The Impact of Amino Acid Metabolism on Recombinant Protein Production in *Pichia pastoris*

Dissertation

zur Erlangung des Doktorgrades an der Universität für Bodenkultur Wien

eingereicht von

Dipl. Ing. Hannes Rußmayer

Betreuer:

Univ. Prof. DI Dr. Diethard Mattanovich

Department für Biotechnologie

Universität für Bodenkultur Wien

Wien, im September 2015

Danksagung

Mein Dank gilt meinen Eltern und meiner Familie, die mich während des gesamten Studiums unterstützt haben.

Ich danke Herrn Univ. Prof. Dr. Diethard Mattanovich für die Chance diese Dissertation in seiner Arbeitsgruppe durchzuführen. Danke, für deine Unterstützung während der gesamten Zeit.

Ich danke Frau Priv. Doz. Dr. Brigitte Gasser und Herrn Priv. Doz. Dr. Michael Sauer für die Betreuung meiner Dissertation. Danke, für das Vertrauen in mich und für die vielen Ratschläge und zielführenden Hinweise.

Ich danke auch Herrn Dr. Hans Marx für die Hilfe bei der Beantwortung so manch kniffliger Frage. Danke, dass der oberösterreichische Assistent so viel vom burgenländischen Erfinder lernen durfte.

Mein Dank gilt auch der Arbeitsgruppe Instrumentelle Analytische Chemie und Metabolomics. Im Speziellen möchte ich mich bei Stephan Hann, Gunda Köllensberger, Christina Haberhauer-Troyer, Stefan Neubauer und Kristaps Klavins bedanken.

Ich möchte mich auch beim gesamten Team 107 für das super Arbeitsklima bedanken. Ich möchte keinen einzigen Arbeitstag missen. Besonders bedanken möchte ich mich bei Frau Marlene Thierrichter und Frau Stefanie Müller für die Hilfe bei den unzähligen Experimenten. Mein spezieller Dank gilt auch Stefan Pflügl für die lustige Zeit im und außerhalb des Labors. TANTRUM !!!!!! Es war eine super Zeit.

Contents

Zusammenfassung

Die methylotrophe Hefe *Pichia pastoris* (syn. *Komagataella sp.*) ist ein häufig verwendeter Wirtsorganismus für die Produktion von rekombinanten Proteinen. Eine Vielzahl von rekombinanten Proteinen wurde bereits in *P. pastoris* mit hohen Ausbeuten produziert. Einhergehend mit der erfolgreichen Expression einiger Proteine konnte aber auch eine Limitierung in der Synthesekapazität für insbesondere komplexe Proteine (z. B. Antikörper) festgestellt werden. Ist dieses Syntheselimit erreicht oder auch überschritten kommt es oftmals zu einer extremen Stresssituation für die Zelle. Solche Stresssituation manifestiert sich oft in einer reduzierte Zellfitness und einer verringerten Ausbeute des rekombinanten Proteins. Meist werden solch Situationen durch Limitierungen in spezifischen zellulären Prozessen verursacht. In Hefen wird die Ursache solcher Situationen einer begrenzten Kapazität in der Proteinfaltung und Sekretion zugeschrieben. Es kann aber des Weiteren auch eine Limitierung in metabolischen Prozessen nicht ausgeschlossen werden. Spricht man von Limitierung innerhalb des Metabolismus handelt es sich meist um eine limitierte Versorgung mit Aminosäuren, den Bausteinen der Proteine, oder eine Limitierung im Engeriestoffwechsel der Zelle. Das Ziel dieser Doktorarbeit war es nun, den Zusammenhang zwischen dem zentralen Kohlenstoffwechsels und rekombinante Protein Produktion zu analysieren. Um nun den Zusammenhang zwischen Metabolismus und rekombinanter Protein Produktion zu verstehen wurde das Metabolom und einzelnen metabolischen Reaktionsraten des zentralen Kohlenstoffwechsels von verschiedenen *P. pastoris* Produktionsstämmen analysiert. Die Analyse ergab, dass intrazellulären Aminosäurepools durch die Produktion eines rekombinanten Proteins beeinflusst werden. Um nun zu verstehen, ob eine Limitierung in der Synthesekapazität für manch Aminosäuren die rekombinante Proteinproduktion einschränkt wurden verschiedene Ansätze verfolgt. In einem ersten Experiment wurden Aminosäuren und Mischungen von Aminosäuren zum Kultivierungsmedium hinzugefügt. Durch die Zugabe der Aminosäuren konnte die Produktion eines Antikörperfragments (HyHEL-Fab) verbessert werden. Es konnte aber auch festgestellt werden, dass die Zugabe von Aminosäuren einen negativen Einfluss auf die rekombinante Proteinproduktion hat. Die Zugabe von Serin führte zu einer Verringerung der Produktion eines Antikörperfragments. Der zweite Ansatz beschäftigte sich mit der Erhöhung intrazellulärer Aminosäurekonzentrationen durch Überexpression von spezifischen Genen involviert in die Aminosäurebiosynthese. Die Überexpression von 9 aus 14 ausgewählten Genen führte zu einer Steigerung der Produktion einer Carboxylesterase um 20 %.

Abstract

The methylotrophic yeast *Pichia pastoris* (syn. *Komagataella sp.*) is a well-established host for recombinant protein production. The outstanding efficiency in recombinant protein production in terms of high product titers was already shown several times. Nevertheless, there are still limits in productivity observed. The production of a recombinant protein at a high level causes additional stress for the cell. The additional cellular stress decreases cellular fitness, which in the end decreases the productivity of the protein of interest. Such cellular stress situations are caused by limitation within specific cellular processes. In yeast expression systems the major bottlenecks for recombinant protein production are attributed to protein folding and secretion. However, there is strong evidence that limitations within metabolism constitute also a major bottleneck for recombinant protein production. Limitations within metabolism are mainly connected to limited supply of precursor metabolites (e.g. amino acids) or a limit in energy needed for the formation of the desired product. The focus of this PhD project was on the analysis of the impact of the metabolism on recombinant protein production in *P. pastoris*. In a first experiment the metabolome and the intracellular flux distribution of *P. pastoris* strains expressing different proteins were analyzed. The analyses showed that recombinant protein production directly affects intracellular amino acid pools. These findings led to the hypothesis that insufficient supply of certain amino acids may hamper protein production. To prove if amino acids are the limiting factor for recombinant protein production different approaches were conducted. In a first experiment specific amino acids and amino acid mixtures were added to the cultivation media. The addition unburdened cellular metabolism during production of an antibody fragment and increased the production of the antibody fragment if glucose was used as sole carbon source. However, the production of the same antibody fragment decreased if methanol was used as carbon source and serine was added to the cultivation media. This shows that addition of amino acids is not in general applicable for improving recombinant protein production. Therefore, a second approach was followed to increase intracellular amino acid levels via the overexpression of genes involved in amino acid metabolism. The overexpression of 9 out of 14 engineering targets succeeded in improving the production of a recombinant carboxylesterase up to 20%.

Recombinant Protein Production

The development of recombinant DNA technology in the 1970´s laid the cornerstone for today´s multi-billion dollar business of recombinant protein production. With regard to the production of recombinant proteins the largest groups are biopharmaceuticals and industrial enzymes. The group of biopharmaceuticals constitutes by far the most important one in terms of market value. In 2009, the market value for recombinant proteins was 99 billion dollars, representing the fastest growing segment of the pharmaceutical industry (Walsh, 2010). Up to now about 200 biopharmaceuticals have been approved by the Food and Drug Association. The approved biopharmaceuticals were expressed in *Escherichia coli* (31%), in yeast (15%) and in mammalian cells (43%) (Berlec & Strukelj, 2013). The segment of industrial enzymes spans different enzymes class (e.g. proteases, lipases). The market value for industrial enzymes is more than 5 billion dollars (Porro *et al.*, 2011). Especially, yeasts are often used as expression host for the production of industrial enzymes.

Recombinant Protein Production in Yeasts

Yeasts comprise several advantages over other expression hosts. Compared to *Escherichia coli*, yeasts are able to perform posttranslational modification and allow a proper folding and secretion of the recombinant protein. If set side by side with mammalian cells, yeasts allow a stable expression of the protein, higher product titers, the use of cheap media and do not contain pyrogens and viruses harmful for humans (Mattanovich *et al.*, 2012). These advantages made yeasts important hosts for recombinant protein production. The major workhorse for recombinant protein production concerning yeast expression systems is *Saccharomyces cerevisiae*. The most prominent examples is the recombinant production of insulin in *S. cerevisiae* by Novo Nordisk. Over the past years other yeast species, including *Pichia pastoris*, *Hansenula polymorpha* and *Yarrowia lipolytica*, showed to have great potential as expression host for recombinant proteins (Mattanovich *et al.*, 2012). Among these yeasts, *Pichia pastoris* (syn. *Komagataella sp.*) turned out to be the most promising candidate. Several features of the methylotrophic yeast support its important role for recombinant protein production (e.g. growth up to high cell densities accompanied with high product titers) (Mattanovich *et al.*, 2009). The availability of genetic tools and the genome sequence (Mattanovich *et al.*, 2009) allow an easy manipulation to further improve the performance of *P. pastoris* for recombinant protein production. In 2009, the FDA approved the first recombinant protein produced by *P. pastoris* and many other products are already in clinical studies. Beside the expression of therapeutical proteins *P. pastoris* turned out to be an excellent host for the production of industrial enzymes (e.g.

lipases, proteases) (Porro *et al.*, 2011). Due to the fast development of the *P. pastoris* expression system, it emerged as a serious competitor for *S. cerevisiae* on an industrial scale. Despite, the several hundred times shown high level protein production, limitations in synthesis for certain recombinant proteins were observed (e.g. for complex antibody fragments). Therefore, the focus of strain engineering is now on identification of cellular processes causing this limitation in recombinant protein production and development of strategies to overcome these limitations.

Bottlenecks for Recombinant Protein Production in Yeasts

Protein Folding and Secretion

In general, the production of recombinant proteins exerts additional cellular stress for the expression host, which is often observed by reduced cellular growth, decreased cellular fitness or increased cellular lysis (Gonzalez et al., 2002). For yeast expression systems the major limiting factor for recombinant protein production appears to be protein folding and secretion (Delic et al., 2014). The process of protein production is a series of consecutive steps (i.e. transcription, translation, protein folding and maturation and transport via the secretory pathway) involving several hundreds of proteins (Delic *et al.*, 2013). Each of these steps can be rate limiting for recombinant protein production. Engineering of the secretory pathway proved to be a valuable strategy for improving recombinant protein production (Idiris et al., 2010; Delic et al., 2014). Typical processes within the secretory pathway constituting a bottleneck for recombinant protein production are depicted in Fig.1. In principle, strain engineering with focus on protein folding and secretion concentrates on following topics: 1) engineering of the protein folding and ER quality control machinery, 2) engineering of intracellular transport, 3) minimization of intracellular proteolytic activity (Idiris *et al.*, 2010).

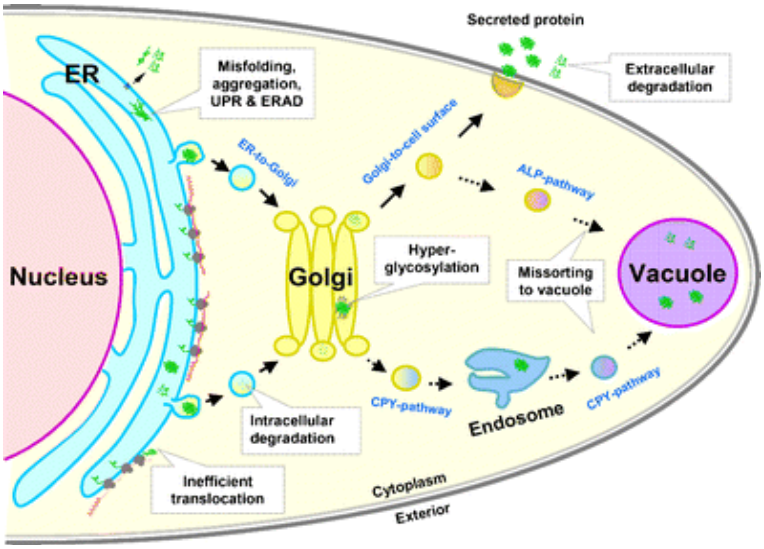

Fig.1: A schematic overview of potential bottlenecks for recombinant protein production in the secretory pathway of yeast. (Idiris et al., 2010)

The most promising and widespread strategy among the aforementioned is the modification of protein folding and ER quality control. The overexpression of multiple chaperones (e.g. Pdi1, Kar2) already showed to improve recombinant protein production in *S. cerevisiae*. For example, the overexpression of the chaperone Kar2, a member of the Hsp70 family, improved human erythropoietin in *S. cerevisiae* up to 5-fold (Idiris *et al.*, 2010). Furthermore, the overexpression of Pdi1, a protein disulfide isomerase involved in formation and breakage of disulfide bonds, showed to increase secretion of various recombinant proteins (Gasser et al., 2007; Robinson et al., 1994). However, over-expression of chaperones (e.g. Kar2) is not general applicable for improving recombinant protein secretion. It was demonstrated that the effect of Kar2 expression is protein and host specific. For example, the overexpression of Kar2 caused a decrease in glucose oxidase secretion in *H. polymorpha* (van der Heide et al., 2002). The authors explain this decrease in secretion by a prolonged binding of Kar2 to the target protein, which triggers ERAD (ER associated protein degradation) rather than recombinant protein secretion. As the protein folding machinery is rather complex and involves various chaperones, a simultaneous over-expression of chaperones was done. The joint over-expression of Kar2 and Pdi1 had a synergistic effect on secretion of a recombinant antibody fragments in *S. cerevisiae*. Furthermore, overexpression of co-chaperones (Jem1, Sil1, Lhs1, and Scj1) increased secretion of recombinant human albumin in *S. cerevisiae* (Idiris *et al.*, 2010). Other approaches targeted *HAC1*, encoding the master regulator of UPR (unfolded protein response). UPR is activated by accumulation of unfolded proteins in the ER and further induces the expression of other chaperones. The overexpression of *HAC1* in *S. cerevisiae* and *P. pastoris* already showed

improvement in recombinant protein secretion (Gasser et al., 2007; Idiris et al., 2010; Valkonen et al., 2003). However, the overexpression of *HAC1* is not general applicable for improving recombinant protein secretion. Furthermore, it is highly dependent on the protein of interest and host organism. Another strategy towards improving recombinant protein production is the engineering of protein transport through the secretory pathway. The first step in the early secretory pathway is the translocation of the protein into the ER lumen. This translocation is accomplished via a translocation pore and a signaling peptide. The efficient translocation of the protein is determined by the signaling peptide. Therefore, much effort was put on the development of efficient signaling peptides for secretion (e.g. mating pheromone α-factor signal) (Idiris *et al.*, 2010). After correct folding the next step is the transport of the protein to post ER compartments (e.g. Golgi apparatus, vacuole). The engineering of the two most important components (SNARE and SM proteins) for vesicular transport of proteins to post ER compartments, had a beneficial effect on recombinant protein production (Delic *et al.*, 2014). Other processes within the late secretory pathway to be mentioned, influencing recombinant protein production, are the miss-targeting of proteins to vacuole/lysosome and intracellular proteolytic degradation. The way of transport of vacuolar proteins is similar to the one for secreted proteins. In the late secretory pathway vacuolar proteins are recognized by *VPS10*, a transmembrane sorting receptor, which targets carboxypeptidase Y and other proteases to the vacuole. A knockout of *VSP10* in *Schizosaccharomyces pombe* increased the recombinant production of human growth hormone (hGH) (Idiris *et al.*, 2010). Furthermore, the intracellular degradation of proteins is a major issue concerning efficient recombinant protein production. The knockout of several proteases (intracellular and extracellular) revealed a positive effect for recombinant protein production (Delic et al., 2014). The engineering of the secretory pathway of yeasts showed to improve production of various recombinant proteins. The engineering of the secretory pathway is a difficult process, because still the complex interplay of processes is not known and furthermore necessitates a systems level understanding of the cell. Recent developments in high throughput technologies and other systems biotechnological tools now allow the analysis of the different processes involved in protein secretion.

Metabolism

Recent studies gave strong evidence that recombinant protein production in yeasts is also limited by cellular metabolism (Heyland et al., 2010, 2011; Klein et al., 2014). The production of a recombinant protein needs additional cellular resources, which have to be provided by the host cell. The additional synthesis effort may lead to the overburden of cellular metabolism and causes increased cellular stress. This additional stress due to recombinant production is also referred to as metabolic burden or metabolic load. The metabolic burden can be defined as a portion of the host cell´s resources, in form of either energy or precursor metabolites (e.g. amino acids, nucleotides), which are required for production of a recombinant protein (Fig.2) (Glick, 1995). Therefore, to improve recombinant protein production the metabolic burden exhibited on the host organism must be minimized. The extent of the metabolic burden is influenced by the protein of interest itself, bioprocess parameters (e.g. temperature, dissolved oxygen content, substrates) and gene dosage (Baumann et al., 2010; Dragosits et al., 2009; Hohenblum et al., 2004).

The production of a recombinant protein exhibits an additional need for amino acids. Certain limitations in amino acid biosynthetic capacity, or the supply of amino acid biosynthetic pathways can lead to the observed burden (Fig.2). Recombinant protein production is an energy expensive process and therefore limited supply of energy in form of ATP, GTP or redox equivalents may restrict recombinant protein production (Fig.2). For example, during glucose limited chemostats, Heyland et al (2011) detected a higher TCA-cycle flux for the recombinant strain producing β-aminopeptidase compared to the control strain. The increased TCA-cycle flux indicates a higher need for energy due to recombinant protein production. Additionally, the calculated ATP generation rates were slightly increased for the recombinant strain. A second study done in *P. pastoris* showed that over-expression of human superoxide dismutase increased the TCA-cycle flux, which again indicated the increased need for energy due to recombinant protein production (Nocon *et al.*, 2014). Both studies showed that metabolism adapted in a way to meet the demand of recombinant protein production. However, it seems that the observed metabolic burden is not explained by an increased drain of precursors or energy for recombinant protein production.

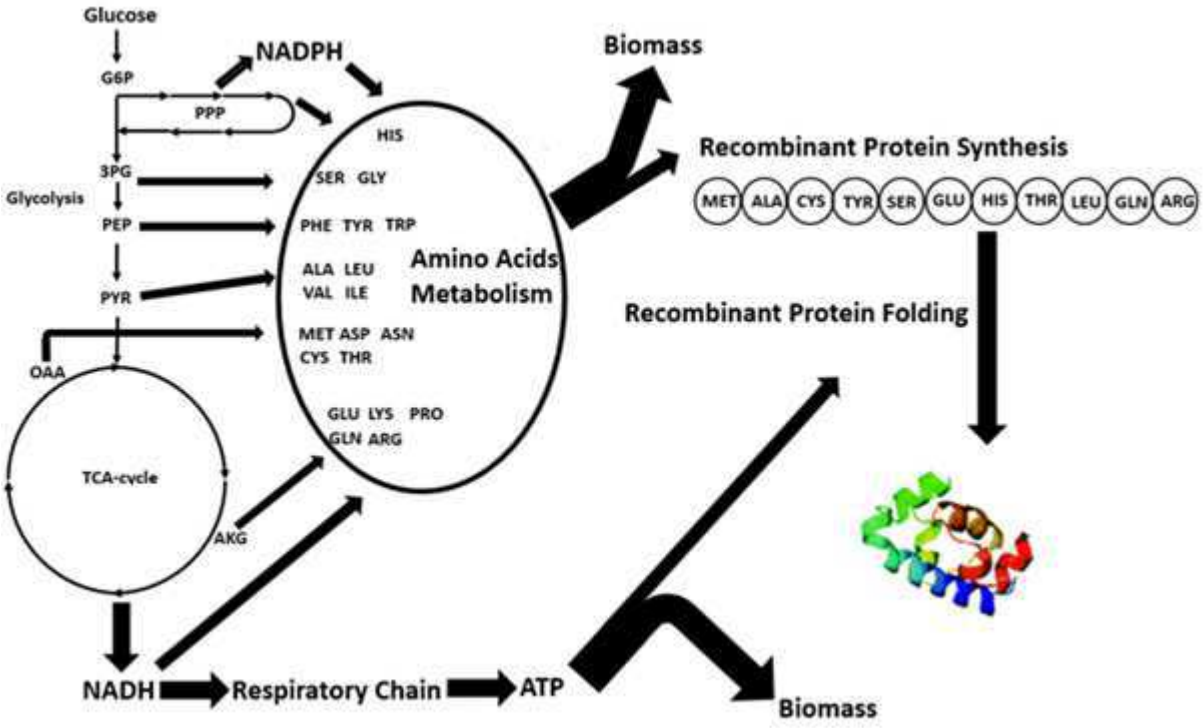

Fig.2: Limitation within cellular metabolism with respect to recombinant protein production.
(adapted from Piškur & Compagno, 2014)

Recent strain engineering strategies focused on the analysis of the cellular phenotype on global scale to identify limiting processes for recombinant protein production. As the production of recombinant protein necessitates a complex interplay of different cellular processes and components a global analysis using a systems biotechnological approach is necessary to identify limiting processes.

Systems Biotechnology

First attempts in strain engineering towards improved recombinant protein production followed the concept of metabolic engineering, which uses targeted engineering strategies to improve production formation. Metabolic engineering was defined as “The improvement of cellular activities by manipulations of enzymatic, transport, and regulatory functions of the cell with the use of recombinant DNA technology” (Bailey, 1991). Metabolic engineering necessitates the detailed knowledge about the metabolic system used. To gain this knowledge, methods from chemical engineering, computational sciences, biochemistry, and molecular biology are used. Metabolic engineering can be distinguished into two parts (1) the analysis part and (2) the synthesis part (Nielsen, 2001; Ostergaard et al., 2000). The analysis part deals with the careful analysis of the constructed strain (e.g. expression analysis, metabolite concentrations, metabolic flux analysis). In the synthesis part, a new recombinant strain is constructed with the use of the gained knowledge within the analysis part. In most cases, several rounds of analysis and synthesis steps are necessary to obtain a recombinant strain with the desired properties. In the field of recombinant protein production, several engineering approaches were made to increase the production of the recombinant protein. (such as optimization of bioprocess parameter, protein engineering, engineering of the host organism for improved folding, posttranslational modification and secretion) (Nielsen, 2001). Despite being rational and targeted the used approaches faced some problems. The engineering of the strain improves recombinant protein production, but the extent of improvement might still be limited due to the fact that limitations are shifted to processes not directly connected to product formation. These processes are not considered in the used local approach. Therefore, a more detailed analysis of the cellular phenotype on global scale is necessary to further increase productivity.

In order to get a more global view of the cellular phenotype during recombinant protein production systems biotechnology evolved as a powerful tool. A systems biotechnological approach uses methods already know from systems biology. The difference to systems biology is the focus on processes involved in formation of the desired product. A typical systems biotechnological approach is an iterative process consisting of consecutive cycles of strain analysis and strain engineering (Fig.3). This cycle is repeated until the desired goal is reached (Graf et al., 2009). The final aim is the development of general models, which later are used for the design of strategies for engineering the host organism towards improved product formation (Hou et al., 2012). To obtain the information necessary for the development of such models, systems biotechnology uses different tools. On the one hand mathematical models are applied to simulate or predict cellular behavior due to genetic or

environmental perturbations. On the other hand high throughput technologies are used to analyze the different cellular compounds on global scale (e.g. proteins, metabolites).

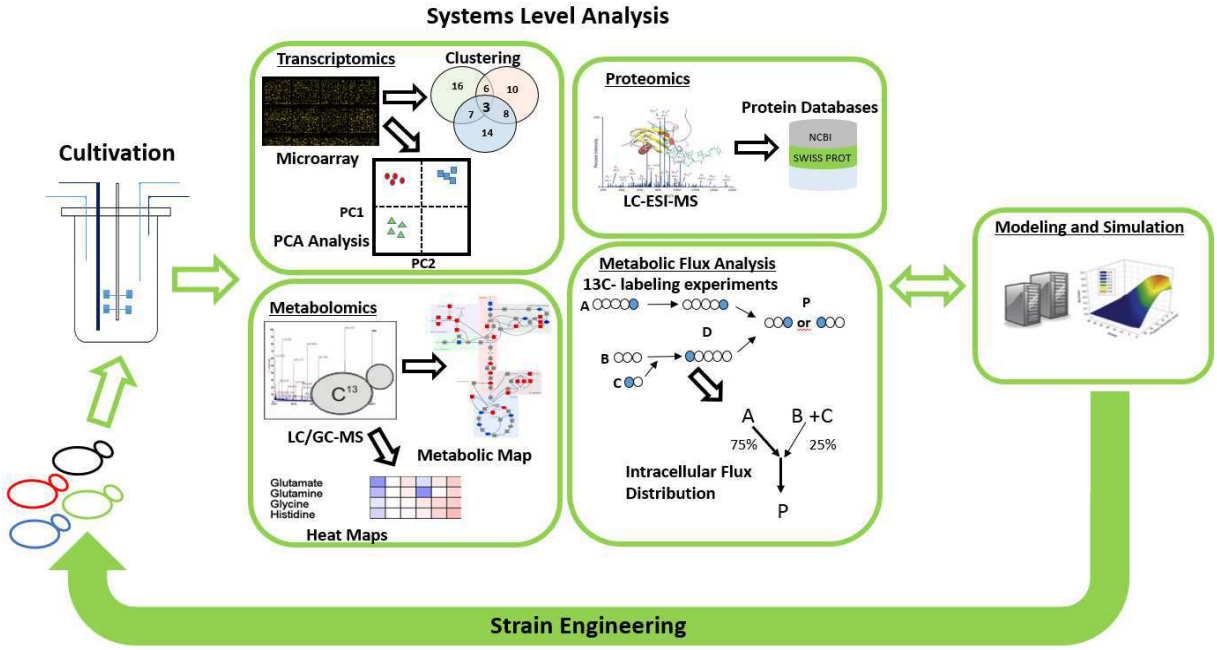

Fig.3: A typical systems biotechnological cycle. The desired strain is cultivated and high throughput technologies are used for systems level analysis. Computational models are predicting cellular behavior and to design engineering strategies.

High Throughput Technologies used in Systems Biotechnology

The phenotype of a cell is determined by the interplay of different cellular components (e.g. proteins, metabolites). The improvement in genome sequencing and the first available genome were the first step towards understanding the cellular phenotype. However, analyzing just simply the genome is not enough to get a closer picture of the cellular phenotype. Therefore, other high throughput technologies aiming at the global quantification of different cellular components (e.g. mRNA, proteins, metabolite, and intracellular fluxes) were developed.

Transcriptomics

Transcriptomics is the study of the transcriptome, the whole set of transcripts present in a given cell, under defined conditions. In contrast to the genome, which is rather static, the transcriptome is extremely dynamic and changes after environmental perturbations can be detected. Therefore, transcriptomics refers to the global assessment of gene expression profiles through quantification of

all RNA molecules. For the analysis of the transcriptome several methods were developed. In the past years the most widespread method used for transcriptomic analysis was the DNA microarray. The big advantages of microarrays are the possibilities to simultaneously determine the expression levels of many genes in multiple samples. In the beginning of transcriptomics analyses, the use of microarrays was limited to well-studied organism (e.g. *S. cerevisiae*) as genome sequences were only available for those. However, with the increase in the number of available genome sequences, and bioinformatic tools for oligo design, the analysis of the transcriptome for almost every organism was possible. Microarray analysis of the transcriptome helped tremendously in understanding the transcriptional machinery of cells, but in the past years limitations in microarray analysis were observed. Examples are, the high background noise due to cross hybridization, or the limited comparability of microarray data due to differences in normalization (van Vliet, 2010; Wang et al., 2010). To overcome the limitations of microarrays other methods for analyzing the transcriptome were developed (Wang et al., 2010). One of the most promising methods is RNA sequencing (RNAseq). A typical workflow for a RNAseq experiment is shown in Fig.4. RNA sequencing is based on the already well-established next generation sequencing methods and starts with the generation of cDNA libraries. After generation of the cDNA-library adaptors are ligated onto each single RNA molecule and high throughput sequencing is performed. The obtained reads are mapped onto a reference genome or reference transcriptome. If no reference sequence is available, a de novo assembly of the sequences can be done to obtain a genome scale transcription map. RNAseq has several advantages over the DNA microarrays based systems. RNAseq is applicable for organisms where no genome sequence is available, has lower background noise and a detailed study of the structure of the transcriptome is possible. For example alternative splicing of transcripts, post-transcriptional modification and the different population of RNA transcripts (miRNAs, tRNAs) can be studied (van Vliet, 2010; Wang et al., 2010). Transcriptomics analyses using RNAseq were already performed for several yeast species (Grabherr *et al.*, 2011; Liang *et al.*, 2012; Nookaew *et al.*, 2012).

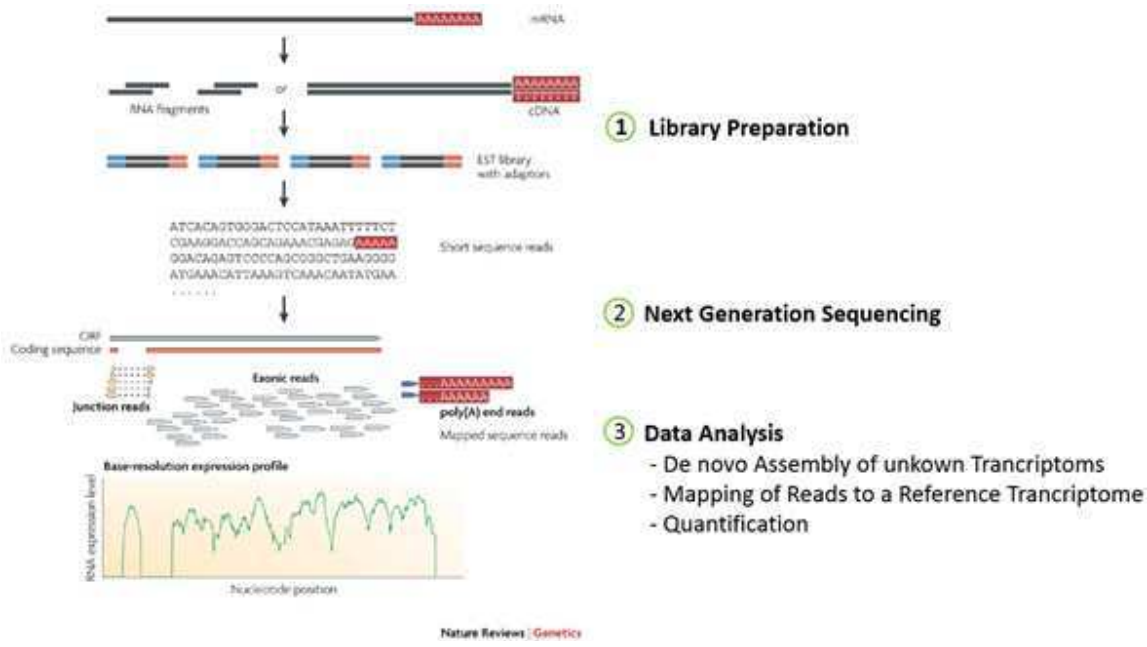

Fig.4: Workflow for a RNAseq experiment (Wang et al., 2010)

In the past years, transcriptomics analyses focused on the understanding of the impact on cell physiology due to different stress situations (e.g. temperature, pH, oxygen availability, recombinant protein production). Baumann and co-workers (2011) quantified the impact of different levels of oxygen availability on the transcriptome of *S. cerevisiae* and *P. pastoris*. Transcriptome analysis showed differences in the adaptive process for Crabtree positive and Crabtree negative yeast. For *P. pastoris* transcriptomic studies under recombinant protein production conditions were conducted. The gained knowledge was used for the knowledge-based engineering of cellular processes to further improve the performance of product formation. The transcriptome of *P. pastoris* producing human trypsinogen was analyzed and compared to the transcriptome of a non-expressing strain, which led to the identification of thirteen novel helper factors for protein secretion. These helper factors were overexpressed in a recombinant *P. pastoris* strain producing an antibody fragment. The overexpression of 12 out of 13 identified novel helper factors increased the production of the antibody fragment. Furthermore, six novel helper factors improved the specific productivity and the space time yield of the antibody fragment up to ~ 2.5 fold in lab scale fed batch cultivations (Gasser *et al.*, 2007). A similar approach was used to improve membrane protein production in *S. cerevisiae*. Transcriptomics analysis of *S. cerevisiae* strains producing a eukaryotic glycerol facilitator (Fps1) led to the identification of novel engineering targets (Bonander *et al.*, 2005). The knockout of three identified transcription factors (*GCN5*, *STP3* and *STP5*) improved production of Fps1. Additionally, all analyzed knockout strains showed increased transcript levels of *BMS1*, a gene connected to

ribosomal subunit assembly, compared to the wild type strain. A follow up study showed that engineering of the expression profile of *BMS1* improves production of Fps1 (Bonander *et al.*, 2009).

However, a transcriptomics analysis provides no information about regulatory processes on the protein level. For example, post translational modification of proteins greatly influences the activity of the protein itself, which is not detectable on a transcriptomic levels. Furthermore, mRNA expression levels often do not correspond with the observed protein abundance in the cell. However, Kim et al. (2012) stated that at least for genes and proteins having the same cellular function a good correlation between gene/protein levels was found. In order to obtain a more global view of regulatory processes of cells under specific conditions the knowledge about abundance of proteins is necessary. A proteomic analysis can provide this information.

Proteomics

Proteomics aims at the determination of the whole set of proteins produced by an organism. The proteome itself is highly dynamic and changes with time and different stresses acting on the organism. This indicates that a comprehensive overview about protein abundance, posttranslational modifications and regulatory processes is necessary. Due to this dynamic structure of the proteome, proteomic analysis faces several challenges. 1) A single gene can code for more than one protein. Detection limits regarding the identification of proteins lead to an incomplete picture of the proteome. 2) Proteins are dynamic. They can change their conformation and furthermore form complexes with other proteins. 3) Posttranslational modifications change the properties of the proteins and add difficulties to the detection of the proteins. To overcome the aforementioned limitations different methods for the analysis of the proteome were developed (Otto et al., 2014).

The first method used for proteomics analysis was 2D gel electrophoresis. Proteomics analysis using 2D gels ran into several limitations. For example, only a small amount of proteins per gel (around 1000 proteins per gel) can be analyzed and most of the proteome stays unknown. Furthermore, difficulties in detection of low abundant proteins exist. Also, the comparison of proteomic data obtained from 2D gel studies is often not possible, because different data processing methods and analysis platforms are used (Sechi & Oda, 2003).

Due to limitations of 2D gel analysis quantitative MS based methods for proteomics studies were developed. The quantitative MS based methods are high throughput and allow a global view of the proteome. Fig.5 shows the most prominent approaches for MS based quantitative proteomics

analysis. Quantitative MS based methods are based on tagging or labeling of specific amino acids. For example, ICAT (isotope coded affinity Tag) tags specifically cysteine residues and allows the quantitative comparison of protein levels. ICAT has a higher sensitivity than 2D gels and is the most used chemical isotope labeling method. This method is extremely helpful for the detection of low abundance peptides. However, the ICAT method is limited to proteins with cysteine residues (Otto et al., 2014). A second MS based method uses the stable isotope labeling with amino acids in cell cultures (SILAC). The SILAC method allows the detection of the whole proteome through feeding of stable isotope labelled amino acids. SILAC is most commonly used for differential proteomics analyses and studying of the posttranslational modifications. A third method used for quantitative proteomics is iTRAQ (isobaric Tag for Relative and Absolute Quantitation). This method is used for the relative and absolute quantification of proteins. Furthermore, a parallel analysis of different samples is possible (Otto et al., 2014).

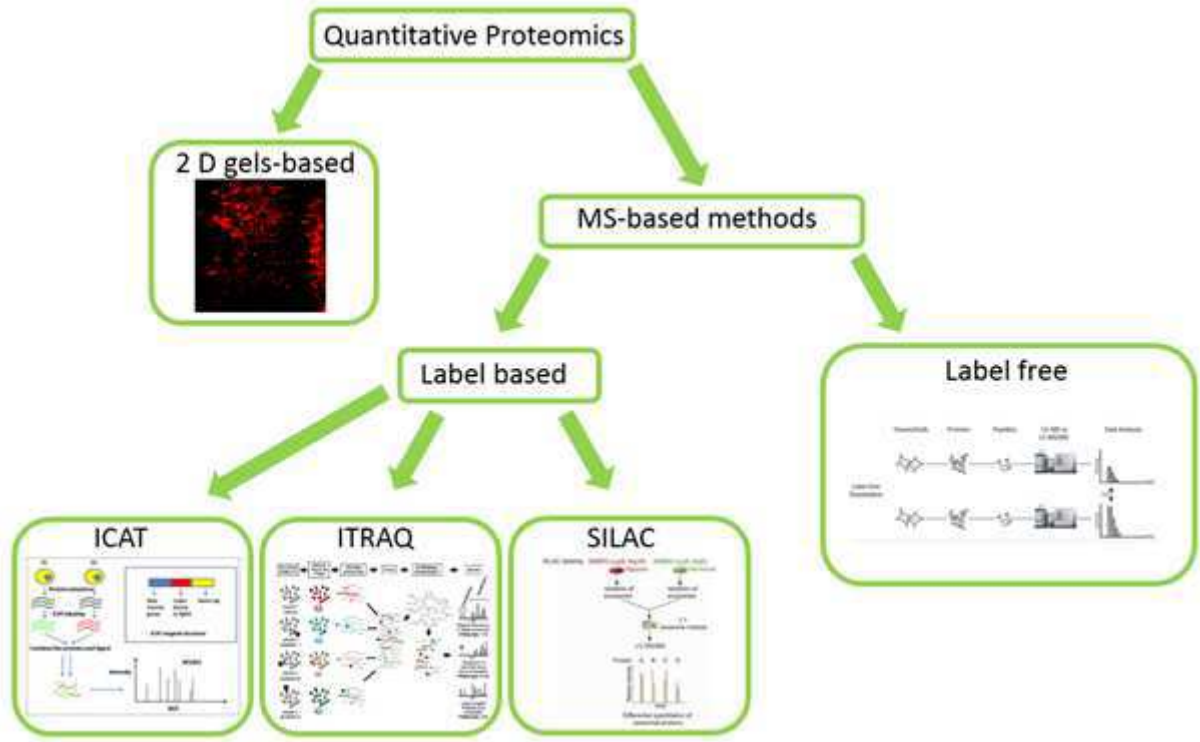

Fig.5: Overview of the different approaches used for proteomics

However, proteomics studies on comparing strains overproducing recombinant proteins with non-producers are scarce compared to other omics disciplines. Hypoxic conditions showed to be beneficial for the production of an antibody fragment in *P. pastoris*. Therefore, relative protein abundance in *P. pastoris* grown in glucose limited conditions under normoxic (21 % O_2) and hypoxic

(8% O_2) condition was compared, in order to understand the difference in regulatory processes due to a change in oxygen availability. Using 2D DIGE gel electrophoresis 85 proteins spots with a significant difference for normoxic and hypoxic condition were detected. Proteins with higher abundance on hypoxic condition were mainly connected to glycolysis, amino acid metabolism and general stress response. Significantly lower abundance under hypoxic conditions was found for proteins dedicated to TCA-cycle, vitamin metabolism, and oxidative stress response (Baumann *et al.*, 2010). A second study revealed that the cultivation temperature significantly influences the production of an antibody fragment. If the cultivation temperature is decreased to 20°C the productivity for an antibody fragment increased. This was attributed to the lower protein production rate, which allows a proper folding and secretion of proteins. The proteomic analysis using 2D DIGE gels revealed a lower abundance for proteins connected to the oxidative stress response and TCA-cycle when shifting cells from 30°C to 20°C (Dragosits *et al.*, 2009). In addition, the impact of recombinant protein production on cellular physiology using the strong inducible *AOX1*-promoter system was investigated on a proteomic level. Switching the carbon source to methanol led to drastic changes on the proteome level. Genes involved in the metabolism of methanol became highly abundant. Furthermore, it was observed that the methanol dissimilation pathway dominates over the assimilation of methanol into biomass. The proteomics analysis of *P. pastoris* overexpressing a Hepatitis B surface antigen showed an induction of UPR (unfolded protein response) and ERAD (ER-associated degradation) due to accumulation of the recombinant protein in the ER (Vanz *et al.*, 2012).

Proteins are the final catalysts of most reactions within a cell. Their activities directly influence the metabolic state of the cell. Nevertheless, proteomics analyses provide no information about metabolites, which are the final product of most cellular reactions. The intracellular concentration of metabolites can directly influence the rate of a cellular reaction, which later on defines the physiological state of the cell. Therefore, a comprehensive overview of intracellular metabolite concentrations is necessary to take one step further to fully understand cellular behavior.

Metabolomics

Metabolomics is the systemic study of all measurable intracellular metabolites produced by a cell at a certain time and given conditions. Beside the absolute quantification of intracellular metabolite pools, metabolomics aims at the determination of changes in metabolite levels due to genetic or environmental modifications. In general, the analysis of the metabolome comprises three different steps: (1) metabolomics sample preparation (including sampling, quenching, and cell extraction) (2) metabolite quantification using different analytical platforms and (3) analysis and interpretation of the obtained metabolomic data.

A prerequisite for an accurate determination of intracellular metabolite concentrations is a well-established sample preparation method. Unfortunately, there is no unique sample preparation method. For example, the standard cold methanol quenching protocol is not applicable for bacteria, because of the increased metabolite leakage during quenching (Bolten et al., 2007). In contrast, for yeasts quenching into cold methanol turned out to be the best suited method for metabolomics sampling (Carnicer et al., 2012a; de Koning & van Dam, 1992; Villas-Bôas et al., 2005). Therefore, it is necessary to develop or adapt already existing sample preparation procedures for the investigated microorganism. Every single step of the sampling procedure has to be evaluated for potential metabolite loss. Metabolite loss is a sum parameter comprising the diffusion of metabolites into the quenching solution and metabolite degradation due to chemical reactions (e.g. oxidative processes, insufficient stop of metabolism). Metabolite loss increases with prolonged sampling time. Therefore, for the preservation of the metabolic profile of a cell throughout the whole sampling process a fast sampling and efficient stop of metabolism is necessary. Typical steps of a metabolomics sampling preparation procedure are depicted in Fig 6.

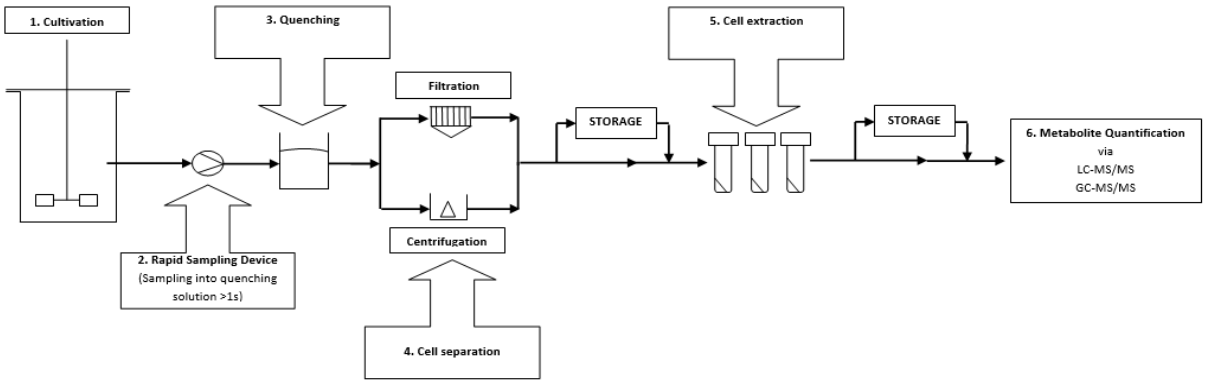

Fig.6: A typical workflow for metabolomics sample preparation (Russmayer et al., 2015, Manuscript 1)

After, rapid sampling of the cultivated cells within sub seconds the cells are quenched. For yeasts, quenching in cold methanol is the most applied quenching procedure. The first cold methanol quenching protocol was developed for *S. cerevisiae* by de Koning & van Dam (1992) and stayed almost unchanged over the past years. The most important parameters influencing metabolite loss during quenching are the methanol content of the quenching solution and the temperature. Therefore, several studies addressed this problem and optimized the quenching procedure in order to minimize the metabolite loss. Metabolite loss for *S. cerevisiae* was found to be at a minimum if buffered 60 % (v/v) methanol at -40°C is used (Villas-Bôas et al., 2005). Quenching protocols were also optimized for other yeast species. For *P. pastoris* quenching parameters were optimized and quenching turned out to be best in terms of minimal metabolite leakage at -27°C in 60 % (v/v) methanol (Carnicer et al., 2012a). After quenching, the fast removal of the quenching solution must be ensured. For the removal of the quenching solution, the most frequently used method is centrifugation. However, a disadvantage of centrifugation is the long exposure time of quenched cells to the quenching solution. Moreover, prolonged exposure of quenched cells to the quenching solution increases metabolite degradation. Therefore, other cell separation methods (e.g. filtration) were used for metabolomic sample preparation. Nevertheless, up to now in literature no consensus is found which method is the better performing in terms of metabolite loss. After this first steps, a storage step for the metabolomic samples is possible, as it is needed for parallel sampling (sampling of parallel cultures at different time points). In order to determine intracellular metabolite levels an internal standard can be added to the stored sample. The internal standard used is a uniformly ^{13}C labeled cell extract containing all desired metabolites (Mashego et al., 2004). To correct for metabolite loss during extraction the internal standard is added prior to the extraction procedure. The next step within the sampling procedure is cell extraction. The metabolome consists of several hundred metabolites with different chemical properties. The great diversity in chemical properties causes additional difficulties for cell extraction. Therefore, several cell extraction methods were developed to assess as many metabolites as possible (Mashego et al., 2005; Tredwell et al., 2011). The most important parameters for a cell extraction method are the efficiency of extraction and the recoveries for every metabolite. Both parameters must be evaluated for every investigated organism and measured metabolite (Neubauer *et al.*, 2012).

The quantification of metabolites is performed using different analytical platforms. Depending on the metabolites measured (e.g. different chemical properties) and the complexity of the sample the

appropriate platform must be chosen. Traditionally, the quantification of metabolites was based on enzymatic assays. The enzyme based methods had several drawbacks (e.g. high sample volume needed, high detection limits and matrix effects) (Mashego et al., 2007). The development of analytical systems with improved sensitivity and precision (e.g. mass spectroscopy based systems) now allows an accurate quantification of metabolites. Furthermore, with increased sensitivity of the analytical systems the number of detectable metabolites increased. The most frequently used analytical platforms combine a chromatographic step (liquid chromatography (LC) or gas chromatography (GC)) with a mass spectrometer (Mashego et al., 2007). To obtain a comprehensive view of the metabolome a combined analysis with LC-MS and GC-MS is necessary, as with a single method only specific metabolite classes are detectable. GC-MS based methods have a higher chromatographic resolution, but are restricted to volatile metabolites. With LC-MS based systems a broader mass range can be accessed, therefore with LC-MS more metabolite species are measurable. Therefore, LC-MS is the method used for metabolite profiling and exploration of unknown metabolites (Kell, 2004). However, LC-MS based methods suffer of matrix effects during electrospray ionization (ESI) (Kell, 2004). Matrix effects refer to the co-elution of components with the metabolite of interest and enhance or suppress ionization during ESI. Due to the matrix effect an accurate quantification of intracellular metabolite concentrations is not possible. To minimize the matrix effect a further optimization of the sample preparation and instrumental settings is necessary.

The metabolome can be analyzed using two different strategies: (1) targeted analysis for specific metabolites or (2) metabolite profiling (Villas-Bôas et al., 2005) (Fig.7). The targeted analysis approach focuses on the quantification of a small set of predefined metabolites. The quantification of the metabolites is done using internal or external reference components. Due to high accuracy of the metabolome data, via detection of peak integrals, the obtained data can be used for pathway analysis or calculation of intracellular fluxes using metabolic models. Metabolite profiling focuses on the rapid analysis of a larger number of metabolites using one or more analytical techniques. The aim is to obtain a metabolic profile of the cell at a specific time point. Metabolite profiling can be divided into metabolite fingerprinting and metabolite footprinting. Metabolite fingerprinting aims at the comprehensive study of changes in intracellular metabolite levels using different analytical platforms. The obtained data set can be used for quantification of metabolites or for comparison of metabolites profiles for different samples. Metabolite footprinting uses the same technologies as fingerprinting. The difference in both approaches is that footprinting focuses on the analysis of extracellular metabolites in the cultivation media. Extracellular metabolites are mainly secreted cellular

metabolites or media components converted by cellular components (Villas-Bôas et al., 2005). The metabolite profiling approach allows a high throughput analysis of different samples, which leads to vast accumulation of data. To extract biological meaningful data powerful statistical methods are used. Most commonly used are cluster analysis, principal component analysis, and partial least square discriminant analysis (Kim *et al.*, 2012).

