towards a group with a more pronounced regulatory pattern.

**Table 8. IDs with different regulation patterns in the AOX HyHEL production strain in comparison against AOX control.** IDs are listed for their respective sub-group according to differential regulation on the protein and/or transcript level. Additional IDs are marked in grey when they did not meet the chosen thresholds for significant regulation but appear closer to this sub-group than to the group of unregulated IDs.They are also shown in Figure 9 in dashed lined boxes.

| sub-group | IDs | functional information about gene/protein |
|---|---|---|
| protein up, transcript up | 4 | recombinant heterodimeric product: HyHEL-LC, HyHEL-HC<br>antibiotic resistance (*Zeo*)<br>1 secretion: calnexin, an ER chaperone for folding and quality control of glycoproteins (*CNE1*) |
| only protein up | - | 7 additional genes/proteins:<br>2 mitochondrial (*FMP52, MRS4*), 1 peroxisomal (*PMP47*), 1 ribosomal (*RPS29A*), 1 translation (*SUP35*), 1 proteasomal (*PRE4*) and 1 possible chaperone (*HSP31*) |
| only transcript up | 14 | 7 secretion related: 2 for glycosylation (*SEC53, WBP1*), secretory vesicles (*SSO2*), retrograde transport (*SEC28*), folding (disulfide isomerase, *PAS_chr1-1_0160*) and 2 transport to ER (subunits of the Sec63 complex *SEC62* and *SEC72*)<br>7 other: mRNP complexes component (mRNA protein complexes *BFR1*), respiration (*EMI5*), response to osmotic stress (*YPD1*), mitochondrial ribosomal protein (*MRPL8*), lipid: fatty acid synthesis (*ETR1*), cell wall stability (*HSP150*) and unknown (*PAS_chr3_0837*) |

| | | |
|---|---|---|
| only transcript down | 31 | 11 carbon metabolism: TCA (*KGD1, SDH2*), glyoxylate cycle (*ICL1, MLS1*), acetyl-CoA transport (*YAT2*), high affinity glucose transporter glucose transporter (*GTH1*), glucose phosphorylation (*GLK1*), glycogene accumulation (*GLC3*), carboxylic acid transporters (*JEN1* and *THI73*), glycerol degradation (*GUT2*)<br>5 lipid metabolism: ergosterol biosynthesis (*ERG5*) and 3 fatty acid metabolism (*FAA1, FAA2, PAS_chr2-1_0701*) and myo-inositol transporter ( *PAS_chr4_0828*)<br>3 amino acid metabolism (*SHM1, ARO9, GDH3*)<br>3 vacuole (*PEP4, APE3, VPS70-1*)<br>2 cell wall proteins (*ROT2, PST1*)<br>2 histone (histone *HHO1* and for histone acetylation *ACS1*)<br>5 other: methanol metabolism (*FBP1*), microtubules (*TUB1*), mitochondrial proteins (*YKR070W-2*), transporter for ammonia (*ATO2*) and unknown (*PAS_chr4_0947*)<br>40 Additional genes/proteins:<br>7 carbon metabolism (*GDB1, ACH1, ADH2, PDC1, ALD4-1, PGI1, TKL1*), 3 TCA (*KGD2, ACO1, SDH1*) and 1 ATP production (*ATP2*)<br>5 lipid biosynthesis (*ELO2, OLE1, FAS1, GPT2, HMG1*)<br>1 membrane lipid organization (*PDR5-1*)<br>4 vacuole (*VPH1, VTC2, VMA2, CPS1*)<br>2 cell wall (*SCW10, GAS1-1*)<br>2 mRNA processing (*DHH1, PAB1*) and 1 ribosome biogenesis (*TMA108*)<br>3 DNA (*RNR1, CDC9, HHF1*)<br>3 amino acids metabolism (*GLT1, CAR1-2, ALT1*)<br>2 mitochondrial: mitochondrial NADH:ubiquinone oxidoreductase (*NUAM*) and Mitochondrial protein may be involved in meiotic recombination (*MSC1*)<br>6 other: protein folding: HSP70 family (*SSA1*), secretion: vesicle trafficking between the ER and Golgi (*SLY1*), negative regulation of pseudohyphal differentiation (*SOK2*), Plasma membrane H+-ATPase (*PMA1*), methanol metabolism (*FDH1*), unknown (*PHM7*) |
| only protein down | 2 | 2 carbon metabolism: production of 2-oxoglutarate (TCA intermediate), but also amino acid metabolism (*GDH2*) and glyoxylate cycle (*MDH3*)<br>1 additional: mRNA 3'-end polyadenylation (*PAP1*) |
| protein down, transcript down | 3 | 2 carbon metabolism: aldehyde dehydrogenase, production of acetyl-CoA (*ALD4-2*) and production of 2-oxoglutarate (TCA intermediate) (*IDP2*)<br>1 methanol metabolism (*AOX2*) |

94% of the quantified proteins show no change in intracellular concentrations at both the transcript and protein level between the two strains. Not surprisingly, there is an accumulation of transcript and protein levels of the HC and LC of the HyHEL Fab and of the gene/protein conferring resistance to the antibiotic Zeocin. As previously explained for the GAP HyHEL strain, higher levels (protein and transcript) of the ER chaperone calnexin (important for the folding and quality control of glycoprotein) and higher transcript abundance of two proteins involved in glycosylation can be related to the fact that HyHEL Fab secretion is guided by the glycosylated alpha mating factor leader. Transcript levels of other proteins involved in the secretory pathway are also increased in the producing strain.

GO terms analysis of the group of 810 transcript/proteins that are present at the same level in the two strains, show that the most important cellular processes are not

affected by the expression of HyHEL in the AOX background.

Reduced levels of transcript and/or proteins involved in TCA and glyoxylate cycle were observed. Reduced TCA cycle flux in recombinant protein producing strains of *P. pastoris* has also been reported in literature by us and others previously for the GAP based systems (Dragosits et al. 2009; Heyland et al. 2010; Nocon et al. 2014). Interestingly, in both HyHEL producing strains, reduced level of the transcript of the plasma membrane ATPase (*PMA1*) were observed compared to the respective control strains. Considering this protein as essential to maintain intracellular pH homeostasis, a possible explanation could be that during the production of the HyHEL the cells do not have to pump out as much $H^+$ as in the control strain, or they have enough protein to keep the intracellular pH and less *PMA1* gene has to be transcribed.

#### 4.4.2.4 AOX control vs. GAP control

The two non-producing control strains were also compared to each other. They were cultivated differently, using a glucose-based medium for the GAP control strain and a methanol/glycerol-based medium for the AOX control strain, a strain that also has the $mut^S$ (methanol utilization slow) phenotype, where the *AOX1* gene is deleted so that the native *AOX2* gene (same gene product but under a weaker promoter) is solely responsible for methanol utilization. The comparison of the two non-producing strains shows that 17% of the quantified proteins have different levels in the AOX strain compared to GAP. In agreement with the transcriptomics data of the glycerol/methanol growth condition, there is an accumulation of proteins involved in the methanol metabolism, glyoxylate cycle, vitamin metabolism, proteins related to amino acid biosynthesis, translation, peroxisomes and the metabolism of cofactors (especially for methanol utilization proteins). As expected, there are reduced level of both transcript and protein of the *P. pastoris* high affinity glucose transporter (Gth1) and the hexokinase I (Hxk1) during growth in glycerol/methanol.