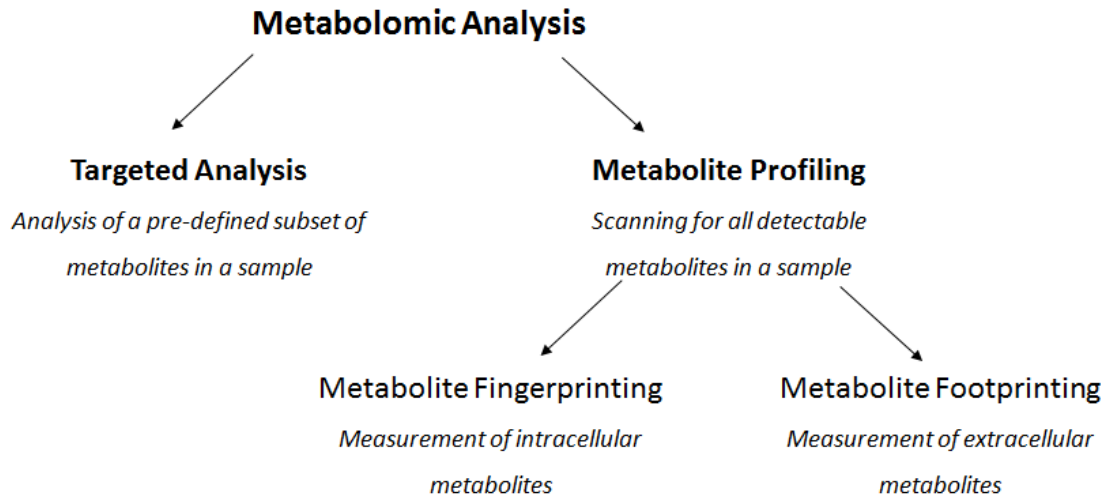

Fig.7: A schematic overview of the approaches used for the analysis of the metabolome.
According to Villas-Bôas et al., (2005)

A metabolomics analysis is performed to determine the impact of genetic modifications or environmental changes (e.g. change in carbon source, oxygen availability) on the metabolic profile of the cell or subset of intracellular metabolite pools (Carnicer et al., 2012b; Celton et al., 2012). Furthermore, quantitative data about intracellular metabolite pools is integrated into metabolic models and used for calculation of intracellular fluxes. The obtained data is used as an additional constraint and allows a more precise calculation of intracellular fluxes. This approach was used by Jordà and co-workers to gain novel insights in the methanol metabolism of *P. pastoris* when grown on mixtures of glucose and methanol (Jordà et al., 2013). However, up to now only little information about intracellular metabolite levels during recombinant protein production is available. Carnicer et al. (2012b) analyzed the impact of the production of an antibody fragment on intracellular amino acids levels.

Metabolites are the end products of every cellular reaction and knowledge about intracellular metabolite levels can give novel insights in regulatory processes of a cell. Furthermore, this knowledge helps to bridge the gap from genotype to phenotype. However, in most cases the knowledge of intracellular metabolite levels is not enough to fully describe the phenotype of a cell. In

this case, the knowledge about the intracellular flux distribution is necessary to depict the phenotype.

Metabolic Flux Analysis

Metabolic flux analysis attempts to predict rates of intracellular reactions using mathematical approaches in combination with experimental data. Metabolic fluxes are defined as the material that is processed through a sequence of feasible and observable biochemical reaction steps (Stephanopoulos, 1999). The flux of a reaction is the functional output of the interplay of the different cellular layers. Therefore, the knowledge about intracellular fluxes of a cell is crucial for the assessment of the physiological state of a cell. For the calculation of *in vivo* fluxes different approaches are used: 1) Flux balance analysis (FBA) uses stoichiometric properties of the metabolic network while 2) ^{13}C metabolic flux analysis combines the stoichiometric properties with labeling patterns of amino acids or metabolites as constraints.

Flux balance analysis (FBA)

The calculation of intracellular fluxes using FBA is based on the knowledge of the stoichiometric properties of the metabolic network. A prerequisite for FBA is the assumption of a steady state for the biological system, which means that the intracellular pools of metabolites do not change over time. Steady state conditions are reached via cultivation in continuous culture, in exponential fed batch cultivations or during the exponential phase in batch cultivations. Due to a metabolic steady state no information about kinetic parameters is necessary for the calculation and the problem is simplified to solving a set of linear algebraic equations. For the calculation of *in vivo* fluxes the set of linear equations are solved to find a flux distribution satisfying the stoichiometric constraints in steady state while optimizing an objective function relevant to the problem studied (i.e. for prediction of growth on different substrates biomass formation is maximized) (Wiechert, 2001). To reduce the number of possible solutions and to get a more realistic picture of the metabolism constraints are added. Possible constraints are limits in substrate uptake rates or constraints limiting the flux through certain conditions. FBA necessitates the knowledge about the stoichiometric properties of a metabolic network and information about extracellular fluxes (uptake rates for substrates, secretion rates for products or by products) (Wiechert, 2001).

FBA developed as a powerful tool for studying cellular physiology during growth on different substrates or different environmental conditions (e.g. temperature, pH) (Dragosits et al., 2009;

Simeonidis et al., 2010; Varma & Palsson, 1994). Furthermore, FBA is used for the prediction of putative gene knock outs or over-expression targets to further improve the performance of a host organism in terms of product formation (Bordbar et al., 2014; Burgard et al., 2003; Lee et al., 2012; Nocon et al., 2014)

Nevertheless, FBA is strongly limited in resolving complex networks. For example if two parallel metabolic pathways exist and none of the pathways is connected to extracellular fluxes it is not possible to resolve this situation without any additional information. Furthermore, metabolic cycles without connection to measurable fluxes and bidirectional reactions within metabolic pathways are impossible to resolve just using FBA (Wiechert, 2001).

To overcome the above mentioned limitations more information about the metabolic network is needed, which is provided by carbon labeling experiments.

^{13}C-Metabolic Flux analysis (^{13}C-MFA)

In order to overcome the limitations of FBA and to further understand the structure of metabolic networks ^{13}C-MFA was developed. The basis for all ^{13}C-MFA is carbon labeling experiments (Wiechert, 2001). For carbon labeling experiments ^{13}C labeled substrate (uniformly or specifically labeled) is fed to the culture. Due to the activity of the metabolic reaction the labeled carbon atoms are distributed throughout the metabolic network of a cell. After reaching a stable isotopic labeling pattern of intracellular metabolite pools, the labeling pattern of certain metabolites can be detected by NMR or MS based systems (Wiechert, 2001). The labeling pattern of amino acids are most frequently used as constraints for intracellular flux calculation, as the carbon backbone of the measured amino acids code for 8 key intermediates of the central carbon metabolism (Sauer, 2006). Using the uptake and secretion rates, and the obtained labelling pattern the steady state intracellular flux distribution is calculated (Sauer, 2006; Wiechert, 2001).

Two different methods are used to extract the intracellular flux distribution from labeling patterns of amino acids (Fig.8). The first method uses extracellular fluxes, biomass requirements and ^{13}C labeling patterns. This information is later integrated into a metabolic model. The model contains all reactions for the investigated network and atom transitions for each reaction. Furthermore, the model includes reactions mapping the labeling pattern of the amino acids to their precursors. The flux distribution is determined by iterative fitting of fluxes to experimental data, whereby the residual of measured and simulated labeling pattern is minimized (Quek et al., 2009; Wiechert, 2001). The second flux analysis method is a more local approach and focuses on the interpretation of labeling patterns of specific

nodes (e.g. pyruvate node, anapleurotic reactions) in the metabolic network, and is termed metabolic flux ratio analysis (Blank et al. 2005; Zamboni et al., 2005). The flux ratio calculation is done using probabilistic equations, which are derived from the relative contribution of each converging pathway for the specific metabolite. Furthermore, the obtained flux ratios can be used as constraints for the calculations of absolute fluxes (Sauer, 2006; Zamboni et al., 2005). The drawback of this method is that it is restricted to the labeling pattern of 10-15 amino acids. The information gained from the labeling patterns of amino acids allows only the calculation of fluxes connected to the central carbon metabolism. To overcome this restriction, the measurement of the labeling patterns of intracellular metabolite pools came into focus of research. The knowledge about the labeling patterns of metabolites allows a more precise calculation of intracellular fluxes. Furthermore, fluxes beyond the central carbon metabolism can be calculated (Sauer, 2006). Due to low metabolite concentrations and fast conversion times, the assessment of metabolite labeling patterns was hampered in the past. Recent developments in metabolomics sampling procedures and analytical methods now allow the determination of metabolite labelling patterns. However, several methods were already developed for the calculation of intracellular fluxes using metabolite labelling patterns (Jordà et al., 2013; Nöh & Wiechert, 2006; Wiechert & Nöh, 2013).

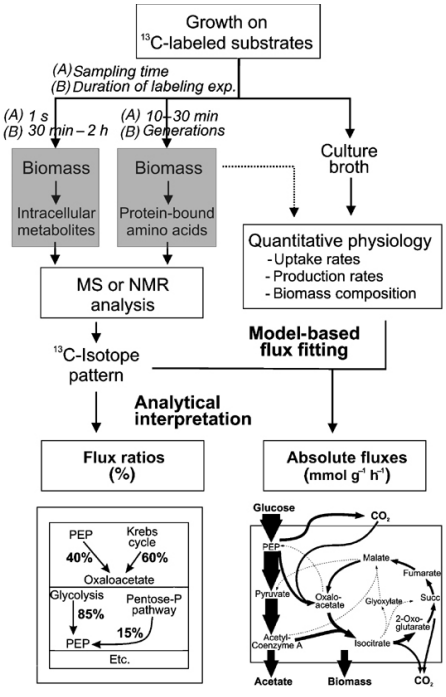

Fig.8: Schematic overview of the different approaches used ^{13}C-based metabolic flux analysis (Sauer, 2006)

^{13}C-MFA proved to be a powerful tool for the exploration of cellular physiology. Blank and co-workers (2005) quantified the intracellular flux distribution within the central carbon metabolism of fourteen *hemiascomycetous* yeast species. Furthermore, the influence of temperature and pH on the TCA-cycle activity in *S. cerevisiae* was determined via ^{13}C flux analysis (Blank & Sauer, 2004). In another study Solà et al. (2007) quantified the intracellular fluxes of *P. pastoris* of through the central carbon metabolism during growth on glycerol and methanol/glycerol mixtures. Recently, studies quantifying the impact of recombinant protein production on the central carbon metabolism were conducted. For example Baumann et al (2010) determined intracellular flux distributions of *P. pastoris* expressing an antibody fragment under normoxic and hypoxic conditions. In general, a shift from respirative to fermentative metabolism was observed when shifting oxygen availability from normoxic to hypoxic conditions. Jorda et al. (2012) could show a significant impact of recombinant protein production on central carbon metabolism when *P. pastoris* was grown on glucose/methanol mixtures. Recombinant protein production led to higher glycolytic, TCA-cycle and NADH regeneration fluxes. A third study trying to quantify the impact of recombinant protein production on metabolism was conducted by Nocon et al. (2014). In this study intracellular flux distributions of a strain producing human superoxide dismutase (hSOD) and a non-producer were compared. For the hSOD over producing strain an increased flux through the TCA-cycle was detected. The increased TCA-cycle flux was attributed to the higher need of energy due to recombinant protein production.

In the past years ^{13}C-MFA developed as a standard method for intracellular flux calculation. ^{13}C-MFA is constantly improved and applied to more complex problems. However, for more complex biological systems ^{13}C-MFA showed to have limitations. These limitations are caused by the basic steady state assumptions of ^{13}C-MFA: (1) metabolic steady state, pool size of all intracellular metabolites stays constant and (2) isotopic steady state, means that the labeling pattern of all biomass pools is constant. However, usually long labeling times are necessary to reach an isotopic steady state. Labeling times can be shortened by the use of instationary ^{13}C-MFA. Instationary ^{13}C-MFA is based on the measurement of labeling enrichment of intracellular metabolite pools. As turnover rates of metabolites are high, the labelling time is shortened too few minutes and no isotopic steady state must be reached. However, the experimental design, analytical methods and the computational models are much more demanding (Nöh & Wiechert, 2006)

In general, fluxes calculated by instationary ^{13}C-MFA are more precise, because much more information about the metabolic network is gained during the experiment. Instationary ^{13}C-MFA helps to resolve fluxes, which are not resolvable with ^{13}C-MFA (e.g. anapleurotic reactions). A first study using instationary ^{13}C-MFA was performed by Jorda et al. in order to get a deeper insight in the impact of methanol assimilation on *P. pastoris* central carbon metabolism (Jordà et al., 2013).

Mathematical Models used in Systems Biotechnology

In addition to the experimental high throughput technologies, mathematical models are used to increase the understanding of cellular behavior under specific conditions. A mathematical model comprises basic features of a cellular network (e.g. cellular components and their interactions), which are later translated in a set of mathematical equations (Stelling, 2004). Especially, mathematical models, modeling the metabolic network of organism, turned out to be powerful tools for metabolic engineering and subsequent strain improvement.

In principle, two different types of metabolic models are used, on the one hand the steady state stoichiometric models and on the other hand the time dependent dynamic models (Varner & Ramkrishna, 1999; Gombert & Nielsen, 2000). Stoichiometric models are based on the quantitative relationship between substrates and reaction. Therefore, the metabolic network of an organism is translated into a set of stoichiometric equations. The model consists of all reactions supplying essential precursors for cell growth. In fact, stoichiometric models are powerful tools, but the steady state assumption limits the predictive strength of such models. To improve the predictive power the implementation of time dependent concentrations of metabolites and enzymes is necessary.

The combination of kinetic parameters and stoichiometry is realized in dynamic models. However, for the development of dynamic models a detailed knowledge of kinetic parameters, interacting compounds, and regulatory structures is necessary. Despite the higher predictive power, dynamic modeling is restricted to small parts of metabolism, due to a lack of kinetic parameters for most metabolic reactions. As an example, Rizz et al. (1997) simulated the behavior of glycolysis in *S. cerevisiae* after a glucose pulse over short time scale.

Aim of the Work

Recombinant protein production in *P. pastoris* often results in an additional metabolic burden for the cell. Accompanied with the metabolic burden a decreased cellular fitness and more importantly a decreased productivity for the protein of interest is observed. Therefore, research focused on the understanding of the processes causing the observed metabolic burden. Metabolic limitations responsible for the observed burden are ascribed to insufficient supply of energy (in form of ATP, GTP or redox equivalents) or precursors (e.g. amino acids) needed for production of a recombinant protein. The supplementation of amino acids already showed to unburden cellular metabolism for recombinant protein production (Gorgens et al., 2005; Heyland et al., 2011; van Rensburg et al., 2012). For example, Heyland et al. (2011) revealed that the addition of energetically expensive amino acids (His, Ile, Leu, Lys, Met, Phe and Tyr) increases recombinant production of a β-aminopeptidase in *P. pastoris*. Furthermore, the uptake of supplemented amino acid was highly selective and correlated with the energy costs of amino acids. In a second study Gorgens et al. (2005) showed that addition of a mixture of Ala, Arg, Asn, Glu, Gln and Gly increases recombinant xylanase production in *S. cerevisiae*. Furthermore, both authors mentioned that the increased productivity is not simply achieved by incorporation of the supplemented amino acids into the recombinant protein. Moreover, Heyland et al. (2011) showed that added glutamate is co-metabolized with the carbon source and used for the synthesis of other amino acids. Additionally, Gorgens et al. (2005) showed that addition of the TCA-cycle intermediate succinate increased recombinant protein production in *S. cerevisiae*. These results indicate that supplemented amino acids or metabolites are used as additional carbon- or nitrogen sources, which fuel the production of other metabolites possibly limiting recombinant protein production. However, the observed metabolic burden cannot be explained by the increased drain of precursor metabolites used for recombinant protein production, as the amount of protein produced is often neglectable compared to the total synthesized cellular protein. In fact, changes in intracellular amino acid concentrations trigger several cellular processes. Most prominent is the feedback regulation caused by amino acids on amino acid biosynthesis pathways. Furthermore, changes in intracellular amino acid concentration influence other metabolic pathways (e.g. shuttling of redox equivalents with the malate/aspartate shuttle). To obtain a deeper insight in the connection between intracellular metabolite levels and the observed metabolic burden a comprehensive overview of intracellular metabolite levels during recombinant production is necessary. Up to now only little information is available on intracellular metabolite levels during recombinant protein production. Therefore, the aim of this PhD work was on the one hand to obtain a bigger dataset

about intracellular metabolite levels during recombinant protein production. On the other hand the aim was to specifically analyze the datasets to increase the understanding of the connection between metabolism and recombinant protein production and to use this knowledge for subsequent strain improvement engineering.

The first part of the work aimed at the development of methods for preparation of samples for metabolomics analysis or metabolic flux analysis. In the second part of the work the newly established methods were used to analyze cellular processes responsible for the differences in recombinant protein productivity for the two major carbon sources (methanol/glycerol vs. glucose) used for cultivation of *P. pastoris*. In the third part of this PhD work the established methods were used for the analysis of the metabolome and intracellular flux distribution during recombinant protein production. *P. pastoris* strains producing three different recombinant model proteins were cultivated in chemostats in order to analyze the metabolome and the intracellular flux distribution. Both the methanol based and glucose based expression system were used for recombinant protein production, in order to analyze carbon source specific effects on recombinant protein production. The metabolomics analyze showed that recombinant protein production clearly influenced intracellular amino acid pools. Based on this finding, the working hypothesis was to increase intracellular amino acid levels via engineering of amino acid metabolism. The effect of increased intracellular amino acid levels was tested for a recombinant carboxylesterase secreting strain. Increased supply with intracellular amino acids improved recombinant carboxylesterase production. In order to examine if the applied strategy for improved protein production is specific for carboxylesterase or generally applicable for improving production of recombinant proteins, the same strategy was applied for an antibody fragment.

Results

Metabolomics for *P. pastoris* – How to minimize metabolite loss during cell separation?

Manuscript 1

The accurate determination of intracellular metabolite concentrations relies on a well-established sample preparation method. For *P. pastoris* the procedure for metabolomics sample preparation was already well evaluated. However, some steps within the whole process chain received little attention and therefore needed some more optimization. Furthermore, for some steps different methods are available (i.e. centrifugation and filtration for cell separation). However, no comparison of both methods in terms of minimal metabolite loss during cell separation has been done. Therefore, the first question addressed in Manuscript 1 was the determination of the better performing cell separation method in terms of minimization of metabolite loss. The obtained metabolomics data sets clearly showed that filtration used for cell separation is the better performing method. The metabolite loss within the first 2 minutes after quenching was less than 10 %, which is acceptable for accurate measurement of intracellular metabolite levels.

In summary, the most efficient sample preparation method in terms of minimizing metabolite loss for *P. pastoris* comprises 1) rapid sampling within sub seconds 2) quenching in 60 % v/v methanol at -27°C 3) cell separation done by filtration 4) storage of cell pellets less than 3 weeks 5) metabolite analysis with LC-MS/MS.

FEMS Yeast Research, 15, 2015, fov049

doi: 10.1093/femsyr/fov049
Advance Access Publication Date: 19 June 2015
Short Communication

SHORT COMMUNICATION

Metabolomics sampling of *Pichia pastoris* revisited: rapid filtration prevents metabolite loss during quenching

Hannes Russmayer[1,2], Christina Troyer[2,3], Stefan Neubauer[2,3], Matthias G. Steiger[1,2], Brigitte Gasser[1,2], Stephan Hann[2,3], Gunda Koellensperger[2,4], Michael Sauer[1,2,*] and Diethard Mattanovich[1,2]

[1]Department of Biotechnology, University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria, [2]Austrian Centre of Industrial Biotechnology, Muthgasse 11, 1190 Vienna, Austria, [3]Department of Chemistry, Division of Analytical Chemistry, University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria and [4]Institute of Analytical Chemistry, University of Vienna, Waehringer Str. 42, 1090 Vienna, Austria

*Corresponding author: Department of Biotechnology, University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria. Tel/Fax: +43-1-3180078; E-mail: michael.sauer@boku.ac.at
One sentence summary: Rapid filtration within 2 min after quenching is the preferred method for metabolomics sampling of *Pichia pastoris* because it leads only to minor metabolite losses below 2%.
Editor: Jens Nielsen

ABSTRACT

Metabolomics can be defined as the quantitative assessment of a large number of metabolites of a biological system. A prerequisite for the accurate determination of intracellular metabolite concentrations is a reliable and reproducible sample preparation method, which needs to be optimized for each organism individually. Here, we compare the performance of rapid filtration and centrifugation after quenching of *Pichia pastoris* cells in cold methanol. During incubation in the quenching solution, metabolites are lost from the cells with a half-life of 70–180 min. Metabolites with lower molecular weights showed lower half-lifes compared to metabolites with higher molecular weight. Rapid filtration within 2 min after quenching leads to only minor losses below 2%, and is thus the preferred method for cell separation.

Keywords: metabolomics; quenching; filtration; centrifugation; metabolite loss; *Pichia pastoris*

Systems biology became an important tool for comprehensive and quantitative analysis of biological systems (Lee, Lee and Kim 2005; Graf *et al.* 2009; Kim *et al.* 2012). Among -omics technologies used for systems level analyses, metabolomics and fluxomics underwent a strong development, because metabolite concentrations and intracellular fluxes constitute the major determinants of cellular physiology (Fiehn 2002). The fast development of metabolomics necessitates suitable sample preparation procedures for accurate and robust quantification of metabolites. Unfortunately, a universal sample preparation procedure does not exist. Appropriate sampling for metabolomics analyses depends upon the organism to be investigated (Bolten *et al.* 2007); therefore, the best conditions for sampling of a specific organism must be evaluated individually (Villas-Bôas and Bruheim

Received: 24 February 2015; **Accepted:** 12 June 2015

2007; Bolten and Wittmann 2008; Canelas *et al.* 2008). Several metabolomics sampling procedures for different organisms were developed using centrifugation as cell separation method. The most widespread method using centrifugation was developed by de Koning and van Dam (1992), and remained mostly unchanged over the years. However, centrifugation has several disadvantages, among them a long exposure time of the cells to the quenching solution which is potentially detrimental (Villas-Bôas *et al.* 2005). Therefore, the protocol was adapted with regard to shorter contact times (shorter centrifugation times). Additionally, methods employing solely fast filtration instead of quenching have been reported to minimize cold-shock phenomena and related metabolite leakage that occurs particularly in bacterial cells (Wittmann *et al.* 2004). Recently, a combination of methanol quenching and fast cell separation using filtration has been described, but was not evaluated in comparison to centrifugation (Carnicer *et al.* 2012a). A critical parameter for the evaluation of a sample preparation method is the determination of potential metabolite loss during quenching and cell separation. Metabolite loss is a sum parameter, which is composed of unwanted diffusion of metabolites into the quenching solution and the degradation of metabolites due to chemical reactions. To our knowledge, the present work is the first comparative assessment of filtration and centrifugation of methanol quenched cells reported in the literature.

Over the last two decades, the yeast *Pichia pastoris* (syn. *Komagataella* sp.) became a well-established host for recombinant protein production (Gasser *et al.* 2013). Nevertheless, for complex recombinant proteins limitations in productivity were observed. To identify potential bottlenecks within the cellular protein synthesis machinery, detailed knowledge about cellular physiology during recombinant protein production must be available. For a metabolomic analysis of intracellular metabolites, an immediate stop of metabolism is essential to get a completed snapshot of the yeast metabolome. For yeasts, spraying of cells into cold 60% methanol was shown to efficiently stop cellular metabolism (de Koning and van Dam 1992). In the last years, much effort has been directed towards optimization of quenching protocols for metabolomics analyses of *P. pastoris* with the aim of minimization of metabolite loss (Tredwell *et al.* 2011; Carnicer *et al.* 2012a). Thereby, direct sampling into 60% (v/v) cold methanol (temperature less than −20°C) as quenching solution and subsequent extraction using boiling ethanol was found to be the best approach. However, methods used for the separation of cells from the quenching solution have not yet been directly compared.

The aim of this study therefore was to compare the performance of centrifugation and filtration as cell separation methods for *P. pastoris* in terms of metabolite loss. To assess the extent of metabolite loss for both cell separation methods, cells from a chemostat cultivation were sampled and quenched into cold methanol (60% methanol at −27°C) using the parameters previously optimized by Carnicer *et al.* (2012a) for quenching of *P. pastoris* cells. After selected time points of incubation in the quenching solution (2, 14, 34, 58 and 85 min), filtration was performed. A total of 2 mL of quenched cells (corresponding to approximately 10 mg dry biomass) were filtered by a filter device (Satorius Biolab, Polycarbonate Filter Holder) equipped with a cellulose acetate filter (0.45 μm, Satorius Biolab Products) using a vacuum pump. After filtration, the cells were washed once with 10 mL cold 60% (v/v) methanol and the filter with cells was kept on dry ice until cell extraction. The processing time for cell separation using filtration was below 1 min. To determine the sampling time span until which metabolite loss is acceptable, quenched cells were incubated in the quenching solution and samples for intracellular metabolite analysis were taken at certain time points.

The sample processing time for centrifugation (15 min) is considerably longer than for filtration. For centrifugation, quenched cells (corresponding to approximately 10 mg of dry biomass) were transferred to cooled pre-weighed tubes and centrifuged for 5 min at 4000 *g*. After the first centrifugation step, the cells were resuspended in 5 mL cold 60% (v/v) methanol and centrifuged again for 5 min at 4000 *g*. During the whole centrifugation process, the centrifuge (SORVAL RC6 PLUS, Cat.No.: 46915, Rotor: FIBER LITE, F10–6×500y) was cooled to −20°C.The washed cell pellets were kept on dry ice until cell extraction. Due to the prolonged sample processing time, only one sampling point after 2-min incubation in the quenching solution was used. After sampling and washing, the cell pellets were extracted using boiling ethanol extraction (Neubauer *et al.* 2012). Boiling ethanol extraction allows only an accurate quantification of a specific subclass of metabolites (e.g. amino acids, TCA-cycle intermediates). Therefore, the focus for analyzing metabolite loss was mainly laid on this specific metabolite class. Subsequently, metabolites were analyzed with LC-MS/MS according to Klavins *et al.* (2013).

The data presented in Table 1 clearly show that prolonged exposure of cells to the quenching solution causes a significant decrease in metabolite levels in the cell extracts. As metabolite loss can occur from the onset of incubation, the contact time with the quenching solution should be kept to a minimum. With one exception, 80–88% of the measured amino acid amounts were retained in the cells after 14 min of incubation in cold methanol (Table 1). Only methionine, which is known to be unstable and prone to oxidative degradation (Chace *et al.* 2010; Klavins *et al.* 2013), showed a loss of more than 20%. Therefore, the drop in intracellular concentration of methionine is possibly not only caused by diffusion into the quenching solution. Alanine, serine and valine concentrations dropped by 18–20% after 14-min incubation. These amino acids are not known to be unstable, but considering their low molecular weight it is plausible that the loss inversely correlates with the molecular weight. Indeed, amino acid loss correlated with the molecular weight (Fig. 1). A dependence of loss on molecular weight was also observed by Canelas *et al.* (2008). Furthermore, charged amino acids such as glutamate and aspartate showed higher intracellular stability. This charge dependence became even more obvious when comparing metabolite loss after centrifugation, where only potentially charged amino acids were retained within the cell at levels of more than 30%. Similarly, charged amino acids also show higher intracellular concentrations after 58 min of incubation, indicating that charged metabolites are less likely to pass the hydrophobic cell membrane. Consequently, small uncharged and hydrophobic amino acids turned to be more prone to diffusion than charged and hydrophilic amino acids. A slightly different behavior was observed for organic acids: only about 74–81% of the compounds could still be detected in the cells after 14 min of incubation (Table 1); however, a correlation trend between net charge and retention behavior was observed. For this class of metabolites, a fast removal of the quenching solution appears even more important to avoid extensive losses. When centrifugation was used for the removal of the quenching solution, increased metabolite losses were observed relative to the 2-min sampling time point of filtration. Depending on the respective compound, the metabolite loss for centrifugation amounted to 40–90%. Again, amino acids with a lower molecular weight showed a higher metabolite loss compared to amino acids with a higher molecular weight. The obtained data

Table 1. Intracellular metabolite concentrations (IC [μmol gCDW^{-1}]) of all measured metabolites for filtration and centrifugation.

	FILTRATION										CENTRIFUGATION	
Time in quenching solution	2 min		14 min		34 min		58 min		87 min		2 min	
Cell separation time	1 min		1 min		1 min		1 min		1 min		15min	
	IC	[%][a]	IC	[%][a]	IC	[%][a]	IC	[%][a]	IC	[%][a]	IC	[%][a]
Alanine	16.1	100	11.1	80.3	10.9	76.4	9.72	58.9	7.74	50	2.75	9.3
Serine	6.89	100	4.88	81.7	6.43	76.1	4.74	61.4	4.4	53.3	2.2	16.1
Proline	33.7	100	27.9	85	26.1	81.1	19.2	66.8	15.5	57.1	1.6	10.4
Valine	1.44	100	1.19	81.9	0.94	78.1	0.77	63.4	0.86	57.8	–	–
Threonine	2.36	100	2.14	84	2.21	80.7	2.03	68.9	1.92	62.2	0.48	21.3
Isoleucine	0.34	100	0.28	81.9	0.25	73.9	0.19	57.5	0.19	56.6	0.06	14.1
Leucine	0.84	100	0.7	83.8	0.63	75.9	0.5	61	0.51	61.5	0.16	17.5
Asparagine	4.46	100	4.26	83.8	2.5	79.6	2.5	66.1	2.58	60.5	–	–
Aspartate	33.2	100	25.2	87.3	24.6	83.6	17.4	73.1	17	67.6	8.5	36.8
Glutamine	143	100	116	84.3	107	80.2	88.4	68.1	79	61.7	30.5	26
Lysine	7.76	100	6.86	87.7	7.53	82	6.21	67.5	6.41	64.2	3.39	61.6
Glutamate	172	100	147	87.9	138	84.6	117	75.5	109	70.3	50.1	37.1
Methionine	1.12	100	0.89	78.8	0.75	67	0.57	50.8	0.47	41.9	0.09	6.4
Histidine	6.57	100	5.53	87.1	5.35	83.4	5.53	66	4.93	63.2	3.67	54.6
Phenylalanine	0.19	100	0.17	85.6	0.16	81.6	0.12	61	0.12	65.4	0.05	23.1
Arginine	94.9	100	98.9	87.6	84	83.6	74.7	69.2	70.6	65.3	55.8	64.5
Malate	4.66	100	3.72	79.3	3.78	79.5	3.04	64.9	3.11	66.8	0.64	11.4
Isocitrate	0.12	100	0.09	74.4	0.08	63.5	0.07	61.1	0.07	56.6	0.05	40.9
Fumarate	1.74	100	1.3	71.4	1.08	71.2	1.32	68.6	1.01	55.6	0.25	11.2
Succinate	2.25	100	1.85	81.4	1.82	83.2	1.59	69.7	1.39	61.8	0.16	6.1

[a] relative abundance of intracellular metabolites compared to the first sampling point (2 min).

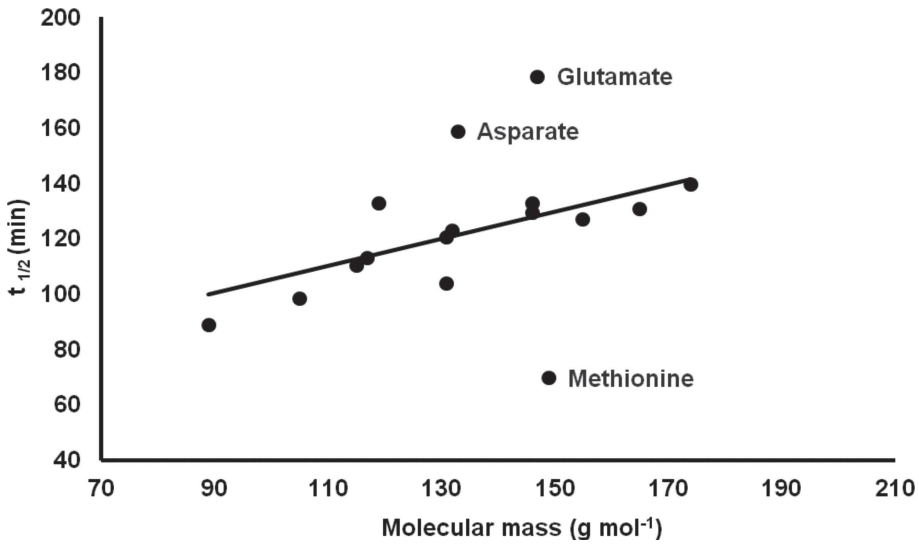

Figure 1. Correlation of loss of amino acids in quenching solution to their molecular mass. Retention of amino acids expressed as half-life ($t_{1/2}$) is plotted vs. molecular mass. Negatively charged amino acids (aspartate and glutamate) are better retained in the cells, while overall retention correlates with molecular mass (indicated by the solid line). Methionine is unstable due to oxidative degradation and thus more rapidly lost during quenching.

set (Table 1) for filtration and centrifugation clearly shows that filtration performs better in terms of minimization of metabolite loss. The lower metabolite loss for filtration may be attributed to the shorter contact time of the cells to the quenching solution. However, also compared to longer contact times (17 and 34 min) prior to filtration, centrifugation led to higher metabolites loss than filtration, indicating that additional stress or constraints (e.g. shear forces) are exerted on the cells during centrifugation, making them more susceptible to metabolite loss. It has to be mentioned that the observed correlation of metabolite loss and contact time is only valid for the analyzed metabolite subclass (such as amino acids, organic acids). For other metabolite classes this correlation has to be verified.

Besides the contact time of the cells with the quenching solution, which we confirmed to be a crucial parameter influencing metabolite loss also for *P. pastoris*, another factor previously shown to influence metabolite loss in *Saccharomyces cerevisiae* is the methanol content of the quenching solution (Canelas *et al.* 2008). As this topic was addressed by the two previous studies optimizing metabolomics sampling for *P. pas-*

toris, we used the previously optimized methanol concentration (Carnicer *et al.* 2012a; Tredwell *et al.* 2011). It has to be taken into account, however, that in the study by Tredwell *et al.* (2011) which used centrifugation for cell separation, the quenching solution was buffered and cooled to an even lower temperature of −40°C. While one could argue that different quenching conditions might be better compatible with different cell separation methods, both previous studies reported that *P. pastoris* was rather robust to different variations of the cold methanol quenching method, all yielding very similar results.

When aiming at the absolute quantification of metabolites in cellular samples, one needs to estimate the entire extent of the metabolite loss even if the exact value is experimentally not amenable. To calculate the metabolite loss within the first 2 min of quenching (shortest experimental time span), a first-order kinetics regression analysis of the obtained time course data showed that metabolite loss within the first 2 min was 2% or less of the original value, which indicates high quality of the data allowing to determine the true values under biological conditions. The measured intracellular amino acid pools were compared to previously published data from *P. pastoris* cultivated under the same conditions. In general, the comparison of the obtained data in this study with published data showed a similar trend for all free amino acid pools (Carnicer *et al.* 2012b) with the exception of proline (which is approximately 3-fold higher in our measured data set). The largest pools were found for amino acids deriving from the TCA-cycle intermediate α-ketoglutarate (Glu, Gln, Arg, Pro). In summary, the obtained results clearly show that the choice of the cell separation method (such as centrifugation or filtration) has direct influence on metabolite loss during sample preparation. To avoid extensive metabolite loss, and allow for an accurate determination of intracellular metabolite concentrations, it is suggested to use filtration for cell separation and to keep the contact time of cells to quenching and washing solution to a minimum.

In future, the storage stability of metabolites in prepared samples is a further point to address. Some kind of sample storage is usually required before the samples can be analyzed. However, almost no information is available on the actual storage stability of metabolites in quenched cells. Furthermore, it has to be understood that the sampling procedure needs to be reevaluated for each new class of metabolites, which shall be analyzed. It is clear by now that cell extraction methods directly impact the determination of the intracellular concentration of metabolites. However, other factors such as quenching or sample storage may very well have an influence.

ACKNOWLEDGEMENTS

The authors thank Kristaps Klavins for support in MS analytics and data handling, and Matthias Mattanovich for help in data evaluation.

FUNDING

This project was financially supported by Biomin Research Center, Boehringer-Ingelheim RCV, Lonza AG, Biocrates Life Sciences AG, VTU Technology GmbH, Sandoz GmbH, 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. EQ BOKU VIBT GmbH is acknowledged for providing LC-MS/MS instrumentation.

Conflict of interest. None declared.

REFERENCES

Bolten CJ, Kiefer P, Letisse F, *et al.* Sampling for metabolome analysis of microorganisms. *Anal Chem* 2007;**79**:3843–9.

Bolten CJ, Wittmann C. Appropriate sampling for intracellular amino acid analysis in five phylogenetically different yeasts. *Biotechnol Lett* 2008;**30**:1993–2000.

Canelas AB, Ras C, Pierick A, *et al.* Leakage-free rapid quenching technique for yeast metabolomics. *Metabolomics* 2008;**4**:226–39.

Carnicer M, Canelas AB, Ten Pierick A, *et al.* Development of quantitative metabolomics for *Pichia pastoris*. *Metabolomics* 2012a;**8**:284–98.

Carnicer M, Ten Pierick A, van Dam J, *et al.* Quantitative metabolomics analysis of amino acid metabolism in recombinant *Pichia pastoris* under different oxygen availability conditions. *Microb Cell Fact* 2012b;**11**:83.

Chace DH, Luo Z, De Jesús VR, *et al.* Potential loss of methionine following extended storage of newborn screening samples prepared for tandem mass spectrometry analysis. *Clin Chim Acta* 2010;**411**:1284–6.

De Koning W, van Dam K. A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. *Anal Biochem* 1992;**204**:118–23.

Fiehn O. Metabolomics–the link between genotypes and phenotypes. *Plant Mol Biol* 2002;**48**:155–71.

Gasser B, Prielhofer R, Marx H, *et al. Pichia pastoris*: protein production host and model organism for biomedical research. *Future Microbiol* 2013;**8**:191–208.

Graf A, Dragosits M, Gasser B, *et al.* Yeast systems biotechnology for the production of heterologous proteins. *FEMS Yeast Res* 2009;**9**:335–48.

Kim I-K, Roldão A, Siewers V, *et al.* A systems-level approach for metabolic engineering of yeast cell factories. *FEMS Yeast Res* 2012;**12**:228–48.

Klavins K, Neubauer S, Al Chalabi A, *et al.* Interlaboratory comparison for quantitative primary metabolite profiling in *Pichia pastoris*. *Anal Bioanal Chem* 2013;**405**:5159–69.

Lee SY, Lee D-Y, Kim TY. Systems biotechnology for strain improvement. *Trends Biotechnol* 2005;**23**:349–58.

Neubauer S, Haberhauer-Troyer C, Klavins K, *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.

Tredwell GD, Edwards-Jones B, Leak DJ, *et al.* The development of metabolomic sampling procedures for *Pichia pastoris*, and baseline metabolome data. *PLoS One* 2011;**6**:e16286.

Villas-Bôas SG, Bruheim P. Cold glycerol-saline: the promising quenching solution for accurate intracellular metabolite analysis of microbial cells. *Anal Biochem* 2007;**370**:87–97.

Villas-Bôas SG, Højer-Pedersen J, Akesson M, *et al.* Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast* 2005;**22**:1155–69.

Wittmann C, Krömer JO, Kiefer P, *et al.* Impact of the cold shock phenomenon on quantification of intracellular metabolites in bacteria. *Anal Biochem* 2004;**327**:135–13

The effect of methanol induction on cell physiology of *P. pastoris*

Manuscript 2

The number of promoters used for recombinant protein production in *P. pastoris* experienced a fast increase in the past years. Nevertheless, most commonly used are the strong inducible P_{AOX1} and the strong constitutive P_{GAP}. The P_{AOX1} regulates the transcription of the *AOX1* gene in *P. pastoris*, which encodes an alcohol oxidase (*AOX*) catalyzing the conversion of methanol into formaldehyde. The enzyme shows a rather low substrate affinity, therefore strong expression of the alcohol oxidase gene is important to maintain high amounts of enzyme allowing growth on methanol (Daly & Hearn, 2005). Beside the *AOX1*, a second alcohol oxidase *AOX2* is encoded in the *P. pastoris* genome. However, during growth on methanol the *AOX1* gene is more actively transcripted than *AOX2* (Cregg et al., 1989). P_{AOX1} is an attractive promoter for recombinant protein production, because of the high promoter strength in combination with an easy induction by switching the carbon source to methanol.

However, several drawbacks are associated with methanol based protein production. Recent studies showed an increased release of host cell protein due to increased cell lysis (Jahic et al., 2006). The released host cell protein directly contaminates the secreted recombinant protein and therefore adds difficulties to subsequent downstream processes. Furthermore, at industrial scale cooling is an important issue, because the utilization of methanol is an exothermic process and the produced heat has to be discharged. For larger scale the heat exchange causes problems and may lead to a drop in productivity. Furthermore, as methanol is a flammable and a rather toxic chemical the storage of methanol necessitates additional safety requirements for industrial processes.

To avoid above mentioned problems big effort is put on the development of alternative promoters, allowing efficient recombinant protein production on other carbon sources. Nevertheless, most commonly the glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter is used for recombinant protein production. The P_{GAP} is a strong constitutive promoter and allows a constant expression of recombinant protein. Furthermore, no induction with toxic inducers is necessary and process times are shortened due to an easier cultivation strategy. The strength of P_{GAP} directly depends on the used substrate. For cells grown on glucose, the promoter strength is equal to the strong P_{AOX1}. A drop in promoter strength of about 30% is observed if cells are grown on glycerol (Waterham et al., 1997). Despite, almost equal promoter strength for P_{AOX1} and P_{GAP} often higher productivities for P_{AOX1} are observed. In order to understand if the difference in productivity is caused by the use of different carbon source a systems level analysis of *P. pastoris* wild type strains grown on methanol/glycerol and

on glucose was done. Therefore, chemostat cultivations of the *P. pastoris* wild type strain on both carbon sources were conducted (triplicates for each carbon sources). After reaching a metabolic steady state samples for transcriptomics, proteomics, metabolomics and determination of intracellular fluxes were taken. The combined analysis of all cellular layers should increase the understanding of processes responsible for the difference in productivity.

The analysis of the gene expression pattern for both conditions showed significant changes for genes connected to methanol metabolism, peroxisome biogenesis, amino acid biosynthesis, vitamin biosynthesis and oxidative stress. Vitamin biosynthesis was up regulated for riboflavin and thiamine. The up-regulation of these two vitamin-biosynthesis pathways is related to the increased need as co-factors for enzymes involved in methanol metabolism, which were highly induced after induction with methanol. Fatty acid beta oxidation was among the pathways, which were down-regulated on a transcriptomic level. Furthermore, the analysis of the transcriptome and proteome led to the identification of unknown peroxisomal isoforms of genes involved in methanol assimilation.

Proteomics analysis revealed a higher abundance of proteins involved in translation (ribosomes), cytoskeleton biosynthesis, amino acid metabolism, vitamin biosynthesis, and methanol metabolism for cells grown on methanol/glycerol mixtures. From 141 significantly regulated proteins, only 4 proteins showed lower abundance in methanol/glycerol grown cells.

The analysis of the metabolome showed 18 significantly changed intracellular metabolite levels within amino acid and central carbon metabolism. Interestingly, metabolites connected to the pentose phosphate pathway (6-phosphogluconate, ribulose-5-phosphate and sedoheptulose-7-phosphate) were at lower abundance in methanol/glycerol grown cells. The drain of xylulose-5-phosphate used for methanol metabolism can explain the lower abundance of these intermediates. Furthermore, TCA-cycle associated metabolites had lower abundance in methanol/glycerol grown cells. A pathway with significantly increased metabolite levels was the lower part of glycolysis. The analysis of intracellular amino acid levels showed that in general intracellular levels are slightly lower for methanol/glycerol grown cells. However, a higher abundance of proteinogenic amino acids per g biomass for methanol/glycerol grown cells was observed. The measurement of the protein content and the subsequent calculation of the protein synthesis rate for both conditions showed increased protein synthesis rate for methanol/glycerol grown cells.

The analysis of the intracellular flux distribution for both conditions revealed changes in the intracellular flux patterns of the central carbon metabolism. A higher inverse flux through the upper

part of glycolysis (starting from glyceraldehyde-3-phosphate to glucose-6-phosphate), and an increased pentose phosphate pathway flux were detected. Active gluconeogenesis is needed for the synthesis of C6-sugars from the C3 backbone of glycerol and C1 backbone of methanol. The assimilation of methanol had a significant impact on the pentose phosphate pathway flux. For methanol induced cells more carbon is channeled through the oxidative branch of the pentose phosphate pathway, thus enabling an increased NADPH production rate. NADPH is the major electron donor for amino acid biosynthesis. Calculation of the protein synthesis rate revealed that methanol/glycerol grown cells have a higher protein synthesis rate (0.054 g(protein) $g(CDW)^{-1}$ h^{-1} vs. 0.040 g(protein) $g(CDW)^{-1}$ h^{-1} on glucose). Therefore, higher a NADPH production rate is needed to meet the demand for the increased protein synthesis for methanol induced cells.

The calculated intracellular flux through the TCA-cycle was lower for cells grown on methanol/glycerol. This indicates that the major route for NADH generation is not the TCA-cycle for methanol/glycerol grown cells. For these cells, alternative NADH generation routes are more important. An alternative route for NADH generation for methanol/glycerol grown cells is the metabolism of methanol. The metabolism of one mole of methanol produces two moles of NADH via the dissimilatory pathway. The NADH is later transported by shuttle systems into the mitochondrion, where it is re-oxidized. Indeed, for methanol/glycerol grown cells both mitochondrial transporters *AGC1* (glutamate-aspartate transporter) and *ODC1* (malate-α-ketoglutarate transporter) were up-regulated on transcript and protein level further supporting the hypothesis that shuttling of cytosolic NADH for ATP production is of relevance for methanol induced cells. Furthermore, a carbon flux through the glyoxylate shunt was detected for cells grown on methanol/glycerol.

In summary, for all cellular layers (transcriptome, proteome, metabolome and intracellular flux distribution) significant changes were observed. Most strikingly, methanol as carbon source directly influenced free intracellular amino acid pool sizes. For cells grown on methanol/glycerol mixtures free intracellular amino acid pools were decreased, whereas a higher content of protein bound amino acids were found. Furthermore, genes connected to amino acid biosynthesis were up-regulated on transcript and protein levels. Taking all facts together, this indicates a re-adjustment of the amino acid metabolism due to the change in carbon source. However, this re-adjustment may cause increased supply of intracellular amino acids and possibly is responsible for the observed higher productivities on methanol. Furthermore, an increased translational activity (up-regulation of ribosomal genes) for methanol/glycerol grown cells was found, which may also explains the observed higher productivity of recombinant proteins on methanol.