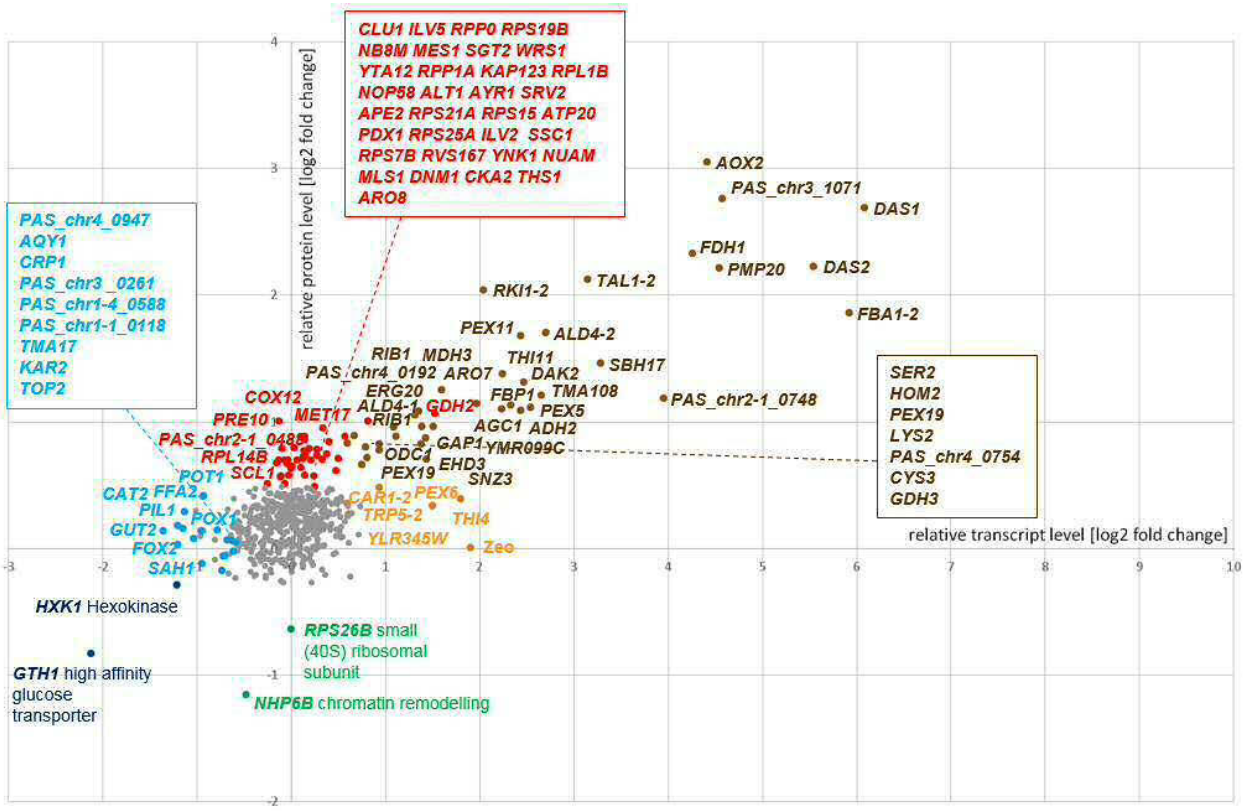


**Figure 10**. **IDs with different regulation patterns in the AOX control strain in comparison against GAP control strain.** IDs are listed for their respective sub-group according to differential regulation on the protein and/or transcript level.

**Table 9. IDs with different regulation patterns in the AOX control strain in comparison against GAP control.** IDs are listed for their respective sub-group according to differential regulation on the protein and/or transcript level.

| sub-group | IDs | functional information about gene/protein |
|---|---|---|
| protein up, transcript up | 38 | 9 amino acid metabolism (*ODC1, ARO7, GDH3, CYS3, HOM2, LYS2, SER2, GAP1, AGC1*)<br>6 carbon metabolism: Glucose-6-phosphate 1-epimerase (*YMR099C*), glyoxylate cycle (*MDH3*), 2 aldehyde dehydrogenases (*ALD4-2, ALD4-1*), alcohol dehydrogenase (*ADH2*) and mannitol dehydrogenase (*PAS_chr4_0754*)<br>10 methanol metabolism (*DAS1, FBA1-2, DAS2, AOX2, FDH1, SHB17, TAL1-2, FBP1, RKI1-2, DAK2*)<br>3 peroxisomal: peroxiredoxin (*PMP20*), Peroxisomal membrane PTS1 signal receptor (PEX5), peroxisomal chaperone (*PEX19*)<br>2 lipid metabolism: (sterol biosynthesis (*ERG20*), beta oxidation and peroxisome proliferation (*PEX11*),<br>2 oxidative stress response: Thioredoxin (*PAS_chr2-1_0748*) and ER redox buffer (*PAS_chr4_0192*),<br>3 vitamin biosynthesis (*RIB1, SNZ3, THI11*)<br>1 translation (PAS_chr3_1071) and 1 ribosome (*TMA108*)<br>other: 3-hydroxyisobutyryl-CoA hydrolase (*EHD3*) |
| only protein up | 40 | 7 amino acid metabolism (*GDH2, MET17, ILV5, ALT1, ILV2, ARO8, APE2*)<br>8 ribosomal: 5 40S subunits, 2 60S subunits, 1 assembly (*KAP123*), 1 rRNA processing (*NOP58*)<br>6 translation (*CLU1, RPP0, WRS1, RPP1A, THS1, MES1*)<br>4 mitochondria (*ATP20, DNM1, PDX1, SSC1*) and 3 respiration (*COX12, NB8M, NUAM*)<br>5 protein folding/stability: 2 proteasome (*PRE10, SCL1*) and 1 degradation of misfolded proteins (*YTA12*), 1 protein folding (*SGT2*)<br>2 actin related (*SRV2, RVS167*)<br>other: cell growth (*CKA2*), glyoxylate cycle (*MLS1*), lipid (*AYR1*), DNA (*YNK1*), unknown (*PAS_chr2-1_0488*) |
| only transcript up | 6 | 2 amino acid metabolism (*CAR1-2, TRP5-2*)<br>other: antibiotic resistance Zeo, carbon metabolism *YLR345W*, recycling of peroxisomal signal receptor *PEX6*), vitamin biosynthesis (*THI4*) |
| only transcript down | 17 | 4 fatty acid beta oxidation (*FAA2, FOX2, POX1, POT1*), 1 Carnitine acetyl-CoA transferase *CAT2*) and 1 amino acid-lipid metabolism (*SAH1*)<br>2 DNA structures (*CRP1, TOP2*)<br>other: transcription (*PAS_chr3_0261*), spore maturation (*AQY1*), proteasome homeostatic levels (*TMA17*), endocytosis (*PIL1*), chaperone folding (*KAR2*), glycerol degradation (*GUT2*) and 3 unknown (*PAS_chr4_0947, PAS_chr1-4_0588, PAS_chr1-1_0118*) |
| only protein down | 2 | small (40S) ribosomal subunit (*RPS26B*), chromatin remodeling (*NHP6B*) |
| protein down, transcript down | 2 | 2 carbon metabolism (high affinity glucose transporter *GTH1*, hexokinase *HXK1*) |

Comparing methanol-grown cells with glucose-grown cells, 82% of the quantified proteins show unaltered transcript and protein level.

Apart from the peculiar results of the GAP HyHEL strain carrying the two nuclei, the

carbon sources apparently has a bigger impact on the cell physiology than the recombinant protein product (CpB under GAP promoter and HyHEL under GAP or AOX promoter). The results and findings from this analysis were published by Rußmayer and Buchetics (System-level organization of yeast methylotrophic lifestyle; 2015).

Methanol-grown cells are in a state of higher intracellular stress. Their need to better control the quality of their translated proteins could be an explanation for the observed accumulation of proteins involved in folding, degradation of misfolded protein and of the proteasome in the AOX strain. This induced “quality control system” might contribute together with the increased translational activity to the higher productivity obtained in methanol based expression systems.

Lower levels of transcript for gene/proteins involved in fatty acid beta oxidation have been detected in methanol growing cells. The observed changes in fatty acid beta oxidation come somehow unexpected as the cells were not cultivated on fatty acids. However, these are not followed by changes at the protein levels suggesting that we do not have higher activity of beta fatty oxidation in glucose growing cells. Differences in transcript stability or on the promoter strength might be responsible for these results. As expected there are higher transcript/protein levels of two proteins specifically required for growth on glucose but not on methanol: a high affinity glucose transporter and hexokinase.

### 4.4.3 Conclusion

Our transcriptomics data generally correlate well with proteomics data so the combination of both technologies allow for the detection of changes on a wider range of transcripts and/or proteins involved in a defined process. However, in all comparisons, more transcripts were found to be less abundant compared to only a small number of proteins with decreased amounts. Less transcripts but unchanged protein levels indicates a change in translation efficiency and we assume a faster mRNA turnover compared to the turnover of proteins. In addition, we assume that the amount of present protein does not fully correlate with the amount of functional protein, which is dependent on correct protein folding, correct post-translational modifications and stability of the functional form.

It is generally assumed that changes in transcript levels substantially translate to the protein level. The interplay of all modes of regulation are poorly understood. Studies on human and mouse have already revealed a quite complex relationship between transcriptome and proteome data. Post-translational processing and inherent variations in protein stability affect protein levels independently of transcript levels. The rather poor correlations are likely to be explained by molecular events such as translational efficiency, alternative splicing, protein folding, modifications, assembly into complexes, transport and localization, secretion or degradation (Ghazalpour et al. 2011; Glaesener et al. 2008).

In conclusion, no significant changes were observed as a consequence of CpB production (in the GAP strain) and only few changes when HyHEL is expressed in the AOX strain. In contrast, a high number of changes were detected for the strain GAP HyHEL compared to GAP which correlate very well with the high number of changes already observed at transcriptome level. These results clearly confirm the altered cellular state of the clone analyzed as the clone GAP HyHEL#3 was found to carry two nuclei. Additionally, the analysis of the strain AOX vs. GAP confirmed the results anticipated by transcriptomics and metabolomics data. Generally, we observed that during growth on glycerol/methanol compared to glucose there are more proteins involved directly and indirectly in methanol metabolism. There is a good correlation between proteome and transcriptome, however, the expression of the two heterologous proteins *per se* does not alter the proteome level of the cell.

### 4.4.4 Contributions

For this part of the thesis, the author performed the following tasks: Cloning of all expression vectors, generation of production clones, clone screenings, design of chemostat media, bioreactor cultivations and sampling, microarray analysis and data interpretation.

Hannes Rußmayer assisted in chemostat cultivation.

Clemens Gruber performed proteomic measurements of the chemostat samples.

Minoska Valli and Brigitte Gasser contributed to this work by data mining, setting up a transcript/protein database and data interpretation.