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 μ_{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

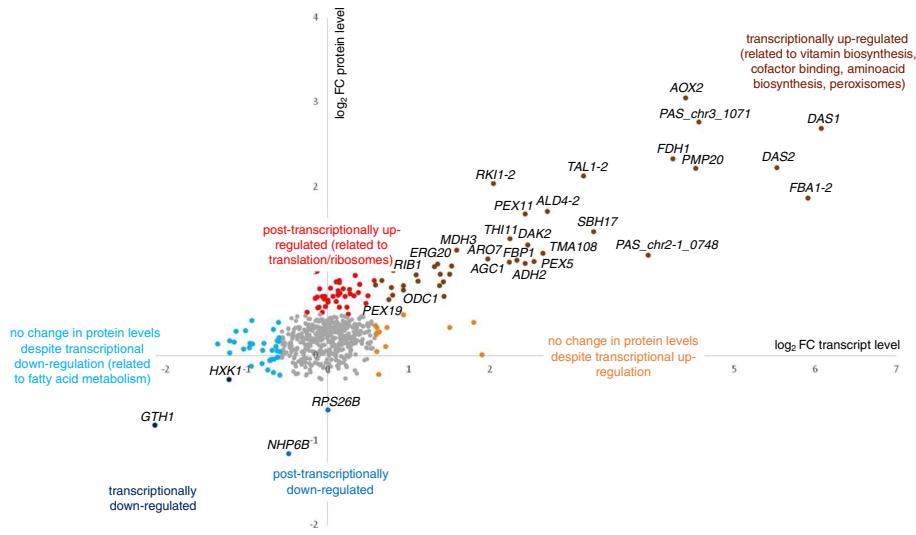

Fig. 1 Co-regulation of transcripts and proteins of *P. pastoris* grown on methanol/glycerol compared to glucose in chemostats at $\mu = 0.1\ h^{-1}$. Log_2 fold changes of transcripts versus log_2 fold changes of proteins for the 575 data pairs are shown which have been grouped based on their regulation 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_2FC +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

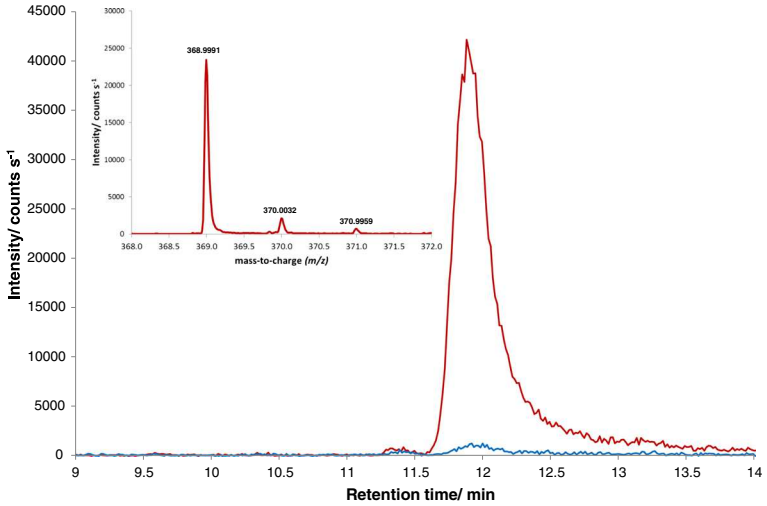

Fig. 2 Extracted ion chromatogram ($m/z = 368.9993 \pm 10$ ppm) of a sample grown on methanol (red) or glucose (blue) showing the sedoheptulose-1,7-bisphosphate peak at a retention time of 12 min. In the upper left corner the mass spectrum (m/z) of the peak of the methanol-grown sample after background 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

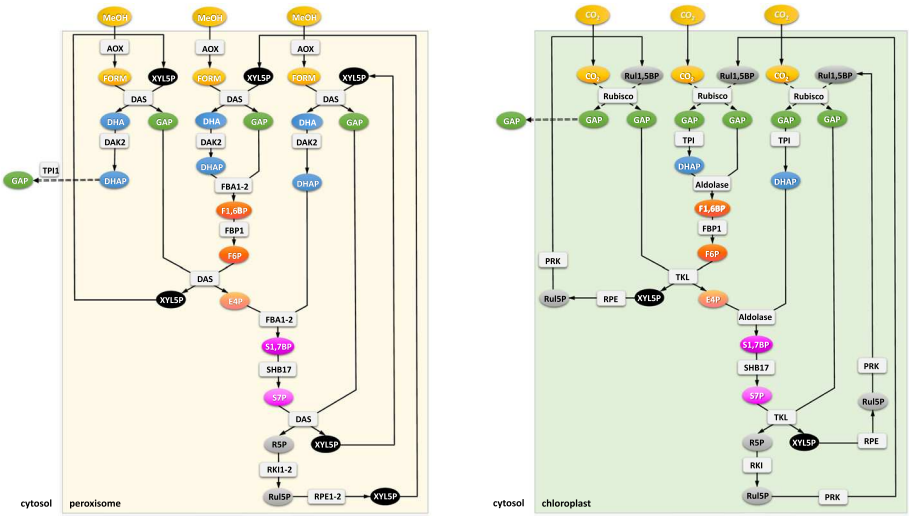

Fig. 3 Regeneration of pentose phosphates. Left: Methanol assimilation through the xylulose-monophosphate cycle: proposed rearrangements employing an alternative XYL5P forming pathway via S1,7BP. The net reaction of methanol assimilation is the formation of one GAP molecule from three methanol molecules. 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

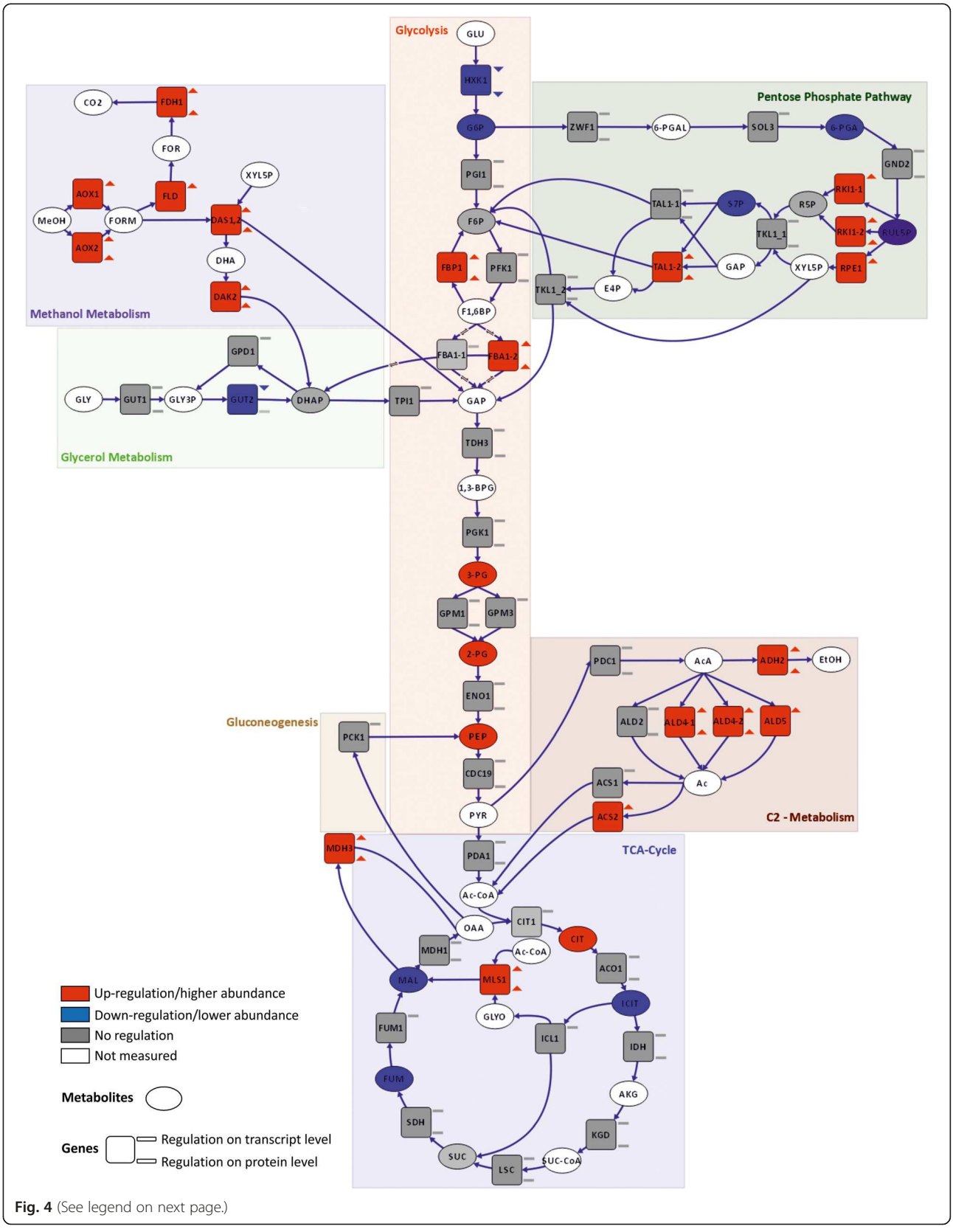

Fig. 4 (See legend on next page.)

(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

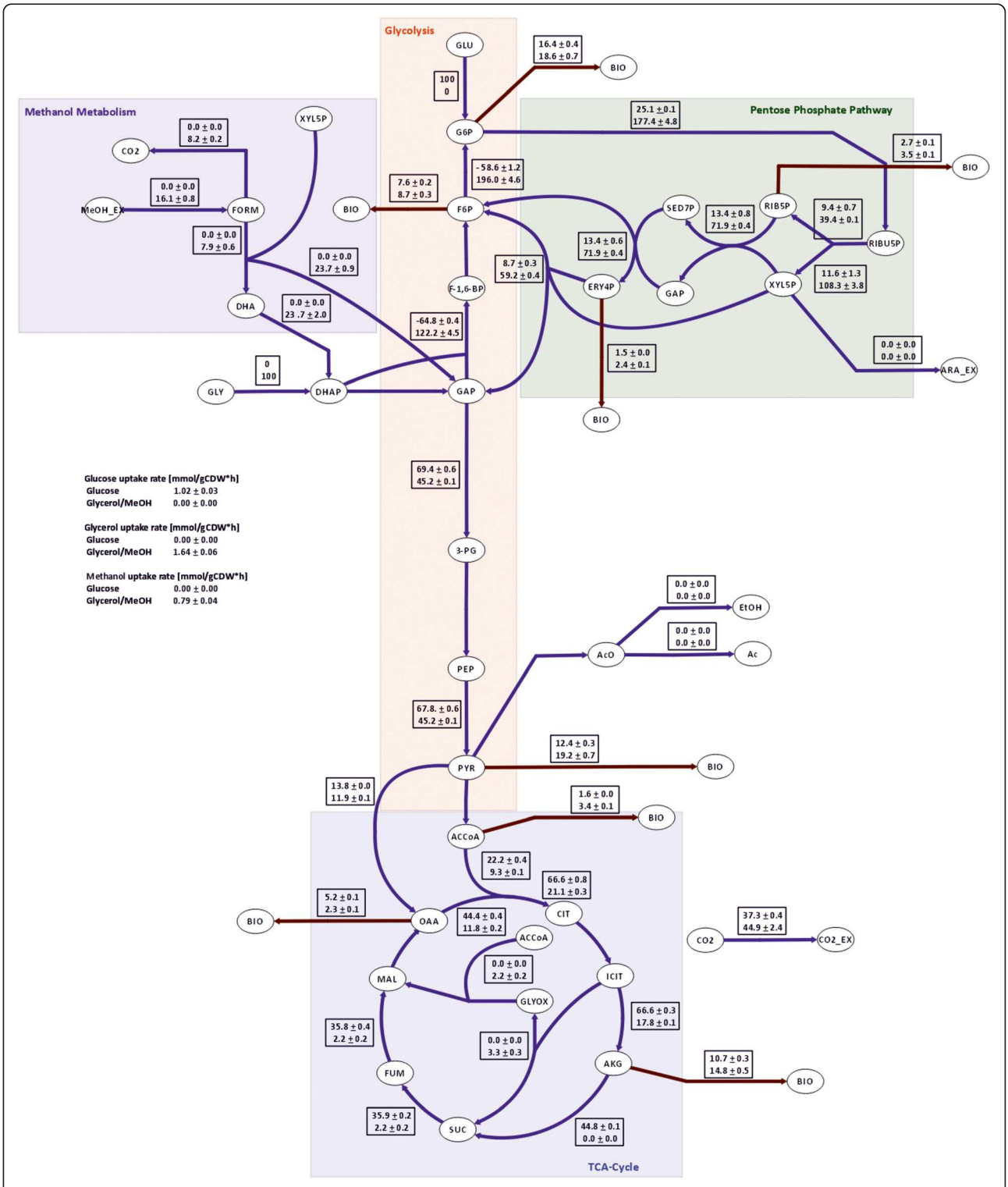

Fig. 5 Intracellular metabolic flux distributions of methanol/glycerol- and glucose-grown cells. The flux values are normalized to glycerol or glucose uptake, respectively, and presented in [% Cmol]. The upper value in the rectangular boxes represents the flux distribution on glucose and the lower value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

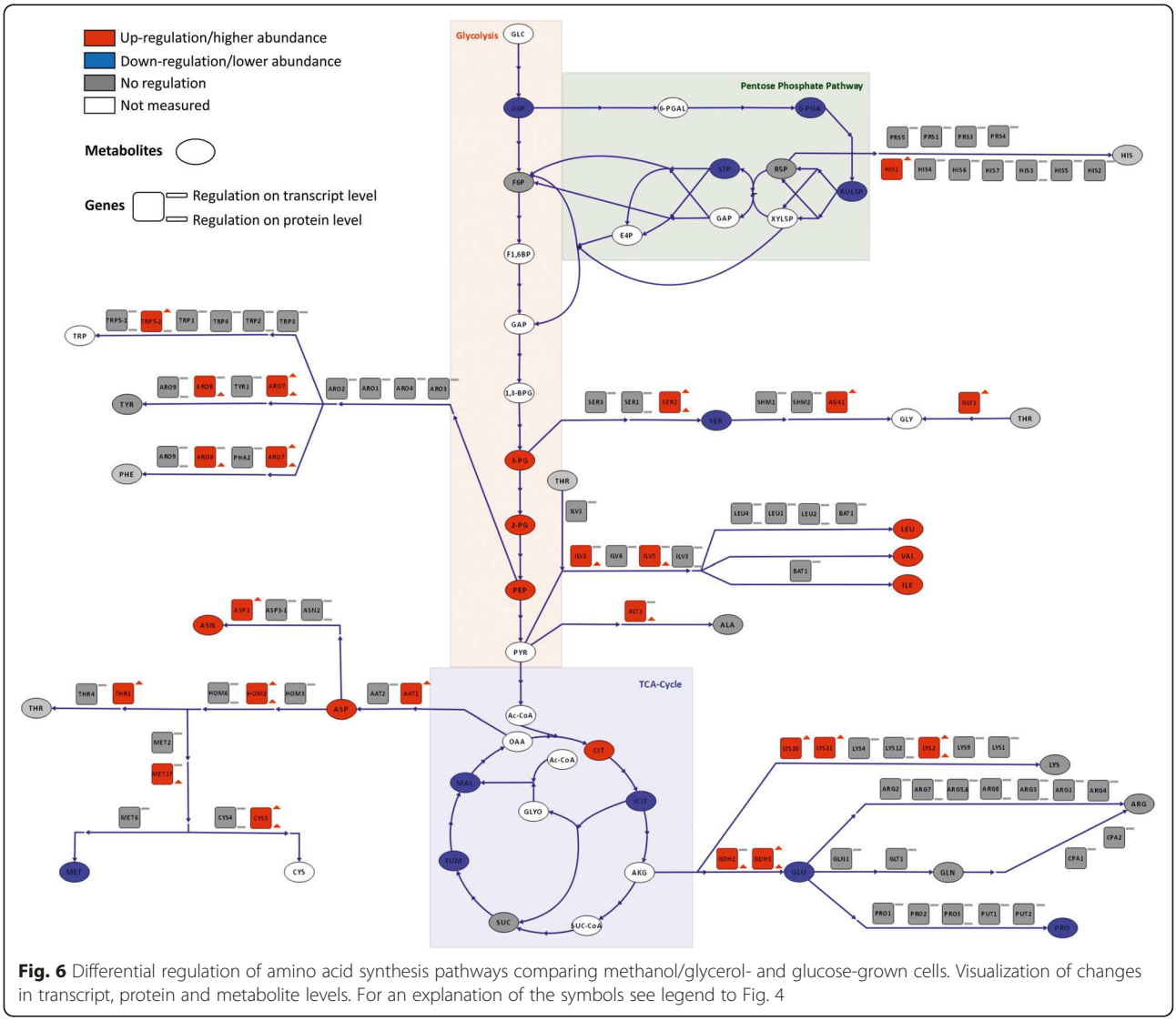

Fig. 6 Differential regulation of amino acid synthesis pathways comparing methanol/glycerol- and glucose-grown cells. Visualization of changes in transcript, protein and metabolite levels. For an explanation of the symbols see legend to Fig. 4

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

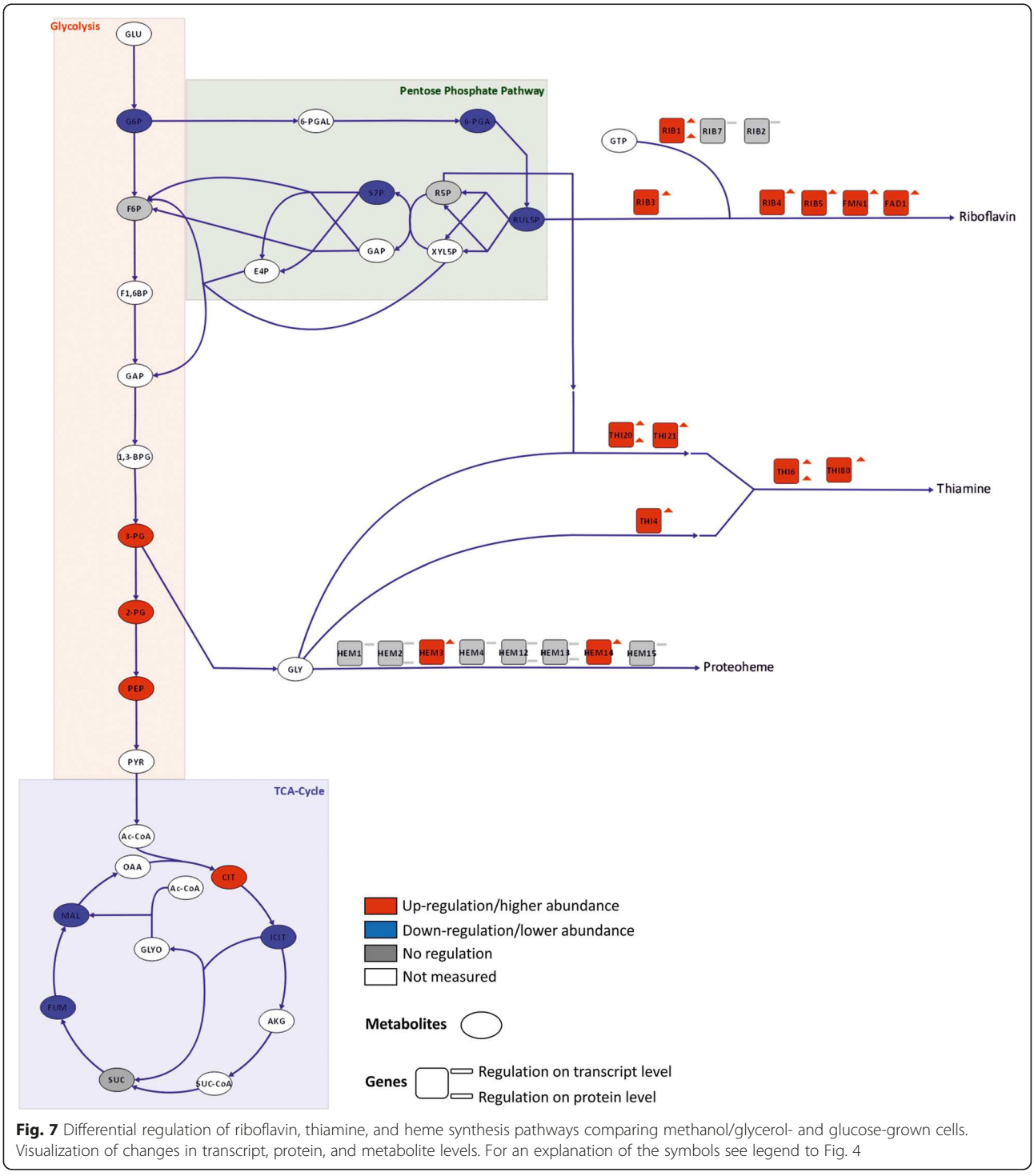

Fig. 7 Differential regulation of riboflavin, thiamine, and heme synthesis pathways comparing methanol/glycerol- and glucose-grown cells. Visualization of changes in transcript, protein, and metabolite levels. For an explanation of the symbols see legend to Fig. 4

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_2FC 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_2FC −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.

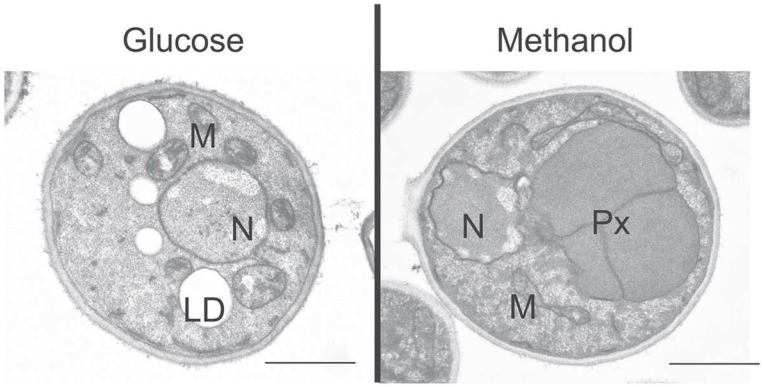

Fig. 8 Electron microscopy of *P. pastoris* grown on glucose or methanol. Cells were cultivated in complex media containing either glucose or methanol as the sole carbon source until they reached the late exponential growth phase. N, Nucleus; M, Mitochondria; LD, Lipid droplet; Px, Peroxisome. Scale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

Metabolomics and Fluxomics analysis of production strains and respective control strains

The metabolomics analysis of free intracellular amino acid levels and intermediates of the central carbon metabolism revealed that recombinant protein production has an impact on certain metabolite levels (Appendix, Fig.11-20). The impact on metabolite levels was more pronounced for the complex HyHEL-Fab protein. Furthermore, we could also show that a change in carbon source also can influence free intracellular amino acid levels. Interestingly, for the carboxylesterase and carboxypeptidase almost no changes in free metabolite levels (for both carbon sources) were detected. A completely different picture was observed for the HyHEL-Fab. During recombinant protein production (for both carbon sources) increased free intracellular amino acid levels were detected. The same trend was observed for the fluxomics data. No change in the intracellular flux distribution was detected for carboxylesterase and carboxypeptidase compared to the respective control. For the HyHEL-Fab an increased flux through the TCA-cycle was detected, which may points at an increased energy demand during recombinant protein production. A more detailed analysis of the metabolomics and fluxomics data can be found in the Appendix.

Customizing the amino acid metabolism of *Pichia pastoris* for recombinant protein production

Manuscript 3

The analysis of the metabolomics data from chemostat cultivation of production strains and amino acid supplementation experiments indicated a connection between amino acid metabolism and recombinant protein production. Additionally, Heyland et al. (2010) revealed that amino acid uptake is highly selective during recombinant protein production. In other words, energy expensive amino acids (i.e. aromatic amino acids) were highly up taken, whereas amino acids with lower energetic costs were synthesized by the cells. Another study analyzing the intracellular amino acid levels during recombinant protein production revealed also a de-regulation of intracellular amino acid levels due to recombinant protein production. Furthermore, the authors stated the hypothesis that the cells might adjust amino acid metabolism to reduce energetic burden caused by recombinant protein production (Carnicer et al., 2012b).

Therefore, the working hypothesis in this study was to decouple amino acid metabolism from energy metabolism to avoid such an energy dependent re-adjustment of amino acid biosynthesis. Therefore, 14 genes connected to amino acid biosynthesis were selected for overexpression.

Genes connected to alanine, serine, proline, methionine, cysteine, and aromatic amino acid biosynthesis were chosen. Except for alanine, at least two genes per pathway were selected as engineering targets. In order to simultaneously de-regulate more than one amino acid synthesis pathways a gene involved in general amino acid control (*GCN4*) was selected for overexpression. *Gcn4* is a transcriptional activator of amino acid biosynthetic genes from 19 out of 20 synthesis pathways. Additionally, *Gcn4* is involved in purine biosynthesis, autophagy and multiple stress response (Hinnebusch & Natarajan, 2002).

Furthermore, a gene connected to NADPH metabolism (*MAE1*) was chosen (Boles et al., 1998). The engineering of NADPH via over-expression of *MAE1* should potentially increase intracellular NADPH level. NADPH is the major electron donor for amino acid biosynthesis. Therefore, a limited supply of NADPH may restrict amino acid biosynthesis, which later on limits recombinant protein production. The effect of overexpression of the metabolic targets on recombinant protein production was tested for two model proteins. On the one hand, a recombinant carboxylesterase (CES) from *Sphingopyxis macrogoltabida* and on the other hand an antibody fragment (HyHEL-Fab) was chosen.

The expression of the recombinant carboxylesterase was under control of the GAP-promoter, whereas the expression of the HyHEL-Fab was under control of the AOX-promoter. Both promoter systems use different carbon sources for growth and production. As the choice of carbon source directly influences intracellular amino acid levels (Manuscript 2) it was interesting to assess the effect of over-expression of the engineering targets for both promoter systems.

The production of recombinant carboxylesterase could be improved by over-expression of genes connected to amino acid metabolism (Fig.9). The strongest improvement in recombinant CES production was observed for the overexpression of *GCN4*. On the contrary, overexpression of genes connected to proline and methionine, cysteine biosynthesis had no effect on CES production (data not shown).

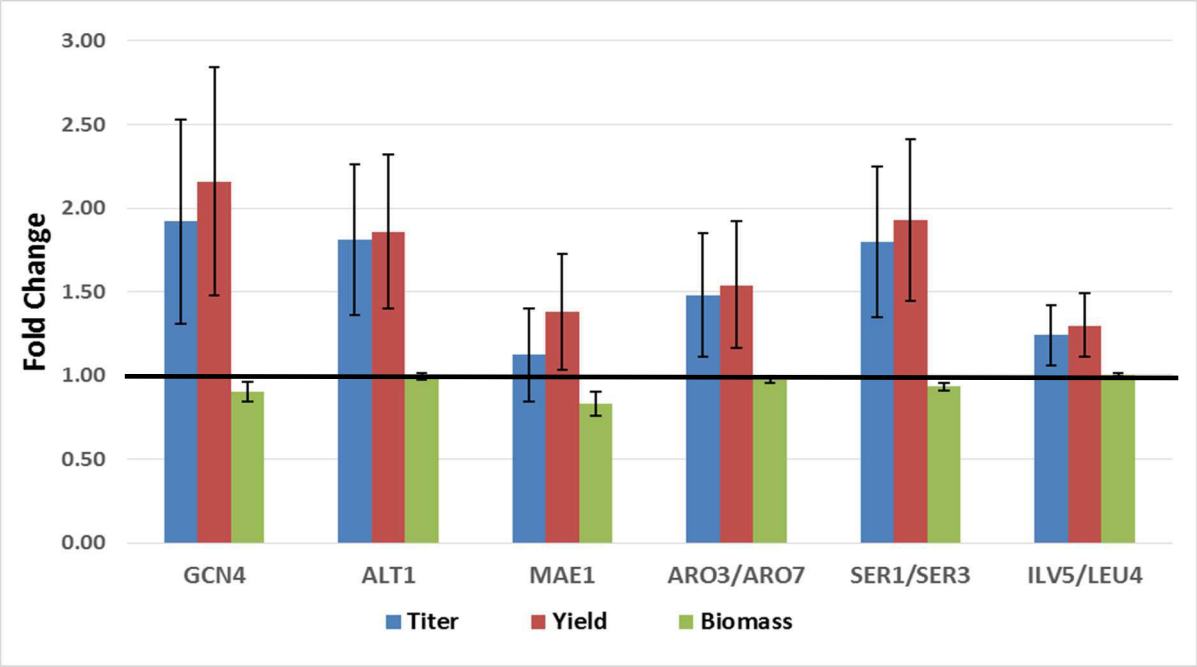

Fig.9: Evaluation of the effect of selected engineering targets on production of a carboxylesterase (CES) under control of P_{GAP}. Relative changes in titer (blue), yield (red) and biomass (green) are shown. Error bars indicate the standard errors of the mean. (10 clones were screened per construct, except for GCN4 5 clones were screened)

For the second model protein the HyHEL-Fab the picture changed completely. None of the tested construct showed an improvement in HyHEL-Fab production (Fig.10)

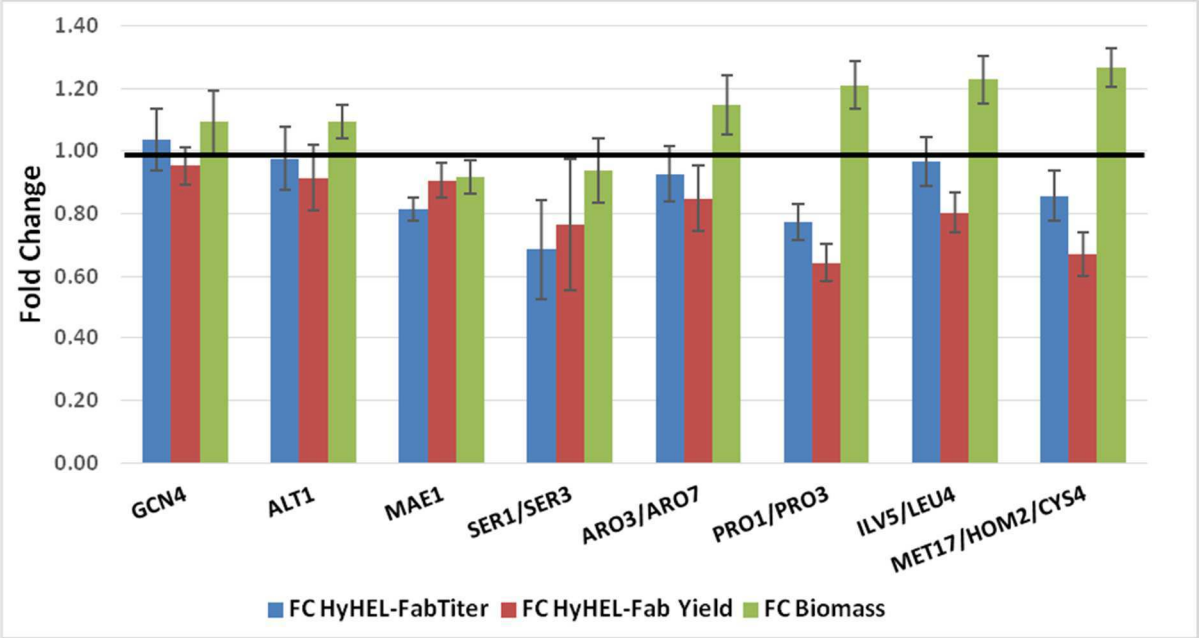

Fig.10: Evaluation of the effect of selected engineering targets on production of a HyHEL-Fab fragment under the control of the P_{AOX}. Relative changes in titer (blue), yield (red) and biomass (green) are shown. Error bars indicate the standard errors of the mean. (10 clones per constructs were screened)

An explanation for the observed picture could be the increased complexity of the recombinant protein. The production of a HyHEL-Fab fragment requires the proper folding and assembly of both antibody chains. Therefore, the major bottleneck for the production of such complex proteins could be attributed to a limitation in protein folding capacity rather than to limitations within metabolism. Another aspect to be considered is the change in carbon source. In Manuscript 2 analysis of growth on methanol/glycerol caused a readjustment of the amino acid metabolism in *P. pastoris*. For methanol/glycerol grown cells mainly up-regulation of genes connected amino acid metabolism was found. Furthermore, free intracellular amino acid pools were increased for methanol induced cells. The observed readjustment of the amino acid metabolism may suppress the effect of the overexpression of the chosen engineering targets.

Customizing Amino Acid Metabolism of *Pichia pastoris* for Recombinant Protein Production

Hannes Rußmayer [1,2], Markus Buchetics [1,2], Stefan Neubauer [2,3], Kristaps Klavins [2,3], Matthias Steiger [1,2], Alexandra B. Graf [2,4], Stephan Hann [2,3], Michael Sauer [1,2], Diethard Mattanovich [1,2] and Brigitte Gasser [1,2]

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

[2] Austrian Centre of Industrial Biotechnology

[3] Department of Chemistry, BOKU - University of Natural Resources and Life Sciences Vienna, Austria

[4] School of Bioengineering, University of Applied Sciences FH Campus Vienna

Abstract

The efficient production of recombinant proteins is often hampered by cellular processes, such as transcription, translation, folding. A major bottleneck concerning efficient protein production in *Pichia pastoris* (syn. *Komagataella sp.*) is attributed to protein folding and secretion. However, recently it was shown that high level recombinant protein production is also limited by cellular metabolism. In order to elucidate cellular processes constituting bottlenecks for recombinant proteins a more detailed view of regulatory patterns during recombinant protein production is necessary. In this study the focus was on the analysis of the impact of metabolism on recombinant protein production. A first comprehensive overview of this impact was obtained by the analysis of the metabolome of different *P. pastoris strains* during recombinant protein production. A deregulation of intracellular amino acid levels was detected for strains producing an antibody fragment. Upon these results, the working hypothesis was to increase intracellular amino acid levels to unburden cellular metabolism and thus improve recombinant protein production. The increased intracellular amino acid level was achieved by overexpressing genes connected to amino acid metabolism. Twelve out of the 14 engineering targets were genes involved in specific amino acid biosynthetic pathways. Furthermore, a transcription factor (*GCN4*) responsible for regulation of amino acid biosynthetic genes during amino acid starvation, and a gene connected to NADPH metabolism (*MAE1*) were over expressed. The over expression of 9 out of 14 engineering targets increased production of a carboxylesterase from *Sphingopyxis macrogoltabida*. Furthermore, the analysis of intracellular amino acid levels for selected clones indicated a direct linkage of improved recombinant carboxylesterase production to the increased availability of intracellular amino acids.

Introduction

Over the past two decades yeasts emerged as efficient production host for recombinant protein production. At the end of 2011 about 15% of all approved biopharmaceutical proteins were produced using yeasts as expression system (Berlec & Strukelj, 2013). Among the yeast expression systems, *Saccharomyces cerevisiae* and *Pichia pastoris* (syn. *Komagataella sp.*) are the major players (Gasser *et al.*, 2013). *P. pastoris* already proved to allow high level recombinant protein production. In most cases the overexpression of the recombinant protein negatively effects cellular growth and thus productivity of the recombinant protein of interest (Mattanovich *et al.*, 2012). From a metabolic point of view the observed phenomenon is possibly caused either by extra energy needed for the production of the recombinant protein or by a limited supply with precursors for recombinant protein synthesis (such as amino acids). The additional demand of cellular resources needed for recombinant protein production is summarized as metabolic burden (Glick, 1995). Indeed, physiological studies of *P. pastoris* during recombinant protein production indicated a re-distribution of intracellular carbon fluxes towards the TCA-cycle and ATP production (Dragosits *et al.*, 2009; Jordà *et al.*, 2012; Nocon *et al.*, 2014). An explanation is that the additional energy demand is created due to bottlenecks in the posttranslational machinery (folding and secretion). Furthermore, re-folding and/or degradation of unfolded or misfolded proteins are energy expensive processes increasing the energy demand during recombinant protein production (Mattanovich et al., 2004). This effect is also observed at rather low recombinant protein production levels and even when the produced amount of recombinant protein is neglectable compared to the total cellular protein. Therefore, a simple drain of precursor metabolites may not explain the observed metabolic burden. However, the addition of amino acids to the cultivation media proved to be a valuable strategy to unburden cellular metabolism for recombinant protein production (Görgens et al., 2005; Heyland et al., 2011). Heyland and coworkers (2011) revealed that addition of energetically expensive amino acids or glutamate increased recombinant production of β-aminopeptidase in *P. pastoris*. Interestingly, the amino acid uptake for supplemented cultures directly correlated with the energy costs for synthesis of the amino acid. Carnicer et al. (2012) specifically analyzed the intracellular amino acids levels during recombinant Fab production in *P. pastoris*. The comparison of intracellular amino acid levels between the control and the Fab production strain revealed an increase for some intracellular amino acid levels. In general, the authors stated that a re-adjustment of amino acid metabolism happened to compensate for the additional resources needed for recombinant protein production. Furthermore, a comparison of fold increase in intracellular amino acid levels and the energy costs of their synthesis revealed lower increase of energy expensive amino acids under conditions of higher Fab production. The results of

both studies point at a general readjustment of the amino acid metabolism to minimize synthesis costs of amino acids. Therefore, the approach followed in this study was to decouple amino acid biosynthesis from the observed energy demand adjustment via overexpression of specific genes connected to amino acid metabolism. The genes for engineering the amino acid metabolism of *P. pastoris* were chosen according to previously obtained metabolomics data of intracellular amino acid pools of *P. pastoris* during recombinant protein production and the amino acid composition of the protein of interest. The effect of increased intracellular amino acid pools due to overexpression of selected amino acid biosynthetis genes was tested for a carboxylesterase (CES) from *Sphingopyxis macrogoltabida.*

Material and Methods

Strains and Vectors

P. pastoris wild type CBS 7435 (Centraalbureau voor Schimmelcultures, NL) was used as host organism. For the expression of the carboxylesterase from *Sphingopyxis macrogoltabida* the plasmid pPM2dZ30-PGAPα, a derivate of the pPUZZLE vector (Stadlmayr *et al.*, 2010), was used. For transformation, the plasmid was linearized using the restriction enzyme AvrII for homologous integration into the native *GAP* promoter. The generation of the strain expressing the carboxylesterase was described in Ruth *et al.* (2014). The strain expressing the HyHEL-Fab heavy chain and light chain was generated using the strategy described in Gasser *et al.*(2006).

Overexpression of engineering targets

The coding sequence of the defined engineering targets were amplified from CBS7435 genomic DNA and cloned into the pPUZZLE derived plasmid pPM2aK21, containing a KanMX4 cassette conferring resistance to kanamycin/geneticin (G418) and the GAP promoter. For transformation, the plasmid was linearized using the restriction enzyme *AscI* and integrated into the AOX1 terminator locus of *P. pastoris*. Plasmid containing a combination of expression cassettes (for example combination of *ARO3* and *ARO7*) were constructed by double digestion of the *ARO7* vector with *BglII* and *BamHI*, and subsequent insertion of this construct in the unique *BglII* of the vector containing *ARO3*. The same strategy was used for the construction of the plasmid containing the different combinations of engineering targets.

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.

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 400 ml batch medium in 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 condition, three independent chemostat cultivations were performed. Temperature, pH and dissolved oxygen were maintained at 25°C, 5.0 (with 8M KOH) and 20% dissolved oxygen (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 $MgSO_4.7H_2O$, 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 $CaCl_2.2H_2O$, 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.

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 $CaCl_2.2H_2O$, 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).

Cultivation of engineered strains

10 mL YP-medium (per liter: 10 g yeast extract, 20 g peptone) containing 10 g/L glycerol and 50 µg/mL Zeocin were inoculated with a single colony of engineered *P. pastoris* strains and grown overnight at 25 °C. For the main culture 10 mL of freshly prepared M2-Medium (pH 5.0, containing 20 g/L glucose as carbon source) was filled into a 100 mL shake flask. The main culture was inoculated with an OD of 0.1 with the washed pre culture. After inoculation the main culture was cultivated for

52 h at 25°C and 180 rpm on a shaker. Cells were fed with 0.5 % Glucose (100 µL of a 50% glucose stock solution) after 22 h, 28 h, 42 h and 48 h. After 52 h samples for the determination of carboxylesterase production were taken. The M2-medium contained per liter: 22.0 g citric acid monohydrate 3.15 g $(NH_4)_2PO_4$, 0.49 g $MgSO_4*7H_2O$, 0.80 g KCl, 0.0268 g $CaCl_2*2H_2O$, 1.47 mL PTM1 trace metals, 4 mg Biotin; pH was set to 5 with KOH (solid).

Quantification of carboxylesterase production with ELISA

Quantification of CES by ELISA was done using rabbit-Anti-CES antiserum 6287 (produced in house) as coating antibody, guinea pig-anti-CES antiserum 12206 (produced in house) and goat-anti-guinea pig-IgG-alkaline phosphatase conjugate (Sigma, A-5062) as detection antibody. Purified CES was used as standard with a starting concentration of 100 ng/mL, supernatant samples are diluted accordingly. Detection was done with pNPP (Sigma S0942). Coating-, Dilution- and Washing buffer were based on PBS (2 mM KH_2PO_4, 10 mM $Na_2HPO_4.2\ H_2O$, 2.7 mM g KCl, 8 mM NaCl, pH 7.4) and completed with BSA (1% (w/v)) and/or Tween 20 (0.1% (v/v)) accordingly.

Sampling and Quenching

Samples for analysis of intracellular metabolites were taken immediately by using a pump. Approximately 50 mL fermentation broth was 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 through a cellulose acetate filter (0.45 µm, Sartorius 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, Sartorius 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 pre-weight beakers.

Analysis of intracellular amino acid levels

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 sec. 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 seconds, 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, 4000g). The ethanolic extract was decanted into a fresh cooled 15 mL tube and kept on dry ice until LC-MS/MS analysis. The sample preparation for LC-MS/MS analysis of free intracellular metabolites was done according to Klavins et al. (2013). For the comparison of intracellular amino acid levels fold changes and p-values (for statistical significance) were calculated. Intracellular amino acids were considered as higher abundant when following criteria were met FC >1.20 and p-value < 0.05. In case of lower abundance the criteria FC <-1.20 and p-value < 0.05 were applied.

Microarrays and data analysis

The RNA was isolated from chemostat sample cells using TRI reagent according to the supplier's instructions (Ambion, US). 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 done according to the Agilent protocol for 2-color expression arrays. Each sample was hybridized against a 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. For identifying 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 done using the R software package http://www.rproject.org, and the limma package.

Results

Identification of metabolic engineering targets

Amino acids constitute the building blocks for the synthesis of recombinant proteins. Therefore, the question to answer was if a limited supply with intracellular amino acids may restrict the synthesis of the recombinant protein. In order to test if amino acids are the limiting factor, firstly amino acid supplementation was tested on two different substrates (glucose, methanol). The following media with supplemented amino acids were chosen to analyze their effect on production of a recombinant antibody fragment (1) addition of proline, (2) addition of glutamate (main amino donor), (3) addition of cysteine (in general low levels in *P. pastoris*), (4) addition of methionine, (5) addition of proline,

glutamate and methionine, (6) addition of all amino acids. The supplementation of all selected amino acids to the cultivation medium in combination with glucose as carbon source increased production of the antibody fragment (Fig.1). The addition of proline alone increased the productivity 1.5 fold. The increase in productivity for the other amino acid combinations was almost at the same level and yielded a 1.4 fold increase. Thus, the results were consistent in the fact that addition of amino acids is beneficial for recombinant protein production. For growth on methanol addition of proline again proved to increase production of antibody fragments (1.4 fold and 1.1 fold). Interestingly, addition of a combination of all amino acids decreased productivity for both tested antibody fragments (Fig.2). This indicates that amino acids are not simply the building blocks for protein synthesis. Furthermore, intracellular amino acid concentrations seem to be involved in other processes influencing the synthesis of the recombinant protein. In order to more specifically analyze the connection between intracellular amino acid levels and recombinant protein production a metabolomics analysis of intracellular amino acid levels was performed, comparing a HyHEL-Fab producer and a non-expression control. The analysis for potential limiting amino acids was done on two different carbon sources (glucose and a mixture of methanol/glycerol), since the amino acid supplementation experiments indicated that the used carbon source impacts the effect of the added amino acids on recombinant protein production. The relative changes in intracellular amino acid levels for the HyHEL-Fab expressing strain compared to the control were mapped onto a metabolic map (Fig.3). Intracellular amino acid levels with a significant difference in abundance were marked in either red (higher abundance in the HyHEL-Fab expression strain) or blue (lower abundance in the HyHEL-Fab expression strain). If glucose was used as carbon sources for the production of the antibody fragment 12 intracellular amino acid pools were differentially abundant compared to the respective control strain (3 amino acid pools with significantly lower abundance, 9 amino acid pools with significantly higher abundance) (Fig.3A). Interestingly, amino acids belonging to the group of branched chain amino acids and aromatic amino acids were higher abundant in the antibody expressing strain, whereas amino acids derived from TCA-cycle intermediates showed a lower abundance. Cultivation of a strain producing the same Fab fragment using glycerol/methanol as carbon source showed that 9 out of 16 measured intracellular amino acid pools were significantly altered (3 amino acid pools with significantly higher abundance and 6 amino acid pools with significantly lower abundance) (Fig. 3B). Additionally, for both conditions (cultivation of Fab expression strain on glucose and Fab expression strain on methanol/methanol) an analysis of the intracellular flux distribution and transcript level for genes associated with amino acid metabolism was conducted. The comparison of the intracellular flux distribution in the Fab expressing strain vs. control strain revealed a redistribution of intracellular

fluxes towards TCA-cycle and subsequent energy production during recombinant protein production for both conditions (glucose and methanol/glycerol, data not shown). The observed adjustment seems to be necessary to meet the increased energy demand for recombinant Fab production. Transcript level analysis revealed significant changes for genes involved in amino acid biosynthesis pathways during recombinant protein production in both conditions (Fig.3). Therefore, the working hypothesis was to decouple amino acid biosynthesis from regulatory mechanisms to increase recombinant protein production. Based on the obtained data fourteen genes connected to amino acid metabolisms were chosen as engineering targets (Tab.1). Twelve genes are directly connected to specific amino acid biosynthetic pathways. *GCN4*, a transcription factor involved in general amino acid control and regulation of different amino acid biosynthetic genes was also selected. Beside to up-regulated amino acid biosynthesis in general also cofactor availability for biosynthesis plays an important role. With regard to amino acid biosynthesis, NADPH constitutes the major electron donor. To increase intracellular NADPH pools malic enzyme (Mae1) was chosen as engineering target. *MAE1* was selected because of the direct connection to NADPH metabolism (Boles et al., 1998).

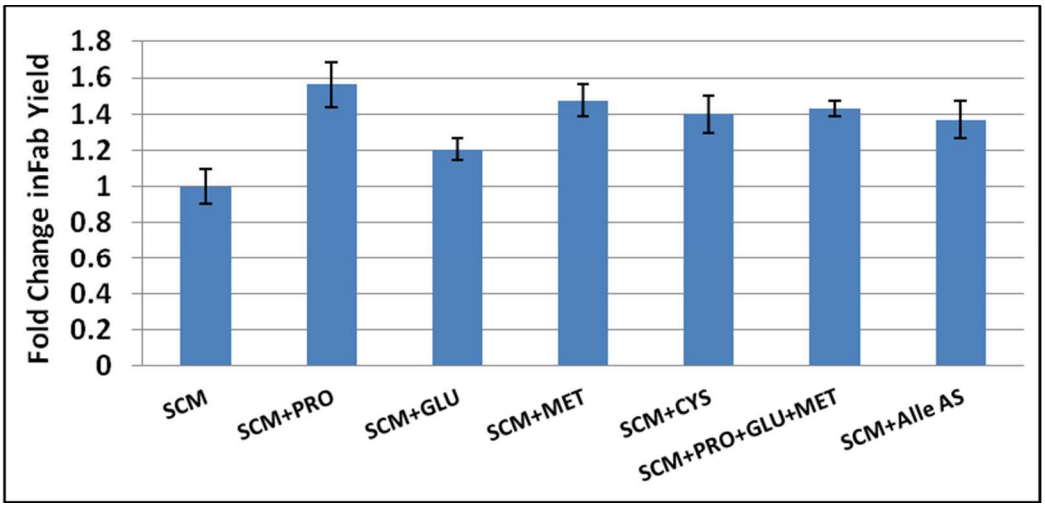

Fig.1: Effect of amino acids supplementation on HyHEL-Fab production with glucose as carbon source. Relative changes of HyHEL-Fab- yield are shown. Error bars indicate the standard errors of the mean. (mg HyHEL-Fab/gCDW on media supplemented with different amino acids relative to the strain cultivated on screening media with no amino acids supplemented)

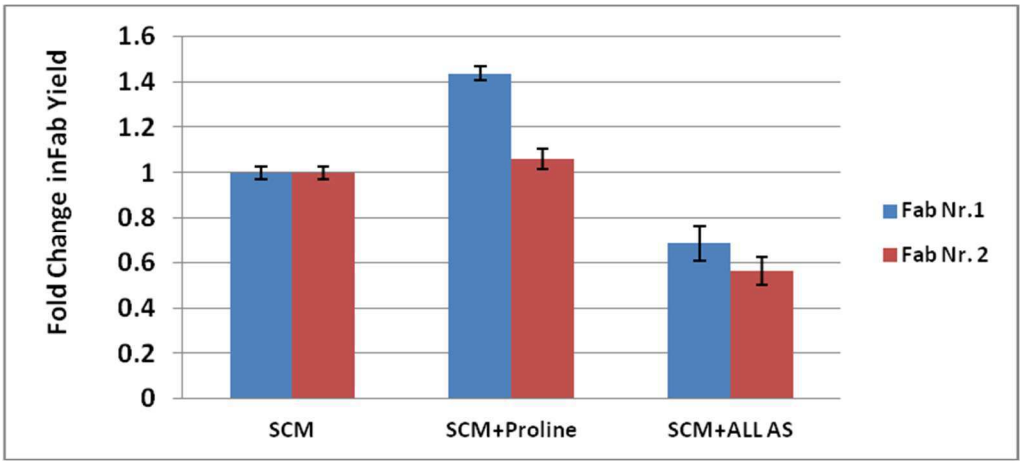

Fig.2: Effect of amino acids supplementation on Fab production with methanol as carbon source. Relative changes of Fab yield are shown. Error bars indicate the standard errors of the mean. (mg Fab/gCDW on media supplemented with different amino acids relative to the strain cultivated on screening media with no amino acids supplemented)

Short Name	*S. cerevisiae* ORF name	Description (taken from *Saccharomyces* Genome Database)
General Amino Acid Control		
GCN4	PP7435_Chr1-1124	bZIP transcriptional activator of amino acid biosynthetic genes; activator responds to amino acid starvation; expression is tightly regulated at both the transcriptional and translational levels
Alanine biosynthesis		
ALT1	PP7435_Chr3-0727	Alanine transaminase (glutamic pyruvic transaminase); involved in alanine biosynthesis and catabolism
Phenylalanine & Tyrosine and Tryptophane biosynsthesis		
ARO3	PP7435_Chr3-0936	3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase; catalyzes the first step in aromatic amino acid biosynthesis and is feedback-inhibited by phenylalanine or high concentration of tyrosine or tryptophan
ARO7	PP7435_Chr4-0965	Chorismate mutase; catalyzes the conversion of chorismate to prephenate to initiate the tyrosine/phenylalanine-specific branch of aromatic amino acid biosynthesis
Isoleucine & Leucine and Valine biosynthesis		
ILV5	PP7435_Chr1-0750	Acetohydroxyacid reductoisomerase and mtDNA binding protein; involved in branched-chain amino acid biosynthesis
LEU4	PP7435_Chr2-0884	Alpha-isopropylmalate synthase (2-isopropylmalate synthase); the main isozyme responsible for the first step in the leucine biosynthesis pathway
Serine biosynthesis		
SER1	PP7435_Chr3-0640	3-phosphoserine aminotransferase; catalyzes the formation of phosphoserine from 3-phosphohydroxypyruvate, required for serine and glycine biosynthesis
SER3	PP7435_Chr2-0624	3-phosphoglycerate dehydrogenase; catalyzes the first step in serine and glycine biosynthesis
Proline biosynthesis		
PRO1	PP7435_Chr3-0925	Gamma-glutamyl kinase; catalyzes the first step in proline biosynthesis
PRO3	PP7435_Chr4-0587	Delta 1-pyrroline-5-carboxylate reductase; catalyzes the last step in proline biosynthesis
Methionine & cysteine biosynthesis		
MET17	PP7435_Chr4-0665	O-acetyl homoserine-O-acetyl serine sulfhydrylase; required for methionine and cysteine biosynthesis
HOM2	PP7435_Chr2-0144	Aspartic beta semi-aldehyde dehydrogenase; catalyzes the second step in the common pathway for methionine and threonine biosynthesis
CYS4	PP7435_Chr3-0736	Cystathionine beta-synthase; catalyzes synthesis of cystathionine from serine and homocysteine, the first committed step in cysteine biosynthesis; responsible for hydrogen sulfide generation
Energy metabolism		
MAE1	PP7435_Chr3-1050	Mitochondrial malic enzyme; catalyzes the oxidative decarboxylation of malate to pyruvate, which is a key intermediate in sugar metabolism and a precursor for synthesis of several amino acids

Tab. 1: Summary of defined metabolic engineering targets

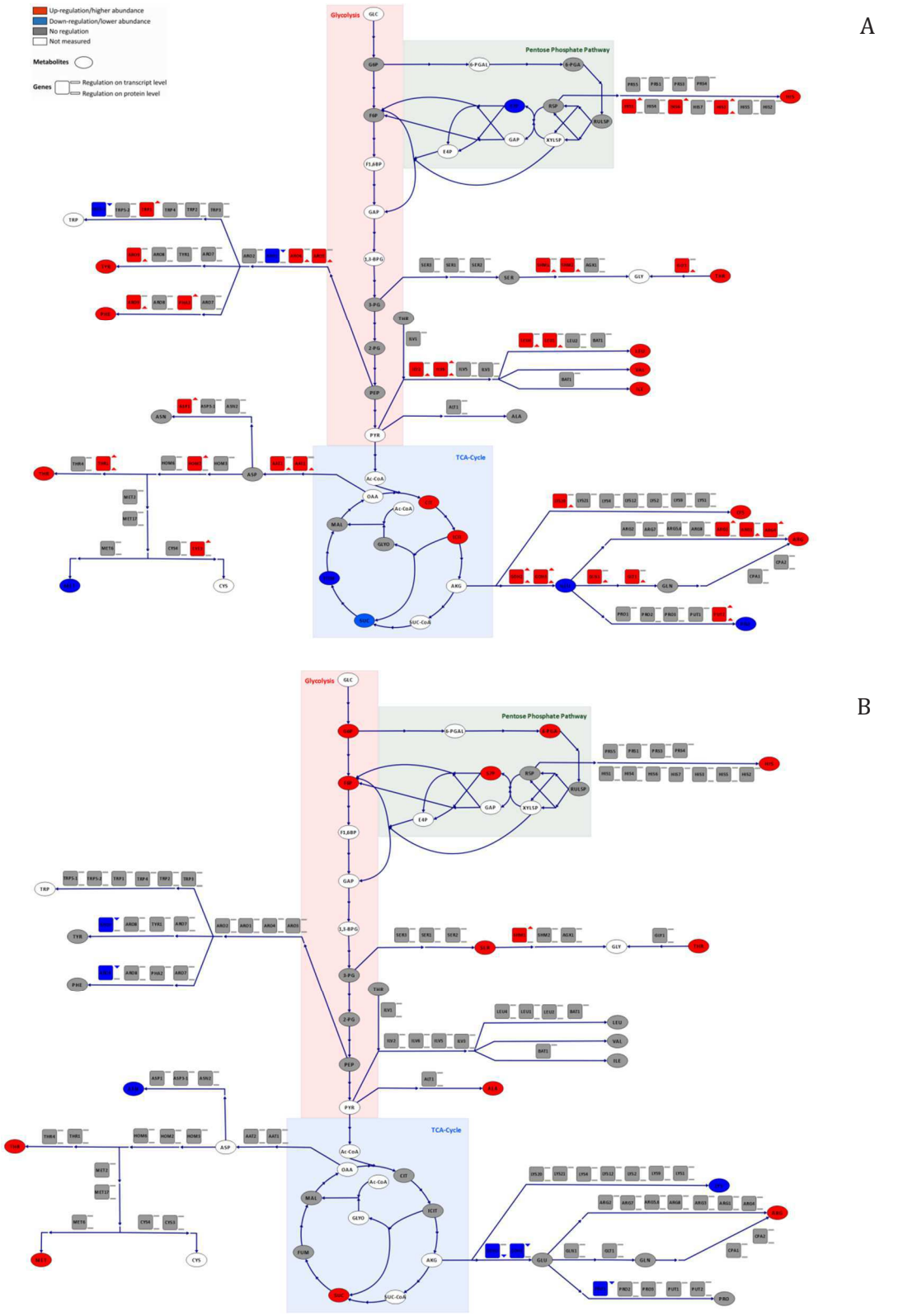

Fig.3: Visualization of metabolite abundance and regulated genes for a *P. pastoris* strain producing an antibody fragment on different carbon sources compared to the respective controls. (A) Glucose as carbon source (B) Methanol/glycerol mixture as carbon source. (Legend: blue colour: lower levels in Fab expressing strain, red colour: higher levels in Fab expressing strain, grey colour: no regulation, white colour: not measured, rectangular symbols: genes, elliptical symbols: metabolites)

Evaluation of the effect of engineering of amino acid metabolism on recombinant carboxylesterase production in *P. pastoris*

The obtained metabolomics data indicated the direct impact of recombinant antibody production on intracellular amino acid levels, even when the amount of recombinant protein is small compared to the cellular protein produced. The observed effect should be even more pronounced when more recombinant protein is produced. Therefore, a less complex recombinant protein, a carboxylesterase (CES) from *S. macrogoltabida* with higher expression levels was used as model protein. The results of the screening rounds showed that 5 out of 8 tested constructs improved recombinant production at least by 20 % in CES yield (fold change > 1.2)(Fig.4). The highest improvement (FC 1.92 for titer and FC 2.16 for yield) in CES production was achieved by over expression of *GCN4*, encoding the transcriptional activator of amino acid biosynthesis genes (Hinnebusch & Natarajan, 2002). Over expression of genes connected to alanine, serine, and branched chain amino acid biosynthesis improved titer and yield about 1.8 fold. The average improvement for clones engineered in aromatic amino acid biosynthesis was 1.5 fold in titer and yield. For strains over expressing the malic enzyme (*MAE1*) an improvement in yield of about 10% was observed. Overexpression of genes involved in proline and methionine/cysteine biosynthesis did not show a positive effect on product yields. Overall, the obtained data suggests that increased fluxes through certain amino acid biosynthetic pathways are beneficial for recombinant protein production. Furthermore, the results show that production is limited by cellular metabolism. Next we analysed, if engineering of amino acid biosynthetic pathways increased free intracellular amino acid levels. Therefore, for selected engineered clones a metabolomics analysis of intracellular amino acid levels was done.