# 5 Conclusion

The microbial host for recombinant protein production, *Pichia pastoris*, has been subject of research and engineering (especially since the availability of its genome) in the past decade. The work conducted for this thesis increased our knowledge of strain characteristics, especially of *P. pastoris´*s methanol metabolism. It also expanded our repertoire of methods and tools for strain engineering. Hence, the output of this thesis is multifaceted:

**Slow-releasing glucose tablets serve a different purpose in methanol induced cultures and improve reproducibility of small scale screenings**

The small-scale cultivation protocols, which we introduced for both major carbon source/promoter strategies (glucose/$P_{GAP}$ and methanol/$P_{AOX1}$), take into account some key aspects of fed batch cultivation. In the bioreactor scale, a multi-step approach is needed when using methanol for *AOX1* promoter driven protein production (Cos et al. 2006). We transferred the phases for the generation of biomass, de-repression/induction of methanol utilization enzymes and heterologous protein production to the small-scale, where monitoring of oxygen consumption or methanol accumulation is not possible. A combined glucose batch and fed batch phase for fast biomass accumulation with automatic transition into a glucose-limited slow growth phase enables uniform handling of non-starved cultures with de-repressed *AOX1* promoter (Lin-Cereghino et al. 2006) before the cultures are induced with methanol. Unlike for glycerol, slow continuous release of glucose into the growth medium is made possible by glucose feed tablets. With their help, the continued slow growth after the initial batch phase not only keeps the cultures in a uniform and reproducible state but also serves as de-repression phase for the *AOX1* promoter.

**Aft1 is involved in carbon source response rather than iron regulation and can positively affect secretion**

We found a transcription factor (Aft1) to regulate several genes involved in protein

secretion. By regulatory sequence analysis, common motifs for putative binding sites were found in the promoter regions of the two ER chaperones Pdi1 and BiP/Kar2, the chaperone Ssa4, the serine/threonine kinase Kin2 and Ncp1, a protein involved in lipid metabolism. These proteins all had been shown in previous work to improve secretion of a recombinant product when co-overexpressed in *P. pastoris* (Gasser et al. 2008; Idiris et al. 2010). The binding sites turned out to correspond to the transcription factor Aft1 and were found in a total of 561 annotated genes. Only 6 of these genes are involved in iron regulation, despite the fact that Aft1 was originally annotated as regulator of iron uptake and homeostasis due to sequence homology to *S. cerevisiae´*s Aft1/2. On the other hand, many other genes with Aft1 binding sites in the promoter regions were found to play a role in the secretory path, namely protein targeting (23 genes), Golgi to vesicle transport (19), oxidative stress response (12 genes), protein folding (11), glycosylation and vesicle organization (10), vacuole organization and regulation of transport (9), and exocytosis (6).

Aft1 was then overexpressed and knocked out in the wildtype strain. Both modified strains showed no different behavior in iron-depleted conditions. We additionally performed microarray analysis of the Δaft1 strain and confirmed the unaltered state of the iron-regulatory pathway as transcription levels of *FEP1* and other genes involved in iron uptake and homeostasis were not affected. This observation fitted to the absence of Aft1 binding motifs in the promoter regions of iron regulatory genes in *P. pastoris*. We found out that expression levels of *AFT1* depend on the carbon source: There is transcriptional up-regulation during glucose-limited growth and during growth on methanol. We also present strong indications that Aft1 is involved in the regulation of glucose-repressed genes as we found its binding sites in the promoter regions of several carbon-source responsive genes.

We did not observe differential regulation of the previously mentioned genes involved in secretion improvement (which have the Aft1 binding site in their promoters). Nevertheless, the secretion enhancing effect of *AFT1* overexpression was confirmed: Co-overexpression in a *P. pastoris* strain producing a carboxylesterase increased titers up to 2.5-fold in bioreactor production.

## *P. pastoris* possesses a duplicated methanol inducible enzyme set targeted to peroxisomes

The choice for the *AOX1* promoter system for heterologous gene expression in *P. pastoris* is also a choice for the carbon source methanol. While promoter activities have been investigated extensively for various growth parameters, there was a lack of data describing the physiological and metabolic state of the cell on a systems-level. We therefore compared *P. pastoris* cultures grown on methanol/glycerol with cultures grown on glucose and performed analyses on the transcript, protein and metabolite level. As expected, we observed a strong induction of peroxisome proliferation on methanol. In fact, all known enzymes of the methanol utilization pathway had significantly higher transcript and protein levels when methanol was present. *P. pastoris* has second isoforms of some pentose phosphate pathway (PPP) enzymes and the (for cells grown in absence of glucose) gluconeogenetic enzyme fructose-1,6-bisphosphate-aldolase. Not only were these isoforms found to have peroxisomal targeting signals, but they were also found to have higher transcript and protein levels on methanol, in contrast to the cytosolic isoforms. The analysis of isolated peroxisomal fractions confirmed that the entire methanol assimilation pathway is localized to peroxisomes. Another interesting finding was that during growth on methanol, xylulose-5-phosphate is not regenerated by the cytosolic, oxidative branch of the PPP but rather by a set of peroxisomal enzymes for the rearrangement of sugar phosphates, netting one molecule of dihydroxyacetone (DHA, then converted to glyceraldehyde-3-phosphate, GAP) per three molecules of methanol. In this so-called xylose-monophosphate cycle, sedoheptulose-1,7-bisphosphate is used as intermediate, similar to the Calvin cycle for $CO_2$ assimilation in the chloroplasts of plants. We also found out that an active glyoxylate cycle in the peroxisomes results in lower carbon fluxes through the TCA cycle when the cells are grown on methanol/glycerol.

## Methanol-grown *P. pastoris* increase the flux towards amino acid and protein synthesis

The change in carbon source led to significant changes in the amino acid

metabolism: The higher overall protein content per cell grown on methanol can be explained by the strong induction of the methanol assimilation enzymes localized to the peroxisomes and the respective cofactors that are needed for these enzymes. On transcript and protein level, we also saw increased ribosome biogenesis and translation and amino acid biosynthesis, which fits to the observed high specific production rates of reduced NADPH, the main electron donor in amino acid biosynthesis. Probably due to their immediate need, less free amino acids were measured. Overall, methanol-grown cells display a higher translational capacity. This finding could be the key to understanding the often observed high productivities during recombinant protein production in this host.

**The carbon source strongly affects the cellular proteome, recombinant gene expression does not**

The combined analysis of the transcriptome and the proteome of recombinant protein producing cultures compared to non-producing controls delivered no significant changes as a consequence of CpB production (in the GAP strain) and only few changes caused by HyHEL is production in the AOX strain. Interpretation of the data for HyHEL production in the GAP strain was hampered by the fact, that the (highly productive) clone used for chemostat cultivation turned out to possess two nuclei. However, higher transcript and protein levels of the ER chaperone calnexin (important for the folding and quality control of glycoprotein) and higher transcript levels of two genes involved in glycosylation were found in both HyHEL Fab secreting strains. We assume that the glycosylated alpha mating factor leader attached to the HyHEL is responsible for this response. Transcript levels of other proteins involved in the secretory pathway were also increased in the HyHEL strains.

The versatile outcome of this thesis can be summarized as:

In an educated down-scaling approach, we developed small scale cultivation protocols for screening and/or scientific purposes based on our understanding of the host´s physiology and its behavior in (fed batch) bioreactor cultivation.

By browsing the genome for transcription factor binding sites in the promoter regions of genes that had previously been shown to enhance secretion, a novel transcription

factor was found to benefit recombinant protein production.

The combined analysis of the transcriptome and the proteome of different *P. pastoris* production strains draws a complex picture of the regulation of protein abundance. Nevertheless, heterologous protein production has a lesser impact on the proteome as expected. The carbon source was shown to have a much more pronounced effect on the transcriptome and proteome of the cell.

The impact of the carbon source was investigated in more detail, incorporating the measurement of metabolites. This led to new insights into the methylotrophic lifestyle of *P. pastoris*, which was found to assimilate methanol entirely in its peroxisomes. In addition to that, we suggest a highly active translation apparatus and amino acid metabolism to be a key factor in the often-observed higher productivities during growth on methanol.

# 6 References


Ahmad M, Hirz M, Pichler H, Schwab H. 2014. Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. Appl Microbiol Biotechnol 98(12):5301-17.

Bantscheff M, Lemeer S, Savitski MM, Kuster B. 2012. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal Bioanal Chem 404(4):939-65.

Benjamini Y, Heller R, Yekutieli D. 2009. Selective inference in complex research. Philos Trans A Math Phys Eng Sci 367(1906):4255-71.

Beyer A, Hollunder J, Nasheuer HP, Wilhelm T. 2004. Post-transcriptional expression regulation in the yeast Saccharomyces cerevisiae on a genomic scale. Mol Cell Proteomics 3(11):1083-92.

Caspeta L, Shoaie S, Agren R, Nookaew I, Nielsen J. 2012. Genome-scale metabolic reconstructions of Pichia stipitis and Pichia pastoris and in silico evaluation of their potentials. BMC Syst Biol 6:24.

Chung BK, Lakshmanan M, Klement M, Ching CB, Lee DY. 2013. Metabolic reconstruction and flux analysis of industrial Pichia yeasts. Appl Microbiol Biotechnol 97(5):1865-73.

Chung BK, Selvarasu S, Andrea C, Ryu J, Lee H, Ahn J, Lee DY. 2010. Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast Pichia pastoris for strain improvement. Microb Cell Fact 9:50.