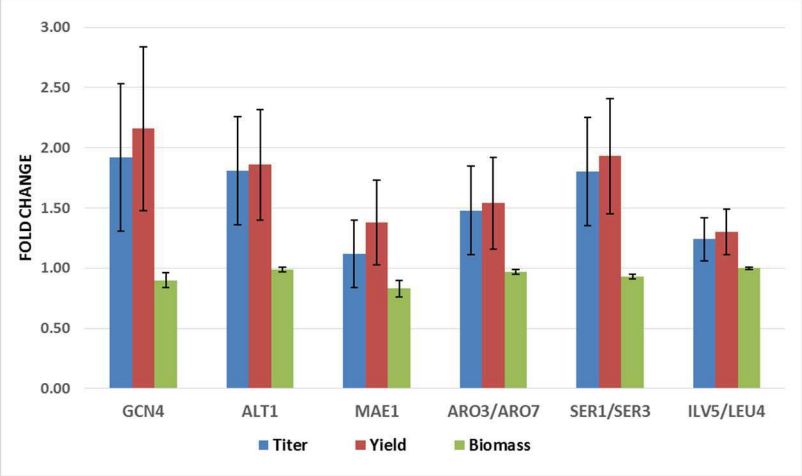

Fig.4 Engineering of amino acid metabolism and its effect on production of recombinant CES. Relative changes in titer (blue), yield (red) and biomass (green) are shown. Error bars indicate the standard errors of the mean. (10 clones per construct were screened, except for GCN4 5 clones were analyzed)

Analysis of intracellular amino acid levels for selected engineered clones

For each construct, at least two engineered clones and two control strains were analyzed. In order to understand, if differences in intracellular amino acid pools also possibly explain clonal differences in CES productivity, at least one clone with improved CES production and one clone with low or no improvement of CES production were selected. For the comparison of intracellular amino acid levels of engineered strains with the non-engineered empty vector control (EVC) FC and p-values were calculated.

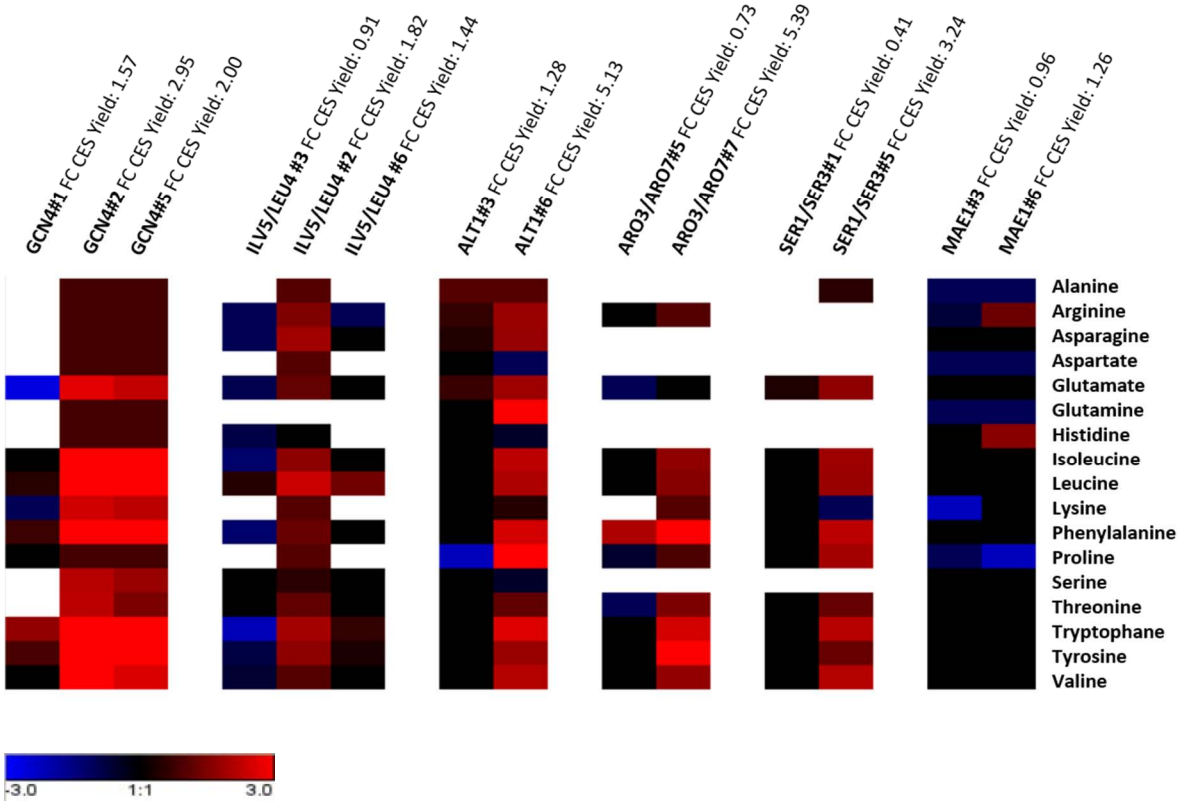

Fig.5: Impact of over-expression of engineering targets on intracellular amino acids levels. (Relative changes in intracellular amino acid levels are depicted, selected clones vs. control strains, values shown are Log2FC)

The detailed analysis of intracellular amino acid levels showed that over-expression of selected genes within a specific amino acid biosynthetic pathway increased the intracellular levels of the respective amino acid (Fig.5). For example strains with engineered alanine biosynthesis showed increased intracellular alanine levels compared to the respective control. Furthermore, for the construct overexpressing genes involved in aromatic amino acid biosynthesis increased levels for phenylalanine, tyrosine and tryptophan were detected for clone ARO3/ARO7#7 (Fig.5). However, for ARO3/ARO7#5

only increased levels for phenylalanine and tyrosine were detected. The over expression of the malic enzyme had almost no effect on intracellular amino acid levels.

A second issue during analysis of the metabolomics data was the comparison of the regulatory pattern for intracellular amino acid pools for engineered clones with different CES productivity. For the engineering target, *GCN4* three different clones were analyzed (Fig.5). Two clones (GCN4#2 and GCN4#5) showed highly increased CES production compared to the control strain, whereas GCN4#1 was less improved in CES production. The analysis of intracellular amino acid levels showed that for improved CES producers intracellular amino acid levels were drastically increased (30 fold increased for engineered clones). However, for the clone with less improvement in CES production the increase in intracellular amino acid levels was less pronounced (Fig.5). For GCN4#2 and GCN4#5 all intracellular amino acid levels were significantly higher (up to 30 fold increased for Trp, Try). However, for GCN4#1 only 5 intracellular amino acid pools had significantly higher abundance (Leu, Phe, Try, Trp). Additionally, three intracellular amino acid pools were detected with a lower abundance (Glu, Lys, Thr). Furthermore, the order of magnitude in regulation of intracellular amino acid pools for GCN4#1 was far smaller, than for both highly improved CES producer clones. A similar trend was observed for the other constructs tested. Unexpectedly, the engineering of a specific amino acid biosynthetic pathway led also to a deregulation of other intracellular amino acid levels. This effect was more pronounced for clones with improved CES production, where an increase in almost all intracellular amino acid levels was detected (Fig. 5). Interestingly, the regulatory pattern in intracellular amino acid concentrations was completely different for clones with less improved CES production. The obtained data indicated that an improvement in recombinant production of CES is directly connected with increased amino acid biosynthesis and that metabolism limits production of CES.

Discussion

Recombinant protein production in *P. pastoris* was reported to increase the intracellular TCA-cycle flux trying to compensate the increased energy demand during recombinant protein production (Dragosits et al., 2009; Heyland et al., 2010; Nocon et al., 2014). Beside the changes in the intracellular flux distribution also a change in uptake and secretion rate of substrates and by products was detected. For example, the flux through fermentative pathways (e.g. ethanol production) decreased during recombinant protein production (Heyland *et al.*, 2011; Nocon *et al.*, 2014). Supplementation of amino acids to the cultivation media already proved to be a valuable approach to partially unburden cellular metabolism for recombinant protein production (Görgens et al. 2005, Heyland et al. 2011). Interestingly, the uptake rate for energy expensive amino acids was higher than for amino acids with less energetic synthesis costs. In other words *P. pastoris* tries to compensate for the increased energy demand by simply avoiding additional synthesis costs of energy expensive amino acids. However, we showed that amino acid supplementation was even detrimental to product yield. Addition of all amino acids to the Fab producer grown on methanol as carbon source is not always beneficial for recombinant protein production and highly dependent on the used carbon sources. Therefore, intracellular amino acid concentration also may trigger several other metabolic processes within a cell, which influence recombinant protein production. For growth on methanol it is known the productivity of the recombinant protein is highly dependent on the nitrogen source in the media (Rumjantsev et al., 2013). Most amino acids are known as rather poor nitrogen sources and therefore may negatively influence productivies. Our analysis of intracellular amino acid levels showed a direct impact of recombinant protein production on intracellular amino acid levels. The metabolomics analysis revealed that 9 out of 18 free intracellular amino acids pools were increased when glucose is used as carbon source. The same trend was also observed for expression of other Fab antibody fragment, where 10 out of 16 free amino acids pools were increased (Carnicer *et al.*, 2012). Interestingly, in both cases free intracellular proline pools were down regulated. Furthermore, Carnicer et al. (2012) also showed a dependence of the increase in intracellular amino acid pools on the energy cost of their synthesis during conditions with higher recombinant protein productivity. The increased energy demand during recombinant protein production possibly triggers amino acid metabolism in a way to decrease fluxes through energetically costly amino acids, which in the end can limit the synthesis of the recombinant protein. Therefore, the working hypothesis in this study was to decouple amino acid metabolism from this energy dependent shutdown via overexpression of genes connected to amino acid metabolism. Furthermore, as energy is proposed as a limiting factor also a gene involved in energy metabolism (*MAE1*) was chosen for overexpression. The highest

improvement in recombinant carboxylesterase production was observed, when the transcription factor *GCN4* was overexpressed. Accompanied with the increased productivity the highest increase in free intracellular amino acid levels was detected. Most strikingly, for the best producer clones all intracellular amino acid levels were increased compared to the respective control, whereas a clone with less improved production showed a completely distinct pattern in regulation of intracellular amino acid pools. The increase in all intracellular pools is possibly explained by the fact that *GCN4* regulates genes within 19 out of 20 amino acid biosynthetic pathways (Hinnebusch, 1988). A direct linkage between increased intracellular amino acid levels and increased transcript levels for *GCN4* was shown for *S. cerevisiae* during recombinant insulin production (Kazemi et al., 2013). Furthermore, Gcn4p was found to interact with Hac1p triggering the UPR response, which in the end could be beneficial for recombinant protein production (Patil et al., 2004). A second target for overexpression was *ALT1*, a gene connected to alanine biosynthesis. This amino acid was potentially interesting as the amino acid composition of the model protein showed a higher frequency for alanine compared to the *P. pastoris* proteome. The higher frequency of alanine in the model protein may cause an increased need for this specific amino acid, which should be compensated by higher alanine biosynthesis. For *Escherichia coli*, it was recently shown that the amino acid composition of the recombinant protein directly influences the product yield (Rahmen *et al.*, 2015). Indeed, intracellular alanine levels for the engineered clones were increased compared to the respective control. Furthermore, the size of the intracellular alanine pools seems to trigger recombinant protein production, as for improved producer clones pool sizes were up to 20 fold increase compared to clones with less improvement. Also for *S. cerevisiae* it was observed that overexpression of *ALT1* directly influences intracellular alanine levels (Peñalosa-Ruiz *et al.*, 2012; Yu *et al.*, 2013). Amino acid supplementation, especially aromatic amino acids, was shown to be beneficial for recombinant protein production. Furthermore, the uptake of these amino acids was favoured over *de novo* synthesis, which is possibly explained by the fact that synthesis of aromatic amino acids is energetically expensive (Heyland et al. 2011). Branched chain amino acids constitute another group of energetically expensive amino acids and therefore the biosynthesis of these amino acids was shown to limit recombinant protein production (Raiford *et al.*, 2008). In order to assure sufficient supply of these amino acids despite energy limitation during recombinant protein production, genes connected to aromatic and branched chain amino acid biosynthesis were over expressed. Furthermore, the amino acid composition of the model protein indicated a higher frequency for aromatic amino acids compared to the *P. pastoris* proteome, which also emphasises the higher need for such amino acids. The engineering of aromatic amino acid biosynthesis and branched chain amino

acid biosynthesis was shown to have a beneficial effect on the production of recombinant carboxylesterase. The comparison of intracellular amino acid levels for clones with engineered aromatic amino acid biosynthesis (producer vs. EVC) showed that the improved producer clones again had an increased level for all measured intracellular amino acids, whereas the intracellular amino acid levels for the clones with less improved CES production stayed almost unchanged. Furthermore, for improved clones the order of magnitude in increase of intracellular amino acid pools was higher than in clones with less improvement in CES production. The same picture was observed for clones simultaneously overexpressing genes connected to branched chain amino acids and serine biosynthesis. The low cysteine levels in P. pastoris correlate with limited synthesis capacity for this amino acid (Sohn *et al.*, 2010). Despite the low cysteine levels in *P. pastoris* (Carnicer et al. 2012), no improvement in carboxylesterase productivity was detected when genes connected to methionine and cysteine biosynthesis were over expressed. In summary, increasing fluxes through tightly regulated amino acid biosynthetic pathways showed to be beneficial for recombinant protein production. A deregulation of a specific amino acid biosynthetic pathway caused not only a change in the respective intracellular amino acid levels, but also in many other amino acid levels. The same effect was observed for *S. cerevisiae* with an engineered aromatic biosynthetic pathway. Besides increased levels of aromatic amino acid also an increase in several other non-aromatic amino acids was detected (Luttik *et al.*, 2008). The authors hypothesized that this effect is caused by transaminases, which transaminate different amino acids. It seems that a similar mechanism is active in *P. pastoris*.

Conclusion

It was successfully demonstrated that engineering of the amino acid metabolism in *P. pastoris* increased the production of a recombinant carboxylesterase. Most strikingly, increased productivity for the recombinant enzyme was directly connected to increased intracellular amino acid levels. In general, the results of this study indicate that recombinant protein production in *P. pastoris* is limited by precursor availability.

References

Benjamini Y, Heller R & Yekutieli D (2009) Selective inference in complex research. *Philos Trans A Math Phys Eng Sci* **367**: 4255–4271.

Berlec A & Strukelj B (2013) Current state and recent advances in biopharmaceutical production in Escherichia coli, yeasts and mammalian cells. *J Ind Microbiol Biotechnol* **40**: 257–274.

Carnicer M, Ten Pierick A, van Dam J, Heijnen JJ, Albiol J, van Gulik W & Ferrer P (2012) Quantitative metabolomics analysis of amino acid metabolism in recombinant Pichia pastoris under different oxygen availability conditions. *Microb Cell Fact* **11**: 83.

Dragosits M, Stadlmann J, Albiol J, *et al.* (2009) The Effect of Temperature on the Proteome of Recombinant Pichia pastoris. *J. Proteome Res.* 1380–1392.

Gasser B, Maurer M, Gach J, Kunert R & Mattanovich D (2006) Engineering ofPichia pastoris for improved production of antibody fragments. *Biotechnol Bioeng* **94**: 353–361.

Gasser B, Prielhofer R, Marx H, Maurer M, Nocon J, Steiger M, Puxbaum V, Sauer M & Mattanovich D (2013) Pichia pastoris: protein production host and model organism for biomedical research. *Future Microbiol.* **8(2)** : 191–208.

Glick BR (1995) Metabolic load and heterologous gene expression. *Biotechnol Adv* **13**: 247–261.

Görgens JF, van Zyl WH, Knoetze JH & Hahn-Hägerdal B (2005) Amino acid supplementation improves heterologous protein production by Saccharomyces cerevisiae in defined medium. *Appl Microbiol Biotechnol* **67**: 684–691.

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

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**: 357–368.

Heyland J, Fu J, Blank LM & Schmid A (2011) Carbon metabolism limits recombinant protein production in Pichia pastoris. *Biotechnol Bioeng* **108**: 1942–1953.

Hinnebusch AG (1988) Mechanisms of Gene Regulation in the General Control of Amino Acid Biosynthesis in Saccharomyces cerevisiae. *Microbiol Rev.* **52**: 248–273.

Hinnebusch AG & Natarajan K (2002) Gcn4p , a Master Regulator of Gene Expression , Is Controlled at Multiple Levels by Diverse Signals of Starvation and Stress. *Eukaryot Cell.* **1**: 22–32.

Jordà J, Jouhten P, Cámara E, Maaheimo H, Albiol J & Ferrer P (2012) Metabolic flux profiling of recombinant protein secreting Pichia pastoris growing on glucose:methanol mixtures. *Microb Cell Fact* **11**: 57.

Kazemi Seresht A, Cruz AL, de Hulster E, Hebly M, Palmqvist EA, van Gulik W, Daran J-M, Pronk J & Olsson L (2013) Long-term adaptation of Saccharomyces cerevisiae to the burden of recombinant insulin production. *Biotechnol Bioeng* **110**: 2749–2763.

Luttik MAH, Vuralhan Z, Suir E, Braus GH, Pronk JT & Daran JM (2008) Alleviation of feedback inhibition in Saccharomyces cerevisiae aromatic amino acid biosynthesis: Quantification of metabolic impact. *Metab Eng* **10**: 141–153.

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

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

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

Patil KR, Akesson M & Nielsen J (2004) Use of genome-scale microbial models for metabolic engineering. *Curr Opin Biotechnol* **15**: 64–69.

Peñalosa-Ruiz G, Aranda C, Ongay-Larios L, Colon M, Quezada H & Gonzalez A (2012) Paralogous ALT1 and ALT2 retention and diversification have generated catalytically active and inactive aminotransferases in Saccharomyces cerevisiae. *PLoS One* **7**: e45702.

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.

Rahmen N, Fulton A, Ihling N, Magni M, Jaeger K-E & Büchs J (2015) Exchange of single amino acids at different positions of a recombinant protein affects metabolic burden in Escherichia coli. *Microb Cell Fact* **14**: 1–18.

Raiford DW, Heizer EM, Miller R V., Akashi H, Raymer ML & Krane DE (2008) Do amino acid biosynthetic costs constrain protein evolution in Saccharomyces cerevisiae? *J Mol Evol* **67**: 621–630.

Rumjantsev a. M, Padkina M V. & Sambuk E V. (2013) Effect of nitrogen source on gene expression of first steps of methanol utilization pathway in Pichia pastoris. *Russ J Genet* **49**: 394–400.

Ruth C, Buchetics M, Vidimce V, Kotz D, Naschberger S, Mattanovich D, Pichler H & Gasser B (2014) Pichia pastoris Aft1 - a novel transcription factor, enhancing recombinant protein secretion. *Microb Cell Fact* **13**: 1–15.

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**: 705–715.

Stadlmayr G, Mecklenbräuker A, Rothmüller M, Maurer M, Sauer M, Mattanovich D & Gasser B (2010) Identification and characterisation of novel Pichia pastoris promoters for heterologous protein production. *J Biotechnol* **150**: 519–529.

Yu S-L, An YJ, Yang H-J, *et al.* (2013) Alanine-metabolizing enzyme Alt1 is critical in determining yeast life span, as revealed by combined metabolomic and genetic studies. *J Proteome Res* **12**: 1619–1627.

Conclusion

The production of a secreted recombinant protein is a complex process involving different cellular pathways. Up to now, engineering strategies for improving recombinant protein production in yeasts focused on specific cellular processes (e.g. transcription, translation, folding and secretion, metabolism). Nevertheless, the applied strategies improve production only for a single recombinant protein or a smaller group of recombinant proteins. Recent developments in high throughput technologies and bioinformatics now allow a simultaneous in depth analysis of the different cellular components under production conditions. High throughput technologies already proved to provide useful data, which later can be used for the design of engineering strategies. The major challenge nowadays is to decipher the correlation between different omics data sets, which is essential to understand cellular processes on a systems level. Some attempts of integrated analysis of multi omics data have been made and showed that this approach is a powerful tool for understanding regulatory patterns of complex cellular processes. Nevertheless, analysis of cellular processes using multi omics approaches are in an early stage and improvements in high throughput technologies (sensitivity and specificity) and statistical analysis methods are necessary to increase the predictive power of such approaches.

The aim of this PhD thesis was to elucidate the impact of primary metabolism on recombinant protein production in *P. pastoris* using a multi omics approach. Metabolomics and the intracellular flux distribution of *P. pastoris* cells during recombinant protein production were used to find potential bottlenecks within metabolism, which may limit recombinant protein production. This project gave novel insights to the connection of metabolism and recombinant protein production. First of all, the metabolomics analysis showed that for strains producing a HyHEL-Fab fragment almost all intracellular amino acid levels were increased compared to the respective control. Furthermore, this analysis was done on two different substrates (glucose and methanol/glycerol) and in both cases increased intracellular amino acids levels were detected. This finding may indicate that the production of the Fab fragment necessitates the increased supply with precursors (amino acids) for recombinant protein production. In addition to the comparison of intracellular amino acid levels of producer and nonproducer strains also a comparison of intracellular amino acid levels on two different substrates (methanol/glycerol vs. glucose) was done. In general, the comparison of the intracellular amino acid levels for both conditions showed a decrease in free intracellular amino acid pools for methanol/glycerol grown cells. However, for methanol/glycerol grown cells a higher amount of amino acids incorporated in proteins was found. This result clearly shows that also the used

substrate causes significant changes within cellular metabolism. Furthermore, the observed changes might directly influence productivity of the recombinant protein.

Based on the results from the metabolomic analysis of producers and non-producers two different approaches were followed to elucidate if insufficient supply with amino acids limits recombinant protein production. In the first approach amino acid supplementation experiments were conducted to analyze if additional supply with amino acids unburdens cellular metabolism for recombinant protein production. The effect of amino acid supplementation was tested for two antibody fragments and two different substrates (glucose and methanol). The addition of amino acids increased productivity for the antibody fragment if glucose was used as carbon source. However, if methanol was used as carbon source addition of the amino acid serine negatively affected the production of the antibody fragment. The obtained results clearly showed that amino acid supplementation is on the one hand beneficial for recombinant protein production, but is not generally applicable for improving recombinant protein production. Furthermore, it is also likely that not the direct incorporation of supplemented amino acids into the recombinant protein increases productivity, as the improvement in productivity somehow correlated with the amount of amino acids added. Therefore, the use as additional nitrogen source or for other metabolic processes is possible. With regard to industrial processes, the addition of amino acids is problematic (downstream processes, costs). Therefore, the second approach dealt with metabolic engineering of the amino acid metabolism in *P. pastoris* to overcome possible synthesis limitations in specific amino acid biosynthetic pathways. The obtained metabolomics data and the amino acid composition of the model proteins were used for the definition of metabolic engineering targets connected to specific amino acid biosynthesis pathways. The overexpression of genes connected to amino acid metabolism improved the production of a carboxylesterase. On the contrary, for the second model protein (the antibody fragment, HyHEL-Fab cultivated on methanol as carbon source) no improvement in production was observed. A reason for the observed picture could be the increased complexity of the model protein. The production of an antibody fragment necessitates correct folding and assembly of the light and heavy chain, which may overburdens the folding and secretion machinery. Therefore, the major bottleneck for the production of the antibody fragment is constituted by limits in folding and secretion of the antibody fragment and possibly not by metabolism.
In summary, the results of this thesis show that recombinant protein production is limited by metabolism. To avoid this metabolic limitation, amino acid supplementation and metabolic engineering of amino acid metabolism turned out to be valuable strategies. However, the improvement for recombinant protein production was highly dependent on the used carbon source

and expressed recombinant protein. In manuscript 2, we showed that the used carbon source has a significant impact on amino acid metabolism, which could influence the effect of the chosen strategy (amino acid supplementation or engineering of amino acid metabolism) on recombinant protein production. To evaluate the effect of limitation of metabolism on recombinant protein production of more complex proteins (antibody fragments), a first engineering of bottlenecks within protein folding, and secretion is necessary. A subsequent engineering of amino acid or energy metabolism could further increase recombinant protein titers.

References

Bailey J (1991) Toward a science of metabolic engineering. *Science (80-)* **252**: 1668–1675.

Baumann K, Carnicer M, Dragosits M, *et al.* (2010) A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions. *BMC Syst Biol* **4**: 141.

Baumann K, Dato L, Graf AB, Frascotti G, Dragosits M, Porro D, Mattanovich D, Ferrer P & Branduardi P (2011) The impact of oxygen on the transcriptome of recombinant *S. cerevisiae* and *P. pastoris* - a comparative analysis. *BMC Genomics* **12**: 218.

Berlec A & Strukelj B (2013) Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells. *J Ind Microbiol Biotechnol* **40**: 257–274.

Blank LM (2004) TCA cycle activity in *Saccharomyces cerevisiae* is a function of the environmentally determined specific growth and glucose uptake rates. *Microbiology* **150**: 1085–1093.

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

Boles E, Jong-gubbels P De & Pronk JT (1998) Identification and Characterization of MAE1 , the *Saccharomyces cerevisiae* Structural Gene Encoding Mitochondrial Malic Enzyme Identification and Characterization of MAE1 , the Saccharomyces cerevisiae Structural Gene Encoding Mitochondrial Malic Enzyme. **180**.

Bolten CJ, Kiefer P, Letisse F, Portais J-C & Wittmann C (2007) Sampling for metabolome analysis of microorganisms. *Anal Chem* **79**: 3843–3849.

Bonander N, Hedfalk K, Larsson C, Mostad P, Chang C, Gustafsson L & Bill RM (2005) Design of improved membrane protein production experiments : Quantitation of the host response. Protein Sci. **14**:1729–1740.

Bonander N, Darby RA, Grgic L, Bora N, Wen J, Brogna S, Poyner DR, O'Neill MA & Bill RM (2009) Altering the ribosomal subunit ratio in yeast maximizes recombinant protein yield. *Microb Cell Fact* **8**: 10.

Bordbar A, Monk JM, King Z a & Palsson BO (2014) Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet* **15**: 107–120.

Burgard AP, Pharkya P & Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* **84**: 647–657.

Carnicer M, Canelas AB, Ten Pierick A, Zeng Z, van Dam J, Albiol J, Ferrer P, Heijnen JJ & van Gulik W (2012a) Development of quantitative metabolomics for *Pichia pastoris*. *Metabolomics* **8**: 284–298.

Carnicer M, Ten Pierick A, van Dam J, Heijnen JJ, Albiol J, van Gulik W & Ferrer P (2012b) Quantitative metabolomics analysis of amino acid metabolism in recombinant *Pichia pastoris* under different oxygen availability conditions. *Microb Cell Fact* **11**: 83.

Celton M, Sanchez I, Goelzer A, Fromion V, Camarasa C & Dequin S (2012) A comparative transcriptomic, fluxomic and metabolomic analysis of the response of *Saccharomyces cerevisiae* to increases in NADPH oxidation. *BMC Genomics* **13**: 317.

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.

Daly R & Hearn MTW (2005) Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J Mol Recognit* **18**: 119–138.

Delic M, Valli M, Graf AB, Pfeffer M, Mattanovich D & Gasser B (2013) The secretory pathway: exploring yeast diversity. *FEMS Microbiol Rev* 1–43.

Delic M, Göngrich R, Mattanovich D & Gasser B (2014) Engineering of protein folding and secretion-strategies to overcome bottlenecks for efficient production of recombinant proteins. *Antioxid Redox Signal* **21**: 414–437.

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**: 1380–1392.

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

Glick BR (1995) Metabolic load and heterologous gene expression. *Biotechnol Adv* **13**: 247–261.

Gombert a K & Nielsen J (2000) Mathematical modelling of metabolism. *Curr Opin Biotechnol* **11**: 180–186.

Gonzalez R, Andrews BA, Molitor J & Asenjo JA (2002) Metabolic analysis of the synthesis of high levels of intracellular human SOD in *Saccharomyces cerevisiae* rhSOD 2060 411 SGA122. *Biotechnol Bioeng* **82**: 152–169.

Gorgens JF, van Zyl WH, Knoetze JH & Hahn-Hägerdal B (2005) Amino acid supplementation improves heterologous protein production by *Saccharomyces cerevisiae* in defined medium. *Appl Microbiol Biotechnol* **67**: 684–691.

Grabherr MG, Haas BJ, Yassour M, *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**: 644–652.

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

Van der Heide M, Hollenberg CP, van der Klei IJ & Veenhuis M (2002) Overproduction of BiP negatively affects the secretion of Aspergillus niger glucose oxidase by the yeast Hansenula polymorpha. *Appl Microbiol Biotechnol* **58**: 487–494.

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**: 357–368.

Heyland J, Fu J, Blank LM & Schmid A (2011) Carbon metabolism limits recombinant protein production in *Pichia pastoris*. *Biotechnol Bioeng* **108**: 1942–1953.

Hinnebusch AG & Natarajan K (2002) Gcn4p , a Master Regulator of Gene Expression , Is Controlled at Multiple Levels by Diverse Signals of Starvation and Stress. **1**: 22–32.

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**: 367–375.

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

Idiris A, Tohda H, Kumagai H & Takegawa K (2010) Engineering of protein secretion in yeast: strategies and impact on protein production. *Appl Microbiol Biotechnol* **86**: 403–417.

Jahic M, Veide A, Charoenrat T, Teeri T & Enfors S-O (2006) Process technology for production and recovery of heterologous proteins with *Pichia pastoris*. *Biotechnol Prog* **22**: 1465–1473.

Jordà J, Jouhten P, Cámara E, Maaheimo H, Albiol J & Ferrer P (2012) Metabolic flux profiling of recombinant protein secreting *Pichia pastoris* growing on glucose:methanol mixtures. *Microb Cell Fact* **11**: 57.

Jordà J, Suarez C & Carnicer M (2013) Glucose-methanol co-utilization in *Pichia pastoris* studied by metabolomics and instationary 13C flux analysis. *BMC Syst Biol* ***7:17***

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

Kim I-K, Roldão A, Siewers V & Nielsen J (2012) A systems-level approach for metabolic engineering of yeast cell factories. *FEMS Yeast Res* **12**: 228–248.

Klein T, Lange S, Wilhelm N, Bureik M, Yang T-H, Heinzle E & Schneider K (2014) Overcoming the metabolic burden of protein secretion in *Schizosaccharomyces pombe*--a quantitative approach using 13C-based metabolic flux analysis. *Metab Eng* **21**: 34–45.

De Koning W & van Dam K (1992) A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. *Anal Biochem* **204**: 118–123.

Lee J, Jung M-Y, Oh M-K & Ng CY (2012) Production of 2,3-butanediol in *Saccharomyces cerevisiae* by in silico aided metabolic engineering. *Microb Cell Fact* **11**: 68.

Liang S, Wang B, Pan L, Ye Y, He M, Han S, Zheng S, Wang X & Lin Y (2012) Comprehensive structural annotation of *Pichia pastoris* transcriptome and the response to various carbon sources using deep paired-end RNA sequencing. *BMC Genomics* **13**: 738.

Mashego MR, Wu L, Van Dam JC, Ras C, Vinke JL, Van Winden W a, Van Gulik WM & Heijnen JJ (2004) MIRACLE: mass isotopomer ratio analysis of U-13C-labeled extracts. A new method for accurate quantification of changes in concentrations of intracellular metabolites. *Biotechnol Bioeng* **85**: 620–628.

Mashego MR, Jansen ML a, Vinke JL, van Gulik WM & Heijnen JJ (2005) Changes in the metabolome of *Saccharomyces cerevisiae* associated with evolution in aerobic glucose-limited chemostats. *FEMS Yeast Res* **5**: 419–430.

Mashego MR, Rumbold K, De Mey M, Vandamme E, Soetaert W & Heijnen JJ (2007) Microbial metabolomics: past, present and future methodologies. *Biotechnol Lett* **29**: 1–16.

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

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

Neubauer S, Haberhauer-Troyer C, Klavins K, Russmayer H, Steiger MG, Gasser B, Sauer M, Mattanovich D, Hann S & Koellensperger G (2012) U13C cell extract of *Pichia pastoris*--a powerful tool for evaluation of sample preparation in metabolomics. *J Sep Sci* **35**: 3091–3105.

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

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

Nöh K & Wiechert W (2006) Experimental design principles for isotopically instationary 13C labeling experiments. *Biotechnol Bioeng* **94**: 234–251.

Nookaew I, Papini M, Pornputtapong N, Scalcinati G, Fagerberg L, Uhlén M & Nielsen J (2012) A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in Saccharomyces cerevisiae. *Nucleic Acids Res* **40**: 10084–10097.

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

Otto A, Becher D & Schmidt F (2014) Quantitative proteomics in the field of microbiology. *Proteomics* **14**: 547–565.

Piškur J & Compagno C (2014) *Molecular Mechanisms in Yeast Carbon Metabolism*. Piškur J & Compagno C, eds. Springer Verlag, ISBN: 3642550134

Porro D, Gasser B, Fossati T, Maurer M, Branduardi P, Sauer M & Mattanovich D (2011) Production of recombinant proteins and metabolites in yeasts. *Appl Microbiol Biotechnol* **89**: 939–948.

Quek L-E, Wittmann C, Nielsen LK & Krömer JO (2009) OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis. *Microb Cell Fact* **8**: 25.

Van Rensburg E, den Haan R, Smith J, van Zyl WH & Görgens JF (2012) The metabolic burden of cellulase expression by recombinant *Saccharomyces cerevisiae* Y294 in aerobic batch culture. *Appl Microbiol Biotechnol* **96**: 197–209.

Rizzi M, Baltes M, Theobald U & Reuss M (1997) In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. Mathematical model. *Biotechnol Bioeng* **55**: 592–608.

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

Sauer U (2006) Metabolic networks in motion: 13C-based flux analysis. *Mol Syst Biol* **2**: 62. Sechi S & Oda Y (2003) Quantitative proteomics using mass spectrometry. *Curr Opin Chem Biol* **7**: 70–77.

Simeonidis E, Murabito E, Smallbone K & Westerhoff H V (2010) Why does yeast ferment? A flux balance analysis study. *Biochem Soc Trans* **38**: 1225–1229.

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

Stelling J (2004) Mathematical models in microbial systems biology. *Curr Opin Microbiol* **7**: 513–518. Stephanopoulos G (1999) Metabolic fluxes and metabolic engineering. *Metab Eng* **1**: 1–11.

Tredwell GD, Edwards-Jones B, Leak DJ & Bundy JG (2011) The development of metabolomic sampling procedures for *Pichia pastoris*, and baseline metabolome data. *PLoS One* **6**: e16286.

Valkonen M, Penttilä M, Saloheimo M & Penttila M (2003) Effects of Inactivation and Constitutive Expression of the Unfolded- Protein Response Pathway on Protein Production in the Yeast *Saccharomyces cerevisiae*. *Mol Microbiol* **69**: 2065–2072.

Vanz AL, Lünsdorf H, Adnan A, Nimtz M, Gurramkonda C, Khanna N & Rinas U (2012) 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* **11**: 103.

Varma A & Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol* **60**: 3724–3731.

Varner J & Ramkrishna D (1999) Mathematical models of metabolic pathways. *Curr Opin Biotechnol* **10**: 146–150.

Villas-Bôas SG, Højer-Pedersen J, Akesson M, Smedsgaard J & Nielsen J (2005) Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast* **22**: 1155–1169.

Van Vliet AHM (2010) Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS Microbiol Lett* **302**: 1–7.

Walsh G (2010) Biopharmaceutical benchmarks 2010. *Nat Biotechnol* **28**: 917–924.

Wang Z, Gerstein M & Snyder M (2010) RNA-Seq : a revolutionary tool for transcriptomics. Nat. Rev. Genet. **10**: 57–63.

Waterham HR, Digan ME, Koutz PJ, Lair S V & Cregg JM (1997) Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene* **186**: 37–44.

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

Wiechert W & Nöh K (2013) Isotopically non-stationary metabolic flux analysis: complex yet highly informative. *Curr Opin Biotechnol* **24**: 979–986.

Zamboni N, Fischer E & Sauer U (2005) FiatFlux--a software for metabolic flux analysis from 13C-glucose experiments. *BMC Bioinformatics* **6**: 209.

Appendix

A) Metabolomics analysis of *P. pastoris* strains during recombinant protein production

A key issue of this PhD work was the analysis of the metabolome of *P. pastoris* during recombinant protein production. The impact of three different model proteins on the intracellular metabolite levels of *P. pastoris* was quantified. The impact of two recombinant produced enzymes, a carboxylesterase and a carboxypeptidase was tested. The third model protein was a more complex antibody fragment (HyHEL-Fab). Furthermore, for the carboxypeptidase and the antibody fragment (HyHEL-Fab) the effect of recombinant protein production on metabolism using two different carbon sources was tested.

In order to detected significant changes in intracellular metabolite levels between production and control strain p-values and fold changes were calculated. For the calculation of the p-values the student´s test was used. The criteria for significant higher abundance of metabolites was FC > 1.20 and p-value < 0.05 and for a lower abundance of metabolites was FC < -1.20 and p-value < 0.05. For each production and control strain three biological replicates were used for calculation of fold changes and p-values.

Effect of recombinant protein production controlled by the strong constitutive GAP promotor on free intracellular amino acid pools (glucose as carbon source)

The analysis of the metabolomics data for the two recombinant enzymes revealed almost no changes in measured intracellular metabolite levels. For the carboxypeptidase increased levels of phosphoenolpyruvate were detected (Appendix Fig.13-14). The impact of HyHEL-Fab production on intracellular metabolite levels was higher (with glucose as carbon source). Out of 30 measured intracellular metabolite pools 17 were compared to the respective control. Measurement of 16 intracellular amino acid pools showed that nine amino acid pools were significantly increased, and three significantly decreased. It has to be mentioned that this strain has severe growth problems and had a reduced viability during chemostat cultivation (~75%) (Appendix Fig.15-16). Furthermore, this strain was later identified to have a peculiar binuclear phenotype.

Effect of recombinant protein production controlled by the strong and inducible AOX promotor on free intracellular amino acid pools (methanol as carbon sources)

The effect of carboxypeptidase expression on intracellular metabolite levels was even higher when cell were grown on a mixture of methanol/glycerol and the AOX expression system was used. Out of the 30 measured intracellular metabolites 4 were significantly different compared to the control strain (2-phosphoglycerate, alanine, asparagine and isoleucine) (Appendix Fig.17-18). For the production strain expressing the HyHEL-Fab on methanol/glycerol as carbon sources again a higher number of significant changes in intracellular metabolite pools were detected. Thirteen intracellular metabolite levels were increased, whereas only two were significantly decreased. With regard to the intracellular amino acid pools 7 out of 16 measured pools were increased and two were decreased (Appendix Fig.19-20).

For both HyHEL-Fab expression strains (cultivated on glucose or on methanol/glycerol) a de-regulation of intracellular amino acid pools was detected. Most of the intracellular amino acid pools were increased in the strains producing the recombinant Fab protein. This may indicate an increased need for amino acids due to recombinant protein production. Also, for other *P. pastoris* strain producing an antibody fragment increased free intracellular amino acid pools were observed (Carnicer et al., 2012b).

B) Visualization of metabolomic and transcriptomic data for all production strains

Visualization of changes in free intracellular metabolite levels and transcript levels for selected genes, when expression of model proteins is controlled by the strong constitutive GAP-promoter

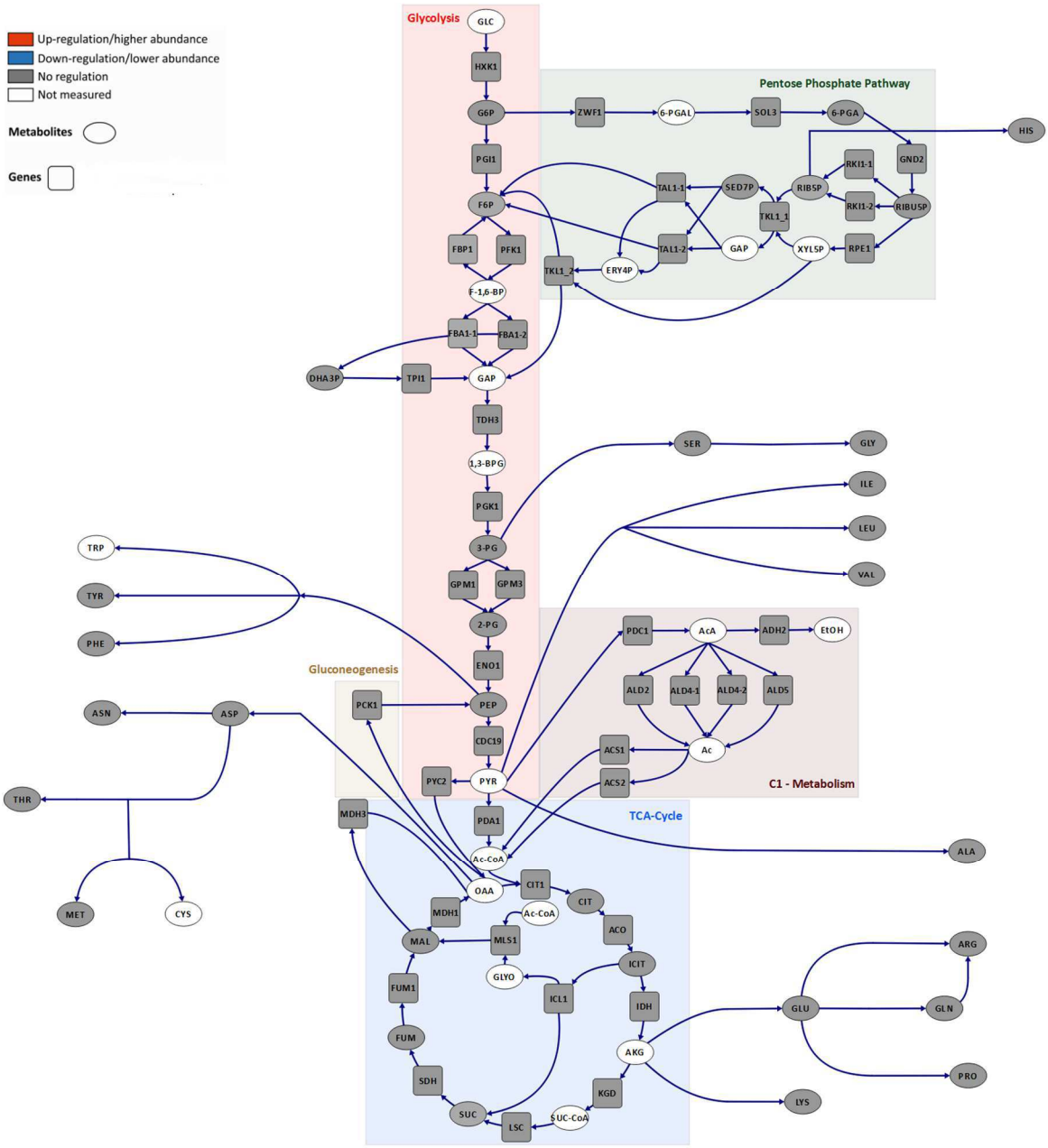

Fig.11: Visualization of significant changes in free intracellular metabolite pools and significant changes of transcript levels for genes connected to central carbon metabolism for a *P. pastoris* strain expressing a **carboxylesterase** compared to the respective control. The expression of the carboxylesterase was under the control of the **GAP-promoter**.