Coghlan A, Wolfe KH. 2000. Relationship of codon bias to mRNA concentration and protein length in Saccharomyces cerevisiae. Yeast 16(12):1131-45.

Cos O, Ramón R, Montesinos J, Valero F. 2006. Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast Pichia pastoris under different promoters: a review. Microb Cell Fact 5:17.

Cregg JM, Vedvick TS, Raschke WC. 1993. Recent advances in the expression of foreign genes in Pichia pastoris. Biotechnology (N Y) 11(8):905-10.

Damasceno LM, Anderson KA, Ritter G, Cregg JM, Old LJ, Batt CA. 2007. Cooverexpression of chaperones for enhanced secretion of a single-chain antibody fragment in Pichia pastoris. Appl Microbiol Biotechnol 74(2):381-9.

De Schutter K, Lin Y, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouzé P, Van de Peer Y, Callewaert N. 2009. Genome sequence of the recombinant protein production host Pichia pastoris. Nat Biotechnol 27(6):561-6.

de Sousa Abreu R, Penalva LO, Marcotte EM, Vogel C. 2009. Global signatures of protein and mRNA expression levels. Mol Biosyst 5(12):1512-26.

Delmotte N, Ahrens CH, Knief C, Qeli E, Koch M, Fischer HM, Vorholt JA, Hennecke H, Pessi G. 2010. An integrated proteomics and transcriptomics reference data set provides new insights into the Bradyrhizobium japonicum bacteroid

metabolism in soybean root nodules. Proteomics 10(7):1391-400.

Demain AL, Vaishnav P. 2009. Production of recombinant proteins by microbes and higher organisms. Biotechnol Adv 27(3):297-306.

Dimitrov DS. 2012. Therapeutic proteins. Methods Mol Biol 899:1-26.

Dragosits M, Stadlmann J, Albiol J, Baumann K, Maurer M, Gasser B, Sauer M, Altmann F, Ferrer P, Mattanovich D. 2009. The effect of temperature on the proteome of recombinant Pichia pastoris. J Proteome Res 8(3):1380-92.

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

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):353-61.

Gasser B, Saloheimo M, Rinas U, Dragosits M, Rodríguez-Carmona E, Baumann K, Giuliani M, Parrilli E, Branduardi P, Lang C and others. 2008. Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microb Cell Fact 7:11.

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):6499-507.

Ghazalpour A, Bennett B, Petyuk VA, Orozco L, Hagopian R, Mungrue IN, Farber CR, Sinsheimer J, Kang HM, Furlotte N and others. 2011. Comparative analysis of proteome and transcriptome variation in mouse. PLoS Genet 7(6):e1001393.

Glaesener S, Honecker F, Veltman IM, Gillis AJ, Rohlfing T, Streichert T, Otto B, Brummendorf TH, Looijenga LH, Bokemeyer C and others. 2008. Comparative proteome, transcriptome, and genome analysis of a gonadal and an extragonadal germ cell tumor cell line. J Proteome Res 7(9):3890-9.

Glisovic T, Bachorik JL, Yong J, Dreyfuss G. 2008. RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett 582(14):1977-86.

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

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

Haider S, Pal R. 2013. Integrated analysis of transcriptomic and proteomic data. Curr Genomics 14(2):91-110.

Hamilton SR, Gerngross TU. 2007. Glycosylation engineering in yeast: the advent of fully humanized yeast. Curr Opin Biotechnol 18(5):387-92.

Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, Lin-Cereghino J, Kovar K, Cregg JM, Glieder A. 2008. Promoter library designed for fine-tuned gene expression in Pichia pastoris. Nucleic Acids Res 36(12):e76.

Hegde PS, White IR, Debouck C. 2004. Interplay of transcriptomics and proteomics. Drug Discov Today 9(2 Suppl):S53-6.

Heyland J, Fu J, Blank LM, Schmid A. 2010. Quantitative physiology of Pichia pastoris during glucose-limited high-cell density fed-batch cultivation for recombinant protein production. Biotechnol Bioeng 107(2):357-68.

Hohenblum H, Gasser B, Maurer M, Borth N, Mattanovich D. 2004. Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant Pichia pastoris. Biotechnol Bioeng 85(4):367-75.

Idiris A, Tohda H, Kumagai H, Takegawa K. 2010. Engineering of protein secretion in yeast: strategies and impact on protein production. Appl Microbiol Biotechnol 86(2):403-17.

Irani ZA, Kerkhoven EJ, Shojaosadati SA, Nielsen J. 2016. Genome-scale metabolic model of Pichia pastoris with native and humanized glycosylation of recombinant proteins. Biotechnol Bioeng 113(5):961-9.

Kell DB. 2004. Metabolomics and systems biology: making sense of the soup. Curr Opin Microbiol 7(3):296-307.

Kreil DP, Russell RR, Russell S. 2006. Microarray oligonucleotide probes. Methods Enzymol 410:73-98.

Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, Jaenicke S, Brinkrolf K, Goesmann A, Szczepanowski R and others. 2011. High-quality genome sequence of Pichia pastoris CBS7435. J Biotechnol 154(4):312-20.

Laurent JM, Vogel C, Kwon T, Craig SA, Boutz DR, Huse HK, Nozue K, Walia H, Whiteley M, Ronald PC and others. 2010. Protein abundances are more conserved than mRNA abundances across diverse taxa. Proteomics 10(23):4209-12.

Lee MV, Topper SE, Hubler SL, Hose J, Wenger CD, Coon JJ, Gasch AP. 2011. A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. Mol Syst Biol 7:514.

Li H, d'Anjou M. 2009. Pharmacological significance of glycosylation in therapeutic proteins. Curr Opin Biotechnol 20(6):678-84.

Li H, Sethuraman N, Stadheim T, Zha D, Prinz B, Ballew N, Bobrowicz P, Choi B, Cook W, Cukan M and others. 2006. Optimization of humanized IgGs in glycoengineered Pichia pastoris. Nat Biotechnol 24(2):210-5.

Lin S, Houston-Cummings NR, Prinz B, Moore R, Bobrowicz B, Davidson RC, Wildt S, Stadheim TA, Zha D. 2012. A novel fragment of antigen binding (Fab) surface display platform using glycoengineered Pichia pastoris. J Immunol Methods 375(1-2):159-65.

Lin-Cereghino GP, Godfrey L, de la Cruz BJ, Johnson S, Khuongsathiene S, Tolstorukov I, Yan M, Lin-Cereghino J, Veenhuis M, Subramani S and others. 2006. Mxr1p, a key regulator of the methanol utilization pathway and peroxisomal genes in Pichia pastoris. Mol Cell Biol 26(3):883-97.

Lin-Cereghino J, Hashimoto MD, Moy A, Castelo J, Orazem CC, Kuo P, Xiong S, Gandhi V, Hatae CT, Chan A and others. 2008. Direct selection of Pichia pastoris expression strains using new G418 resistance vectors. Yeast 25(4):293-9.

Maier T, Güell M, Serrano L. 2009. Correlation of mRNA and protein in complex biological samples. FEBS Lett 583(24):3966-73.

Marx H, Mecklenbräuker A, Gasser B, Sauer M, Mattanovich D. 2009. Directed gene copy number amplification in Pichia pastoris by vector integration into the ribosomal DNA locus. FEMS Yeast Res 9(8):1260-70.

Mattanovich D, Callewaert N, Rouzé P, Lin Y, Graf A, Redl A, Tiels P, Gasser B, De Schutter K. 2009b. Open access to sequence: browsing the Pichia pastoris genome. Microb Cell Fact 8:53.

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.

Meehl MA, Stadheim TA. 2014. Biopharmaceutical discovery and production in yeast. Curr Opin Biotechnol 30:120-7.

Morrow K. 2007. Improving protein production processes. Gen. Eng. News. p 50-54.

Neubauer P, Cruz N, Glauche F, Junne S, Knepper A, Raven M. 2013. Consistent development of bioprocesses from microliter cultures to the industrial scale. Engineering in Life Sciences 13(3):224-238.

Nielsen J, Jewett MC. 2008. Impact of systems biology on metabolic engineering of Saccharomyces cerevisiae. FEMS Yeast Res 8(1):122-31.

Nocon J, Steiger MG, Pfeffer M, Sohn SB, Kim TY, Maurer M, Rußmayer H, Pflügl S, Ask M, Haberhauer-Troyer C and others. 2014. Model based engineering of Pichia pastoris central metabolism enhances recombinant protein production. Metab Eng 24:129-38.

Noorman H. 2011. An industrial perspective on bioreactor scale-down: what we can learn from combined large-scale bioprocess and model fluid studies. Biotechnol J 6(8):934-43.

Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A. 2012. Deletion of the Pichia pastoris KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. PLoS One 7(6):e39720.

Ohi H, Okazaki N, Uno S, Miura M, Hiramatsu R. 1998. Chromosomal DNA patterns and gene stability of Pichia pastoris. Yeast 14(10):895-903.

Orth JD, Thiele I, Palsson B. 2010. What is flux balance analysis? Nat Biotechnol

28(3):245-8.