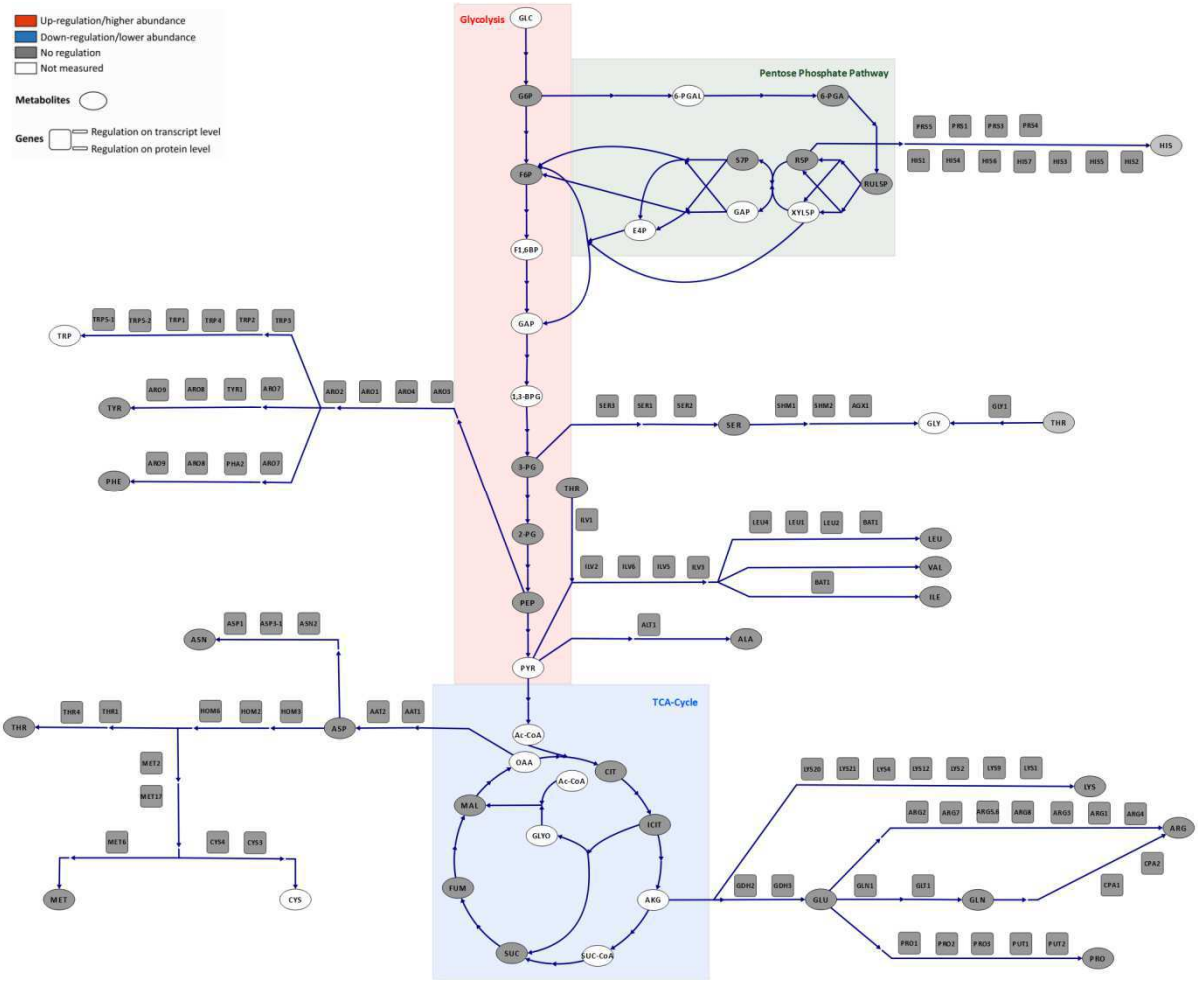

Fig.12: Visualization of significant changes in free intracellular metabolite pools and significant changes of transcript levels for genes connected to amino acid metabolism for a *P. pastoris* strain expressing a **carboxylesterase** compared to the respective control. The expression of the carboxylesterase was under the control of the **GAP-promoter**.

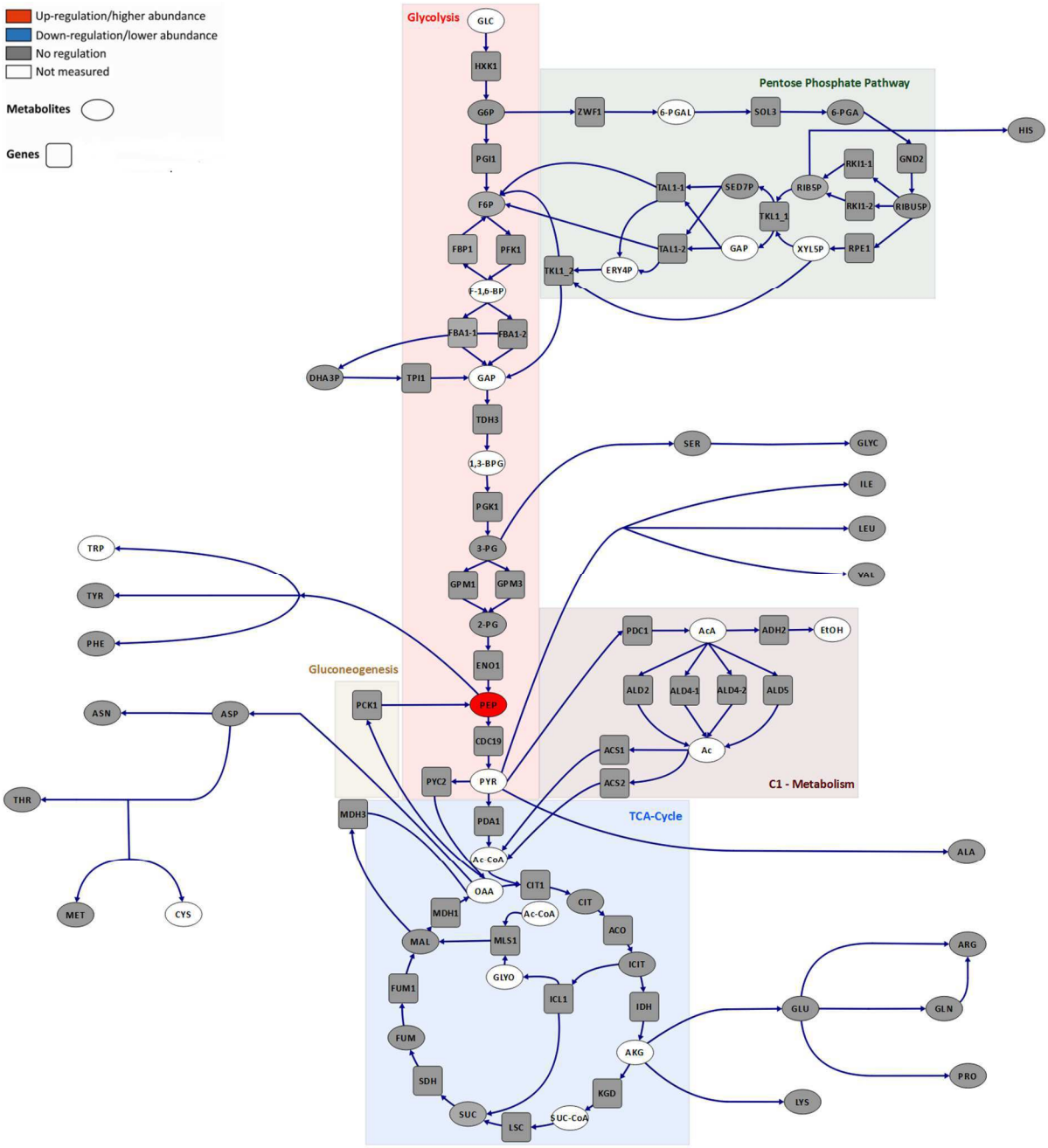

Fig.13: Visualization of significant changes in free intracellular metabolite pools and significant changes of transcript levels for genes connected to central carbon metabolism for a *P. pastoris* strain expressing a **carboxypeptitase** compared to respective control. The expression of the carboxypeptitase was under the control of the **GAP-promoter**.

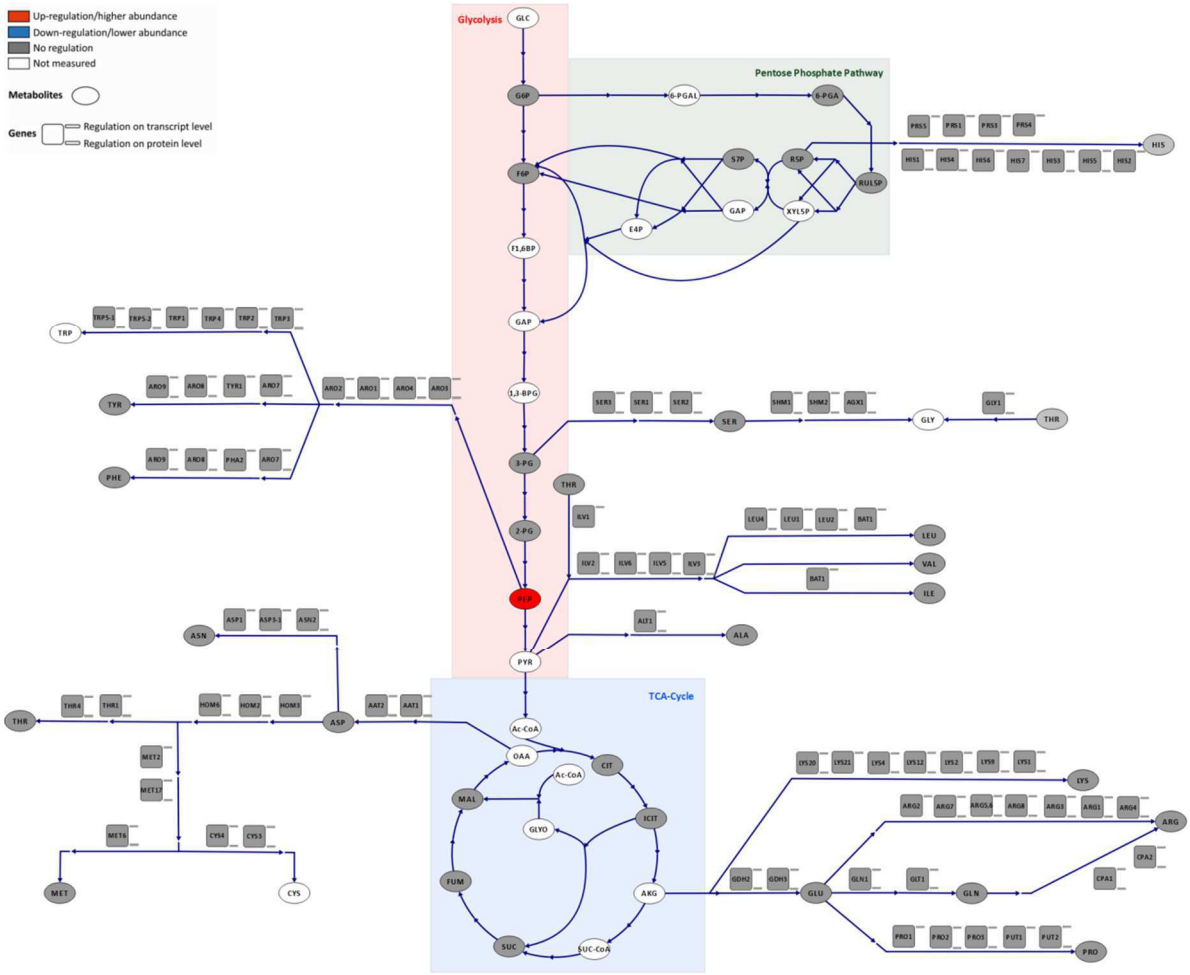

Fig.14: Visualization of significant changes in free intracellular metabolite pools and significant changes of transcript and protein levels for genes connected to amino acid metabolism for a *P. pastoris* strain expressing a **carboxypeptitase** compared to respective control. The expression of the carboxypeptitase was under the control of the **GAP-promoter**.

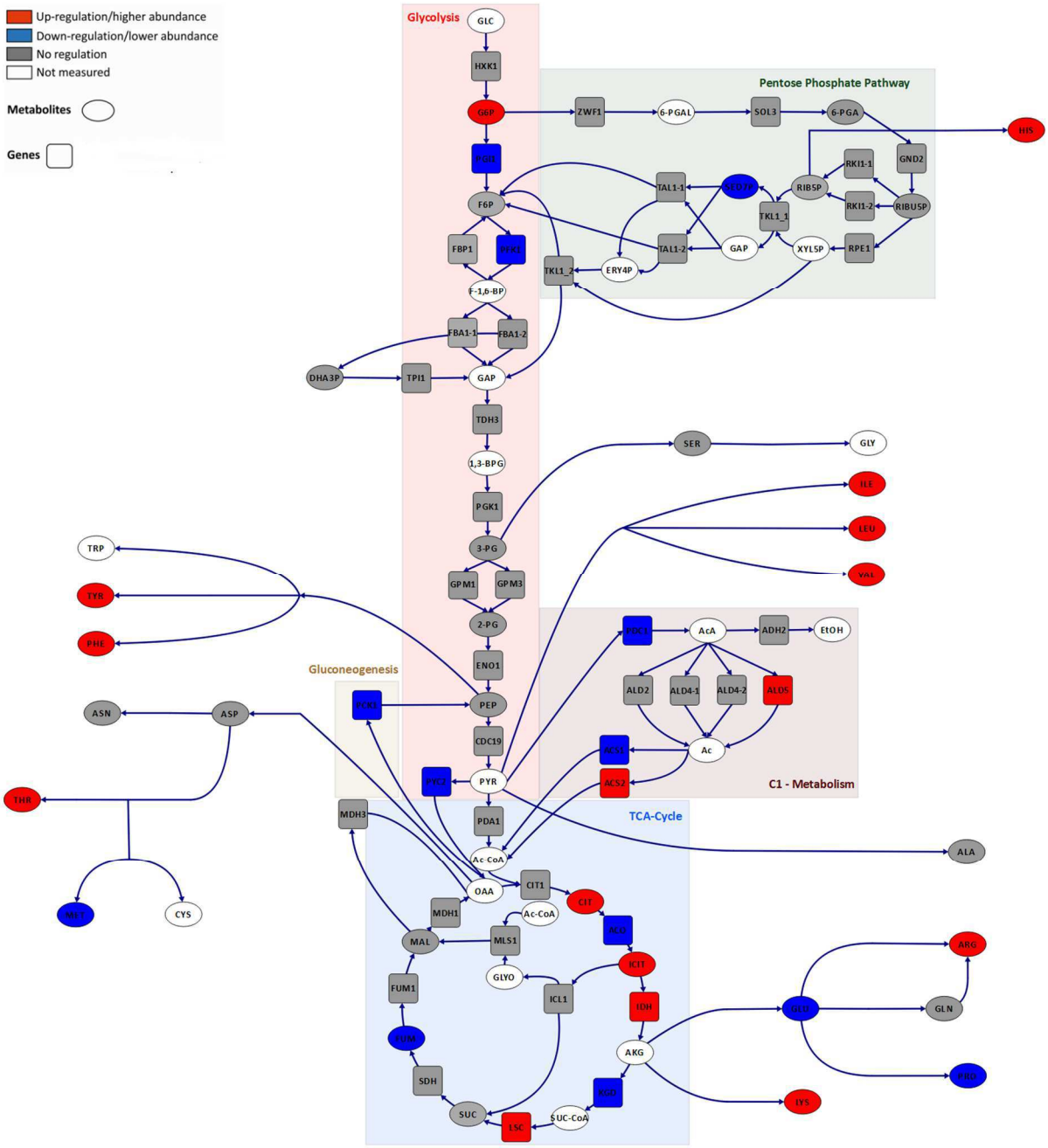

Fig.15: Visualization of significant changes in free intracellular metabolite pools and significant changes of transcript levels for genes connected to central carbon metabolism for a *P. pastoris* strain expressing a **HyHEL-Fab** compared to respective control. The expression of the HyHEL-Fab was under the control of the **GAP-promoter**.

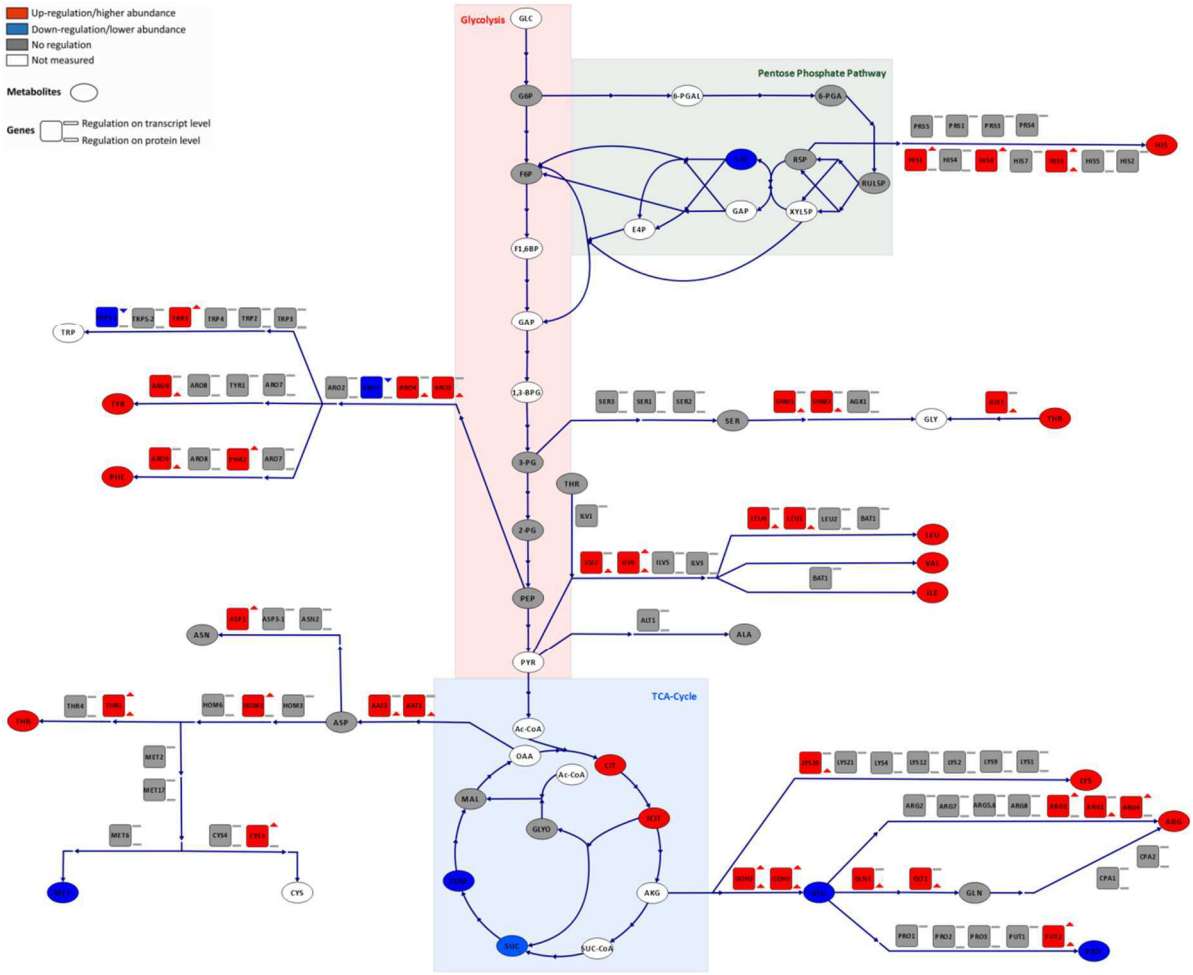

Fig.16: Visualization of significant changes in free intracellular metabolite pools and significant changes of transcript and protein levels for genes connected to central carbon metabolism for a *P. pastoris* strain expressing a **HyHEL-Fab** compared to respective control. The expression of the HyHEL-Fab was under the control of the **GAP-promoter**.

Visualization of changes in free intracellular metabolite levels and transcript levels for selected genes, when expression of model proteins is controlled by the strong inducible AOX-promoter

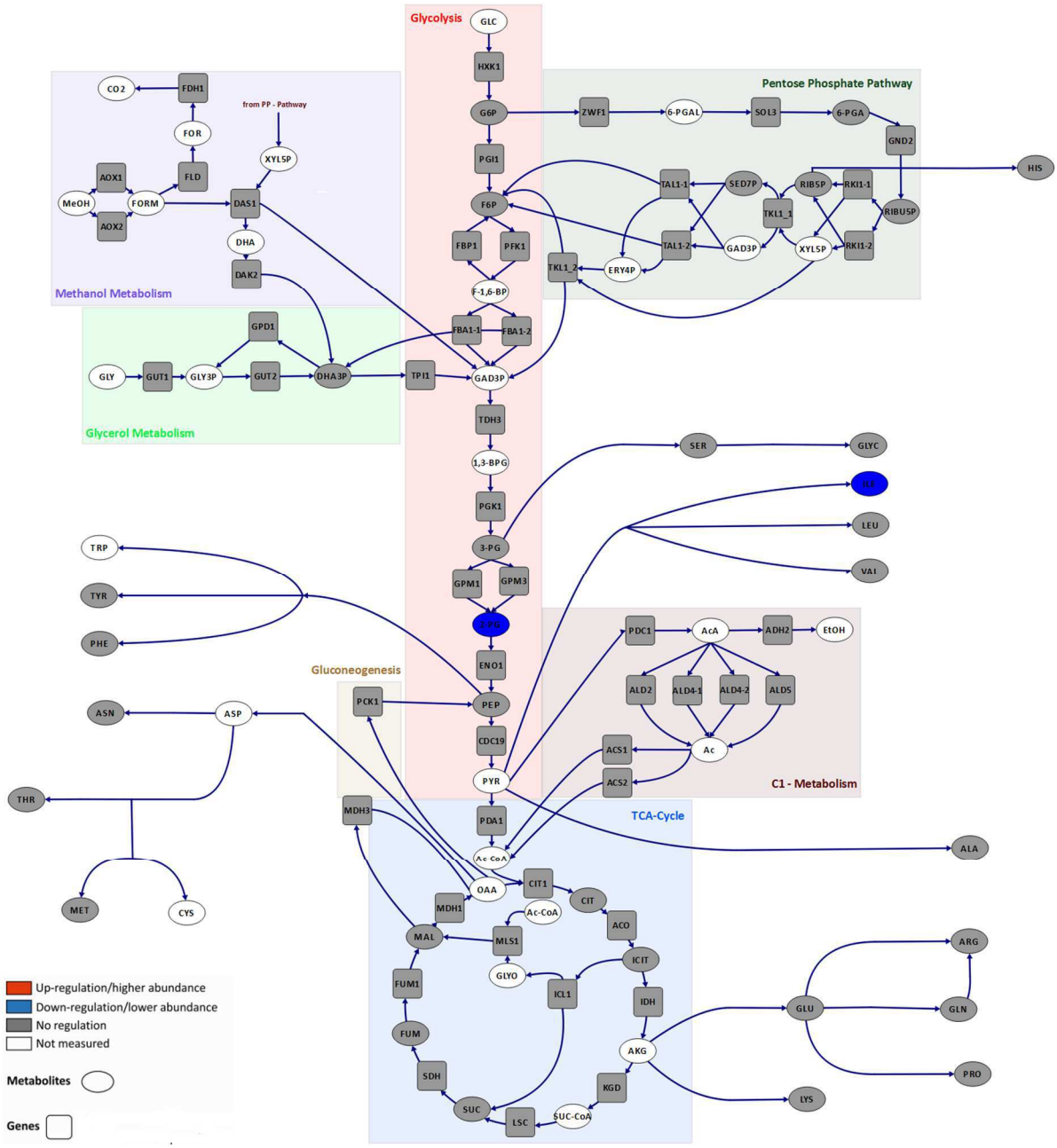

Fig.17: Visualization of significant changes in free intracellular metabolite pools and significant changes of transcript levels for genes connected to central carbon metabolism for a *P. pastoris* strain expressing a **carboxypeptitase** to the respective control. The expression of the carboxypeptitase was under the control of the **AOX-promoter**.

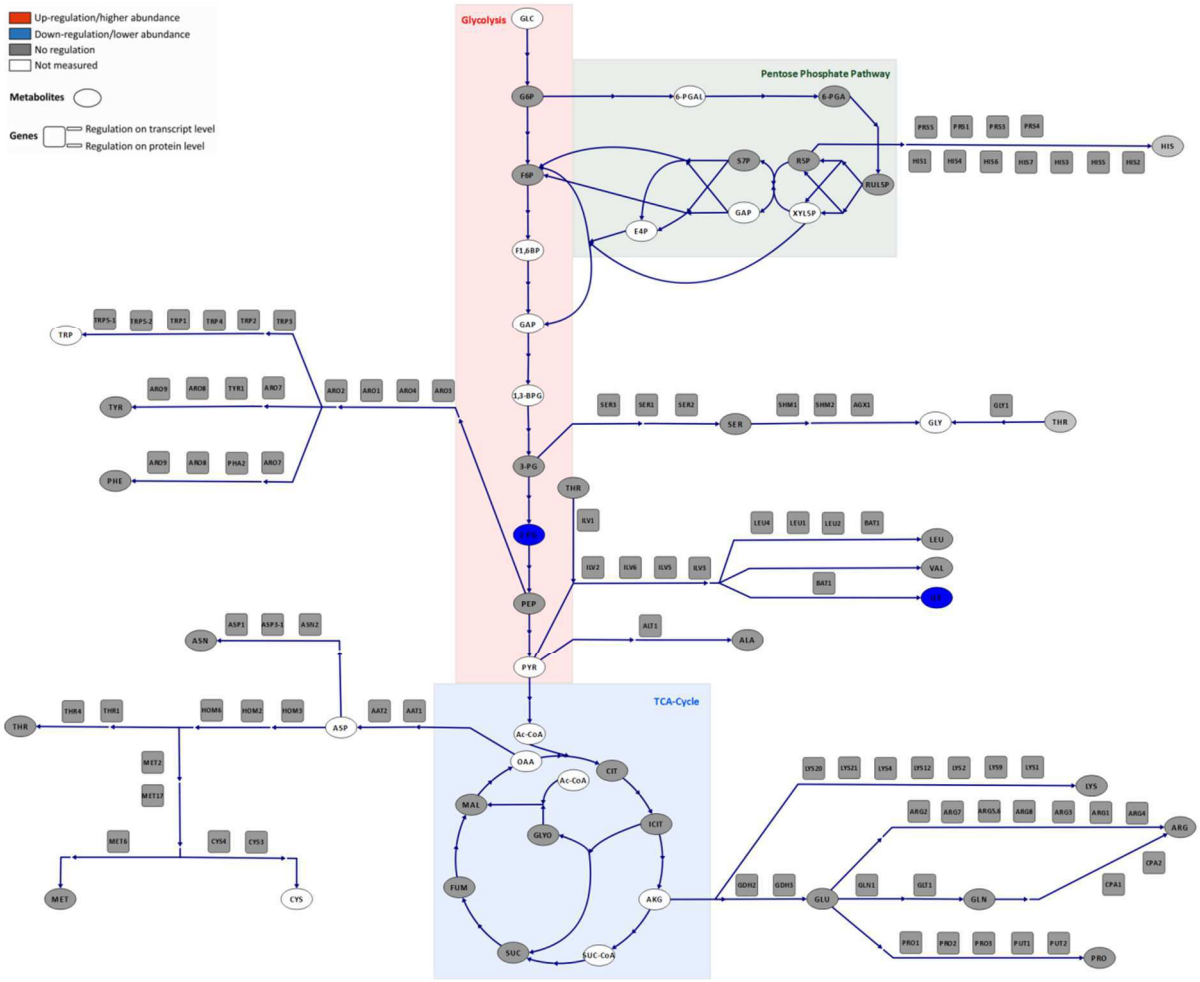

Fig.18: Visualization of significant changes in free intracellular metabolite pools and significant changes of transcript levels for genes connected to amino acid metabolism for a *P. pastoris* strain expressing a **carboxypeptitase** to the respective control. The expression of the carboxypeptitase was under the control of the **AOX-promoter**.

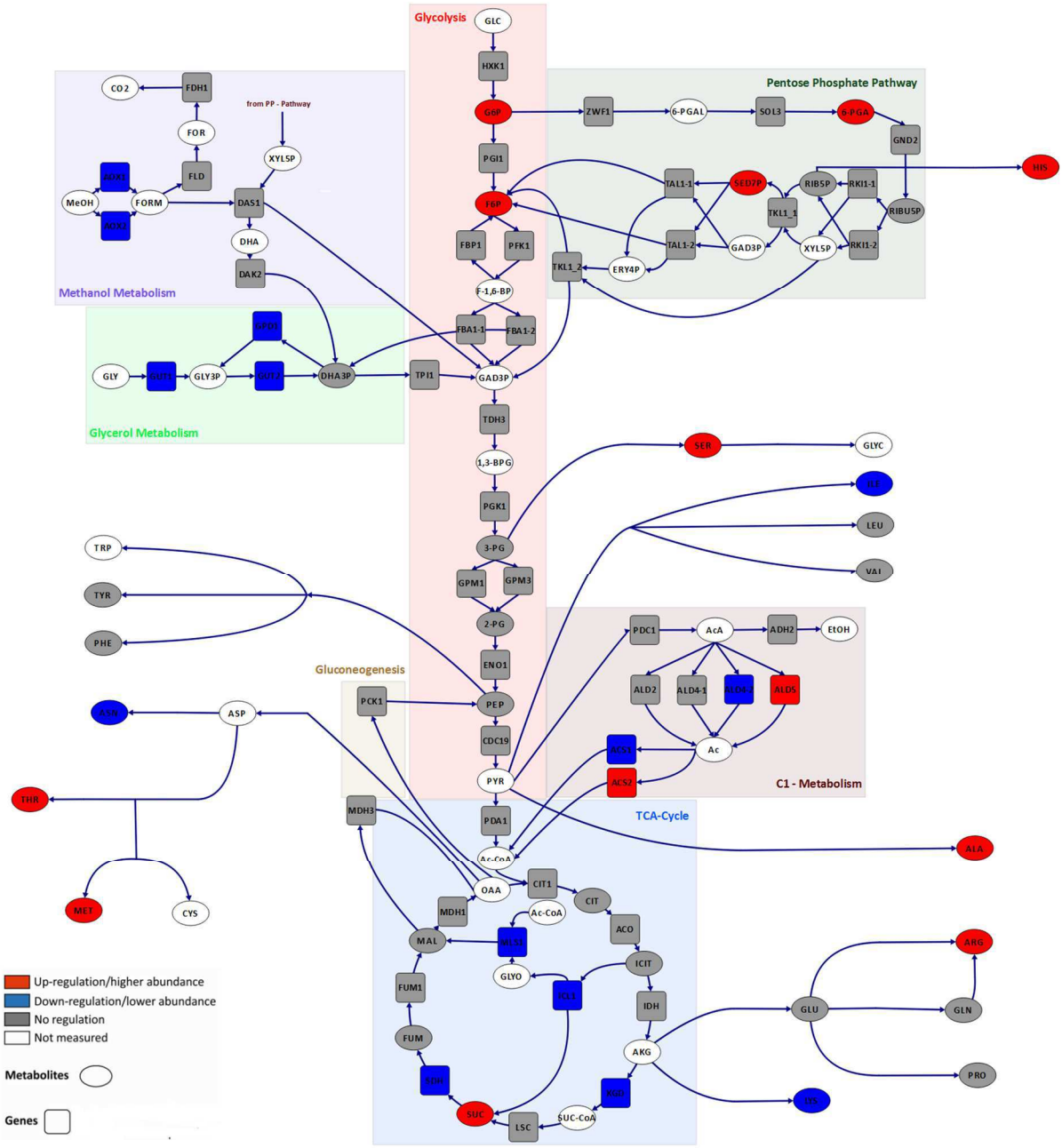

Fig.19: Visualization of significant changes in free intracellular metabolite pools and significant changes of transcript levels for genes connected to central carbon metabolism for a *P. pastoris* strain expressing a **HyHEL-Fab** to the respective. The expression of the HyHEL-Fab was under the control of the **AOX-promoter**.

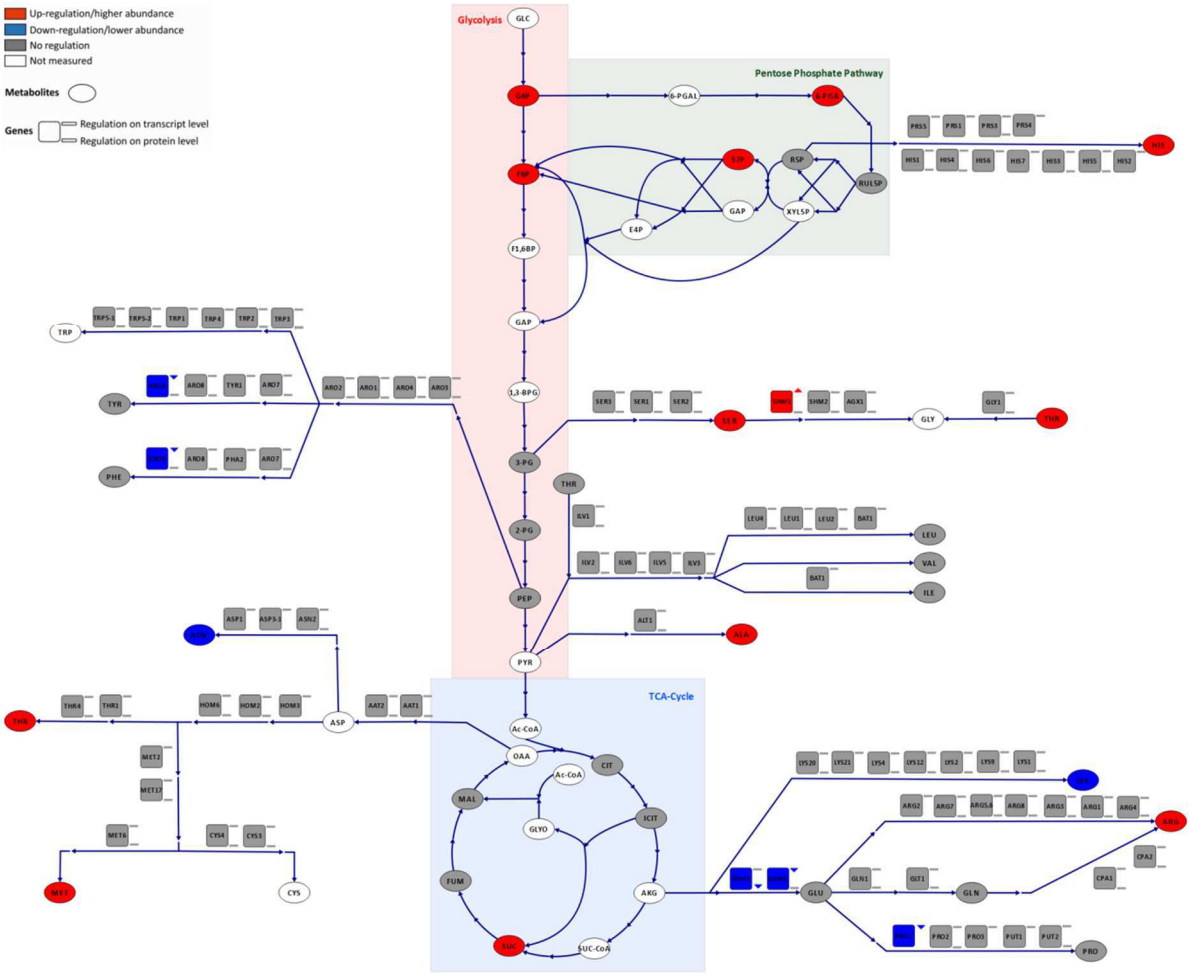

Fig.20: Visualization of significant changes in free intracellular metabolite pools and significant changes of transcript and protein levels for genes connected to amino acid metabolism for a *P. pastoris* strain expressing a **HyHEL-Fab** to the respective. The expression of the HyHEL-Fab was under the control of the **AOX-promoter**.

D) Visualization of the intracellular flux distribution for all production and control strains

Intracellular flux distribution for *P. pastoris* production strains when expression of the recombinant protein is controlled by the strong constitutive GAP-promoter (Fig. 21)

The intracellular flux distribution for all production and control strains was calculated using the software Openflux. As constraint the labelling pattern of proteinogenic amino acids were used. Furthermore, uptake and secretion rates were used for the calculation of the flux distribution. For production strains secreting the carboxylesterase or the carboxypeptitase no changes in the intracellular flux distribution were observed (Fig. 21). The expression of the HyHEL-Fab clearly had an impact on the flux distribution. For this strain an increased flux through the glycolysis and TCA-cycle was observed compared to the respective control. The increased TCA-cycle might point at an increased energy demand due to the production of the recombinant protein (Fig. 21). In case of the HyHEL-Fab a complex protein is produced, which has to be folded correctly. Folding of proteins needs high amounts of ATP. Therefore, the cell adjusts the intracellular fluxes in way to produce more ATP. Under this condition most of the ATP is produced via oxidative phosphorylation, where NADH is consumed to produce ATP. A higher flux through the TCA-cycle produces more NADH and subsequently more ATP can be produced. It has to be stated that the GAP-HyHEL-Fab#3 strain had a peculiar binucleate phenotype, which may also contribute to these changes.

Intracellular flux distribution for *P. pastoris* production strains when expression of the recombinant protein is controlled by the strong constitutive AOX-promoter (Fig. 22)

Again no changes in the intracellular flux distribution for the production strain secreting the recombinant carboxypeptidase was observed (Fig. 22). The flux data of the AOX-HyHEL-Fab showed that more substrate is channelled to the TCA-cycle and less went through the pentose phosphate pathway compared to the control. Again this could indicate a higher need of NADH due to recombinant protein production (Fig. 22). The cell adjusts the central carbon metabolism in a way that more NADH is produced by rerouting the flux towards the TCA-cycle. The folding of the HyHEL-Fab is a complex process and needs ATP. A higher flux through the TCA-cycle produces more NADH, which later is converted to ATP via oxidative phosphorylation.

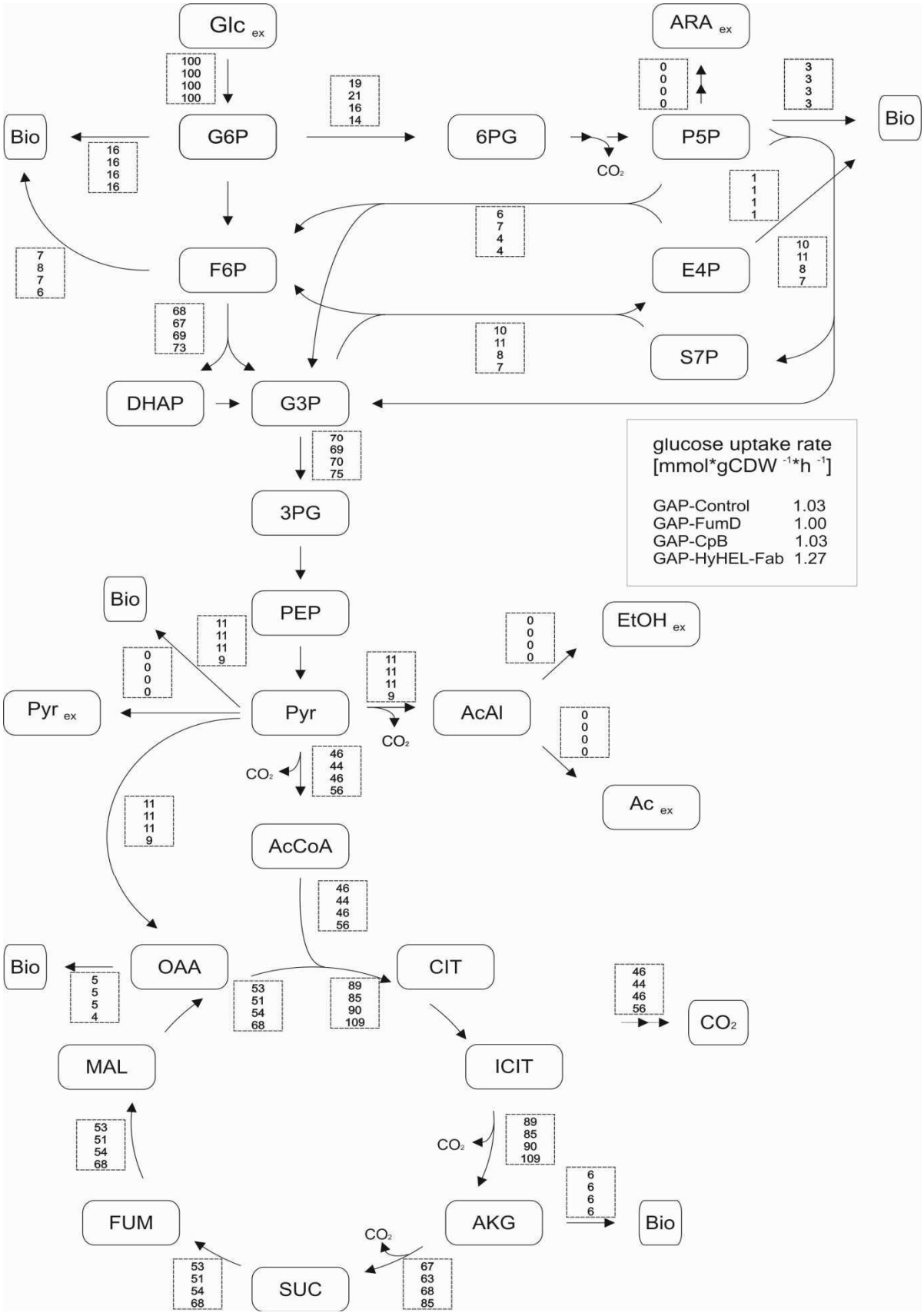

Fig.21: Comparison of intracellular flux distribution for alll GAP-strains

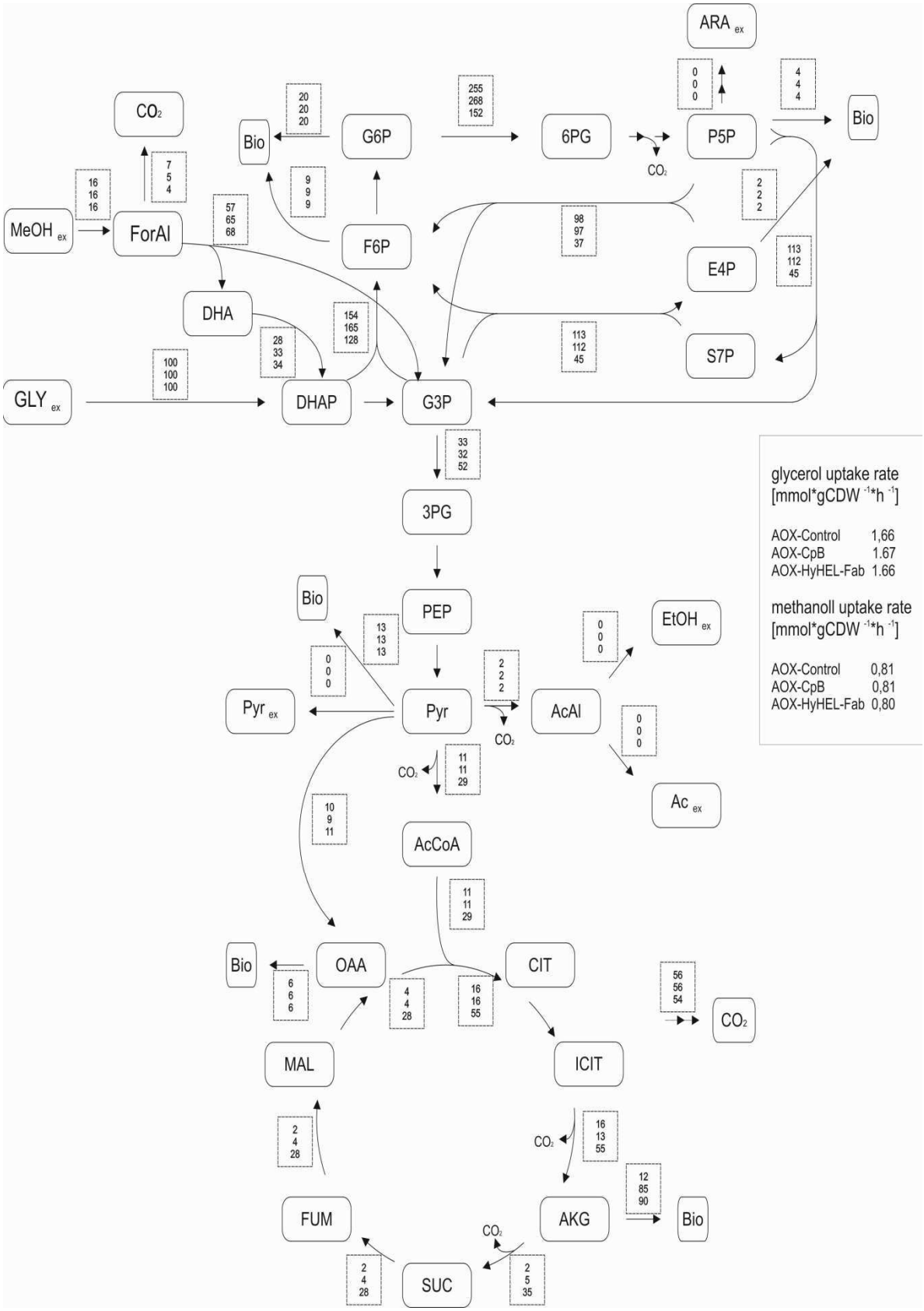

Fig.22: Comparison of the intracellular flux distribution for all AOX-strains

C) Amino Acid Supplementation Experiments

Material and Methods

Amino Acid Stock preparation:

50 mL of the respective amino acid stock was prepared. The calculated amount of amino acids was weighted in and dissolved in 50 mL of water. For the preparation of the tyrosine 1mL of 1 M NaOH was added. Stock concentration of the amino acids was always 10g/L, except for serine with 40 g/L.

Screening of pAOX-HyHEL-Fab#8 and SDZ-Fab#9

Pre-Culture:

50 mL of YPD + Zeocine was inoculated with a single colony of a freshly prepared YPD + Zeocine plate. The pre culture was cultivated at 25°C, 180 rpm for 22h.After 22h of pre-culture the cells are harvested by centrifugation (3000 g, 5 min) and washed once with 10 mL of sterile water. For inoculation of the main culture the cells are resuspended in 3 mL of sterile water. After resuspending of the cell pellet the OD600 of the cell suspension is measured.

Main Culture:

For preparation of the cultivation media 25 mL of 2 times concentrated M2 Medium without glucose was taken and amino acids stocks are added to reach the desired concentration. In a next step the pH of the media is set to a pH of 5.0 with 4 M KOH and diluted with water up to 50 mL end volume. All amino acids, except serine, valine and threonine, had a final concentration of 100 mg/L in the cultivation media. The final concentration of serine was 400 mg/L, for valine 200 mg/L and for threonine 200 mg/L. The M2 -Medium contained per liter: 22.0 g Citric acid monohydrate, 3.15 g $(NH_4)_2PO_4$, 0.49 g $MgSO_4.7H_2O$, 0.80 g KCl, 0.0268 g $CaCl_2.2H_2O$, 1.47 mL PTM1 trace metals, 4 mg Biotin; pH was set to 5 with KOH (solid).

The screening was done in 24 deep well plates. The cultivation volume was 2 mL per well. Therefore, 2 mL of the media containing the different amino acids mixtures were transferred into the wells. Each well was inoculated with an OD 4.0 using the washed pre culture cells. Calculation of the amount of inocula was done according to the equation below. For methanol induction of the cells 0.5 % methanol (10 µl of a 100 % methanol) was directly added after inoculation. For cultivation on the shaker the 24 deep well plates are closed with a sterile oxygen permeable membrane. The cells were

cultivated at 25°C, 320 rpm for 48 h. After 6h, 22h and 30h of cultivations cells were fed with 1% methanol (20µl of 100% methanol).

Inoculum [mL] = (OD [main culture]*Volume [mL main culture])/OD [washed pre culture]

After 48h samples were taken. Therefore, 1mL of culture is transferred into a 2 mL eppendorf tube. The eppendorf tube is centrifuged at 13.000 rpm for 5 min. The supernatant was transferred into a new eppendorf tube and used for determination of HyHEL-Fab and SDZ-Fab by ELISA.

For the determination of the cell dry weight the pellet was washed once with water and centrifuged (13.000 rpm, 5min). After the washing step the cell pellet are transferred into a weighted glass tube and dried for 2 days at 110°C. After drying the glass tubes are weighted again and the cell dry mass was calculated with following formula:

CDW [g/L] = [Glass tube (full) – Glass tube (empty)]*1000

For each cultivation CDW determination was done in duplicates.

Quantification of Fab with ELISA

Quantification of intact Fab by ELISA was done using anti-human IgG antibody (ab7497, Abcam) as coating antibody and a goat anti-human IgG (Fab specific) - alkaline phosphatase conjugated antibody (Sigma A8542) as detection antibody. Human Fab/Kappa, IgG fragment (Bethyl P80-115) was used as standard with a starting concentration of 100 ng/mL, supernatant samples are diluted accordingly. Detection was done with pNPP (Sigma S0942). Coating-, Dilution- and Washing buffer were based on PBS (2 mM KH_2PO_4, 10 mM $Na_2HPO_4.2\ H_2O$, 2.7 mM g KCl, 8 mM NaCl, pH 7.4) and completed with BSA (1% (w/v)) and/or Tween20 (0.1% (v/v)) accordingly.