Paradise WA, Vesper BJ, Babich M, Colvard MD, Radosevich JA. 2011. Methods, Processes and Comparative Yield Economics for the Production of Antibodies and Recombinant Proteins. Immunology Endocrine & Metabolic Agents in Medicinal Chemistry 11(3):199-219.

Petranovic D, Vemuri GN. 2009. Impact of yeast systems biology on industrial biotechnology. J Biotechnol 144(3):204-11.

Plotkin JB. 2010. Transcriptional regulation is only half the story. Mol Syst Biol 6:406.

Prielhofer R, Maurer M, Klein J, Wenger J, Kiziak C, Gasser B, Mattanovich D. 2013. Induction without methanol: novel regulated promoters enable high-level expression in Pichia pastoris. Microb Cell Fact 12:5.

Roberts LD, Souza AL, Gerszten RE, Clish CB. 2012. Targeted metabolomics. Curr Protoc Mol Biol Chapter 30:Unit 30.2.1-24.

Rußmayer H, Buchetics M, Gruber C, Valli M, Grillitsch K, Modarres G, Guerrasio R, Klavins K, Neubauer S, Drexler H and others. 2015. Systems-level organization of yeast methylotrophic lifestyle. BMC Biol 13:80.

Sanchez-Garcia L, Martín L, Mangues R, Ferrer-Miralles N, Vázquez E, Villaverde A. 2016. Recombinant pharmaceuticals from microbial cells: a 2015 update. Microb Cell Fact 15:33.

Sauer U. 2006. Metabolic networks in motion: 13C-based flux analysis. Mol Syst Biol 2:62.

Shalon D, Smith SJ, Brown PO. 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res 6(7):639-45.

Smyth GK, Speed T. 2003. Normalization of cDNA microarray data. Methods 31(4):265-73.

Snyder M, Gallagher JE. 2009. Systems biology from a yeast omics perspective. FEBS Lett 583(24):3895-9.

Sohn SB, Graf AB, Kim TY, Gasser B, Maurer M, Ferrer P, Mattanovich D, Lee SY. 2010. Genome-scale metabolic model of methylotrophic yeast Pichia pastoris and its use for in silico analysis of heterologous protein production. Biotechnol J 5(7):705-15.

Spadiut O, Capone S, Krainer F, Glieder A, Herwig C. 2014. Microbials for the production of monoclonal antibodies and antibody fragments. Trends Biotechnol 32(1):54-60.

Sreekrishana K, Nelles L, Potenz R, Wood P. 1999. Pichia pastoris: A eukaryotic system for the large-scale production of biopharmaceuticals. Biopharm International. p 36-42.

Sreekrishna K, Nelles L, Potenz R, Cruze J, Mazzaferro P, Fish W, Fuke M, Holden K, Phelps D, Wood P. 1989. High-level expression, purification, and

characterization of recombinant human tumor necrosis factor synthesized in the methylotrophic yeast Pichia pastoris. Biochemistry 28(9):4117-25.

Stadlmayr G, Benakovitsch K, Gasser B, Mattanovich D, Sauer M. 2010a. Genome-scale analysis of library sorting (GALibSo): Isolation of secretion enhancing factors for recombinant protein production in Pichia pastoris. Biotechnol Bioeng 105(3):543-55.

Stadlmayr G, Mecklenbräuker A, Rothmüller M, Maurer M, Sauer M, Mattanovich D, Gasser B. 2010b. Identification and characterisation of novel Pichia pastoris promoters for heterologous protein production. J Biotechnol 150(4):519-29.

Takagi S, Tsutsumi N, Terui Y, Kong X. 2008. Method for methanol independent induction from methanol inducible promoters in Pichia. United States patent US 8,143,023.

Tang T, François N, Glatigny A, Agier N, Mucchielli MH, Aggerbeck L, Delacroix H. 2007. Expression ratio evaluation in two-colour microarray experiments is significantly improved by correcting image misalignment. Bioinformatics 23(20):2686-91.

Valli M, Tatto NE, Peymann A, Gruber C, Landes N, Ekker H, Thallinger GG, Mattanovich D, Gasser B, Graf AB. 2016. Curation of the genome annotation of Pichia pastoris (Komagataella phaffii) CBS7435 from gene level to protein function. FEMS Yeast Res 16(6).

Vervecken W, Callewaert N, Kaigorodov V, Geysens S, Contreras R. 2007. Modification of the N-glycosylation pathway to produce homogeneous, human-like glycans using GlycoSwitch plasmids. Methods Mol Biol 389:119-38.

Vervecken W, Kaigorodov V, Callewaert N, Geysens S, De Vusser K, Contreras R. 2004. In vivo synthesis of mammalian-like, hybrid-type N-glycans in Pichia pastoris. Appl Environ Microbiol 70(5):2639-46.

Vogel C, Abreu ReS, Ko D, Le SY, Shapiro BA, Burns SC, Sandhu D, Boutz DR, Marcotte EM, Penalva LO. 2010. Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. Mol Syst Biol 6:400.

Vogel C, Marcotte EM. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet 13(4):227-32.

Wang Y, Liang ZH, Zhang YS, Yao SY, Xu YG, Tang YH, Zhu SQ, Cui DF, Feng YM. 2001. Human insulin from a precursor overexpressed in the methylotrophic yeast Pichia pastoris and a simple procedure for purifying the expression product. Biotechnol Bioeng 73(1):74-9.

Wiechert W. 2002. An introduction to 13C metabolic flux analysis. Genet Eng (N Y) 24:215-38.

Woo Y, Affourtit J, Daigle S, Viale A, Johnson K, Naggert J, Churchill G. 2004. A comparison of cDNA, oligonucleotide, and Affymetrix GeneChip gene expression microarray platforms. J Biomol Tech 15(4):276-84.

Ye RW, Wang T, Bedzyk L, Croker KM. 2001. Applications of DNA microarrays in

microbial systems. J Microbiol Methods 47(3):257-72.

Zahurak M, Parmigiani G, Yu W, Scharpf RB, Berman D, Schaeffer E, Shabbeer S, Cope L. 2007. Pre-processing Agilent microarray data. BMC Bioinformatics 8:142.

# 7 List of publications

**Reverse engineering of protein secretion by uncoupling of cell cycle phases from growth.** **Buchetics M**, Dragosits M, Maurer M, Rebnegger C, Porro D, Sauer M, Gasser B, Mattanovich D. *Biotechnol Bioeng. 2011 Oct; 108(10):2403-12. Published online 2011 May 25.*

***Pichia pastoris* Aft1 - a novel transcription factor, enhancing recombinant protein secretion.** Ruth C*, **Buchetics M***, Vidimce V, Kotz D, Naschberger S, Mattanovich D, Pichler H, Gasser B. *Microb Cell Fact. 2014; 13: 120. Published online 2014 Sep 3.*

**Systems-level organization of yeast methylotrophic lifestyle.** Rußmayer H*, **Buchetics M***, Gruber C, Valli M, Grillitsch K, Modarres G, Guerrasio R, Klavins K, Neubauer S, Drexler H, Steiger M, Troyer C, Al Chalabi A, Krebiehl G, Sonntag D, Zellnig G, Daum G, Graf A, Altmann F, Koellensperger G, Hann S, Sauer M, Mattanovich D, Gasser B. *BMC Biol. 2015; 13: 80. 2015 Sep 23.*

**High-throughput cultivation techniques for *Pichia pastoris.*** **Buchetics M**, Gasser B, Weis R. (paper accepted for publication in Pichia protocols, Third Edition; edited by Harald Pichler, TU Graz; to be published by Springer Science+Business Media, LLC)

## Manuscripts in preparation

**Customizing amino acid metabolism of *Pichia pastoris* for recombinant protein production.** Rußmayer H, **Buchetics M**, Neubauer S, Steiger M, Hann S, Sauer M, Mattanovich D, Gasser B

*equally contributing authors

## Oral presentations

**Reverse engineering of protein secretion by uncoupling of cell cycle phases from growth.** Buchetics M, Dragosits M, Maurer M, Rebnegger C, Porro D, Sauer M, Gasser B, Mattanovich D. *Pichia 2012 Conference. February 29 - March 3, Alpbach, Austria*

## Poster presentations

**Reverse engineering of protein secretion by uncoupling of cell cycle phases from growth.** Buchetics M, Dragosits M, Maurer M, Rebnegger C, Porro D, Sauer M, Gasser B, Mattanovich D. *RPP6 - 6th Conference on Recombinant Protein Production, Vienna, Austria, February 16-19, 2011*

**Reverse engineering of protein secretion by uncoupling of cell cycle phases from growth.** Buchetics M, Dragosits M, Maurer M, Rebnegger C, Porro D, Sauer M, Gasser B, Mattanovich D. *Pichia 2012 Conference. February 29 - March 3*

**Influence of the carbon source on the transcriptome of *Pichia pastoris* cultures.** Buchetics M, Graf A, Valli M, Mattanovich D, Gasser B. *13th International Congress on Yeasts (ICY 2012) – Yeasts for a sustainable future*, *26.8.2012 – 30.8.2012, Madison, Wisconsin, USA*

# 8 Curriculum vitae

| | |
|---|---|
| Name: | Markus Buchetics |
| Date of birth: | 04.07.1984 |
| Place of birth: | Oberwart |
| Nationality: | Austria |
| Adress: | Pflegersiedlung 16, 7551 Stegersbach |