Amino acid supplementation and its effect on recombinant protein production

The obtained metabolomics data showed that recombinant protein production influences intracellular amino acid pools. Amino acids are the building blocks for recombinant proteins and limits in synthesis capacities for certain amino acids may limit recombinant protein production. In order to understand if amino acids are the limiting factors, in a first experiment the cultivation media was supplemented with different amino acids or mixtures of amino acids. The amino acids for supplementation were chosen according to the obtained metabolomics data. The impact of amino acid supplementation was tested for two different antibody fragments (HyHEL-Fab and SDZ-Fab). Furthermore, the analysis of the different cellular layers (e.g. metabolome) described in Manuscript 2

showed that changes in carbon source directly impacts the regulation of amino acid biosynthetic pathways and furthermore intracellular amino acid pools. In order to understand if increased supply of amino acids unburdens cellular metabolism for recombinant protein production, in a first experiment amino acids were supplemented to the cultivation media.

Effect of amino acid supplementation on HyHEL-Fab and SDZ-Fab production on methanol as carbon source

To test if amino acid supplementation is generally applicable (independent of the carbon source) the impact of supplemented amino acids on HyHEL-Fab production during growth on methanol was tested. Therefore, the cultivation media was supplemented with proline and with a mixture all amino acids (results shown in manuscript 3). Both, proline and the mixture of all amino acids, were chosen because they had the highest impact on HyHEL-Fab production when screenings were done with glucose as carbon source (results shown in manuscript 3).

Addition of proline again showed to increase recombinant protein production for both model proteins (HyHEL-Fab and SDZ-Fab). The addition of all amino acids had exactly the opposite effect and caused a drop of Fab titers in the supernatant. In order to understand if this effect is specifically caused by a single amino acid or a combination of amino acids cultivations with different amino acid mixtures were conducted (Fig.23). The concentration of supplemented amino acids was chosen according to Heyland et al (2011).

For the SDZ-Fab producing strain cultivation on the different media showed that only media lacking serine, glycine and threonine had no negative effect on SDZ-Fab titers in the supernatant (Fig.23). To understand if this effect is specifically caused by one of the three amino acids or combinations of the three amino acids are responsible for the observed decrease in SDZ-Fab titers another cultivation round was done. This cultivation round clearly showed that addition of serine caused the drop in SDZ-Fab-titer (Fig. 24). The reason for this is unknown yet.
For the HYHEL-Fab expressing strain the cultivation on different media lacking specific amino acids showed that none of the tested amino acid mixtures specifically caused the observed drop in HyHEL-Fab titers (Fig.25).

Legend:

SCM: Standard Screening Media

Control: Production strain cultivated on standard screening media without amino acid addition

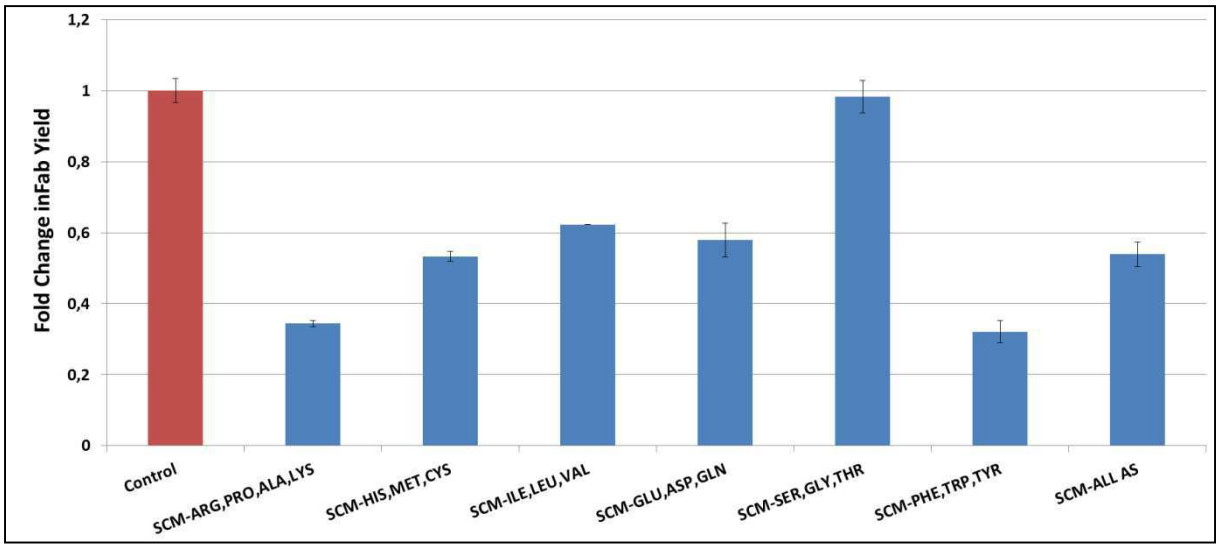

Fig.23: Results for screening of SDZ-Fab production on media lacking different amino acid mixtures. Relative changes of SDZ-Fab yield are shown. Error bars indicate the standard errors of the mean (mg SDZ-Fab/gCDW on media lacking different amino acids relative to the strain cultivated on screening media with no amino acids supplemented)

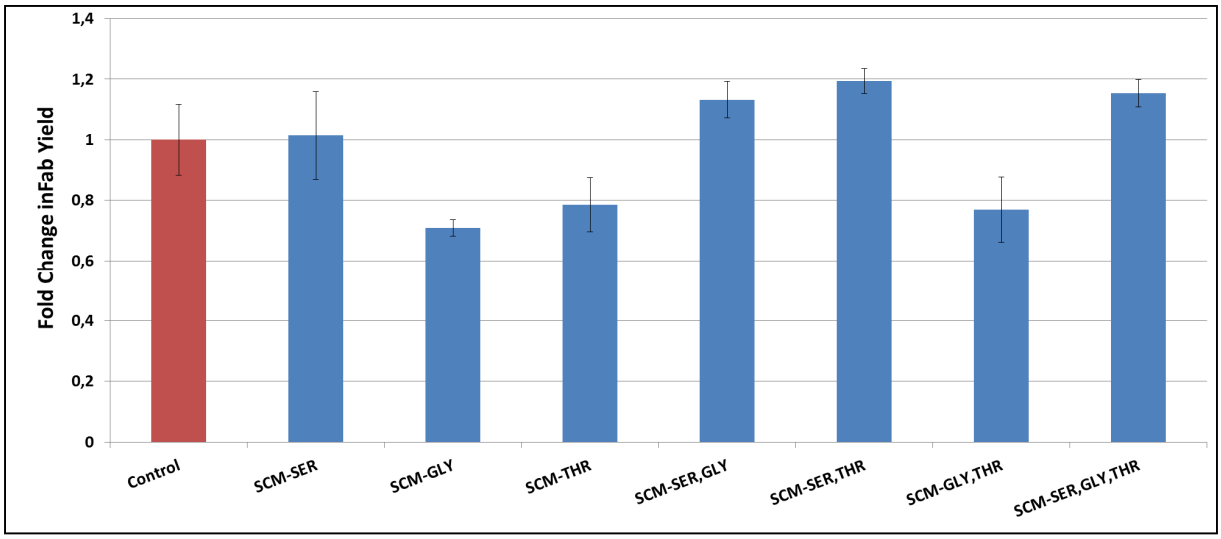

Fig.24: Results for SDZ-Fab production media lacking different amino acids and amino acids mixtures. Relative changes of SDZ-Fab yield are shown. Error bars indicate the standard errors of the mean (mg SDZ-Fab/gCDW on media lacking different amino acids relative to the strain cultivated on screening media with no amino acids supplemented)

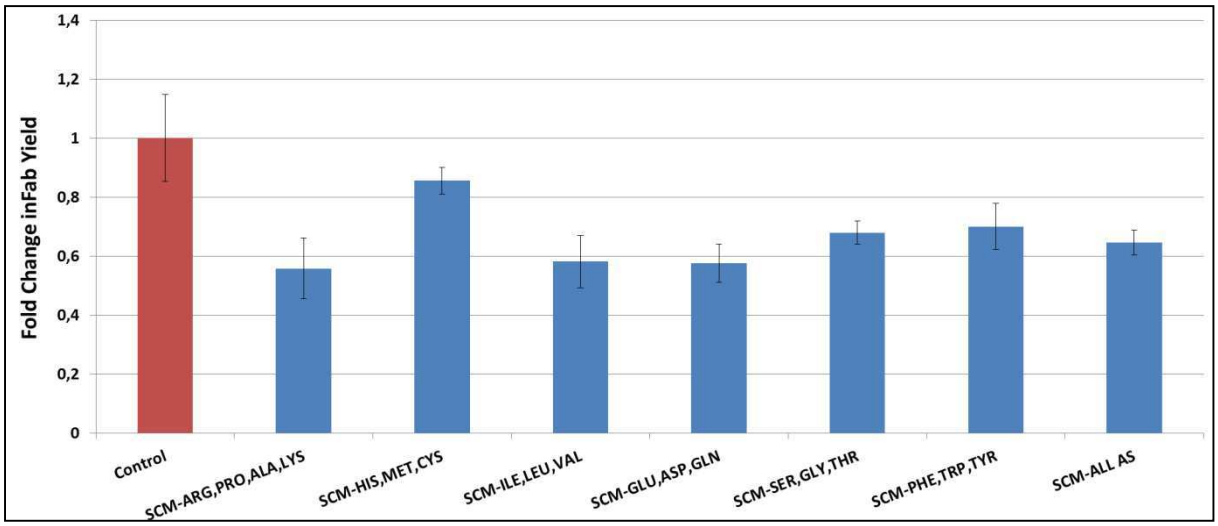

Fig.25: Results for HyHEL-Fab production on media lacking different amino acid mixtures. Relative changes of SDZ-Fab yield are shown. Error bars indicate the standard errors of the mean (mg HyHEL-Fab/gCDW on media lacking different amino acids relative to the strain cultivated on screening media with no amino acids supplemented)

In summary, the addition of specific amino acids (e.g. serine) had a negative effect on the production of both model proteins if methanol was used as sole carbon source. This shows that the approach of amino acid supplementation to unburden cellular metabolism from recombinant protein production is not generally applicable. Furthermore, thinking of an industrial process addition of amino acids causes several problems. First, addition of amino acids in a large scale is expensive, as higher quantities of amino acids must be added. Second, supplemented amino acids in the culture supernatant can influence downstream process, which aim at the purification of the protein of interest. Therefore, a second approach was done to increase intracellular amino acid levels by over-expression of specific amino acid biosynthetic genes (manuscript 3).

D) Figure License Agreements

Figure 1

A schematic overview of potential bottlenecks for recombinant protein production in the secretory pathway of yeast

SPRINGER LICENSE TERMS AND CONDITIONS	
Jan 07, 2015	
This is a License Agreement between Hannes Rußmayer ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.	
All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.	
License Number	3543740554395
License date	Jan 07, 2015
Licensed content publisher	Springer
Licensed content	Applied Microbiology and Biotechnology

publication	
Licensed content title	Engineering of protein secretion in yeast: strategies and impact on protein production
Licensed content author	Alimjan Idiris
Licensed content date	Jan 1, 2010
Volume number	86
Issue number	2
Type of Use	Thesis/Dissertation
Portion	Figures
Author of this Springer article	No
Order reference number	None
Original figure numbers	Fig. 1
Title of your thesis / dissertation	The Impact of Amino Acid Metabolism on Recombinant Protein Production in *Pichia pastoris*
Expected completion date	Oct 2015
Estimated size(pages)	70
Total	0.00 USD
Terms and Conditions	
Figure 5 Schematic overview of the different approaches used ^{13}C-based metabolic flux analysis	

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS	
Jan 07, 2015	
This Agreement between Hannes Rußmayer ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.	
License Number	3543750553476
License date	Jan 07, 2015
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Molecular Systems Biology
Licensed Content Title	Metabolic networks in motion: ^{13}C-based flux analysis
Licensed Content Author	Uwe Sauer
Licensed Content Date	Nov 14, 2006
Licensed Content Volume Number	2
Licensed Content Issue Number	1
Pages	1
Type of use	Dissertation/Thesis

Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	2
Original Wiley figure/table number(s)	Figure 2
Will you be translating?	No
Title of your thesis / dissertation	The Impact of Amino Acid Metabolism on Recombinant Protein Production in *Pichia pastoris*
Expected completion date	Oct 2015
Expected size (number of pages)	70
Requestor Location	Hannes Rußmayer Maderspergerstrasse 23 Kirchdorf, Austria 4560 Attn: Hannes Rußmayer
Billing Type	Invoice
Billing Address	Hannes Rußmayer Maderspergerstrasse 23 Kirchdorf, Austria 4560 Attn: Hannes Rußmayer
Total	0.00 USD

List of publications

B. Timmermann, S. Jarolim, H. Rußmayer, M. Kerick, S. Michel, A. Krüger, K. Bluemlein, P. Laun, J. Grillari, H. Lehrach, M. Breitenbach, M. Ralser; A new dominant peroxiredoxin allele indentified by whole genome resequencing of random mutagenized yeast causes oxidant –resistance and premature aging, *Aging*, **2010**, 2, Vol. 8, 475-486

S. Neubauer, Haberhauer-Troyer C., Klavins K., Russmayer H., Steiger MG., Gasser B., Sauer M., Mattanovich D., Hann S., Koellensperger G.: U(13) cell extract of *Pichia pastoris* – a powerful tool for evaluation of sample preparation in metabolomics, *J. Sep Sci.*, **2012**, 35(22), 3091-105

K. Klavins, S. Neubauer, A. Al Chalabi, D. Sonntag, C. Haberhauer-Troyer, H. Russmayer, M. Sauer, D. Mattanovich, S. Hann, G. Koellensperger: Interlaboratory comparison for quantitative primary metabolite profiling in *Pichia pastoris*, *Anal. Bioanal. Chem.*, **2013**, 405, 5159-5169

M. Sauer, P. Branduardi, H. Russmayer, H. Marx, D. Porro, D. Mattanovich: Molecular Mechanisms in Yeast Carbon Metabolism, Production of Metabolites and Heterologous Proteins, Book chapter *Springer Verlag*, **2014**, ISBN: 3642550134

J. Nocon, M. Steiger, M. Pfeffer, S. Sohn, T. Kim, M. Maurer, H. Rußmayer, C. Haberhauer-Troyer, S. Hann, G. Köllensperger, B. Gasser, SY. Lee, D. Mattanovich: Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production, *Metab. Eng.*, **2014**, 24, 129-138

H. Russmayer, C. Troyer, S. Neubauer, M. Steiger, B. Gasser, Hann S., G. Koellensperger, M. Sauer, D. Mattanovich: Metabolomics sampling of *Pichia pastoris* revisited: rapid filtration prevents metabolite loss during quenching, *FEMS Yeast Res.*, 2015

H. Rußmayer, M. Buchetics, C. Gruber, M. Valli, K. Grillitisch, G. Modarres, R. Guerrasio, K. Klavins, S. Neubauer, H. Drexler, M. Steiger, C. Troyer, A. Al Chalabi, G. Krebiehl, D. Sonntag, G. Zellnig, G. Daum, AB. Graf, F. Altmann, G. Koellensberger, S. Hann, M. Sauer, D. Mattanovich, B. Gasser: Systems-level organization of yeast methylotrophic lifestyle, BMC Biology, 2015

Manuscripts in preparation

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

Oral presentations

H. Rußmayer, M. Buchetics, A. Graf, C. Gruber, G. Modarres, F. Altmann, R. Guerrasio, K. Klavins, S. Neubauer, M. Steiger, A. Al Chalabi, G. Krebiehl, D. Sonntag, C. Haberhauer-Troyer, G. Köllensperger, S. Hann, M. Sauer, B. Gasser, D. Mattanovich: Methanol induced changes on the transcriptome, proteome, metabolome and fluxome of *Pichia pastoris*, *Systems Biology for Metabolic and Cell Engineering*, 27.5.2014 – 28.5.2014, Barcelona

Poster presentations

H. Rußmayer, M. Buchetics, M. Steiger, M. Valli, C. Gruber, F. Altmann, A. Graf, G. Modarres, R. Guerrasio, K. Klavins, S. Neubauer, C. Haberhauer-Troyer, G. Koellensperger, S. Hann, M. Sauer, B. Gasser, D. Mattanovich: Methanol induced changes on the transcriptome, proteome, metabolome and fluxome of *Pichia pastoris*, *Pichia 2014 – Protein Expression Conference*, 2.3.2014 – 5.3.2014, San Diego

H. Rußmayer, M. Buchetics, R. Guerrasio, K. Klavins, S. Neubauer, M. Steiger, C. Haberhauer-Troyer, G. Koellensperger, S. Hann, M. Sauer, B. Gasser, D. Mattanovich: Metabolite analysis in the protein production host *Pichia pastoris*, *ICY2012, Yeast for a sustainable Future* – 26.8.2012 – 30.8.2012, Madison, Wisconsin

Curriculum Vitae

Name: Hannes Rußmayer
Date of Birth: 11 July 1984
Place of birth: Kirchdorf an der Krems, Upper Austria
Nationality: Austria
Adress: Maderspergerstraße 23, 4560 Kirchdorf an der Krems, Austria

Education:

2009 – 2015	PhD thesis at the Institute of Applied Microbiology, Department for Biotechnology, University of Natural Resources and Life Sciences Vienna, in the group of Univ. Prof. DI. Dr. Diethard Mattanovich Field of research: The impact of the amino acid metabolism on recombinant protein production in *P. pastoris*
January 2011	Master studies of Biotechnology finished with the degree of: Diplom-Ingenieur (Master of Science)
2009 - 2010	Diploma thesis at the Institute for Applied Microbiology, Department for Biotechnology, University of Natural Resources and Life Sciences Vienna, in the group of Assoc. Prof. Dr. Johannes Grillari Field of research: Interaction of the Werner gene and SNEV
2008 - 2011	Master studies of Biotechnology at the University of Natural Resources and Life Sciences, Vienna
January2008	Bachelor studies of Food - and Biotechnology at the University of Natural Resources and Life Sciences finished with the degree of: Bachelor of Science
2004 - 2008	Bachelor studies of Food - and Biotechnology at the University of Natural Resources and Life Sciences
2003 - 2004	Community Service
1998- 2003	Höhere technische Bundeslehranstalt für Chemische Betriebstechnik in Wels, finished with Matura
1994 - 1998	High school in Kirchdorf an der Krems
1990 - 1994	Elementary school in Kirchdorf an der Krems

Languages:

German: Native language

English: fluent, written and spoken

Stefan Neubauer[1,2]
Christina Haberhauer-Troyer[1]
Kristaps Klavins[1,3]
Hannes Russmayer[3,4]
Matthias G. Steiger[3,4]
Brigitte Gasser[3,4]
Michael Sauer[2,3,4]
Diethard Mattanovich[3,4]
Stephan Hann[1,3]
Gunda Koellensperger[1,3]

[1]Department of Chemistry Division of Analytical Chemistry, University of Natural Resources and Life Sciences-BOKU, Vienna, Austria
[2]FH Campus Wien, Bioengineering, Vienna, Austria
[3]Austrian Centre of Industrial Biotechnology (ACIB), Vienna, Austria
[4]Department of Biotechnology, University of Natural Resources and Life Sciences-BOKU, Vienna, Austria

Received May 8, 2012
Revised July 12, 2012
Accepted July 15, 2012

Research Article

$U^{13}C$ cell extract of *Pichia pastoris* – a powerful tool for evaluation of sample preparation in metabolomics

Quantitative metabolic profiling is preceded by dedicated sample preparation protocols. These multistep procedures require detailed optimization and thorough validation. In this work, a uniformly ^{13}C-labeled ($U^{13}C$) cell extract was used as a tool to evaluate the recoveries and repeatability precisions of the cell extraction and the extract treatment. A homogenous set of biological replicates ($n = 15$ samples of *Pichia pastoris*) was prepared for these fundamental experiments. A range of less than 30 intracellular metabolites, comprising amino acids, nucleotides, and organic acids were measured both in monoisotopic ^{12}C and $U^{13}C$ form by LC-MS/MS employing triple quadrupole MS, reversed phase chromatography, and HILIC. Recoveries of the sample preparation procedure ranging from 60 to 100% and repeatability precisions below 10% were obtained for most of the investigated metabolites using internal standardization approaches. Uncertainty budget calculations revealed that for this complex quantification task, in the optimum case, total combined uncertainty of 12% could be achieved. The optimum case would be represented by metabolites, easy to extract from yeast with high and precise recovery. In other cases the total combined uncertainty was significantly higher.

Keywords: Extraction / Metabolic profiling / *Pichia pastoris* / Recovery / Sample preparation
DOI 10.1002/jssc.201200447

1 Introduction

The knowledge of the microbial metabolome and its regulations facilitates advanced metabolic engineering and cell factory design [1]. Metabolomics is per definition a global analysis – including both identification and quantification – of the complete metabolome of a given biological system [1–6]. The analytical strategies can be categorized in (i) nontargeted approaches (metabolic fingerprinting, metabolic footprinting) and (ii) targeted approaches. The quantification of metabolite levels, often referred to as quantitative metabolic profiling is based on targeted approaches, e.g. the quantitative determination of selected metabolites sharing the same metabolic pathway or chemical properties. Typical metabolites are intermediates from central carbon metabolism, purine and pyrimidine nucleotides, and amino acids. The aim of quantitative metabolite profiling is to obtain an unbiased in vivo snapshot of the metabolic state of the investigated biological system.

The key points in any analytical quantitative process regard representative sampling, efficient extraction, and subsequent analysis. When quantifying intracellular metabolites in a cell culture, the main focus is on three steps: (i) rapid sampling, (ii) cell quenching, and (iii) cell extraction. The sampling time is defined as the time needed to take cells from their environment (e.g. bioreactor) into quenching solution. Quenching aims at instantly stopping the enzymatic activity of the sample by an abrupt change of sample temperature either to low temperatures (e.g. $<-20°C$) or, like e.g. in the case of whole broth analysis, when quenching and extraction are combined in a single step, at elevated temperatures (e.g. $>80°C$) [7]. And finally, during extraction the cell walls are chemically and/or thermally permeabilized or mechanically disrupted to exhaustively extract the relevant metabolites into the liquid phase.

The sampling time is critical and has to be adapted according to the analyzed metabolites and their enzymatic turnover

Correspondence: Dr. Gunda Koellensperger, Department of Chemistry, Division of Analytical Chemistry, University of Natural Resources and Life Sciences-BOKU, Muthgasse 18, 1190 Wien, Austria
E-mail: gunda.koellensperger@boku.ac.at
Fax: +43147654-6059

Abbreviations: AMP, adenosine monophosphate; **CDW**, cell dry weight; **CE**, collision energy; **CMP**, cytidine monophosphate; **FV**, fragmentor voltage; **GMP**, guanosine monophosphate; **IS**, internal standard; **k**, coverage factor; **k′**, capacity factor; **N**, number of theoretical plates; **NAD**, nicotinamide adenosin dinucleotide; **NADP**, nicotinamide adenosine dinucleotide phosphate; **PEP**, phosphoenol pyruvate; **SRM**, selected reaction monitoring; **SU**, standard uncertainty; $\mathbf{t_R}$, retention time; $\mathbf{U^{12}C}$, monoisotopic signal from cell sample or standard solution; $\mathbf{U^{13}C}$, uniformly carbon thirteen labeled; **UMP**, uridine monophosphate

times [8]. Recently, a comprehensive overview on sampling strategies was published [9]. The development of quenching and extraction methods, which can also be combined in a single step, depends on the investigated metabolites as well as on the investigated sample (organism). Once the cells are quenched, quick separation of the cells from the quenching solution is necessary to avoid leakage of intracellular metabolites, which occurs when metabolites diffuse out of the cell into the quenching solution. Moreover, metabolites present in the extracellular medium need to be efficiently removed by rapid washing. Hence quenching and washing conditions have to be thoroughly optimized for the investigated organism with respect to leakage [8]. After separation of the quenched (and washed) cells from the supernatant, either by centrifugation or filtration, the cells are subjected to extraction in order to release the relevant metabolites, which is besides rapid sampling and quenching one of the most critical steps. Generally, the implemented protocols depend on the investigated microorganism, which show highly variable properties of cell wall or no cell wall at all. In fact, fragile mammalian cells require specialized techniques, which consider the lack of cell walls and the fact, that these cells are often grown adherently [10, 11]. Recently, detailed summaries on developed sample preparation protocols were published [7, 12]. Moreover, quenching and extraction of microbial samples and reported uncertainties were comprehensively summarized [13]. For yeasts, e.g. most commonly, sampling into cold methanol quenching solution was applied [8, 14–17]. Extraction implied both, cold (subambient temperatures) as well as hot methods (e.g. boiling ethanol) [8, 14, 15].

For each investigated metabolite, extraction efficiency, and recovery has to be considered. The extraction efficiency was defined as the ability of a certain extraction method to release metabolites from the cells and could/still can – due to the lack of reference materials – only be obtained by direct comparison of several extraction methods based on different physical and chemical principles [15, 18, 19]. In contrast to this, extraction recovery considers metabolite losses during the application of the extraction method (losses due to adsorption, losses due to degradation, or interconversion of metabolites due to the presence of not fully inactivated enzymes), but does not consider incomplete extraction yields [15].

In this work, the multitarget analysis of a wide range of intracellular metabolites (organic acids, amino acids, nucleotides, and vitamins) was carried out via LC-MS/MS employing triple quadrupole MS. The high number of small metabolites ($m/z < 500$) and the nominal mass resolution of the employed triple quadrupole MS, demanded for the development of powerful chromatographic separations in order to avoid isobaric interferences. Reversed phase liquid chromatography (RP-LC) is a good choice for multitarget quantitative metabolic profiling due to high separation efficiency, high robustness of chromatography and unconfined compatibility with ESI-MS. However, the separation of polar metabolites e.g. polar amino acids is hampered. One way to circumvent this problem is to increase retention of polar analytes on RP-LC by the use of ion-pair reagents like alkyl-ammonium salts or perfluorinated carboxylic acids. Although often applied in metabolic assays [20, 21], the use of an ion pair reagent leads to several severe drawbacks, such as compromised sensitivity by ion suppression, enduring contamination of the HPLC and of the interface may restricting the use of the MS instrument to one polarity and poor robustness regarding retention times. Alternatively, additional derivatization steps were introduced making polar compounds amenable to RP-LC-MS analysis [22]. In other approaches, the metabolite set was divided according to polarity, and the samples were analyzed by complementary separation methods (e.g. RP-LC, adsorption chromatography using porous graphitized carbon, hydrophilic interaction chromatography (HILIC), or weak ion exchange) in separated analytical runs. In this work, a novel C_{18} bonded silica phase having lower ligand density and tolerating highly aqueous conditions and a HILIC column, employing a zwitterionic-bonded silica phase were used for evaluation of cell extraction and sample preparation.

A well-founded evaluation of all critical aspects (e.g. leakage, extraction recovery, storage stability) is mandatory for quantitative analysis of microbial metabolome due to its outstanding challenges. The analyst is confronted with highly reactive target analytes in a complex cellular matrix. Furthermore, the expected intracellular concentrations, in other words the pool sizes of the different metabolites in the observed cells, may range over several orders of magnitude. Hence, the reliability of the analytical data in terms of accuracy and precision has to be verified carefully before making any biological interpretation. Parameters like recoveries and storage stabilities need to be assessed in an early state of method development and determine further steps of optimization. The calculation of total combined uncertainties is, although not yet established in quantitative metabolomics, the most powerful tool to evaluate the introduced analytical process (from sampling to analysis). Comprehensive uncertainty budgeting again pinpoints the major uncertainty contributions.

In this work, we present an approach for evaluation of the sample preparation in microbial metabolomics using the example of the procedure, we optimized for the yeast *Pichia pastoris*. In particular, we assessed metabolite recoveries and repeatability precisions of the cell extraction, of the extract treatment, and of the overall sample preparation using uniformly ^{13}C-labeled ($U^{13}C$) metabolites as tracer. For the present work, a $U^{13}C$ cell extract was produced in our laboratory, it was characterized regarding purity of the metabolite labeling and regarding metabolite pattern and it was spiked to cell samples at three distinct steps of sample preparation procedure.

2 Materials and methods

2.1 Chemicals

For cultivation media, following substances were purchased from Carl Roth (Karlsruhe, Germany): $MgSO_4 \cdot 7\ H_2O$ 99%,

KCl 99.5%, K_2HPO_4 99%, KH_2PO_4 98%, glycerol 99%, glucose monohydrate for microbiology, and citric acid 99.5%. $(NH_4)_2HPO_4$ p.a. from AppliChem (Gatersleben, Germany), $CaCl_2 \cdot 2\ H_2O$ p.a, ammonia (25% v/v), KOH pellets pure and hydrochloric acid (30% ultrapure) from Merck (Darmstadt, Germany) and H_3PO_4 85% (v/v) and biotin 99% from Sigma-Aldrich (Vienna, Austria) were also used for cultivation media. Trace salt stock solution contained per liter: 6.0 g $CuSO_4 \cdot 5\ H_2O$, 0.08 g NaI, 3.0 g $MnSO_4 \cdot H_2O$, 0.2 g $Na_2MoO_4 \cdot 2\ H_2O$, 0.02 g H_3BO_3, 0.5 g $CoCl_2 \cdot 6\ H_2O$, 20.0 g $ZnCl_2$, and 5.0 mL H_2SO_4 (95–97% w/w), all from Merck as well as 5.0 g $FeSO_4 \cdot 7\ H_2O$ 99.5% from Carl Roth. $U^{13}C_6$ D-glucose 98% was used as substrate for the fed batch cultivation medium for production of the $U^{13}C$ cell extract and was purchased from Cambridge Isotope Laboratories (Andover, MA, USA).

For rapid sampling and quenching dry ice pellets from Linde Gas GmbH (Vienna, Austria), ethanol 96% from Merck, ethylene glycol, and methanol (>99%) from Carl Roth were used. For cell extraction were used: ethanol 99.8% and LC-MS grade water from Sigma-Aldrich. Hydrochloric acid (30% ultrapure) and sodium hydroxide monohydrate Suprapure® from Merck and LC-MS grade water from Sigma-Aldrich were used for dissolution of standard substances. For chromatography LC-MS grade water and LC-MS grade acetonitrile from Sigma-Aldrich, LC-MS grade methanol from Fisher Scientific (Loughborough, UK) and formic acid 98–100% Suprapur® from Merck were used.

2.2 Standards

The following standard substances were purchased from Sigma-Aldrich or Fluka (Vienna, Austria): *cis*-aconitic acid 98%, L-alanine 99.5%, adenosine 3′-monophosphoric acid 97%, adenosine 5′-monophosphate sodium salt 99%, L-arginine 98.5%, L-asparagine 98%, biotin 99%, cytidine 5′-monophosphate sodium salt 99%, fumaric acid 99.5%, guanosine 3′-monophosphate sodium salt 99%, guanosine 5′-monophosphate disodium salt 99%, DL-histidine 99%, DL-isocitric acid sodium salt 98%, L-isoleucine 98%, L-lysine 98%, β-nicotinamide adenine dinucleotide sodium salt 95%, β-nicotinamide adenine dinucleotide phosphate disodium salt 97%, phospho(enol)pyruvate 99%, L-proline 99.5%, (-)-riboflavin 98%, DL-serine 99%, succinic acid 99.5%, L-threonine 98%, L-tyrosine 99%, and uridine 5′-monophosphate disodium salt 98%. The following substances were purchased from Merck (Darmstadt, Germany): L-aspartic acid 99%, citric acid 99.5–100.5%, L-glutamic acid 99%, L-glutamine 99%, L-leucine 99%, DL-malic acid 99.5%, L-methionine 99%, L-phenylalanine 99%, and L-valine 99%.

A range of 0.5–2.0 mg of each substance were weighed and 1 mL of LC-MS grade water was added – in the case of amino acids 1 mL of 0.1 mol L^{-1} HCl was added instead of water – and in case of 3′AMP 1 mL of 0.1 mol L^{-1} NaOH was added instead of water – resulting in 5 mmol L^{-1} stock solutions. Further dilution steps were achieved with pipettes with disposable tips from Eppendorf (Vienna, Austria). The final dilution step was conducted by diluting in LC-MS grade water. Stock solutions were kept at −20°C.

2.3 Fully labeled ^{13}C cell extract

The wild-type strain of *P. pastoris* was cultivated in fed batch using a 1.4 L benchtop bioreactor (DASGIP Parallel Bioreactor System, Germany) with a working volume of 400 mL using a minimal medium with $U^{13}C_6$ glucose as sole carbon source. The preculture medium (per liter: 0.5 g $MgSO_4 \cdot 7\ H_2O$, 0.9 g KCl, 0.022 g $CaCl_2 \cdot 2\ H_2O$, 10.98 mL 85% (v/v) H_3PO_4, 4.6 mL trace salt stock solution, 2 mL biotin solution (c = 0.2 g L^{-1}), 10 g $U^{13}C_6$ D-glucose, and 100 mL K_2HPO_4/KH_2PO_4 buffer (1 mol L^{-1}, pH 6), pH set to 5.0 with 25% (v/v) ammonia) was inoculated with 200 μL cryostock of *P. pastoris* CBS 7435. The preculture was grown at 28°C and 150 rpm overnight and was used for inoculation of the bioreactor with an optical density (OD600) of 0.2.

After a batch phase of approximately 35 h the feed was started. The fermentation strategy was a fed batch under carbon-limited conditions with a growth rate of 0.1 h^{-1}. The pH was controlled at 5.0 with 25% (v/v) ammonia, the temperature was controlled at 25°C and the dissolved oxygen was kept constant at 20% by controlling the stirrer speed and inlet gas flow. To ensure complete ^{13}C labeling the CO_2 from the inlet air using an in-house made CO_2 remover with aerosol scrubber were used (two bottles with 4 mol L^{-1} KOH and one bottle with water connected with silicon tubes and frits in series; liquid level is approximately 20 cm and corresponds to the distance the air bubbles through the solution).

Batch and fed batch medium contained per liter: 0.5 g $MgSO_4 \cdot 7\ H_2O$, 0.9 g KCl, 0.022 g $CaCl_2 \cdot 2\ H_2O$, 10.98 mL 85% (v/v) H_3PO_4, 4.6 mL trace salt stock solution, 2 mL biotin solution (c = 0.2 g L^{-1}), and 10 g $U^{13}C_6$ glucose. The pH was adjusted to 5.0 by using 25% (v/v) ammonia.

The cells were grown until a final biomass concentration of approximately 4.5 g_{CDW} L^{-1}. Before the first sampling round the cells were grown at least 2 h to ensure exponential growth. The cultivation broth was then sampled immediately into the four-fold volume of cold quenching and washing solvent (QS = 60% v/v methanol) using a peristaltic pump and silicone tubes (diameter 5 mm, length 81 cm) at a pumping speed of 5 mL s^{-1}. QS and quenched cell suspension were tempered to −30 ± 3°C by a cooling mixture (= ethylene glycol/ethanol = 70:30 + dry ice pellets). Biomass was determined by drying triplicates of 2 mL chemostat culture to constant weight at 105°C in preweight glass tubes and was 4.03 g_{CDW} L^{-1} (cell dry weight) with a standard deviation of 0.08 g_{CDW} L^{-1} ($n = 3$).

The quenched cell suspension was aliquoted in 10 mL portions into 15 mL sample tubes from Greiner (Frickenhausen, Germany) and pelleted via centrifugation at 4000 × *g* and −20°C for 10 min using a Sorvall RC 6+ centrifuge from Thermo Scientific (Waltham, MA, USA). A second aliquot of 10 mL was added on top of each cell pellet and centrifuged

again. The cell pellets were washed twice by resuspension in 5 mL of tempered QS and centrifugation at 4000 × *g* and −20°C. Four milliliters of extraction solvent (= 75% v/v ethanol, tempered to 85°C) was added to each pellet, the cells were suspended by vortexing and the cells were extracted for 3 min at 85°C whereat vortexed a second time after 1.5 min of extraction and vortexed a third time after 3 min of extraction. The extracted pellets were rapidly cooled down but not frozen. The ethanolic extracts were separated from the cell debris by centrifugation at 4000 × *g* and −20°C and decanting; they were pooled and stored at −80°C. The pooled ethanolic extract was evaporated to complete dryness in a vacuum centrifuge (Savant RVT400 from Thermo Scientific) operating at pressures below 1 mbar and resuspended with LC-MS grade water to a final volume that is one-fourth of the primary ethanolic extract. Unsoluble particles were removed via centrifugation at 4000 × *g* at 5°C for 10 min using a table centrifuge from Hettich (Tuttlingen, Germany) and the aqueous extract was pooled. The aqueous $U^{13}C$ cell extract was kept at −80°C.

2.4 Chemostat cultivation for cell samples

The chemostat cultivation was performed in a 1.4 L bentchtop bioreactor (DASGIP Parallel Bioreactor System, Germany) with a working volume of 400 mL. Hundred milliliters preculture medium (per liter: 10 g yeast extract, 20 g peptone, 10 g glycerol) was inoculated from the working cell bank (750 μL cryostock of *P. pastoris* CBS7435). The preculture was grown at 28°C and 150 rpm overnight and 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 feed and harvest for the continuous chemostat cultivation was started. The cells were grown under glucose limited conditions with a dilution rate of 0.1 h^{-1} for at least seven residence times before taking the sample. Temperature, pH and dissolved oxygen were maintained at 25°C, 5.0 (with 8 mol L^{-1} KOH) and 20% (by controlling the stirred speed and inlet gas flow), respectively.

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

Chemostat medium contained per liter: 1.0 g $MgSO_4 \cdot 7\ H_2O$, 2.5 g KCl, 0.04 g $CaCl_2 \cdot 2\ H_2O$, 2.43 g trace salt stock solution, 2 g biotin solution (c = 0.2 g L^{-1}), 2.3 g citric acid, 21.75 g $(NH_4)_2\ HPO_4$, and 55 g glucose monohydrate. The pH was set to 5.0 with 32% (v/v) HCl.

2.5 Rapid sampling and quenching

Samples for analysis of intracellular metabolites were taken immediately using a peristaltic pump and silicone tubes (diameter 5 mm, length 81 cm) at a pumping speed of 5 mL s^{-1}. Approximately 50 mL (47.0 g ± 0.1 g) fermentation broth was quenched in 200 mL of cold quenching and washing solvent (QS = 60% v/v methanol). QS and quenched cell suspension were tempered to −30°C ± 3°C by a cooling mixture (= ethylene glycol/ethanol = 70:30 + dry ice pellets). After quenching, 15 portions of 5 mL quenched cell suspension (approximately 4.50 g ± 0.01 g, corresponding to approximately 5 mg of CDW) were pelleted in 15 mL sample tubes (Greiner, Frickenhausen, Germany) by centrifugation at 4000 × *g* at −20°C for 10 min using a Sorvall RC 6+ centrifuge from Thermo Scientific (Waltham, MA, USA). All tubes were weighed before and after the sampling procedure (± 0.01 g) in order to determine the exact amount of sample taken. The cells were washed two times by adding 5 mL of QS to the cell pellet, resuspended and centrifuged at −20°C and were kept on dry ice until extraction.

Biomass was determined by drying five replicates of 2 mL chemostat culture to constant weight at 105°C in preweight glass tubes and was 5.12 $g_{CDW}\ L^{-1}$ (cell dry weight) with a standard deviation of 0.15 $g_{CDW}\ L^{-1}$ ($n = 5$).

2.6 Boiling ethanol extraction and extract treatment

Prior to extraction, five samples were spiked with 200 μL of aqueous fully ^{13}C-labeled ($U^{13}C$) cell extract and ten samples were spiked with 200 μL of LC-MS grade water. A 3.8 mL of extraction solvent (= 79% v/v ethanol, tempered to 85°C,) was added to each pellet, the cells were suspended by vortexing and the cells were extracted for 3 min at 85°C, vortexed a second time after 1.5 min of extraction and vortexed a third time after 3 min of extraction. The extracted pellets were rapidly cooled down but not frozen. The ethanolic extracts were separated from the cell debris by centrifugation at 4000 × *g* at −20°C for 10 min using a Sorvall RC 6+ centrifuge from Thermo Scientific (Waltham, MA, USA) and decanting. Five unspiked samples were spiked with 200 μL of aqueous $U^{13}C$ cell extract and to ten samples 200 μL of LC-MS grade water was added. The ethanolic extracts were stored on dry ice until they were evaporated to complete dryness in a vacuum centrifuge (Savant RVT400 from Thermo Scientific) operating at pressures below 1 mbar. Five unspiked samples were spiked with 200 μL of aqueous $U^{13}C$ cell extract. All 15 samples were filled up to a final volume to 1000 μL using LC-MS grade water and resuspended via vortexing. Unsoluble particles were removed via centrifugation at 4000 × *g* at 5°C using for 10 min a table centrifuge from Hettich (Tuttlingen, Germany). For HILIC measurement, one additional dilution step of 1:10 was carried out using LC-MS grade water.

2.7 Chromatography

RP-LC was performed using an Atlantis T3® C_{18} reversed phase column (150 × 4.6 mm, 3-μm particle size) from Waters® (Milford, USA), an Atlantis T3® guard column (20 × 4.6 mm, 3-μm particle size), eluent A (water, 0.1% v/v

formic acid), and eluent B (methanol). Following gradient was applied: 0% B was constant for 2 min, then B was increased to 40% within 8 min and was held for 2 min. Subsequent increase to 100%, within 0.1 min and holding for 1.9 min to flush the column, subsequent reconstitution of the starting conditions within 0.1 min and reequilibration with 0% B for 5.9 min resulted in a total analysis time of 20 min. A flow rate of 1.0 mL min^{-1}, an injection volume of 12.5 μL and a column temperature of 40°C were applied.

HILIC was performed using a ZicHILIC© separation column (150 × 4.6 mm, 3.5-μm particle size, 100 Å pore size) from SeQuant (Marl, Germany) a ZicHILIC© guard column (20 × 2.1 mm, 5-μm particle size), eluent A (98% v/v water, 1% v/v acetonitrile, 1% v/v formic acid), and eluent B (98% v/v acetonitrile, 1% v/v water, 1% v/v formic acid) applying the following gradient: 90% B was constant for 2 min, then B was reduced to 10% within 7 min and was held for 1 min. Subsequent reconstitution of the starting conditions within 0.1 min and reequilibration with 90% B for 9.9 min resulted in a total analysis time of 20 min. A flow rate of 0.5 mL min^{-1}, an injection volume of 12.5 μL and a column temperature of 40°C were applied.

2.8 LC-triple quad system

An Agilent G1312A Binary Pump 1200 series from Agilent Technologies (Santa Clara, CA, USA) together with an Agilent G1367B high performance autosampler and an Agilent G1316A column compartment was employed for HPLC. For MS detection an Agilent 6410 Triple Quad LC-MS from Agilent Technologies was used featuring an ESI interface. Source parameters in negative and positive ionization mode for RP-LC using Atlantis T3® were set as follows: drying gas temperature 350°C, drying gas flow 10 L min^{-1}, nebulizer pressure 50 psi and capillary voltage 4000 V. Source parameters in positive ionization mode for HILIC were set as follows: drying gas temperature 300°C, drying gas flow 10 L min^{-1}, nebulizer pressure 25 psi and capillary voltage 4000 V. Selected reaction monitoring (SRM) transitions of the metabolites using the monoisotopic ^{12}C isotopologue ($U^{12}C$) have been determined via flow injection of 5 μmol L^{-1} single standard (isocratic conditions: 0.05% formic acid in methanol/H_2O = 50:50 v/v). For optimization Mass Hunter Optimizer Software from Agilent was applied processing the following four steps in order: (i) optimization of the isolation of the selected precursor ion by varying the fragmentor voltage, (ii) determination of the four most abundant product ions, (iii) optimization of the collision energies for each of these product ions, and (iv) determination of the exact *m/z* value of the product ions. The transition featuring the highest signal to noise ratio was chosen for recovery evaluation (quantifier) and if available, the transition featuring the second highest transition was chosen for identification (qualifier). The optimized SRM transitions of the fully labeled ^{13}C ($U^{13}C$) metabolites were acquired by calculating the number of carbon atoms in the fragments and adopting the fragmentation parameters. Precursor and product ions as well as specific values for fragmentor voltage and collision energy are listed in Table 1. The instrument was set to one polarity mode per run resulting in three methods. For RPLC (+), for RPLC (−) and for HILIC (+), SRM methods were performed setting the retention times of the analytes, a time window of 2 min and a cycle time of 1000 ms. These settings resulted for RPLC and positive ionization in a total of 49 transitions with a minimum dwell time of 39.98 ms (23 concurrent transitions) and a maximum dwell time of 329.83 ms. For RPLC and negative ionization they resulted in a total of 69 transitions with a minimum dwell time of 22.14 ms (39 concurrent transitions) and a maximum dwell time of 996.15 ms and for HILIC and positive ionization they resulted in a total of 43 transitions with a minimum dwell time of 27.75 ms (32 concurrent transitions) and a maximum dwell time of 996.50 ms. LC-ESI-MS/MS data was evaluated employing Agilent Mass Hunter Qualitative and Quantitative Analysis software modules. Peak integration was based on ±0.5 *m/z* peak with for precursor and product ion.

2.9 LC-MS/MS analysis and evaluation of recoveries

From our initial set of 34 metabolites, 19 were measured only on RP-LC-MS/MS (Table 2), ten polar amino acids were measured only on HILIC-MS/MS (Table 3) and five amino acids comprising apolar side chains (met, phe, pro, tyr, val) were measured on both platforms. Low abundant metabolites (S/N $U^{13}C < 10$) were excluded from recovery evaluation (Table 4) and five highly abundant amino acids (arg, gln, glu, his, lys) where measured on HILIC-MS/MS in an additional 1:10 dilution. Metabolite recoveries were evaluated by comparing peak areas of $U^{13}C$ metabolites deriving from three different spikes and the $U^{13}C$ cell extract was characterized from a control sample:

① = addition of 200 μL aqueous $U^{13}C$ cell extract onto cell pellet prior to extraction ($n = 5$)
② = addition of 200 μL aqueous $U^{13}C$ cell extract to cooled (−20°C) ethanolic cell extract ($n = 5$)
③ = addition of 200 μL aqueous $U^{13}C$ cell extract to dried extract after evaporation ($n = 5$)

ctrl = 200 μL aqueous $U^{13}C$ cell extract, 800 μL LC-MS grade water ($n = 4$)

The recoveries of the extraction, of the sample treatment (evaporation) and of the overall sample preparation is then calculated according to Eqs 1, 2, and 3 (see also Fig. 1).

$$R_{ex} = area\ U^{13}C① / area\ U^{13}C② \quad (1)$$

$$R_{ev} = area\ U^{13}C② / area\ U^{13}C③ \quad (2)$$

$$R_{sp} = area\ U^{13}C① / area\ U^{13}C③ \quad (3)$$

Furthermore, the monoisotopic signal ($U^{12}C$) deriving from the cell sample was used as internal standard and $U^{12}C$ corrected recoveries were determined correcting by

Table 1. Measurement conditions and SRM settings of of the LC-MS/MS instrument are shown for all the measured metabolites. Precursor ions ($[M+H]^+$, $[M-H]^-$) and fragment ions (quantifier, qualifier) are shown for the monoisotopic molecule deriving from standard or cell sample ($U^{12}C$) and for the one deriving from fully labeled cell extract ($U^{13}C$)

Compound	Polarity	Chromatography	Dilution	Precursor	Quantifier	Qualifier	FV / V	$CE_{Quantifier}$ / V	$CE_{Qualifier}$ / V	Comp. for corr.
$U^{12}C_6$ cis-Aconitate	–	RP-LC	1:1	173	85	129	– 80	– 5	– 5	–
$U^{13}C_3$ Alanine	+	HILIC	1:1	93	46	–	40	9	–	$U^{12}C_3$ Alanine
$U^{12}C_3$ Alanine	+	HILIC	1:1	90	44	–	40	9	–	–
$U^{13}C_{10}$ 3′AMP	–	RP-LC	1:1	356	216	79	– 114	– 13	– 40	–
$U^{12}C_{10}$ 3′AMP	–	RP-LC	1:1	346	211	79	– 114	– 13	– 40	–
$U^{13}C_{10}$ 5′AMP	–	RP-LC	1:1	356	79	139	– 118	– 40	– 40	$U^{12}C_{10}$ 5′AMP
$U^{12}C_{10}$ 5′AMP	–	RP-LC	1:1	346	79	134	– 118	– 40	– 40	–
$U^{13}C_6$ Arginine	+	HILIC	1:1, 1:10	181	74	–	80	25	–	$U^{12}C_6$ Arginine
$U^{12}C_6$ Arginine	+	HILIC	1:1, 1:10	175	70	60	80	25	13	–
$U^{13}C_4$ Asparagine	+	HILIC	1:1	137	76	90	60	13	5	$U^{12}C_4$ Asparagine
$U^{12}C_4$ Asparagine	+	HILIC	1:1	133	74	87	60	13	5	–
$U^{13}C_4$ Aspartate	+	HILIC	1:1	138	76	73	60	9	13	$U^{12}C_4$ Aspartate
$U^{12}C_4$ Aspartate	+	HILIC	1:1	134	74	70	60	9	13	–
$U^{12}C_{10}$ Biotin	+	RP-LC	1:1	245	227	97	90	9	33	–
$U^{13}C_6$ Citrate	–	RP-LC	1:1	197	116	–	– 90	– 5	–	$U^{12}C_6$ Citrate
$U^{12}C_6$ Citrate	–	RP-LC	1:1	191	111	–	– 90	– 5	–	–
$U^{13}C_9$ 5′CMP	–	RP-LC	1:1	331	79	97	– 124	– 40	– 21	–
$U^{12}C_9$ 5′CMP	–	RP-LC	1:1	322	79	97	– 124	– 40	– 21	–
$U^{13}C_4$ Fumarate	–	RP-LC	1:1	119	74	–	– 80	– 9	–	–
$U^{12}C_4$ Fumarate	–	RP-LC	1:1	115	71	–	– 80	– 9	–	–
$U^{13}C_5$ Glutamate	+	HILIC	1:1, 1:10	153	88	59	70	13	25	$U^{12}C_5$ Glutamate
$U^{12}C_5$ Glutamate	+	HILIC	1:1, 1:10	148	84	56	70	13	25	–
$U^{13}C_5$ Glutamine	+	HILIC	1:1, 1:10	152	88	135	80	17	9	$U^{12}C_5$ Glutamine
$U^{12}C_5$ Glutamine	+	HILIC	1:1, 1:10	147	84	130	80	17	9	–
$U^{13}C_{10}$ 3′GMP	–	RP-LC	1:1	372	216	79	– 118	– 17	– 21	–
$U^{12}C_{10}$ 3′GMP	–	RP-LC	1:1	362	211	79	– 118	– 17	– 21	–
$U^{13}C_{10}$ 5′GMP	–	RP-LC	1:1	372	79	216	– 118	– 21	– 17	–
$U^{12}C_{10}$ 5′GMP	–	RP-LC	1:1	362	79	211	– 118	– 21	– 17	–
$U^{13}C_6$ Histidine	+	HILIC	1:1, 1:10	162	115	87	80	13	25	$U^{12}C_6$ Histidine
$U^{12}C_6$ Histidine	+	HILIC	1:1, 1:10	156	110	83	80	13	25	–
$U^{13}C_6$ Isocitrate	–	RP-LC	1:1	197	116	179	– 90	– 5	– 5	$U^{12}C_6$ Isocitrate
$U^{12}C_6$ Isocitrate	–	RP-LC	1:1	191	111	173	– 90	– 5	– 5	–
$U^{13}C_6$ Isoleucine	+	RP-LC	1:1	138	91	–	70	9	–	$U^{12}C_6$ Isoleucine
				91	–	61	130	–	20	
$U^{12}C_6$ Isoleucine	+	RP-LC	1:1	132	86	–	70	9	–	–
				86	–	57	130	–	20	
$U^{13}C_6$ Leucine	+	RP-LC	1:1	138	91	–	70	9	–	$U^{12}C_6$ Leucine
				91	–	46	130	–	20-	
$U^{12}C_6$ Leucine	+	RP-LC	1:1	132	86	–	70	9	–	–
				86	–	43	130	–	20	
$U^{13}C_6$ Lysine	+	HILIC	1:1, 1:10	153	89	136	60	13	9	$U^{12}C_6$ Lysine
$U^{12}C_6$ Lysine	+	HILIC	1:1, 1:10	147	84	130	60	13	9	–
$U^{13}C_4$ Malate	–	RP-LC	1:1	137	119	–	– 80	– 9	–	$U^{12}C_4$ Malate
$U^{12}C_4$ Malate	–	RP-LC	1:1	133	115	–	– 80	– 9	–	–
$U^{13}C_5$ Methionine	+	RP-LC, HILIC	1:1	155	59	–	70	13	–	–
$U^{12}C_5$ Methionine	+	RP-LC	1:1	150	56	61	70	13	25	–
$U^{13}C_{21}$ NAD^+	–	RP-LC	1:1	683	555	–	– 94	– 13	–	$U^{12}C_{21}$ NAD^+
$U^{12}C_{21}$ NAD^+	–	RP-LC	1:1	662	540	273	– 94	– 13	– 37	–
$U^{13}C_{21}$ $NADP^+$	–	RP-LC	1:1	763	635	418	– 104	– 13	– 33	$U^{12}C_{21}$ $NADP^+$
$U^{12}C_{21}$ $NADP^+$	–	RP-LC	1:1	742	620	408	– 104	– 13	– 33	–
$U^{13}C_3$ PEP	–	RP-LC	1:1	170	79	–	– 70	– 9	–	$U^{12}C_3$ PEP
$U^{12}C_3$ PEP	–	RP-LC	1:1	167	79	–	– 70	– 9	–	–
$U^{13}C_9$ Phenylalanine	+	RP-LC, HILIC	1:1	175	128	111	80	9	29	$U^{12}C_9$ Phenylalanine
$U^{12}C_9$ Phenylalanine	+	RP-LC, HILIC	1:1	166	120	103	80	9	29	–

Table 1. Continued.