**Education**

| | |
|---|---|
| 2010 – 2016 | PhD thesis at the Institute for Applied Microbiology at the University of Natural Resources and Life Sciences under supervision of Univ. Prof. DI Dr. Diethard Mattanovich |
| May 2010 | degree: Diplomingenieur / Master of Science (with distinction) |
| 2009 – 2010 | Diploma thesis at the Institute for Applied Microbiology at the University of Natural Resources and Life Sciences under supervision of Univ. Prof. DI Dr. Diethard Mattanovich |
| 2008 – 2010 | Master studies of Biotechnology at the University of Natural Resources and Life Sciences |
| June 2007 | degree: Bachelor of Science |
| 2003 – 2007 | Bachelor studies of Food and Biotechnology at the University of Natural Resources and Life Sciences |
| 2002 – 2003 | Military service |
| June 2002 | Matura (with distinction) |
| 1994 – 2002 | Bundesgymnasium Oberschützen |
| 1990 – 1994 | Elementary school in Stegersbach |

MICROBIAL CELL FACTORIES

RESEARCH Open Access

# *Pichia pastoris* Aft1 - a novel transcription factor, enhancing recombinant protein secretion

Claudia Ruth[1†], Markus Buchetics[2†], Viktorija Vidimce[1], Daniela Kotz[3], Stefan Naschberger[3], Diethard Mattanovich[2,4], Harald Pichler[1,5*] and Brigitte Gasser[2,4*]

## Abstract

**Background:** The methylotrophic yeast *Pichia pastoris* is frequently used for the production of recombinant proteins. However, expression levels can vary depending on the target protein. Allowing for simultaneous regulation of many genes, which may elicit a desired phenotype like increased protein production, overexpression of transcription factors can be used to overcome expression bottlenecks. Here, we present a novel *P. pastoris* transcription factor currently annotated as Aft1, activator of ferrous transport.

**Results:** The promoter regions of key secretory *P. pastoris* genes were screened for fungal transcription factor binding sites, revealing Aft1 as an interesting candidate for improving secretion. Genome wide analysis of transcription factor binding sites suggested Aft1 to be involved in the regulation of many secretory genes, but also indicated possible novel functions in carbohydrate metabolism. No Aft binding sites were found in promoters of characteristic iron homeostasis genes in *P. pastoris*. Microarrays were used to study the Aft1 regulon in detail, confirming Aft1 involvement in the regulation of carbon-responsive genes, and showing that iron regulation is dependent on *FEP1*, but not *AFT1* expression levels. The positive effect of *AFT1* overexpression on recombinant protein secretion was demonstrated for a carboxylesterase from *Sphingopyxis* sp. MTA144, for which secretion was improved 2.5-fold in fed batch bioreactor cultivations.

**Conclusion:** This study demonstrates that the transcription factor Aft1 can be used to improve recombinant protein secretion in *P. pastoris*. Furthermore, we discovered possible novel functions of Aft1 in carbohydrate metabolism and provide evidence arguing against a direct role of Aft1 in *P. pastoris* iron regulation.

**Keywords:** *Pichia pastoris*, Aft1, Transcription factor, Novel functions, Enhanced secretion

## Background

The methylotrophic yeast *Pichia pastoris* (syn. *Komagataella phaffii*) is among today's most frequently used yeast systems for the production of recombinant proteins [1]. Benefits of this yeast are the capability of high cell density cultivations, eukaryotic posttranslational modifications and good secretion capacity. A low level of endogenously secreted proteins allows for the production of relatively pure, recombinant secretory proteins. The recent availability of the genomic sequence boosted the generation of a versatile *P. pastoris* toolbox, including various strains, plasmids and promoters of different strength. To overcome individual bottlenecks during protein folding and secretion, a variety of helper factors such as the ER foldases Pdi1 or BiP (Kar2) have been studied in recent years [2,3]. The capability of transcription factors (TFs) as expression helpers was demonstrated by Guerfal *et al.* and Gasser *et al.* [4,5], who improved the secretion of the mIL-10 protein and antibody fragments by overexpression of the UPR (unfolded protein response) transcription factor *HAC1*. Also overexpression of the gene encoding the TF Nrg1 was shown to positively influence the secretion of recombinant porcine and human trypsinogen as well as the antibody Fab fragment 2 F5 [6].

Allowing for the simultaneous regulation of different proteins involved in e.g. folding and secretion, TFs have

* Correspondence: harald.pichler@tugraz.at; brigitte.gasser@boku.ac.at
†Equal contributors
[1]Austrian Centre of Industrial Biotechnology (ACIB GmbH), Petersgasse 14, 8010 Graz, Austria
[2]Austrian Centre of Industrial Biotechnology (ACIB GmbH), Muthgasse 11, 1190 Wien, Austria
Full list of author information is available at the end of the article

Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. the Gene Ontology Consortium.** *Nat Genet* 2000, **25:**25–29.
26. Li L, Chen OS, McVey Ward D, Kaplan J: **CCC1 is a transporter that mediates vacuolar iron storage in yeast.** *J Biol Chem* 2001, **276:**29515–29519.
27. Greene JR, Brown NH, DiDomenico BJ, Kaplan J, Eide DJ: **The GEF1 gene of Saccharomyces cerevisiae encodes an integral membrane protein; mutations in which have effects on respiration and iron-limited growth.** *Mol Gen Genet* 1993, **241:**542–553.
28. Kim D, Yukl ET, Moënne-Loccoz P, Montellano PR: **Fungal heme oxygenases: functional expression and characterization of Hmx1 from Saccharomyces cerevisiae and CaHmx1 from Candida albicans.** *Biochemistry* 2006, **45:**14772–14780.
29. Collinson EJ, Wimmer-Kleikamp S, Gerega SK, Yang YH, Parish CR, Dawes IW, Stocker R: **The yeast homolog of heme oxygenase-1 affords cellular antioxidant protection via the transcriptional regulation of known antioxidant genes.** *J Biol Chem* 2011, **286:**2205–2214.
30. Schilke B, Voisine C, Beinert H, Craig E: **Evidence for a conserved system for iron metabolism in the mitochondria of Saccharomyces cerevisiae.** *Proc Natl Acad Sci U S A* 1999, **96:**10206–10211.
31. Lee MCS, Miller EA: **Molecular mechanisms of COPII vesicle formation.** *Semin Cell Dev Biol* 2007, **18:**424–434.
32. Römisch K: **Surfing the Sec61 channel: bidirectional protein translocation across the ER membrane.** *J Cell Sci* 1999, **112**(Pt 2):4185–4191.
33. McNew JA, Coe JG, Søgaard M, Zemelman BV, Wimmer C, Hong W, Söllner TH: **Gos1p, a Saccharomyces cerevisiae SNARE protein involved in Golgi transport.** *FEBS Lett* 1998, **435:**89–95.
34. Kuge S, Jones N: **YAP1 dependent activation of TRX2 is essential for the response of Saccharomyces cerevisiae to oxidative stress by hydroperoxides.** *EMBO J* 1994, **13:**655–664.
35. Vervecken W, Kaigorodov V, Callewaert N, Geysens S, De Vusser K, Contreras R: **In vivo synthesis of mammalian-like, hybrid-type N-glycans in pichia pastoris.** 2004, **70:**2639–2646.
36. Elbert M, Rossi G, Brennwald P: **The yeast par-1 homologs kin1 and kin2 show genetic and physical interactions with components of the exocytic machinery.** *Mol Biol Cell* 2005, **16:**532–549.
37. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23:**2947–2948.
38. Rutherford JC, Jaron S, Ray E, Brown PO, Winge DR: **A second iron-regulatory system in yeast independent of Aft1p.** *Proc Natl Acad Sci U S A* 2001, **98:**14322–14327.
39. Babu MM, Iyer LM, Balaji S, Aravind L: **The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons.** *Nucleic Acids Res* 2006, **34:**6505–6520.
40. Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD: *Molecular Biology of the Cell, 3rd Addition.* New York: Garland Science; 1994.
41. Heiss S, Maurer M, Hahn R, Mattanovich D, Gasser B: **Identification and deletion of the major secreted protein of Pichia pastoris.** *Appl Microbiol Biotechnol* 2013, **97:**1241–1249.
42. 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, **9:**390.
43. Baumann K, Carnicer M, Dragosits M, Graf AB, Stadlmann J, Jouhten P, Maaheimo H, Gasser B, Albiol J, Mattanovich D, Ferrer P: **A multi-level study of recombinant Pichia pastoris in different oxygen conditions.** *BMC Syst Biol* 2010, **4:**141.
44. Nocon J, Steiger MG, Pfeffer M, Sohn SB, Kim TY, Maurer M, Rußmayer H, Pflügl S, Ask M, Haberhauer-Troyer C, Ortmayr K, Hann S, Koellensperger G, Gasser B, Lee SY, Mattanovich D: **Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production.** *Metab Eng* 2014, **24:**129–138. doi:10.1016/j.ymben.2014.05.011.
45. Crameri A, Whitehorn EA, Tate E, Stemmer WP: **Improved green fluorescent protein by molecular evolution using DNA shuffling.** *Nat Biotechnol* 1996, **14:**315–319.
46. Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, Lin-Cereghino J, Kovar K, Cregg JM, Glieder A: **Promoter library designed for fine-tuned gene expression in Pichia pastoris.** *Nucleic Acids Res* 2008, **36:**e76.
47. Hartinger D, Moll W-D: **Fumonisin elimination and prospects for detoxification by enzymatic transformation.** *World Mycotoxin J* 2011, **4:**271–283.
48. Heinl S, Hartinger D, Thamhesl M, Vekiru E, Krska R, Schatzmayr G, Moll W-D, Grabherr R: **Degradation of fumonisin B1 by the consecutive action of two bacterial enzymes.** *J Biotechnol* 2010, **145:**120–129.
49. Baumann K, Maurer M, Dragosits M, Cos O, Ferrer P, Mattanovich D: **Hypoxic fed-batch cultivation ofPichia pastoris increases specific and volumetric productivity of recombinant proteins.** *Biotechnol Bioeng* 2008, **100:**177–183.
50. Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, Jaenicke S, Brinkrolf K, Goesmann A, Szczepanowski R, Pühler A, Schwab H, Glieder A, Pichler H: **High-quality genome sequence of Pichia pastoris CBS7435.** *J Biotechnol* 2011, **154:**312–320.
51. Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A: **Deletion of the Pichia pastoris KU70 homologue facilitates platform strain generation for gene expression and synthetic biology.** *PLoS One* 2012, **7:**e39720.
52. Weis R, Luiten R, Skranc W, Schwab H, Wubbolts M, Glieder A: **Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena.** *FEMS Yeast Res* 2004, **5:**179–189.
53. Zhao W, Wang J, Deng R, Wang X: **Scale-up fermentation of recombinant Candida rugosa lipase expressed in Pichia pastoris using the GAP promoter.** *J Ind Microbiol Biotechnol* 2008, **35:**189–195.
54. Abad S, Kitz K, Hörmann A, Schreiner U, Hartner FS, Glieder A: **Real-time PCR-based determination of gene copy numbers in Pichia pastoris.** *Biotechnol J* 2010, **5:**413–420.
55. Hoffman CS, Winston F: **A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli.** *Gene* 1987, **57:**267–272.

doi:10.1186/s12934-014-0120-5
**Cite this article as:** Ruth *et al.*: *Pichia pastoris* Aft1 - a novel transcription factor, enhancing recombinant protein secretion. *Microbial Cell Factories* 2014 **13**:120.

Rußmayer et al. BMC Biology (2015) 13:80
DOI 10.1186/s12915-015-0186-5

BMC Biology

RESEARCH ARTICLE Open Access

# Systems-level organization of yeast methylotrophic lifestyle

Hannes Rußmayer[1,2†], Markus Buchetics[1,2†], Clemens Gruber[2,3], Minoska Valli[1,2], Karlheinz Grillitsch[4,5], Gerda Modarres[2,6], Raffaele Guerrasio[2,3,10], Kristaps Klavins[2,3,11], Stefan Neubauer[2,3,12], Hedda Drexler[2,3], Matthias Steiger[1,2], Christina Troyer[2,3], Ali Al Chalabi[7], Guido Krebiehl[7], Denise Sonntag[7], Günther Zellnig[8], Günther Daum[4,5], Alexandra B. Graf[2,5], Friedrich Altmann[2,3], Gunda Koellensperger[9], Stephan Hann[2,3], Michael Sauer[1,2], Diethard Mattanovich[1,2*] and Brigitte Gasser[1,2]

## Abstract

**Background:** Some yeasts have evolved a methylotrophic lifestyle enabling them to utilize the single carbon compound methanol as a carbon and energy source. Among them, *Pichia pastoris* (syn. *Komagataella sp.*) is frequently used for the production of heterologous proteins and also serves as a model organism for organelle research. Our current knowledge of methylotrophic lifestyle mainly derives from sophisticated biochemical studies which identified many key methanol utilization enzymes such as alcohol oxidase and dihydroxyacetone synthase and their localization to the peroxisomes. C1 assimilation is supposed to involve the pentose phosphate pathway, but details of these reactions are not known to date.

**Results:** In this work we analyzed the regulation patterns of 5,354 genes, 575 proteins, 141 metabolites, and fluxes through 39 reactions of *P. pastoris* comparing growth on glucose and on a methanol/glycerol mixed medium, respectively. Contrary to previous assumptions, we found that the entire methanol assimilation pathway is localized to peroxisomes rather than employing part of the cytosolic pentose phosphate pathway for xylulose-5-phosphate regeneration. For this purpose, *P. pastoris* (and presumably also other methylotrophic yeasts) have evolved a duplicated methanol inducible enzyme set targeted to peroxisomes. This compartmentalized cyclic C1 assimilation process termed xylose-monophosphate cycle resembles the principle of the Calvin cycle and uses sedoheptulose-1,7-bisphosphate as intermediate. The strong induction of alcohol oxidase, dihydroxyacetone synthase, formaldehyde and formate dehydrogenase, and catalase leads to high demand of their cofactors riboflavin, thiamine, nicotinamide, and heme, respectively, which is reflected in strong up-regulation of the respective synthesis pathways on methanol. Methanol-grown cells have a higher protein but lower free amino acid content, which can be attributed to the high drain towards methanol metabolic enzymes and their cofactors. In context with up-regulation of many amino acid biosynthesis genes or proteins, this visualizes an increased flux towards amino acid and protein synthesis which is reflected also in increased levels of transcripts and/or proteins related to ribosome biogenesis and translation.

**Conclusions:** Taken together, our work illustrates how concerted interpretation of multiple levels of systems biology data can contribute to elucidation of yet unknown cellular pathways and revolutionize our understanding of cellular biology.

**Keywords:** Metabolome, Methanol, Peroxisome, *Pichia pastoris*, Proteome, Transcriptome, Xylulose-monophosphate cycle

* Correspondence: diethard.mattanovich@boku.ac.at
†Equal contributors
[1]Department of Biotechnology, BOKU - University of Natural Resources and Life Sciences Vienna, Muthgasse 18, 1190 Vienna, Austria
[2]Austrian Centre of Industrial Biotechnology, A-1190 Vienna, Austria
Full list of author information is available at the end of the article