Compound	Polarity	Chromatography	Dilution	Precursor	Quantifier	Qualifier	FV / V	$CE_{Quantifier}$ / V	$CE_{Qualifier}$ / V	Comp. for corr.
$U^{13}C_5$ Proline	+	RP-LC, HILIC	1:1	121	74	46	80	13	33	$U^{12}C_5$ Proline
$U^{12}C_5$ Proline	+	RP-LC, HILIC	1:1	116	70	43	80	13	33	–
$U^{12}C_{17}$ Riboflavin	–	RP-LC	1:1	375	255	–	– 124	– 13	–	–
$U^{13}C_3$ Serine	+	HILIC	1:1	109	62	44	60	9	21	$U^{12}C_3$ Serine
$U^{12}C_3$ Serine	+	HILIC	1:1	106	60	42	60	9	21	–
$U^{13}C_4$ Succinate	–	RP-LC	1:1	121	76	–	– 80	– 9	–	–
$U^{12}C_4$ Succinate	–	RP-LC	1:1	117	73	–	– 80	– 9	–	–
$U^{13}C_4$ Threonine	+	HILIC	1:1	124	77	69	60	9	13	$U^{12}C_4$ Threonine
$U^{12}C_4$ Threonine	+	HILIC	1:1	120	74	56	60	9	13	–
$U^{13}C_9$ Tyrosine	+	RP-LC, HILIC	1:1	191	144	98	80	9	29	$U^{12}C_9$ Tyrosine
$U^{12}C_9$ Tyrosine	+	RP-LC, HILIC	1:1	182	135	91	80	9	29	–
$U^{13}C_5$ Valine	+	RP-LC, HILIC	1:1	123	76	59	70	9	21	$U^{12}C_5$ Valine
$U^{12}C_5$ Valine	+	RP-LC, HILIC	1:1	118	72	55	70	9	21	–
$U^{13}C_9$ 5′UMP	–	RP-LC	1:1	332	79	97	– 114	– 40	– 21	–
$U^{12}C_9$ 5′UMP	–	RP-LC	1:1	323	79	97	– 114	– 40	– 21	–

Table 2. Chromatographic separation of 24 metabolites using Atlantis® reversed phase column

Compound	Polarity	t_R/min	k′	N
cis-Aconitate	–	7.4	3.23	38,400
3′AMP	–	7.2	3.09	48,000
5′AMP	–	6.2	2.54	36,000
Biotin	+	13.2	6.54	92,900
Citrate	–	6.3	2.60	17,900
5′CMP	–	3.6	1.06	13,500
Fumarate	–	6.7	2.83	24,400
3′GMP	–	7.8	3.46	44,600
5′GMP	–	6.6	2.77	46,600
Isocitrate	–	3.6	1.04	4700
Isoleucine	+	6.6	2.75	23,800
Leucine	+	6.8	2.90	24,300
Malate	–	3.4	0.95	5700
Methionine	+	4.4	1.52	6300
NAD^+	–	6.6	2.80	50,100
$NADP^+$	–	6.5	2.72	41,800
PEP	–	2.8	0.64	3300
Phenylalanine	+	9.0	7.28	55,800
Proline	+	2.5	0.43	2700
Riboflavin	–	13.9	6.94	12,6700
Succinate	–	6.9	2,94	36,600
Tyrosine	+	6.9	2.91	37,800
5′UMP	–	5.8	2.31	10,900
Valine	+	3.3	0.90	3900

Table 3. Chromatographic separation of 15 metabolites using ZicHILIC© separation column

Compound	Polarits	t_R / min	k′	N
Alanine	+	9.4	2.24	46,100
Arginine	+	11.8	3.06	52,600
Asparagine	+	9.8	2.38	68,900
Aspartate	+	9.7	2.33	55,000
Glutamate	+	9.4	2.26	47,600
Glutamine	+	9.6	2.31	54,400
Histidine	+	11.3	2.89	80,100
Lysine	+	11.6	3.01	76,600
Methionine	+	8.4	1.90	21,200
Phenylalanine	+	7.9	1.71	27,400
Proline	+	8.9	2.07	40,600
Serine	+	9.9	2.40	63,700
Threonine	+	9.5	2.28	52,400
Tyrosine	+	8.9	2.05	36,000
Valine	+	8.5	1.95	34,100

the corresponding monoisotopic metabolite response (area $U^{12}C/m_{sample}$) (Eqs. (4)–(7)).

$$f = area\ U^{13}C * m_{sample}/area\ U^{12}C \quad (4)$$

$$R_{excorr} = f①/f② \quad (5)$$

$$R_{evcorr} = f②/f③ \quad (6)$$

$$R_{spcorr} = f①/f③ \quad (7)$$

The sequence was set up by measuring the first replicate of the spikes and the control sample as block and measuring five blocks consecutively. Individual recoveries were calculated from each block and afterwards averaged.

2.10 Evaluation of total combined uncertainties

The terms measurand, standard uncertainty, and coverage factor are used according to the Guide to the Expression of Uncertainty in Measurement [23]. All combined uncertainties were calculated according to the ISO/GUM guide [24] using the uncertainty propagation procedure. Dedicated software (GUMworkbench software, Metrodata, Grenzach-Wyhlen,

Table 4. Characterization of $U^{13}C$ cell extract by signal to noise ratio and peak area (dilution 1:5 in order to obtain same concentration as in the spike)

Compound	Polarity	Chromatography	S/N ($U^{12}C$, $n = 4$)	S/N ($U^{13}C$, $n = 4$)	RSD of area /% ($U^{13}C$, $n = 4$)	Used for recovery
cis-Aconitate	−	RP-LC	<2	n.a.	n.a.	x
Alanine	+	HILIC	<2	43	26	√
3′AMP	−	RP-LC	<2	8.2	14	x
5′AMP	−	RP-LC	<2	56	9	√
Arginine	+	HILIC	2.2	1400	20	√
Asparagine	+	HILIC	<2	57	17	√
Aspartate	+	HILIC	<2	370	23	√
Biotin	+	RP-LC	6.9	n.a.	n.a.	x
Citrate	−	RP-LC	4.3	67	10	√
5′CMP	−	RP-LC	<2	2.1	20	x
Fumarate	−	RP-LC	<2	2.7	23	x
Glutamate	+	HILIC	3.3	4800	18	√
Glutamine	+	HILIC	<2	2200	17	√
3′GMP	−	RP-LC	<2	5.1	16	x
5′GMP	−	RP-LC	<2	2.4	19	x
Histidine	+	HILIC	4.4	1000	24	√
Isocitrate	−	RP-LC	<2	10	10	√
Isoleucine	+	RP-LC	<2	190	12	√
Leucine	+	RP-LC	3.0	220	7	√
Lysine	+	HILIC	5.6	1800	25	√
Malate	−	RP-LC	<2	110	3	√
Methionine	+	RP-LC	3.8	6.5	18	x
Methionine	+	HILIC	<2	4.2	21	x
NAD^+	−	RP-LC	<2	9400	5	√
$NADP^+$	−	RP-LC	<2	410	3	√
PEP	−	RP-LC	<2	140	21	√
Phenylalanine	+	RP-LC	<2	180	9	√
Phenylalanine	+	HILIC	<2	400	20	√
Proline	+	RP-LC	<2	100	10	√
Proline	+	HILIC	<2	50	8	√
Riboflavin	−	RP-LC	<2	n.a.	n.a.	x
Serine	+	HILIC	5.3	77	21	√
Succinate	−	RP-LC	<2	2.5	23	x
Threonine	+	HILIC	2.1	200	22	√
Tyrosine	+	RP-LC	5.3	960	10	√
Tyrosine	+	HILIC	7.6	120	13	√
5′UMP	−	RP-LC	<2	5.2	7	x
Valine	+	RP-LC	2.2	44	11	√
Valine	+	HILIC	<2	280	21	√

Germany) was employed for the calculations based on the numerical method of differentiation [25]. Accordingly, the uncertainty of the measurement result was determined from quantities through a functional relation called the measurement equation. The final uncertainty arises from the input uncertainties that enter the equation that were categorized in (i) Type A evaluation, i.e. method of evaluation of uncertainty by the statistical analysis of series of observations and (ii) Type B evaluation, i.e. method of evaluation of uncertainty by means other than the statistical analysis of series of observations. Hence, a Type B evaluation of standard uncertainty is usually based on scientific judgment using all of the relevant information available, which may include: previous measurement data, experience with, or general knowledge of, the behavior and property of relevant materials and instruments, manufacturer's specifications, data provided in calibration and other reports, and uncertainties assigned to reference data taken from handbooks.

3 Results and discussion

3.1 Optimization of LC-MS/MS

The assembled set of target analytes comprised a broad range of primary metabolites such as amino acids, organic acids,

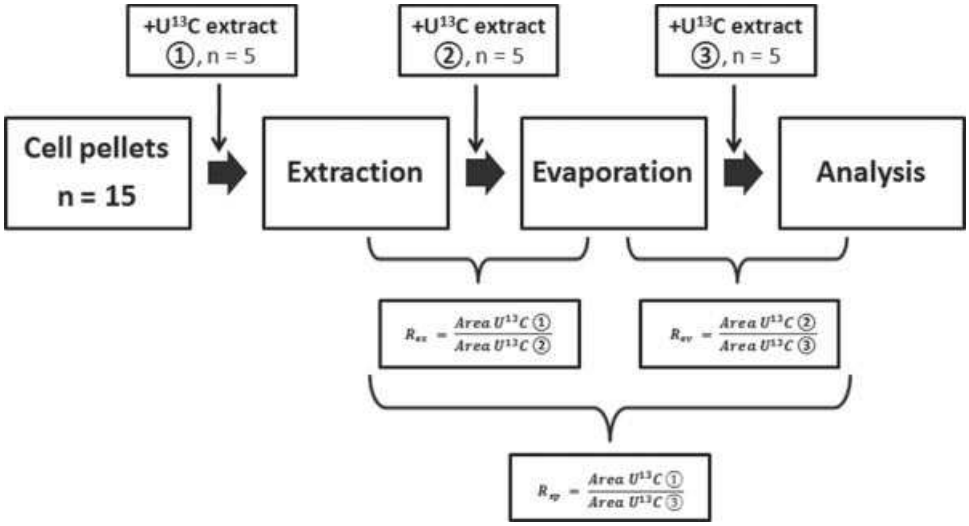

Figure 1. Experimental setup for determination of recoveries of cell extraction, evaporation of the extraction solvent, and overall sample preparation.

nucleotides, cofactors, and vitamins. The complete list of target metabolites is given in Table 4. One objective of the present work was the development of two orthogonal chromatographic separation methods, covering as much of these analytes as possible while avoiding the use of ion pair reagents.

First, the specific SRM transitions for recovery evaluation (quantifier) and for identification (qualifier) were optimized on the LC Triple Quad using single standards and monoisotopic ^{12}C ($U^{12}C$) mass isotopolgues. The transitions for the corresponding uniformly ^{13}C-labeled ($U^{13}C$) metabolites were not optimized individually but calculated from the $U^{12}C$ transitions. For few compounds the $U^{13}C$ transitions ($U^{13}C_6$ cis-aconitate, $U^{13}C_{10}$ biotin, $U^{13}C_{17}$ riboflavin) could not be confirmed because of very low concentrations in the cell extract. All SRM settings and the chosen polarity for the measured metabolites are listed in Table 1 and *m/z* values of $U^{12}C$ and U ^{13}C precursor ions and fragment ions are shown.

A screening procedure of various chromatographic phases was performed. Among the tested stationary phases were silica based HILIC phases and silica-based phases tested under RP gradient conditions as well as phases based on graphitized carbon (Luna NH_2® from Phenomenex®, Mediterranea™ Sea_{18} from Teknokroma, Atlantis T3® from Waters®, ZicHILIC© from SeQuant, and Hypercarb™ from Thermo Scientific). A reversed phase separation using a 100% wettable C_{18} phase (Atlantis T3®) showed promising peak capacity for acidic compounds (organic acids, nucleotides) and for less hydrophilic metabolites (amino acids with apolar side chains, vitamins). A comparable method based on the same chromatographic system was developed in our group for analysis of nucleotides, nucleosides, and nucleobases in feed supplements [26]. Applying the optimized chromatographic condition, and using the gradient program described before, resulted an excellent chromatography for a wide range of small analytes having apolar or acidic nature within a total runtime of 20 min. The retention times, capacity factors, and theoretical plate numbers are listed in Table 2. As an outstanding feature, this method is able to separate many isobaric compounds. Ribonucleotides that differ in their phosphorylation site but lack-specific transitions are baseline separated. This is shown for 5′AMP and 3′AMP – retention times are 6.2 min and 7.2 min, and for 5′GMP and 3′GMP – retention times are 6.6 min and 7.8 min, respectively. One pair of isomer metabolites prominent in the central carbon metabolism comprises citrate and isocitrate. These organic acids are well separated ($t_R = 6.3$ and 3.6) avoiding isobaric interferences. The isomer amino acids leucine and isoleucine can almost be baseline separated ($t_R = 6.8$ and 6.6) that allows accurate peak integration of both amino acids from the same transition. The distinction of the two isomers via monitoring of selective transitions (*m/z* 86.0 → 43 and *m/z* 86 → 57 for leucine and isoleucine, respectively) was not utilized since 86.0 is no $[M+H]^+$-ion but an in-source fragment and loss in sensitivity was observed when compared with transition from the respective $[M+H]^+$-ion. The optimized RP-LC method was finally applied to the aqueous $U^{13}C$-labeled cell extract. The resulting chromatograms are presented in Supporting Material Fig. S1.

Despite of the outstanding performance of the described RP chromatography for metabolic profiling, it cannot be used as a single analysis method due to weak retention of highly polar analytes. The narrow elution window close to the void volume renders analysis susceptible to matrix effects due to ion suppression. Additionally poor selectivity has to be considered due to isobaric interferences in the void volume. This is even more relevant when analyzing in parallel the isotopically labeled metabolome and the metabolome in its natural monoisotopic form (uniformly ^{12}C, ^{1}H, ^{14}N, ^{16}O, ^{32}S, ^{31}P). The screening of the initial set of target analytes on the Atlantis T3® column revealed poor retention (with $k' < 1$) of mainly two substance groups: (i) phosphorylated sugars and phosphorylated hydroxy acids and (ii) polar amino acids. A comprehensive separation of phosphorylated sugars and hydroxy acids – which is required for their investigation – could not be achieved and thus only phosphoenolpyruvate (PEP) was

further surveyed. For amino acid analysis, a HILIC column was used, employing a zwitterionic-bonded silica phase. A comparable method based on the same chromatographic system was developed in our group for analysis for soil solutions [27]. Applying the optimized chromatographic condition, and using the gradient program described above, resulted in superior retention of amino acids, whereupon 15, including the positive charged arginine, leucine, and histidine could be separated within a total runtime of 20 min. The retention times, capacity factors, and theoretical plate numbers are listed in Table 3. HILIC separation of isobaric leucine and isoleucine was not in the focus of our work because of the satisfactory reversed phase separation. The optimized HILIC method was finally applied to the aqueous $U^{13}C$-labeled cell extract. The resulting chromatograms of $U^{13}C$ transitions are presented in Supporting Material Fig. S2.

3.2 Characterization of the $U^{13}C$ cell extract

The in vivo synthesis of $U^{13}C$ metabolites using proper microorganisms cultivated on $U^{13}C$ substrates was introduced in metabolomics within the last decade for use as multifunctional internal standard [28] with the advantageous possibility of internal standardization of any desired metabolite in the proper concentration. In our lab a fully labeled ^{13}C cell extract was produced providing labeled metabolites used as tracer for evaluation of recoveries. Therefore the wild-type strain of the yeast *P. pastoris* was grown in aerobic, glucose limited fed batch culture. The fermentation strategy was an exponential fed batch under carbon-limited conditions with a growth rate of 0.1 h^{-1}. An isotopically labeled glucose ($U^{13}C_6$, purity = 98%) was used as sole carbon source for batch and fed batch medium as well as for the preculture medium to minimize ^{12}C impurities. Furthermore, other sources of ^{12}C contamination were excluded: the citrate used to buffer media was replaced by phosphate buffer and carbon dioxide was removed from inlet air to prevent the carbon dioxide fixation. Cell quenching was carried out by rapid sampling into cold methanol at conditions optimized for *P. pastoris* by Carnicer et al. [14] and cell extraction was carried out by boiling ethanol extraction comparable to the method proposed for *Saccharomyces cerevisiae* by Canelas et al. [15]. The cell extraction via boiling ethanol has become a standard method for investigation of yeasts and was recently applied to *P. pastoris* [14].

The direct addition of ethanolic cell extract to the measurement solution turned out to impede the reversed phase separation starting with 100% aqueous mobile phase. More precisely, the peaks in the retention time window from 5 to 7 min were shifted unpredictably. Therefore the ethanolic extract was evaporated under vacuum to dryness and reconstituted with water. The reconstitution step was also used to concentrate the extract ($f_{conc} = 4$). This aqueous $U^{13}C$ cell extract was used as spike in the recovery evaluation experiment. For quality assessment it was diluted as described above and measured as control sample in replicates. The extract was characterized according to three parameters – the signal to noise ratio (S/N) of the $U^{13}C$ metabolite, the relative standard deviation (RSD) of the $U^{13}C$ area, and the signal to noise ratio of the $U^{12}C$ impurity – which are summarized for all measured metabolites in Table 4. Looking at the $U^{13}C$ S/N values, the high dynamic range of intracellular concentrations can be readily observed (S/N range from <2 to 10^4). In fact the $U^{13}C$ S/N value is a significant criterion to assess the applicability as metabolite standard. For the recovery study, compounds with ratios <10 were excluded (Table 4). A number of ribonucleotides and organic acids did not fulfill the requirement whereas cis-aconitate, biotin, and riboflavin were <LOD. Minor $U^{12}C$ impurities (S/N < 10) are found for citrate, serine, and tyrosine while biotin is found only as unlabeled vitamin (essential for *P. pastoris*, component of the media) (Table 4).

3.3 Evaluation of recoveries of single sample preparation steps and overall sample preparation

A general sample preparation workflow for intracellular metabolic profiling in *P. pastoris* was established in our laboratories, considering the experience of recently published studies. It briefly consisted of rapid sampling from the bioreactor and quenching into cold methanolic solution, separation of quenched cells from supernatant by centrifugation, cell extraction using boiling ethanol, evaporation of the ethanolic extract, and reconstitution in water. Optimized quenching conditions (methanol content and temperature) were adopted from Carnicer et al. [14]. The contact time of cells with the quenching solution was not controlled and was up to 30 min with the applied cold centrifugation method. The boiling ethanol extraction was adapted from Mashego et al. [28] with addition of 4 mL of 75% ethanol instead of 5 mL. Cell extraction via boiling ethanol was already evaluated according to extraction efficiency and extraction recovery for *S. cerevisiae* and rated as one of two best extraction methods (beside cold chloroform–methanol) out of five methods for *S. cerevisiae* [15].

In this work, a fundamental study on extraction recoveries was carried out for *P. pastoris*. The addition of isotopically labeled metabolites at different stages of sample preparation and the comparison of the corresponding signals is a straightforward approach to evaluate metabolite specific recoveries of the sample preparation steps. As mentioned before the fully ^{13}C-labeled cell extract served as an ideal standard in the spiking experiments. This valuable tool provided a full list of labeled metabolites that matched the concentration pattern in the samples. Moreover, a homogeneous set of 15 cell pellets allowed investigating the 3-step procedure (extraction, extract evaporation, and LC-MS/MS analysis) in five replicates. Figure 1 shows the experimental setup used to address the recovery of the cell extraction (R_{ex}), of the extract treatment (evaporation of ethanolic solvent, R_{ev}) and of the overall sample preparation (R_{sp}). The $U^{13}C$ cell extract was spiked to $n = 5$ cell samples at three different steps at sample preparation and the obtained $U^{13}C$ peak areas were compared

Table 5. Observed recoveries and repeatability precisions of cell extraction (R_{ex}), evaporation of the extraction solvent (R_{ev}) and overall sample preparation (R_{sp}) obtained from ratios of $U^{13}C$ responses (no correction) and by correction with the corresponding $U^{12}C$ response and CDW of extracted cell pellet ($U^{12}C$ correction)

Compound	Chromatography	No correction			$U^{12}C$ correction		
		R_{ex} /%	R_{ev} /%	R_{sp} /%	$R_{ex\ corr}$ /%	$R_{ev\ corr}$ /%	$R_{sp\ corr}$ /%
Alanine	HILIC	105 ± 32	123 ± 60	115 ± 24	100 ± 6	115 ± 15	115 ± 18
5′AMP	RP-LC	94 ± 4	111 ± 5	105 ± 6	110 ± 13	123 ± 17	135 ± 19
Arginine	HILIC	116 ± 61	123 ± 71	115 ± 41	76 ± 1	105 ± 3	79 ± 2
Arginine	HILIC, dilution	85 ± 13	99 ± 15	83 ± 6	75 ± 2	105 ± 4	78 ± 2
Asparagine	HILIC	122 ± 56	123 ± 75	119 ± 23	92 ± 3	102 ± 5	94 ± 4
Aspartate	HILIC	99 ± 23	107 ± 37	100 ± 13	94 ± 3	102 ± 3	96 ± 5
Citrate	RP-LC	74 ± 5	100 ± 3	74 ± 5	74 ± 3	102 ± 5	75 ± 1
Glutamate	HILIC	101 ± 23	109 ± 40	103 ± 14	92 ± 2	102 ± 4	94 ± 4
Glutamate	HILIC, dilution	89 ± 16	100 ± 10	88 ± 13	92 ± 2	102 ± 6	94 ± 5
Glutamine	HILIC	107 ± 35	110 ± 54	104 ± 19	88 ± 2	96 ± 9	84 ± 7
Glutamine	HILIC, dilution	85 ± 21	96 ± 14	80 ± 17	87 ± 3	96 ± 10	83 ± 6
Histidine	HILIC	130 ± 76	132 ± 91	125 ± 41	85 ± 3	101 ± 4	85 ± 3
Histidine	HILIC, dilution	90 ± 19	95 ± 20	83 ± 12	87 ± 3	98 ± 8	85 ± 8
Isocitrate	RP-LC	83 ± 10	103 ± 14	86 ± 15	88 ± 11	113 ± 8	98 ± 7
Isoleucine	RP-LC	93 ± 4	99 ± 3	91 ± 4	98 ± 4	100 ± 6	98 ± 8
Leucine	RP-LC	90 ± 5	99 ± 2	89 ± 4	94 ± 3	100 ± 5	94 ± 7
Lysine	HILIC	95 ± 44	117 ± 62	94 ± 27	62 ± 3	107 ± 6	66 ± 5
Lysine	HILIC, dilution	73 ± 12	101 ± 15	72 ± 6	61 ± 3	109 ± 3	67 ± 2
Malate	RP-LC	85 ± 8	108 ± 4	92 ± 9	90 ± 10	109 ± 3	98 ± 13
NAD^+	RP-LC	79 ± 4	100 ± 4	78 ± 3	84 ± 12	92 ± 20	76 ± 10
$NADP^+$	RP-LC	75 ± 4	101 ± 5	76 ± 5	66 ± 5	92 ± 16	61 ± 10
PEP	RP-LC	81 ± 7	96 ± 7	77 ± 4	105 ± 8	90 ± 15	95 ± 22
Phenylalanine	RP-LC	86 ± 7	100 ± 2	86 ± 8	92 ± 1	104 ± 5	95 ± 4
Phenylalanine	HILIC	146 ± 86	133 ± 102	133 ± 34	93 ± 4	101 ± 7	94 ± 5
Proline	RP-LC	81 ± 14	103 ± 5	83 ± 16	94 ± 6	106 ± 7	99 ± 12
Proline	HILIC	112 ± 38	118 ± 66	115 ± 29	91 ± 4	109 ± 13	99 ± 13
Serine	HILIC	122 ± 58	122 ± 70	118 ± 22	90 ± 6	105 ± 8	95 ± 10
Threonine	HILIC	103 ± 40	112 ± 53	99 ± 15	91 ± 2	101 ± 4	92 ± 4
Tyrosine	RP-LC	87 ± 7	99 ± 2	86 ± 8	92 ± 3	103 ± 3	94 ± 3
Tyrosine	HILIC	109 ± 39	120 ± 71	111 ± 26	90 ± 3	106 ± 3	95 ± 2
Valine	RP-LC	83 ± 11	101 ± 2	83 ± 10	91 ± 5	102 ± 8	93 ± 8
Valine	HILIC	114 ± 57	125 ± 82	110 ± 22	94 ± 4	103 ± 8	96 ± 8

(Eqs. (1)–(3)). Furthermore, internal standardization with the monoisotopic signal ($U^{12}C$) deriving from the cell sample was applied (Eqs. (4)–(7)) and recoveries without correction and $U^{12}C$ corrected recoveries were compared. A prerequisite for $U^{12}C$ correction was the availability of homogenous samples and knowledge of the amount of sampled biomass; both requirements were fulfilled in this study. In this way, assessed recoveries could be corrected for instrumental drifts and sample volume losses during sample preparation.

The resulting recoveries – of cell extraction, of evaporation, and of overall sample preparation – are summarized for all the investigated metabolites, without and with $U^{12}C$ correction in Table 5. Several amino acids are listed twice: amino acids comprising apolar side chains (phe, pro, tyr, val) were measured on RP-LC-MS/MS as well as on HILIC-MS/MS; and high abundant amino acids (arg, glu, gln, his, lys) showing signal to noise ratios above 10^3 (Table 4) were measured in an additional 1:10 dilution prior to analysis. As can be readily observed, the investigated metabolites showed excellent recoveries and repeatability precision over the sample preparation procedure. Even the extract treatment step, necessary for subsequent RP-LC-MS/MS measurement showed no significant influence. The values obtained by complementary HILIC and RP-LC-MS/MS analysis were in excellent agreement after internal standardization. Moreover, the relatively poor standard uncertainty obtained by HILIC-MS/MS confirmed the necessity of internal standardization. This regarded especially results obtained from not corrected peak areas measured via HILIC without a further dilution step (see also Table 4, poor precision of $U^{13}C$ areas obtained by HILIC). Higher matrix effects in HILIC compared to RP-LC were the cause. Only five metabolites showed overall recoveries ranging from 60 to 80%: arginine, citrate, lysine, NAD^+, and $NADP^+$. Nevertheless they may be extracted via boiling ethanol and quantified

Table 6. Characterization and quantification of uncertainty sources for sample preparation and quantification of intracellular 5′AMP

Input quantities	Values	RSD	Uncertainty	Unit	Type, distribution	SU	Rel. contr. to comb. unc.
$c_{released}$	0.16	15%[a)]	0.024	μmol g_{CDW}^{-1}	A, normal	0.024	35.4%
CDW	4.5	2.2%[b)]	0.099	g_{CDW} L^{-1}	A, normal	0.099	0.8%
V_{Sample}	0.002		$2\ 10^{-5}$ [c)]	L	B, triangular	$8.16\ 10^{-6}$	0.0%
n_{IS}	0.00156		$3.12\ 10^{-5}$ [c)]	μmol	B, triangular	$1.56\ 10^{-5}$	0.1%
L	1		0.02[d)]		B, rectangular	0.0115	0.2%
R	1.35		0.19		A, normal	0.19	31.1%
R_{IS}	1.35		0.19		A, normal	0.19	32.4%

a) Estimated from published data.
b) Based on the repeatability of the CDW determination, $n = 5$.
c) Uncertainties derived from the manufacturers specification of the pipettes.
d) Experiments in our lab addressing leakage showed that its contribution is negligible (≤2%) if quenching is optimized and cells are separated instantaneously.

correctly when adding the specific internal standard prior to extraction. It has to be underlined here that the sample preparation was not optimized for accurate and simultaneous measurement of redox pairs e.g. NAD/NAD^+ or NADP/$NADP^+$.

3.4 Calculation of sample preparation uncertainties for intracellular metabolite quantification in yeast

The accuracy of LC-MS/MS-based quantitative analysis in metabolomics, though rarely discussed, is a major issue. There is a complete lack of reference materials. Only recently, an overview of uncertainty of measurement in quantitative metabolomics including practical examples was discussed by Guerrasio et al. [13]. Uncertainty calculations are valuable tools for systematically addressing the total combined uncertainty of a quantification task, assessing the uncertainty of each step in the analytical process and its contribution to the total combined uncertainty. Different approaches can be used to calculate the uncertainty of a system: the GUM [24] provides a very detailed bottom-up approach to estimate the combined standard uncertainty of the method. The stepwise procedure starts with the definitions of the measurand (i.e. the particular quantity subject to measurement), of the input quantities, and of the model equation. In this way, all possible sources of uncertainty are identified and the standard uncertainties of each input quantity are evaluated. The value of the measurand and the combined standard uncertainty of the result are calculated, the expanded uncertainty is calculated using a selected coverage factor (k) and finally the result is reported with the expanded uncertainty. Furthermore the uncertainty contributions are analyzed.

In this work, the total combined uncertainty of sample preparation was estimated based on the experimental data obtained in the extraction recovery studies. Accordingly, the compiled model equation took into account the sources of uncertainty of the sample preparation but not of LC-MS/MS measurement. In Eq. (8), the measurand was defined as the ratio of the metabolite concentration versus concentration of the corresponding labeled metabolite in the obtained measurement solution ($c_{final}/c_{final\ IS}$). This ratio is given by the concentration released from the cell upon extraction ($c_{released}$), hence by the extraction efficiency, by the cell dry weight of the quenched cell suspension (CDW), by the volume of sampled quenched cell suspension (V_{Sample}), by the amount of added internal standard (n_{IS}), by leakage (L), and by the recoveries of the sample preparation procedure of the metabolite and the labeled metabolite (R/R_{IS}).

$$\frac{c_{final}}{c_{final\ IS}} = c_{released} \times CDW \times \frac{V_{Sample}}{n_{IS}} \times L \times \frac{R}{R_{IS}} \quad (8)$$

c_{final}	μmol L^{-1}	Final concentration of the metabolite in the measurement solution
$c_{final\ IS}$	μmol L^{-1}	Final concentration of the internal standard in the measurement solution
$c_{released}$	μmol g_{CDW}^{-1}	Intracellular concentration of the metabolite released by extraction
CDW	g_{CDW} L^{-1}	Cell dry weight of quenched cell suspension
V_{Sample}	L	Volume of quenched cell suspension, which is sampled
n_{IS}	μmol	Amount of internal standard spiked before extraction
L		Leakage factor
R		Recovery of sample preparation of the metabolite
R_{IS}		Recovery of sample preparation of the internal standard

Uncertainty budgeting was carried out using the example of three representative metabolites showing different extraction efficiencies and extraction recoveries, namely 5′AMP, citrate, and phenylalanine. The assumed values for $c_{released}$ are the expected intracellular concentrations of the metabolite released by the applied boiling ethanol extraction and the corresponding standard uncertainties (SU) are estimations of the extraction efficiencies of published data [15]. The

Table 7. Characterization and quantification of uncertainty sources for sample preparation and quantification of intracellular citrate

Input quantities	Values	RSD	Uncertainty	Unit	Type, distribution	SU	Rel. contr. to comb. unc.
$c_{released}$	4.35	5%[a)]	0.218	μmol g_{CDW}^{-1}	A, normal	0.218	70.3%
CDW	4.5	2.2%[b)]	0.099	g_{CDW} L^{-1}	A, normal	0.099	13.6%
V_{Sample}	0.002		$2\ 10^{-5}$ c)	L	B, triangular	$8.16\ 10^{-6}$	0.5%
n_{IS}	0.0235		$4.70\ 10^{-4}$ c)	μmol	B, triangular	$1.92\ 10^{-4}$	1.9%
L	1		0.02[d)]		B, rectangular	0.0115	3.7%
R	0.75		0.01		A, normal	0.01	5.0%
R_{IS}	0.75		0.01		A, normal	0.01	5.0%

a) Estimated from published data.
b) Based on the repeatability of the CDW determination, $n = 5$.
c) Uncertainties derived from the manufacturers specification of the pipettes.
d) Experiments in our lab addressing leakage showed that its contribution is negligible (≤2%) if quenching is optimized and cells are separated instantaneously.

assumed values of n_{IS} are the expected concentrations of labeled metabolite in the extract and the SU, as well as the SU of the quenched sample volume, derive from the manufacturers specification of the pipettes. The SU of cell dry weight determination is based on the repeatability of the CDW determination ($n = 5$). Values and SU of recoveries are taken from present experiment and leakage was assumed to be minimal when quenching conditions and contact time of cells and quenching solution is optimized. The identified sources of uncertainty which are present in the model equation are characterized and quantified for the three exemplary metabolites (Tables 6–8) and the resulting absolute and relative standard uncertainties for sample preparation for the three representative intracellular metabolites are summarized in Table 9. In order to illustrate the contribution of extraction efficiency and recovery, we selected exemplary metabolites with reportedly high uncertainty in extraction efficiency and assessed extraction recovery (5′AMP), low uncertainty in both efficiency and recovery (citrate) and moderate uncertainty in extraction efficiency and low uncertainty in recovery (phenylalanine). For 5′AMP a relative high standard uncertainty for extraction efficiency of approximately 15% was reported. Recovery SU was found to range at 20%. The obtained expanded total combined uncertainty of 50% ($k = 2$)

Table 8. Characterization and quantification of uncertainty sources for sample preparation and quantification of intracellular phenylalanine

Input quantities	Values	RSD	Uncertainty	Unit	Type, distribution	SU	Rel. contr. to comb. unc.
$c_{released}$	0.33	10%[a)]	0.033	μmol g_{CDW}^{-1}	A, normal	0.033	70.2%
CDW	4.5	2.2%[b)]	0.099	g_{CDW} L^{-1}	A, normal	0.099	3.4%
V_{Sample}	0.002		$2\ 10^{-5}$ c)	L	B, triangular	$8.16\ 10^{-6}$	0.1%
n_{IS}	0.00226		$4.52\ 10^{-5}$ c)	μmol	B, triangular	$1.85\ 10^{-5}$	0.5%
L	1		0.02[d)]		B, rectangular	0.0115	0.9%
R_{U12C}	0.95		0.04		A, normal	0.04	12.4%
R_{U13C}	0.95		0.04		A, normal	0.04	12.5%

a) Estimated from published data.
b) Based on the repeatability of the CDW determination, $n = 5$.
c) Uncertainties derived from the manufacturers specification of the pipettes.
d) Experiments in our lab addressing leakage showed that its contribution is negligible (≤2%) if quenching is optimized and cells are separated instantaneously.

Table 9. Uncertainties for sample preparation for three representative intracellular metabolites

Compound	C_{final} / $C_{final\ IS}$	Unit	Expanded uncertainty	Expand. unc. (relative)	Coverage factor (k)
5′AMP	0.92	μmol L^{-1}/μmol L^{-1}	0.47	50%	2.0
Citrate	1.67	μmol L^{-1}/μmol L^{-1}	0.20	12%	2.0
Phenylalanine	1.31	μmol L^{-1}/μmol L^{-1}	0.31	24%	2.0

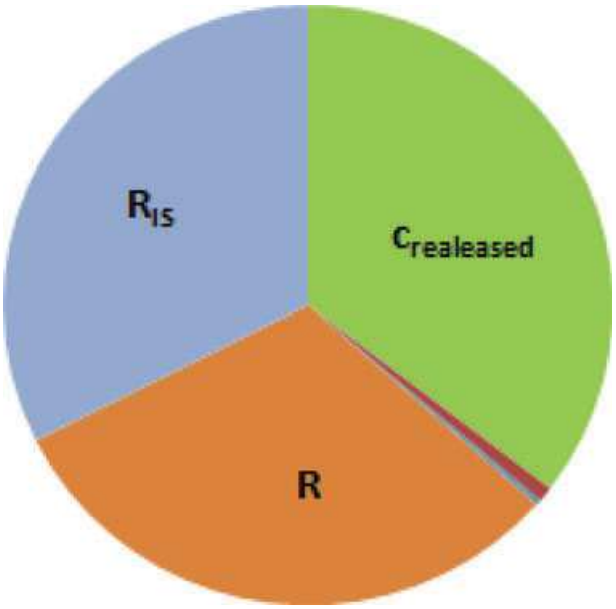

Figure 2. Relative contributions to combined standard uncertainty for 5′AMP.

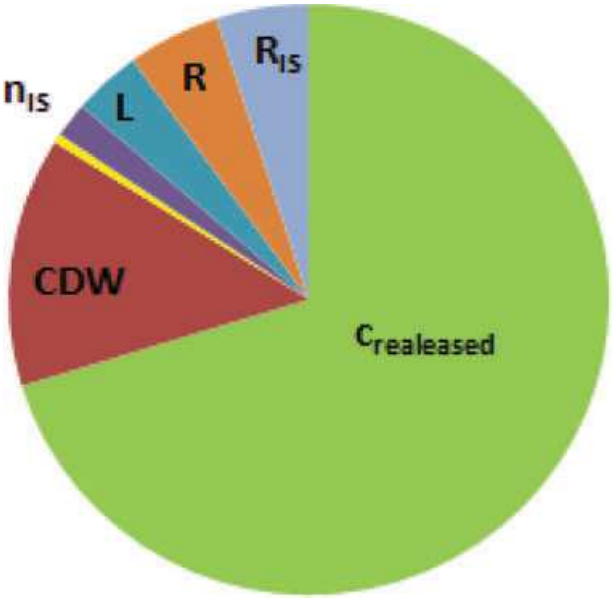

Figure 3. Relative contributions to combined standard uncertainty for citrate.

deriving from sample preparation procedure characterized AMP as a very critical compound in sample preparation. For citrate a SU of approximately 5% was reported for extraction efficiency and the recovery of sample preparation was found to be 75% but with high precision. Accordingly, the obtained expanded total combined uncertainty of 12% corresponded to the minimum contribution to be expected from sample preparation in quantitative metabolic profiling. Finally, approximately 10% SU in extraction efficiency was reported for phenylalanine. The experimentally assessed recovery of sample preparation was found to be 95% ± 4%. Uncertainty calculation revealed a total combined uncertainty of 24% ($k = 2$). This was considered as an average uncertainty stemming from sample in quantitative metabolic profiling. The relative contributions to total combined standard uncertainties are illustrated in Figs. 2–4. In all three cases, the major contribution derives from extraction efficiency ($c_{released}$) whereas recovery is critical for 5′AMP and phenylalanine.

Result (5′AMP): $$c_{final}/c_{final\ IS} = 0.92\ \frac{\mu mol_{metabolite} \times L^{-1}}{\mu mol_{IS} \times L^{-1}}$$

Expanded uncertainty (5′AMP, $k = 2$):

$$U = 0.47\ \frac{\mu mol_{metabolite} \times L^{-1}}{\mu mol_{IS} \times L^{-1}}$$

Result (citrate) : $$c_{final}/c_{final\ IS} = 1.67\ \frac{\mu mol_{metabolite} \times L^{-1}}{\mu mol_{IS} \times L^{-1}}$$

Expanded uncertainty (citrate, $k = 2$):

$$U = 0.20\ \frac{\mu mol_{metabolite} \times L^{-1}}{\mu mol_{IS} \times L^{-1}}$$

Result (phenylalanine) :

$$c_{final}/c_{final\ IS} = 1.31\ \frac{\mu mol_{metabolite} \times L^{-1}}{\mu mol_{IS} \times L^{-1}}$$

Expanded uncertainty (phenylalanine, $k = 2$):

$$U = 0.31\ \frac{\mu mol_{metabolite} \times L^{-1}}{\mu mol_{IS} \times L^{-1}}$$

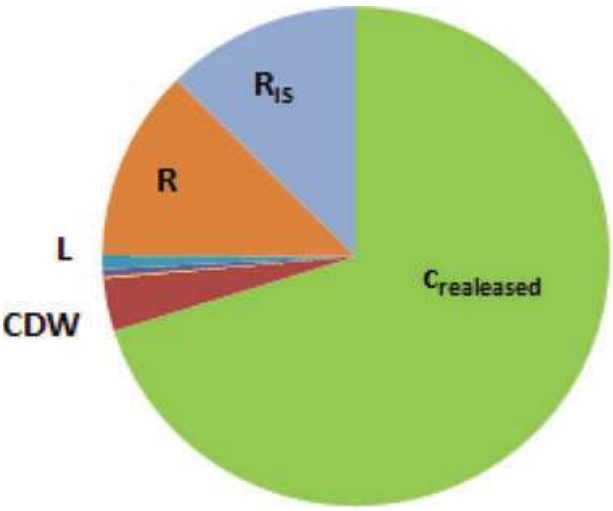

Figure 4. Relative contributions to combined standard uncertainty for phenylalanine.

4 Concluding remarks

The use of two orthogonal chromatographic modes (RP-LC, HILIC) and highly selective triple quadrupole MS enabled the measurement of a wide range of intracellular metabolites (amino acids, organic acids, nucleotides, and cofactors). At the stage of method implementation, $U^{13}C$-labeled cell extract was used as a tool for studying the recovery and standard uncertainty of each step in the sample preparation process. We obtained overall repeatability precisions less than 10% for most of the investigated metabolites using internal standardization approaches. Setting up model equations for uncertainty budgeting, error propagation for all input variables for sample preparation procedure, pinpointed the extraction efficiency, and the extraction recovery as the main sources contributing to total combined uncertainty.

This work has been supported by Austrian Research Promotion Agency FFG (FHplus projekt METORGANIC), Federal Ministry of Economy, Family, and Youth (BMWFJ), 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.

EQ BOKU VIBT GmbH is acknowledged for providing LC-MS/MS instrumentation.

The authors have declared no conflict of interest.

5 References

[1] Sauer, M., Mattanovich, D., *J. Chem. Technol. Biotechnol.* 2012, *87*, 445–450.

[2] Fiehn, O., *Comp. Funct. Genomics* 2001, *2*, 155–168.

[3] Fiehn, O., *Plant Mol. Biol.* 2002, *48*, 155–171.

[4] Villas-Bôas, S. G., Mas, S., Åkesson, M., Smedsgaard, J., Nielsen, J., *Mass Spectrom. Rev.* 2005, *24*, 613–646.

[5] Dettmer, K., Aronov, P. A., Hammock, B. D., *Mass Spectrom. Rev.* 2007, *26*, 51–58.

[6] Allwood, J. W., Ellis, D. I., Goodacre, R., *Physiol. Plantarum* 2008, *132*, 117–135.

[7] Mashego, M. R., Rumbold, K., de Mey, M., Vandamme, E., Soetaert, W., Heijnen, J., *Biotechnol. Lett.* 2007, *29*, 1–16.

[8] Canelas, A. B., Ras, C., ten Pierick, A., van Dam, J. C., Heijnen, J. J., van Gulik, W. M., *Metabolomics* 2008, *4*, 226–239.

[9] Schädel, F., Franco-Lara, E., *Appl. Microbiol. Biotechnol.* 2009, *83*, 199–208.

[10] Sellick, C. A., Hansen, R., Maqsood, A. R., Dunn, W. B., Stephens, G. M., Goodacre, R., Dickson, A. J., *Anal. Chem.* 2009, *81*, 174–183.

[11] Dettmer, K., Nürnberger, N., Kaspar, H., Gruber, M. A., Almstetter, M. F., Oefner, P. J., *Anal. Bioanal. Chem.* 2011, *399*, 1127–1139.

[12] Van Gulik, W., *Curr. Opin. Biotechnol.* 2010, *21*, 27–34.

[13] Guerrasio, R., Haberhauer-Troyer, C., Neubauer, S., Klavins, K., Werneth, M., Koellensperger, G., Hann, S., *Uncertainty of Measurement in Quantitative Metabolomics*, in: Lämmerhofer, M., Weckwerth, W., (Eds.) *Metabolomics in Practice: Successful Strategies to Generate and Analyze Metabolic Data*, Wiley, Weinheim, 2013 (forthcoming).

[14] Carnicer, M., Canelas, A. B., ten Pierick, A., Zeng, Z., van Dam, J., Albiol, J., Ferrer, P., Heijnen, J. J., van Gulik, W., *Metabolomics* 2011, *8*, 284–298.

[15] Canelas, A. B., ten Pierick, A., Ras, C., Seifar, R. M., van Dam, J. C., van Gulik, W. M., Heijnen, J. J., *Anal. Chem.* 2009, *81*, 7379–7389.

[16] Nasution, U., van Gulik, W. M., Kleijn, R. J., van Winden, W. A., Proell, A., Heijnen, J. J., *Biotechnol. Bioeng.* 2006, *94*, 159–166.

[17] Wittmann, C., Krömer, J. O., Kiefer, P., Binz, T., Heinzle, E., *Anal. Biochem.* 2004, *327*, 135–139.

[18] Sellick, C. A., Knight, D., Croxford, A. S., Maqsood, A. R., Stephens, G. M., Goodacre, R., Dickson, A. J., *Metabolomics* 2010, *6*, 427–438.

[19] Ritter, J. B., Genzel, Y., Reichl, U., *Anal. Biochem.* 2008, *373*, 349–369.

[20] Luo, B., Groenke, K., Takers, R., Wandrey, C., Oldiges, M., *J. Chromatogr. A* 2007, *1147*, 153–164.

[21] Coulier, L., Bas, R., Jespersen, S., Verheij, E., van der Werf, M. J., Hankemeier, T., *Anal. Chem.* 2006, *78*, 6573–5682.