40. Elgersma Y, Elgersma-Hooisma M, Wenzel T, McCaffery JM, Farquhar MG, Subramani S. A mobile PTS2 receptor for peroxisomal protein import in *Pichia pastoris*. J Cell Biol. 1998;140:807–20.
41. Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, et al. High-quality genome sequence of *Pichia pastoris* CBS7435. J Biotechnol. 2011;154:312–20.
42. Neuberger G, Maurer-Stroh S, Eisenhaber B, Hartig A, Eisenhaber F. Prediction of peroxisomal targeting signal 1 containing proteins from amino acid sequence. J Mol Biol. 2003;328:581–92.
43. PTS1 Predictor. http://mendel.imp.ac.at/mendeljsp/sat/pts1/PTS1predictor.jsp. Last access April 28, 2015.
44. Wriessnegger T, Gübitz G, Leitner E, Ingolic E, Cregg J, de la Cruz B, et al. Lipid composition of peroxisomes from the yeast *Pichia pastoris* grown on different carbon sources. Biochim Biophys Acta. 1771;2007:455–61.
45. Luers GH, Advani R, Wenzel T, Subramani S. The *Pichia pastoris* dihydroxyacetone kinase is a PTS1-containing, but cytosolic, protein that is essential for growth on methanol. Yeast. 1998;14:759–71.
46. Clasquin MF, Melamud E, Singer A, Gooding JR, Xu X, Dong A, et al. Riboneogenesis in yeast. Cell. 2011;145:969–80.
47. Douma AC, Veenhuis M, de Koning W, Evers M, Harder W. Dihydroxyacetone synthase is localized in the peroxisomal matrix of methanol-grown *Hansenula polymorpha*. Arch Microbiol. 1985;143:237–43.
48. Anderson LE, Carol AA. Enzyme co-localization with rubisco in pea leaf chloroplasts. Photosynth Res. 2004;82:49–58.
49. Rae BD, Long BM, Whitehead LF, Forster B, Badger MR, Price GD. Cyanobacterial carboxysomes: microcompartments that facilitate $CO_2$ fixation. J Mol Microbiol Biotechnol. 2013;23:300–7.
50. Hahn MW. Distinguishing among evolutionary models for the maintenance of gene duplicates. J Hered. 2009;100:605–17.
51. Zhang J. Evolution by gene duplication: an update. Trends Ecol Evol. 2003;18:292–8.
52. Byun-McKay SA, Geeta R. Protein subcellular relocalization: a new perspective on the origin of novel genes. Trends Ecol Evol. 2007;22:338–44.
53. Daran-Lapujade P, Rossell S, van Gulik WM, Luttik MA, de Groot MJ, Slijper M, et al. The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. Proc Natl Acad Sci U S A. 2007;104:15753–8.
54. Tibbetts AS, Sun Y, Lyon NA, Ghrist AC, Trotter PJ. Yeast mitochondrial oxodicarboxylate transporters are important for growth on oleic acid. Arch Biochem Biophys. 2002;406:96–104.
55. Cavero S, Vozza A, del Arco A, Palmieri L, Villa A, Blanco E, et al. Identification and metabolic role of the mitochondrial aspartate-glutamate transporter in *Saccharomyces cerevisiae*. Mol Microbiol. 2003;50:1257–69.
56. Kunze M, Pracharoenwattana I, Smith SM, Hartig A. A central role for the peroxisomal membrane in glyoxylate cycle function. Biochim Biophys Acta. 1763;2006:1441–52.
57. Warner JR, Mitra G, Schwindinger WF, Studeny M, Fried HM. *Saccharomyces cerevisiae* coordinates accumulation of yeast ribosomal proteins by modulating mRNA splicing, translational initiation, and protein turnover. Mol Cell Biol. 1985;5:1512–21.
58. Puxbaum V, Mattanovich D, Gasser B. Quo vadis? The challenges of recombinant protein folding and secretion in *Pichia pastoris*. Appl Microbiol Biotechnol. 2015;99:2925–38.
59. 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;85:367–75.
60. 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.
61. Ozimek P, van Dijk R, Latchev K, Gancedo C, Wang DY, van der Klei IJ, et al. Pyruvate carboxylase is an essential protein in the assembly of yeast peroxisomal oligomeric alcohol oxidase. Mol Biol Cell. 2003;14:786–97.
62. Stewart MQ, Esposito RD, Gowani J, Goodman JM. Alcohol oxidase and dihydroxyacetone synthase, the abundant peroxisomal proteins of methylotrophic yeasts, assemble in different cellular compartments. J Cell Sci. 2001;114:2863–8.
63. Waterham HR, Russell KA, Vries Y, Cregg JM. Peroxisomal targeting, import, and assembly of alcohol oxidase in *Pichia pastoris*. J Cell Biol. 1997;139:1419–31.
64. Gudipati V, Koch K, Lienhart WD, Macheroux P. The flavoproteome of the yeast *Saccharomyces cerevisiae*. Biochim Biophys Acta. 1844;2014:535–44.
65. Marx H, Mattanovich D, Sauer M. Overexpression of the riboflavin biosynthetic pathway in *Pichia pastoris*. Microb Cell Fact. 2008;7:23.
66. Stadlmayr G, Mecklenbrauker A, Rothmuller M, Maurer M, Sauer M, Mattanovich D, et al. Identification and characterisation of novel *Pichia pastoris* promoters for heterologous protein production. J Biotechnol. 2010;150:519–29.
67. Sporty J, Lin SJ, Kato M, Ognibene T, Stewart B, Turteltaub K, et al. Quantitation of $NAD^+$ biosynthesis from the salvage pathway in *Saccharomyces cerevisiae*. Yeast. 2009;26:363–9.
68. Wriessnegger T, Sunga AJ, Cregg JM, Daum G. Identification of phosphatidylserine decarboxylases 1 and 2 from *Pichia pastoris*. FEMS Yeast Res. 2009;9:911–22.
69. Verduyn C, Postma E, Scheffers WA, van Dijken JP. Physiology of *Saccharomyces cerevisiae* in anaerobic glucose-limited chemostat cultures. J Gen Microbiol. 1990;136:395–403.
70. Folch J, Lees M, Sloane Stanley GH. A simple method for the isolation and purification of total lipides from animal tissues. J Biol Chem. 1957;226:497–509.
71. Broekhuyse RM. Phospholipids in tissues of the eye. I. Isolation, characterization and quantitative analysis by two-dimensional thin-layer chromatography of diacyl and vinyl-ether phospholipids. Biochim Biophys Acta. 1968;152:307–15.
72. Benjamini Y, Heller R, Yekutieli D. Selective inference in complex research. Philos Trans A Math Phys Eng Sci. 2009;367:4255–71.
73. R project. http://www.r-project.org. Last access January 31, 2012.
74. Pichler P, Kocher T, Holzmann J, Mazanek M, Taus T, Ammerer G, et al. Peptide labeling with isobaric tags yields higher identification rates using iTRAQ 4-plex compared to TMT 6-plex and iTRAQ 8-plex on LTQ Orbitrap. Anal Chem. 2010;82:6549–58.
75. Breitwieser FP, Muller A, Dayon L, Kocher T, Hainard A, Pichler P, et al. General statistical modeling of data from protein relative expression isobaric tags. J Proteome Res. 2011;10:2758–66.
76. Klavins K, Neubauer S, Al Chalabi A, Sonntag D, Haberhauer-Troyer C, Russmayer H, et al. Interlaboratory comparison for quantitative primary metabolite profiling in *Pichia pastoris*. Anal Bioanal Chem. 2013;405:5159–69.
77. Neubauer S, Haberhauer-Troyer C, Klavins K, Russmayer H, Steiger MG, Gasser B, et al. $U^{13}C$ cell extract of *Pichia pastoris*–a powerful tool for evaluation of sample preparation in metabolomics. J Sep Sci. 2012;35:3091–105.
78. Zamboni N, Fischer E, Sauer U. FiatFlux–a software for metabolic flux analysis from $^{13}C$-glucose experiments. BMC Bioinformatics. 2005;6:209.
79. Quek LE, Wittmann C, Nielsen LK, Kromer JO. OpenFLUX: efficient modelling software for $^{13}C$-based metabolic flux analysis. Microb Cell Fact. 2009;8:25.

## 4.4 Comparison of gene and protein regulations in various *P. pastoris* production strains with different carbon sources

Product titers in recombinant protein production with *P. pastoris* were often observed to vary tremendously, depending on the heterologous gene product (and its copy number), the promoter used for expression thereof, the strain/clone background, or the carbon source used for cultivation and/or induction of protein production.

In order to elucidate the implications of protein production conditions on the proteome of the cell, an experimental setup was created to compare *P. pastoris* production strains with non-producing control strains. In case of the recombinant product HyHEL Fab, this was done under control of the GAP promoter with glucose as carbon source but also under control of the AOX promoter with a glycerol/methanol medium for growth and induction of protein production. Figure 5 shows the different comparisons that were achieved by cultivating five different strains (three production strains, two control strains) in chemostats in three biological replicates in order to obtain samples for transcriptomic and proteomic analysis.

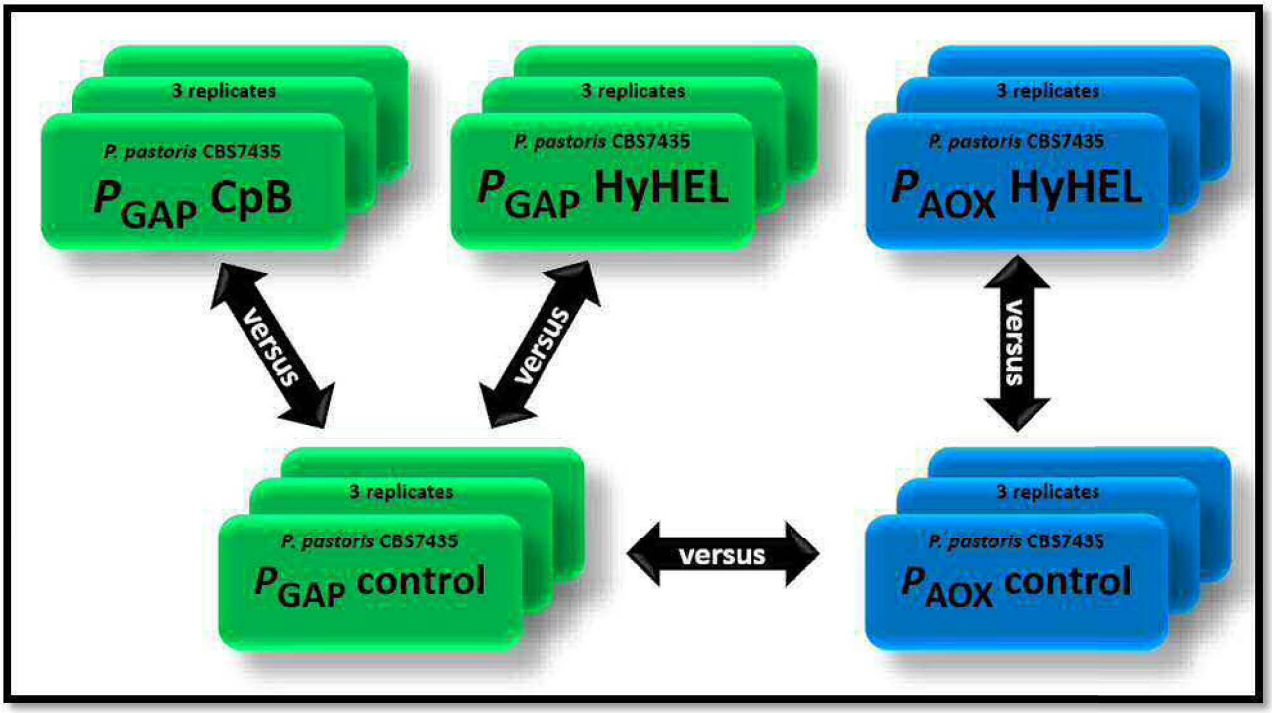


**Figure 5. Strain comparison experiments for the comparative analysis of the proteome and transcriptome of *P. pastoris* production strains.** The strains were cultivated in chemostats in three biological replicates each.