[22] Jaitz, L., Mueller, B., Koellensperger, G., Huber, D., Oburger, E., Puschenreiter, M., Hann, S., *Anal. Bioanal. Chem.* 2011, *400*, 2587–2596.

[23] Ellison, S. L. R., Rosslein, M., Williams, A., *EURACHEM/CITAC Guide: Quantifying Uncertainty in Analytical Measurement*, 2nd Ed., EURACHEM, London 2000.

[24] *Guide to the Expression of Uncertainty in Measurement*, ISO/GUM, Geneva, Switzerland 1995.

[25] Kragten, J., *Analyst* 1994, *119*, 2161.

[26] Neubauer, S., Rugova, A., Dinh, B. C., Drexler, H., Ganner, A., Sauer, M., Mattanovich, D., Hann, S., Koellensperger, G., *Anal. Bioanal. Chem.* 2012, *404*, 799–808.

[27] Dell'mour, M., Jaitz, L., Oburger, E., Puschenreiter, M., Koellensperger, G., Hann, S., *J. Sep. Sci.* 2010, *33*, 1–12.

[28] Mashego, M. R., Wu, L., Van Dam, J. C., Ras, C., Vinke, J. L., Van Winden, W. A., Van Gulik, W. M., Heijnen, J. J., *Biotechnol. Bioeng.* 2004, *85*, 620–628.

Anal Bioanal Chem (2013) 405:5159–5169
DOI 10.1007/s00216-013-6964-4

RESEARCH PAPER

Interlaboratory comparison for quantitative primary metabolite profiling in *Pichia pastoris*

Kristaps Klavins · Stefan Neubauer · Ali Al Chalabi · Denise Sonntag · Christina Haberhauer-Troyer · Hannes Russmayer · Michael Sauer · Diethard Mattanovich · Stephan Hann · Gunda Koellensperger

Received: 8 February 2013 / Revised: 28 March 2013 / Accepted: 2 April 2013 / Published online: 19 April 2013

Abstract For the first time, an interlaboratory comparison was performed in the field of quantitative metabolite profiling in *Pichia pastoris*. The study was designed for the evaluation of different measurement platforms integrating different quantification strategies using internal standardization. Nineteen primary metabolites including amino acids and organic acids were selected for the study. Homogenous samples were obtained from chemostat fermentations after rapid sampling, quenching and filtration, and hot ethanol extraction. Laboratory 1 (BOKU) employed an in vivo-synthesized fully labeled $U^{13}C$ cell extracts of *P. pastoris* for immediate internal standardization upon cell extraction. Quantification was carried out using orthogonal reversed-phase (RP-LC) and hydrophilic interaction chromatography (HILIC) in combination with tandem mass spectrometry. Laboratory 2 (Biocrates) applied a metabolomics kit allowing fully automated, rapid derivatization, solid phase extraction and internal standardization in 96-well plates with immobilized isotopically enriched internal standards in combination with HILIC-MS-MS and RP-LC-MS-MS for organic acids and derivatized amino acids, respectively. In this study, the obtained intracellular concentrations ranged from 0.2 to 108 $\mu mol\,g^{-1}$ cell dry weight. The total combined uncertainty was estimated including uncertainty contributions from the corresponding MS-based measurement and sample preparation for each metabolite. Evidently, the uncertainty contribution of sample preparation was lower for the values obtained by laboratory 1, implementing isotope dilution upon extraction. Total combined uncertainties ($K=2$) ranging from 21 to 48 % and from 30 to 57 % were assessed for the quantitative results obtained in laboratories 1 and 2, respectively. The major contribution arose from sample preparation, hence from repeatability precision of the extraction procedure. Finally, the laboratory intercomparison was successful as most of the investigated metabolites showed concentration levels agreeing within their total combined uncertainty, implying that accurate quantification was given. The application of isotope dilution upon extraction was an absolute prerequisite for the quantification of the redox-sensitive amino acid methionine, where no agreement between the two laboratories could be achieved.

Keywords Metabolomics · Interlaboratory comparison · LC-MS/MS · *Pichia pastoris*

Published in the topical collection *Metabolomics and Metabolite Profiling* with guest editors Rainer Schuhmacher, Rudolf Krska, Roy Goodacre, and Wolfram Weckwerth.

Electronic supplementary material The online version of this article (doi:10.1007/s00216-013-6964-4) contains supplementary material, which is available to authorized users.

K. Klavins · C. Haberhauer-Troyer · H. Russmayer · M. Sauer · D. Mattanovich · S. Hann · G. Koellensperger (✉)
Austrian Centre of Industrial Biotechnology (ACIB), Muthgasse 11, 1190 Vienna, Austria
e-mail: gunda.koellensperger@boku.ac.at

K. Klavins · S. Neubauer · C. Haberhauer-Troyer · S. Hann · G. Koellensperger
Department of Chemistry, Division of Analytical Chemistry, University of Natural Resources and Life Sciences—BOKU Vienna, Muthgasse 18, 1190 Vienna, Austria

A. Al Chalabi · D. Sonntag
BIOCRATES Life Sciences AG, Eduard-Bodem-Gasse 8, 6020 Innsbruck, Austria

H. Russmayer · M. Sauer · D. Mattanovich
Department of Biotechnology, University of Natural Resources and Life Sciences—BOKU Vienna, Muthgasse 18, 1190 Vienna, Austria

Introduction

Targeted analytical strategies in metabolomics refer to either qualitative or quantitative investigation of predefined metabolites. The latter approach, denoted as quantitative metabolite profiling, is the backbone of a cutting-edge strategy in biotechnological production optimization, the metabolic engineering, ultimately aiming at engineering cell fabrics with increased productivity [1]. In this specific application, metabolites of the central carbon metabolism are of key interest since the production rates and concentrations of microorganisms are always linked to the primary metabolism, independent if primary or secondary metabolic products are in demand. Starting from quantitative metabolite profiles, kinetic information regarding metabolite production and consumption rates can be inferred [1]. In this way, bottlenecks in the metabolic reaction network can be identified serving as targets for metabolic engineering.

As comprehensively reviewed elsewhere [2], nowadays, the key analytical platform for metabolite profiling is mass spectrometry in combination with different chromatographic techniques [3–10] covering gas chromatography [9, 10], liquid chromatography [3, 6–8], and capillary electrophoresis [4, 5]. Among these, liquid chromatography was and still is the most versatile technique. As a matter of fact, primary intracellular metabolites show a high variation of chemical and physical properties (charge, polarity etc.); therefore, large coverage of metabolites often is accomplished following two main strategies: ion pairing chromatography or the implementation of at least two orthogonal separations (reversed-phase and hydrophilic interaction chromatography, HILIC).

Well-designed sampling, quenching, and extraction procedures are a prerequisite for successful quantitative metabolite profiling in cellular samples. Complex sample matrix, the high turnover rates of metabolites, and the instability of the extracted metabolites are the major challenges. Many efforts have been directed toward the development and optimization of cell leakage- and degradation-free protocols [11–13]. Optimized procedure for yeast including *Pichia pastoris* [12, 13], which was investigated in the presented analytical exercise here, consisted of (1) rapid sampling in combination with fast centrifugation or fast filtration methods, (2) quenching using cold methanol leading to minimal metabolite leakage in quenching and washing solutions, and (3) extraction using boiling ethanol [3, 11–13]. This hot extraction protocol demonstrated excellent recovery, extraction efficacy, and precision repeatability for a wide range of metabolites [11, 14].

Method validation is an integral part of any analytical method development; however, in this specific application, it is highly challenging to address the accuracy of the obtained quantitative values. As a matter of fact, there is a complete lack of reference materials. Even more, up to now, it is widely unexplored under which conditions and for which metabolites such material could be produced in the future.

The absolute amount of the metabolite, i.e., the best estimate of the true value, can only be obtained by investigating the efficiency and recovery of different extraction methods in a first step [15]. Several groups estimated accuracy by calculations of the mass action ratios from the obtained data and subsequent comparison with published equilibrium constants [12, 16]. Using the first approach, the procedure with the highest extraction efficiency and the best recovery and repeatability precision is then selected for quantification. Evidently, not only the accuracy concerning different sample preparation procedures have to be investigated but also the accuracy of the measurement procedure itself has to be proven. This is in the best case accomplished by a comparison of different analytical platforms either within one laboratory or in different laboratories.

While in other fields interlaboratory comparisons are well-established tools [17–20], this is not the case in targeted metabolomics. The few examples concerned not metabolite quantification but rather the assessment of differential metabolite profiles (relative measurement of different biological status) in order to compare the ability of different analytical platforms and/or laboratories to produce same data sets (metabolite features) by multivariate statistics [21]. Interlaboratory comparison regarding the relative quantification of amino acids using the CE-MS and GC-MS methods has been reported. Using both methods, relative peak areas were calculated employing the internal standard compound, ethionine. The obtained amino acid ratios were used to study cell culture response to stress [22].

In this work, to the best of our knowledge, the first attempt of interlaboratory comparison applied to quantitative metabolite profiling in yeast (*P. pastoris*) is presented. The sampling and the sample preparation were carried out with great care to ensure sample homogeneity. Extraction was performed using the boiling ethanol method that has been thoroughly studied for *P. pastoris* [12, 14]. The quantification of 19 primary metabolites consisting of amino acids and organic acids was studied in two different laboratories. The selection was based on the facts that these compounds were amenable to analysis with different methods in the two participating laboratories. Moreover, for all metabolites, extraction efficiency and recovery had been previously investigated. The requirement for the study was the highest extraction efficiency and stability upon the selected sample preparation procedure.

Experimental

Experiment design (sample collection and handling)

Three biological replicates of *P. pastoris* wild type were grown at BOKU. From each biological replicate, two

samples were obtained using the filtration method on two parallel filtration units, A and B, at exactly the same time after quenching. The sample codes are shown in Table 1. Sample extraction was performed using the boiling ethanol method at BOKU Division of Analytical Chemistry. Extracts were evaporated to complete dryness using the vacuum centrifuge at BOKU. Samples that were obtained with filtration unit A were analyzed according to the protocol established at the BOKU Division of Analytical Chemistry, implementing the internal standardization with $U^{13}C$-labeled yeast extract immediately after sampling. Samples from filtration unit B were sent to Biocrates. For these samples, immediate internal standardization was not possible due to the integrated internal standards on the Biocrates kit. Samples were shipped to Biocrates on dry ice; meanwhile, three samples from filtration unit A where stored in the same conditions as the shipped ones. Metabolite quantification was performed in both laboratories at the same time using two independent methods to exclude errors that can arise from the different sample storage times. Figure 1 shows the overall sample handling procedure.

Cultivation for cell sample

Three independent *P. pastoris* chemostat cultivations were performed as described elsewhere [14].

Sampling and quenching

Samples for the analysis of intracellular metabolites were taken using a peristaltic pump and silicone tubes (diameter, 5 mm; length, 81 cm) at a pumping speed of 5 mL/s. Approximately 50 mL of the fermentation broth was quenched in 200 mL of 60 % (*v*/*v*) methanol at −27 °C. After quenching, 2 mL of cell suspension (corresponding to approximately 10 mg biomass) was filtered using two filtration units (Polycarbonate Filter Holders, Satorius Lab Technologies Product) with cellulose acetate filter (0.45 μm, Satorius Biolab Products). Negative pressure was applied using a vacuum pump. The cells were washed once with cold 60 % (*v*/*v*) methanol and then the filter was kept on dry ice. Biomass was determined by drying five replicates of 2 mL chemostat culture to constant weight at 105 °C in pre-weight glass tubes.

Table 1 Sample codes

Sample code	Fermentation code	Filtration unit code	Participant
A1	M027	A	BOKU (laboratory 1)
A2	M032	A	BOKU (laboratory 1)
A3	M035	A	BOKU (laboratory 1)
B1	M027	B	Biocrates (laboratory 2)
B2	M032	B	Biocrates (laboratory 2)
B3	M035	B	Biocrates (laboratory 2)

Extraction of intracellular metabolites

Quenched and washed cell pellets of *P. Pastoris* on filters were kept in 15-mL tubes on dry ice. Of the internal standard, 200 μL was added to samples A1–A3 and 200 μL of water was added to samples B1–B3 in order to ensure the same sample volume. Four milliliters of boiling ethanol (75 %, *v*/*v*) was poured onto the cell pellets in the tube. The cell pellets were completely resuspended by vortexing for approx. 20 s. The tubes containing the samples were put into the water bath set to 85 °C and heated for 3 min in total. It was vortexed for approx. 10 s after 1.5 min of heating and again for approx. 10 s after a total of 3 min of heating. The hot tube containing the extracted cell pellet was put directly into dry ice for 3 min for rapid cooldown. After cooling, the tube was put back on dry ice. Afterwards, the tubes were centrifuged at 4,000×*g* for 10 min at −20 °C. The supernatant was decanted into a pre-cooled 15-mL tube. The ethanolic extracts were stored on dry ice until they were evaporated to complete dryness in a vacuum centrifuge (Savant RVT400 from Thermo Scientific) operating at pressures below 1 mbar.

Preparation of $U^{13}C$ internal standard

Detailed description of $U^{13}C$ internal standard preparation procedure is available elsewhere [14]. In brief, the yeast cells grown on fully labeled $U^{13}C$ glucose in the fed batch cultivation were quenched and sampled using cold methanol. Metabolite extraction using boiling ethanol method was carried out and the obtained extracts were evaporated and reconstituted with water.

Quantitative analysis performed at laboratory 1 (BOKU) with immediate internal standardization

Sample preparation for LC-MS/MS analysis

Of the LC-MS grade water, 1,000 μL was added to the dried residue using a 1,000-μL piston pipette and disposal pipette tips (Eppendorf). The dried residue was then resuspended in the following steps: (1) vortexing, (2) vortexing again after 10 min, and (3) suspending by drawing–pushing and transferring to the reaction tube, 2 mL (Eppendorf test tube or equivalent), using the piston pipette. Insoluble particles were removed via centrifugation at 4,000×*g* for 10 min at 5 °C using a table centrifuge from Hettich (Tuttlingen, Germany). For HILIC measurement, one additional dilution step of 1:10 (*v*/*v*) was carried out using LC-MS acetonitrile.The clear reconstituted ethanolic extracts were transferred to HPLC glass vials.

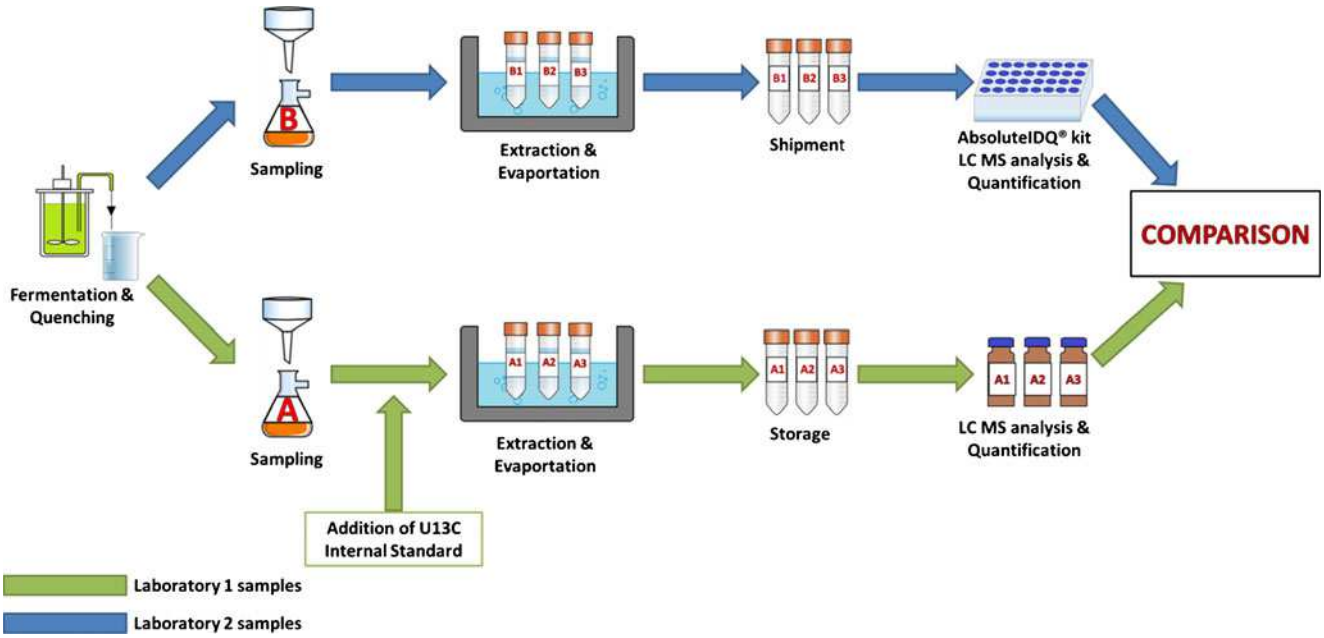

Fig. 1 Sample handling procedure

LC MS/MS analysis

The quantitative analysis of amino acids and organic acids was performed using reversed-phase chromatography coupled with tandem mass spectrometric detection. The analysis was carried out on the LC-MS/MS system consisting of a Thermo Scientific CTC PAL autosampler, Thermo Scientific Accela 1259 pump, and Thermo Scientific TSQ Vantage ESI-MS/MS. Separation was performed on an Atlantis T3® analytical column (150× 4.6 mm, 3-μm particle size, 100-Å pore size) equipped with an Atlantis T3 guard column (20×4.6, 3-μm particle size; Waters, Milford, MA) with eluent A (water, 0.1 % (*v*/*v*) formic acid) and eluent B (methanol). The following gradient was applied in LC-MS/MS: 0 % B was constant for 2 min and then was increased to 40 % within 8 min and was held for 2 min. A subsequent increase to 100 % within 0.1 min and holding for 1.9 min to flush the column, followed by reconstitution of the starting conditions within 0.1 min and re-equilibration with 0 % B for 5.9 min, resulted in a total analysis time of 20 min. For MS detection, TSQ Vantage tandem mass spectrometer from Thermo Scientific was used, featuring a heated ESI interface. The ion source parameters for the positive and negative modes were set as follows: vaporizer temperature, 350 °C; ion transfer tube temperature, 350 °C; aux gas pressure, 15 arbitrary units; sheath gas pressure, 40 arbitrary units; ion sweep gas pressure, 0 arbitrary unit; declustering voltage, 0 V; and spray voltage values for positive polarity and negative polarity, 3,300 and 3,000 V (respectively); collision gas pressure for multiple reaction monitoring (MRM) was set to 1.5 mTorr. MRM transitions of all compounds have been determined via flow injection of 10 μM single standards using a syringe pump coupled to the LC pump using a zero volume T-piece connector. XCalibur tune software was used to optimize the MRM transitions for each compound. The precursor ion, product ion, as well as collision energy values and polarity are listed in Table 2. Quantitative analysis of highly abundant amino acids was performed using HILIC coupled with tandem mass spectrometric detection. An Agilent G1312A Binary Pump 1200 series from Agilent Technologies (Waldbronn, Germany), together with an Agilent G1367B high-performance autosampler and an Agilent G1316A column compartment, was employed for HPLC. For MS detection, an Agilent 6410 Triple Quad LC/MS (Agilent Technologies) was used featuring an ESI interface. Separation was carried out on a ZicHILIC© analytical column (150×4.6 mm, 3.5-μm particle size, 100-Å pore size) from SeQuant (Marl, Germany) and a ZicHILIC© guard column (20×2.1 mm, 5-μm particle size), with eluent A (98 % (*v*/*v*) water, 1 % (*v*/*v*) acetonitrile (CAN), and 1 % (*v*/*v*) formic acid) and eluent B (98 % (*v*/*v*) ACN, 1 % (*v*/*v*) water, and 1 % (*v*/*v*) formic acid) applying the following gradient: 90 % B was constant for 2 min and then was reduced to 10 % within 7 min and was held for 1 min. Subsequent reconstitution of the starting conditions within 0.1 min and re-equilibration with 90 % B for 9.9 min resulted in a total analysis time of 20 min. A flow rate of 0.6 mL min^{-1}, an injection volume of 3 μL, and a column temperature of 40 °C were applied. Mass spectrometer source parameters in positive ionization mode were set as follows: drying gas temperature, 300 °C; drying gas flow,

Table 2 Measurement conditions of LC-MS/MS analysis for all investigated metabolites

Compound	Polarity	HPLC	Dilution	Precursor ion, *m/z*	Product ion, *m/z*	FV (V)[a]	CE (eV)	Comp. for IS
$U^{13}C$ fumarate	−	RP	1:1	119	74		9	–
$U^{12}C$ fumarate	−	RP	1:1	115	71.1		9	$U^{13}C$ fumarate
$U^{13}C$ isoleucine	+	RP	1:1	138	91		9	–
$U^{12}C$ isoleucine	+	RP	1:1	132.1	86		9	$U^{13}C$ isoleucine
$U^{13}C$ leucine	+	RP	1:1	138	91		9	–
$U^{12}C$ leucine	+	RP	1:1	132.1	86		9	$U^{13}C$ leucine
$U^{13}C$ methionine	+	RP	1:1	155	59		13	
$U^{12}C$ methionine	+	RP	1:1	150.1	56.1		13	$U^{13}C$ methionine
$U^{13}C$ phenylalanine	+	RP	1:1	175	128		9	–
$U^{12}C$ phenylalanine	+	RP	1:1	166.1	120		9	$U^{13}C$ phenylalanine
$U^{13}C$ proline	+	RP	1:1	121	74		15	–
$U^{12}C$ proline	+	RP	1:1	116	70		15	$U^{13}C$ proline
$U^{13}C$ succinate	−	RP	1:1	121	76		9	
$U^{12}C$ succinate	−	RP	1:1	117	73.1		9	$U^{13}C$ succinate
$U^{13}C$ tyrosine	+	RP	1:1	191	144		9	–
$U^{12}C$ tyrosine	+	RP	1:1	182.1	135.9		9	$U^{13}C$ tyrosine
$U^{13}C$ valine	+	RP	1:1	123	76		9	–
$U^{12}C$ valine	+	RP	1:1	118.1	72		9	$U^{13}C$ valine
$U^{13}C$ alanine	+	HILIC	1:10	93	46	40	9	–
$U^{12}C$ alanine	+	HILIC	1:10	90	44	40	9	$U^{13}C$ alanine
$U^{13}C$ arginine	+	HILIC	1:10	181	74	80	25	–
$U^{12}C$ arginine	+	HILIC	1:10	175	70	80	25	$U^{13}C$ arginine
$U^{13}C$ asparagine	+	HILIC	1:10	137	76	60	13	–
$U^{12}C$ asparagine	+	HILIC	1:10	133	74	60	13	$U^{13}C$ asparagine
$U^{13}C$ aspartate	+	HILIC	1:10	138	76	60	9	–
$U^{12}C$ aspartate	+	HILIC	1:10	134	74	60	9	$U^{13}C$ aspartate
$U^{13}C$ glutamate	+	HILIC	1:10	153	88	70	13	–
$U^{12}C$ glutamate	+	HILIC	1:10	148	84	70	13	$U^{13}C$ glutamate
$U^{13}C$ glutamine	+	HILIC	1:10	152	88	80	17	–
$U^{12}C$ glutamine	+	HILIC	1:10	147	84	80	17	$U^{13}C$ glutamine
$U^{13}C$ histidine	+	HILIC	1:10	162	115	80	13	–
$U^{12}C$ histidine	+	HILIC	1:10	156	110	80	13	$U^{13}C$ histidine
$U^{13}C$ lysine	+	HILIC	1:10	153	89	60	13	–
$U^{12}C$ lysine	+	HILIC	1:10	147	84	60	13	$U^{13}C$ lysine
$U^{13}C$ serine	+	HILIC	1:10	109	62	60	9	–
$U^{12}C$ serine	+	HILIC	1:10	106	60	60	9	$U^{13}C$ serine
$U^{13}C$ threonine	+	HILIC	1:10	124	77	60	9	–
$U^{12}C$ threonine	+	HILIC	1:10	120	74	60	9	$U^{13}C$ threonine

For MRM transition, collision energy (CE) and fragmenter voltage (FV) are given. Precursor ions and product ions are shown for the monoisotopic molecules deriving from the standard or cell sample ($U^{12}C$) and for the ones deriving from the fully labeled cell extract ($U^{13}C$) used for internal standardization

[a] Only for Agilent 6410 Triple Quad

10 L min^{-1}; nebulizer pressure, 25 psi; and capillary voltage, 4,000 V. MRM transitions were determined via flow injection of 20 μM single standard (isocratic conditions: 0.05 % formic acid in MeOH/H_2O 50:50, *v/v*). For optimization, the Mass Hunter Optimizer Software (Agilent) was applied. The precursor and product ions as well as specific values for fragmentor voltage and collision energy are listed in Table 2. Data processing was performed using XCalibur Quan Browser software (Thermo Scientific) and Agilent MassHunter (Agilent Technologies). Quantification was

performed using a six-point calibration curve with internal standardization. Internal standardization of the calibration solutions were carried out with the same amount of internal standard as for the samples. Internal standardization was performed using the in-house-produced $U^{13}C$ cell extract.

Quantitative analysis performed at laboratory 2 (Biocrates) with internal standardization prior measurements

Sample preparation for LC-MS/MS analysis

Dried frozen extracts of *P. pastoris* cells stored on dry ice were provided by BOKU. According to the agreement, samples were redissolved in 1,000 μL Milli-Q water (Milli-Q Synthesis, Millipore, Molsheim, France), shaken for 20 min at 900 rpm, and placed in an ultrasonic bath on ice for 10 min. The resulting suspensions were centrifuged (5 min, 2,900 rpm, 2 °C) and the supernatant was transferred to another vial. The pellet was discarded. Clear supernatants were diluted 1:5 (*v*/*v*) and 1:10 (*v*/*v*) with Milli-Q water.

LC-MS/MS analysis

Amino acids were quantitatively analyzed using reversed-phase LC-MS/MS to obtain chromatographic separation of isobaric (the same MRM ion pairs) metabolites for individual quantification performed by external calibration and by the use of internal standards. A 10-μL sample volume (processed sample) is required for the analysis using the following sample preparation procedure. Samples were added on filter spots placed in a 96-well Solvinert plate (internal standards were placed and dried down under nitrogen before) and fixed above a 96-deep well plate (capture plate). Twenty microliters of 5 % phenyl–isothiocyanate derivatization reagent was added. The derivatized samples were extracted after incubation by 5 mM ammonium acetate in methanol into the capture plate. Ten-microliter sample extracts were analyzed using LC-ESI-MS/MS in positive MRM detection mode with a 4000 Q Trap® tandem mass spectrometry instrument (AB Sciex, Darmstadt, Germany). Chromatographic separation was performed using an Agilent Zorbax Eclipse XDB C18 column (Agilent Technologies) at a flow rate of 500 μL/min. Mobile phase A consisted of 0.2 % formic acid in water and mobile phase B consisted of 0.2 % formic acid in acetonitrile. A linear gradient from 0 to 95 % B over 5 min was applied. For the quantitative analysis of organic acids, a HILIC-ESI-MS/MS method in highly selective negative MRM detection mode was used. MRM detection was performed using a 4000 Q Trap® tandem mass spectrometry instrument (AB Sciex). Twenty microliters sample volume (processed sample) was protein-precipitated and extracted simultaneously with 80 % methanol (*v*/*v*) in a 96-well plate format. Finally, 20 μL sample extract was used as the injection volume for HILIC-ESI-MS/MS analysis. Chromatographic separation was achieved using an Atlantis HILIC Silica column (Waters GmbH, Eschborn, Germany) at a flow rate of 500 μL/min. Mobile phase A consisted of water/acetonitrile (95:5, *v*/*v*) containing 15 mM ammonium acetate and mobile phase B consisted of 95:5 (*v*/*v*) acetonitrile/water containing 15 mM ammonium acetate. A linear gradient from 7 to 100 % mobile phase A over 10 min was applied. Internal standards (ratio of external to internal standard) and external calibration were used for accurate quantitation. LC-MS/MS data were processed with Analyst 1.4.2 software (AB Sciex). All methods have been validated for human plasma considering FDA Guidance for Industry—Bioanalytical Method Validation [23]. Metabolite concentrations were determined in micromolar units. The concentrations of all independent measurements per sample ($N=6$) were recalculated according to the dilution factor and averaged. Subsequently, in relation to the initial cell concentration in each strain and the fermentation process, the mean values were converted into micromoles per gram cell.

Results and discussion

Design of interlaboratory comparison

The interlaboratory comparison was designed for the comparison of different measurement platforms integrating different quantification strategies for internal standardization. Based on a thoroughly studied sample preparation protocol—with known sample extraction efficiencies and recoveries for the metabolites of the carbon cycle and on the offered analytical methods in the two participating laboratories—19 compounds including amino acids and organic acids were selected for the exercise. For these compounds, Canelas et al. [11] had previously shown that hot ethanol extraction delivered the highest extraction efficiency. Moreover, the sample preparation recovery of metabolites using hot ethanol extraction was assessed in laboratory 1 (BOKU) using in vivo-synthesized $U^{13}C$ yeast extract as the tracer [14]. For the selected metabolite panel, extraction recoveries from 66 to 115 % were assessed. Therefore, for these compounds, both strategies of internal standardization, i.e., before extraction—as established in laboratory 1 (BOKU)—and before measurement—as established in laboratory 2 (Biocrates), could be applied. The different time points of internal standardization had practical reasons. Laboratory 2 applied the metabolomics kit, which is an own analytical development. The kit provided an automated workflow integrating immobilized isotopically enriched metabolite standards and amino acid derivatization and solid phase extraction. In this study, it was used in combination with LC-MS-MS analysis; however, targeted analysis of

biogenic amines, acylcarnitines, phosphatidylcholines, lysophosphatidylcholines, and sphingomyelins in combination with flow injection MS analysis would be possible. Laboratory 1 performed LC-MS-MS analysis implementing orthogonal reversed-phase and hydrophilic interaction chromatography [14]. As internal standard, the in vivo-synthesized $U^{13}C$ yeast extract was used. Using this strategy, potential metabolite loss during sample preparation, storage, and measurement fluctuation was compensated. The time between extraction and measurement was kept as short as possible. The limits of detection (LODs) of both methods are given in Table 3. In the case of BOKU, LODs were calculated using the 3σ criteria of the baseline signal calibrated by the peak height of the lowest calibration point; for Biocrates, LODs were calculated within the measurements with water-based zero samples. Moreover, Table 3 summarizes the sample preparation recoveries for the compounds under investigation.

LC-MS quantification using in vivo-synthesized $U^{13}C$ internal standard (laboratory 1)

Table 4 gives an overview of the quantitative measurements of three biological replicates using internal standardization with the $U^{13}C$-labeled yeast cell extract. As a prerequisite in this study, basic measurement criteria for the in vivo-synthesized internal standard in the resulting yeast extracts were set for all investigated metabolites: a signal-to-noise ratio of $U^{13}C$ metabolite peaks >10 and a signal-to-noise ratio for $U^{12}C$ metabolite impurity peaks <3. As a matter of fact, these criteria were fulfilled for all compounds that were used for interlaboratory comparison. Standard solutions, spiked with the $U^{13}C$ cell extract, i.e., the internal standard (added at the same concentration as in the samples), served as quality control (QC) samples. The measurement of QC samples ($n=5$) was evenly distributed through the measurement sequence (roughly amounting to 20 % of each sequence). It is commonly accepted that QC samples monitor the overall performance of instruments and set criteria for acceptable measurement deviation (relative standard deviation, RSD). As a rule of thumb, the RSD for the repeated sample injections should not exceed the RSD of QC samples. In this study, the RSD values for samples were in the range of 0.5–10 % and were, thus, in good agreement with the QC samples. Typically, standard uncertainties of 5 % deriving from repeated measurements were reported for LC-MS-based quantitative metabolite profiling [8, 24]. Next, the obtained standard uncertainty for the $N=3$ LC-MS determinations were compared to the total combined uncertainties calculated for the LC-MS quantification in order to evaluate the performance of a given LC-MS method. This was considered as an additional test of whether the developed method was under control. Estimation of the total combined uncertainty (TCU) was performed according to the ISO GUM [25]. Calculations were carried out using Kragten's [26] approach via a spreadsheet. The model equation defining the uncertainty

Table 3 LODs for both applied methods given in solution (in micromolars) and sample (in micromoles per gram CDW) and metabolite sample preparation recovery [14]

Compound	Laboratory 1 LOD		Laboratory 2 LOD		Recovery (%) [14]
	µM	µmol $gCDW^{-1}$	µM	µmol $gCDW^{-1}$	
Fumarate	0.07	0.008	4.0	0.45	–
Isoleucine	0.01	0.001	0.73	0.08	98±8
Leucine	0.01	0.001	1.5	0.17	94±7
Methionine	0.03	0.003	0.19	0.02	–
Phenylalanine	0.003	0.000	0.16	0.02	95±4
Proline	0.005	0.001	1.3	0.15	99±12
Succinate	0.38	0.043	5.4	0.61	–
Tyrosine	0.005	0.001	1.6	0.18	95±2
Valine	0.02	0.002	0.28	0.03	93±8
Alanine	0.13	0.015	1.3	0.15	115±18
Arginine	0.12	0.013	0.2	0.02	78±2
Aspargine	0.18	0.020	1.1	0.12	94±4
Aspartate	0.15	0.017	0.21	0.02	96±5
Glutamate	0.29	0.033	0.33	0.04	94±5
Glutamine	0.3	0.034	0.08	0.009	83±6
Histidine	0.09	0.010	0.82	0.09	85±8
Lysine	0.08	0.009	0.28	0.03	67±2
Serine	0.08	0.009	2.6	0.29	95±10
Threonine	0.08	0.009	1.0	0.11	92±4

budget (Eq. 1) gave the metabolite concentration as an extractable intracellular metabolite fraction (in micromolars) per gram cell dry weight (gCDW). R_S is the chromatographic response of the sample that was calculated according to Eq. 2, where A_{Sample} is the peak area of the metabolite in the sample and A_{IS} is the peak area of the corresponding $U^{13}C$-labeled compound. The intercept B_0 and the slope B_1 of the calibration curve were calculated using linear regression (see Eqs. 3 and 4). V_{vial} is the volume of the measured extract; V_{Sample} is the volume of the quenched cell suspension that was sampled. Additionally, the dilution factor of the cell extract (d) and the cell dry weight of the quenched cell suspension (CDW) were taken into account. The experimentally assessed uncertainties for all input quantities are summarized in Electronic supplementary material Tables S1 and S2.

$$C_{measured} = \frac{R_S - B_0}{B_1} \times \frac{V_{vial} \times d}{V_{Sample} \times CDW} \quad (1)$$

$$R_S = \frac{A_{sample}}{A_{IS}} \quad (2)$$

$$B_0 = \frac{\overline{R} \cdot \sum_i C_i^2 - \overline{C} \times \sum_i C_i \times R_{Std\,i}}{\sum_i C_i^2 - n \cdot \overline{C^2}} \quad (3)$$

$$B_1 = \frac{\sum_i C_i \times R_{Std\,i} - n \times \overline{R} \times \overline{C}}{\sum_i C_i^2 - n \times \overline{C^2}} \quad (4)$$

For almost all measurements calculated, the TCU of the LC-MS measurement results was greater than the experimentally assessed measurement repeatability precision RSD, implying good measurement quality and method control. Even though the LODs of an applied reversed-phase-based method were lower than the ones for a HILIC-based method, the measurement deviation (RSD) and the total combined uncertainty (TCU) of LC-MS measurements were comparable for both methods, indicating that TCU was independent of measurement sensitivity.

Comparability of the quantitative measurements

Table 5 summarizes the outcome of the laboratory intercomparison, giving the average values obtained from three biological replicates analyzed in each of the two participating laboratories and the corresponding biological repeatability precision (RSD). Moreover, the total combined uncertainty of the quantitative results for each metabolite obtained in each laboratory was estimated, including uncertainty contributions from sample preparation. Equation 5 gives the simplified model used in this total combined uncertainty calculation. C1, C2, and C3 correspond to single

Table 4 Intracellular metabolite concentrations (IC) and calculated TCU values of LC-MS measurement results for each biological replicate obtained at laboratory 1

Metabolite	IC (μmol $gCDW^{-1}$)			RSD, N=3 (%)			QC RSD, N=5 (%)	TCU, K=2 (%)
	A1	A2	A3	A1	A2	A3		
Fumarate	1.0	1.2	1.5	4	4	1	4	17
Isoleucine	0.31	0.36	0.33	5	4	4	4	8
Leucine	0.60	0.52	0.63	5	5	1	4	7
Methionine	7.3	6.6	7.5	1	4	3	7	4
Phenylalanine	0.19	0.19	0.21	1	0.3	1	1	3
Proline	23	22	24	1	1	3	2	3
Sucinate	1.5	1.4	1.5	9	2	1	8	9
Tyrosine	0.26	0.24	0.29	7	6	4	3	12
Valine	1.2	0.98	1.3	1	1	3	4	6
Alanine	8.5	10	14	2	5	4	5	6
Arginine	51	75	74	2	1	1	1	2
Aspargine	3.2	3.3	4.3	2	2	10	4	9
Aspartate	17	16	19	4	3	3	4	5
Glutamate	78	76	97	0.4	1	1	2	2
Glutamine	95	106	123	2	1	0.2	3	2
Histidine	3.9	4.7	5.2	1	1	2	3	5
Lysine	4.2	7.1	6.8	5	3	1	3	4
Serine	2.9	5.4	5.8	4	4	5	7	9
Threonine	1.4	2.2	2.7	3	5	7	5	14

Relative standard deviations for repeated injections (N=3) and QC sample are given

Table 5 Results of the performed interlaboratory comparison

Metabolites	Laboratory 1				Laboratory 2			
	Average IC, N=3 (μmol $gCDW^{-1}$)	RSD, N=3 (%)	$u(K_{sp})$ (%)	TCU, K=2 (%)	Average IC, N=3 (μmol $gCDW^{-1}$)	RSD, N=3 (%)	$u(K_{sp})$ (%)	TCU, K=2 (%)
Fumarate	1.2	22	25	33	1.1	12	25	52
Isoleucine	0.33	8	8	37	0.39	11	4	36
Leucine	0.58	10	7	35	0.69	26	4	34
Methionine	7.1	6	20	48	0.42	18	25	55
Phenylalanine	0.2	6	4	25	0.27	6	9	30
Proline	23	5	12	32	36	4	19	45
Succinate	1.5	4	25	45	2.7	4	25	57
Tyrosine	0.26	10	2	21	0.29	29	23	51
Valine	1.2	14	9	39	0.9	14	12	44
Alanine	11	28	16	37	12	13	21	50
Arginine	67	21	3	28	46	4	7	31
Asparagine	3.6	17	4	28	5.5	10	19	47
Aspartate	17	11	5	27	20	9	13	36
Glutamate	84	14	5	25	104	24	15	38
Glutamine	108	13	7	27	96	12	21	50
Histidine	4.6	14	9	27	4.9	10	14	39
Lysine	6	26	3	27	4.3	9	8	30
Serine	4.7	34	11	45	6.9	4	19	57
Threonine	2.1	31	4	30	2.3	5	15	40

Average values for intracellular metabolite concentration (IC) in yeast cell extract and TCU obtained from biological replicates (N=3) in both laboratories are shown for all metabolites. Uncertainty of sample preparation ($u(K_{sp})$) and RSD used for TCU assessment are given

concentration values obtained for biological replicates, which are associated with the experimentally assessed repeatability precision. Moreover, an additional factor K_{sp} was introduced accounting for all uncertainties from sample preparation. Accordingly, the K_{sp} value was 1, with the uncertainty derived from sampling and sample preparation procedures. In order to assess this uncertainty, additional experiments were carried out using 15 yeast cell pellets (biological replicates) spiked with the $U^{13}C$ internal standard. For all metabolites, the monoisotopic $U^{12}C$ and $U^{13}C$ signals were measured with the analytical tool set offered by laboratory 1. The recovery of the applied sample preparation was calculated using the $U^{13}C/U^{12}C$ ratio [14], compensating for volume losses. Hence, this experimentally obtained repeatability precision of the metabolite recovery resembled the standard uncertainty of the sample preparation (K_{sp}) in the uncertainty model for laboratory 1. This approach was not valid for laboratory 2 as the volume and metabolite losses during sample preparation were not compensated by an internal standard in this case. Here, the standard uncertainty of the sample preparation and, hence, the uncertainty associated with K_{sp} in the model was deduced from the standard uncertainty (N=15) of the monoisotopic $U^{12}C$ signals. Values for $u(K_{sp})$ of each metabolite are shown in Table 5.

$$C_{average} = \frac{C1 + C2 + C3}{3} \times K_{sp} \quad (5)$$

Accordingly, as can be readily observed in Table 5, TCU values for the results obtained by immediate internal standardization were generally lower compared to the ones obtained with internal standardization prior to LC-MS analysis, especially highly abundant compounds. Hence, not surprisingly, this once again confirms the importance of immediate internal standardization in the field of quantitative metabolite profiling. Moreover, the laboratory intercomparison was successful as most of the investigated metabolites showed concentration levels agreeing within their total combined uncertainty. Hence, interlaboratory reproducibility could be achieved for the first time for metabolite profiling in yeast. At the same time, the agreement implied that accurate quantification was given for those metabolites. It has to be mentioned that for all compounds, the experimentally assessed repeatability was less than than the calculated TCU, implying that the applied model considers all contributions of the total combined uncertainty.

However, it has to be kept in mind that in this uncertainty calculation, sample storage uncertainty was not considered as the study was very tightly organized with minimized time spans between sample preparation and measurement (1 day). This becomes evident only in the case of methionine, known to be a highly unstable compound prone to oxidative degradation. The instability of methionine during sample and standard storage has been comprehensively examined [27, 28]. It was suggested that methionine reacted with the polystyrene and polypropylene surface inside the storage vessel [27]. Accordingly, the results obtained in the two laboratories displayed huge differences: 7.1 μM μmol $gCDW^{-1}$ in the case of laboratory 1 and 0.42 μmol $gCDW^{-1}$ in the case of laboratory 2. Finally, the intercomparison showed that sample shipment is feasible for most of the studied compounds. This situation could be significantly improved in future laboratory intercomparison studies if immediate internal standardization upon extraction was implemented.

Conclusion

In general, it is agreed that such a complex task as quantification metabolite profiling in yeast will not provide total combined uncertainties comparable to other quantitative methods using isotope dilution strategies. The interlaboratory comparison performed here demonstrated that the accurate assessment of metabolic amino acids and organic acids in yeast was feasible. From the obtained results, it can be concluded that, in order to increase the quality of interlaboratory comparison, immediate internal standardization should be applied in all participating laboratories. In several previous studies on the primary metabolome, a minimum twofold difference between the metabolite levels was considered as biologically significant [29–31]. Considering the overall aim of quantitative metabolic profiling, which was to expose those metabolites that show significant biological difference, it could be concluded that the implemented methodological tool sets in both laboratories were fit for the purpose.

Acknowledgments This work has been financially supported by the FHplus Program of the Austrian Research Promotion Agency FFG, Project METORGANIC. Furthermore, this work has been supported by the Federal Ministry of Economy, Family and Youth (BMWFJ), 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. EQ BOKU VIBT GmbH is acknowledged for providing LC-MS/MS instrumentation.

References

1. Buchholz A, Hurlebaus J, Wandrey C, Takors R (2002) Metabolomics: quantification of intracellular metabolite dynamics. Biomol Eng 19:5–15
2. Theodoridis G, Gika H, Want E, Wilson I (2012) Liquid chromatography–mass spectrometry based global metabolite profiling: a review. Anal Chem Acta 711:7–16
3. Wu L, Mashego MR, van Dam JC, Proell AM, Vinke JL, Ras C, van Winden WA, van Gulik WM, Heijnen JJ (2005) Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly ^{13}C-labeled cell extracts as internal standards. Anal Biochem 336:164–171
4. Soga T, Ueno Y, Naraoka H, Ohashi Y, Tomita M, Nishioka T (2002) Simultaneous determination of anionic intermediates for *Bacillus subtilis* metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. Anal Chem 74:2233–2239
5. Sato S, Yanagisawa S (2010) Capillary electrophoresis–electrospray ionization–mass spectrometry using fused-silica capillaries to profile anionic metabolites. Metabolomics 6:529–540
6. Antonio C, Larson T, Gilday A, Graham I, Bergstrom E, Thomas-Oates J (2008) Hydrophilic interaction chromatography/electrospray mass spectrometry analysis of carbohydrate-related metabolites from *Arabidopsis thaliana* leaf tissue. Rapid Commun Mass Spectrom 22:1399–1407
7. Lu W, Clasquin M, Melamud E, Amador-Noguez D, Caudy A, Rabinowitz J (2010) Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer. Anal Chem 82:3212–3221
8. Luo B, Groenke K, Takors R, Wandrey C, Oldiges M (2007) Simultaneous determination of multiple intracellular metabolites in glycolysis, pentose phosphate pathway and tricarboxylic acid cycle by liquid chromatography–mass spectrometry. J Chromatogr A 1147:153–164
9. Vielhauer O, Zakhartsev M, Horn T, Takors R, Reuss M (2011) Simplified absolute metabolite quantification by gas chromatography–isotope dilution mass spectrometry on the basis of commercially available source material. J Chromatogr B 879:3859–3870
10. Kanani H, Chrysanthopoulos PK, Klapa MI (2008) Standardizing GC-MS metabolomics. J Chromatogr B 871:191–201
11. Canelas AB, Pierick A, Ras C, Seifar CR, van Dam J, van Gulik W, Heijnen J (2009) Quantitative evaluation of intercellular metabolite extraction techniques for yeast metabolomics. Anal Chem 81:7379–7389
12. Carnicer M, Canelas AB, Ten Pierick A, Zeng Z, van Dam J, Albiol J, Ferrer P, Heijnen JJ, van Gulik W (2012) Development of quantitative metabolomics for *Pichia pastoris*. Metabolomics 8:284–298
13. Canelas AB, Ras C, Pierick A, Dam JC, Heijnen JJ, Gulik WM (2008) Leakage-free rapid quenching technique for yeast metabolomics. Metabolomics 4:226–239
14. Neubauer S, Haberhauer-Troyer C, Klavins K, Russmayer H, Steiger M, Gasser B, Sauer MG, Mattanovich D, Hann S, Koellensperger G (2012) $U^{13}C$ cell extract of *Pichia pastoris*—a powerful tool for evaluation of sample preparation in metabolomics. J Sep Sci 35:3091–3105
15. Cipollina C, ten Pierick A, Canelas AB, Seifar RM, van Maris AJ, van Dam JC, Heijnen JJ (2009) A comprehensive method for the quantification of the non-oxidative pentose phosphate pathway intermediates in *Saccharomyces cerevisiae* by GC-IDMS. J Chromatogr B 877:3231–3236
16. Taymaz-Nikerel H, de Mey M, Ras C, ten Pierick A, Seifar RM, van Dam JC, Heijnen JJ, van Gulik WM (2009) Development and

application of a differential method for reliable metabolome analysis in *Escherichia coli*. Anal Biochem 386:9–19
17. Hintikka L, Kuuranne T, Leinonen A, Thevis M, Schänzer W, Halket J, Cowan D, Grosse J, Hemmersbach P, Nielen M, Kostiainen R (2008) Liquid chromatographic–mass spectrometric analysis of glucuronide–conjugated anabolic steroid metabolites: method validation and interlaboratory comparison. J Mass Spectrom 43:965–973
18. Farre M, Petrovic M, Gros M, Kosjek T, Martinez E, Heath E, Osvald P, Loos R, Le Menach K, Budzinski H, De Alencastro F, Muller J, Knepper T, Fink G, Ternes TA, Zuccato E, Kormali P, Gans O, Rodil R, Quintana JB, Pastori F, Gentili A, Barcelo D (2008) First interlaboratory exercise on non-steroidal anti-inflammatory drugs analysis in environmental samples. Talanta 76:580–590
19. Pereira E, Rodrigues SM, Otero M, Válega M, Lopes CB, Pato P, Coelho JP, Lillebø AI, Duarte AC, Pardal MA, Rocha R (2008) Evaluation of an interlaboratory proficiency-testing exercise for total mercury in environmental samples of soils, sediments and fish tissue. Trends Anal Chem 27:959–970
20. Bordet F, Inthavong D, Fremy JM (2002) Interlaboratory study of a multiresidue gas chromatographic method for determination of organochlorine and pyrethroid pesticides and polychlorobiphenyls in milk, fish, eggs, and beef fat. J AOAC Int 85:1398–1409
21. Allwood JW, Erban A, Koning S, Dunn WB, Luedemann A, Lommen A, Kay L, Löscher R, Kopka J, Goodacre R (2009) Inter-laboratory reproducibility of fast gas chromatography–electron impact–time of flight mass spectrometry (GC-EI-TOF/MS) based plant metabolomics. Metabolomics 5:479–496
22. Williams BJ, Cameron CJ, Workman R, Broeckling CD, Sumner LW, Smith JT (2007) Amino acid profiling in plant cell cultures: an inter-laboratory comparison of CE-MS and GC-MS. Electrophoresis 28:1371–1379
23. U.S. Department of Health and Human Services, Food and Drug Administration (2001) Guidance for industry: bioanalytical method validation. http://www.fda.gov/downloads/Drugs/…/Guidances/ucm070107.pdf. Accessed 10 August 2012
24. Gika HG, Theodoridis GA, Vrhovsek U, Mattivi F (2012) Quantitative profiling of polar primary metabolites using hydrophilic interaction ultrahigh performance liquid chromatography–tandem mass spectrometry. J Chromatogr A 1259:121–127
25. ISO/GUM (1995) Guide to the expression of uncertainty in measurement. ISO/GUM, Geneva, Switzerland
26. Kragten J (1994) Calculating standard deviations and confidence intervals with a universally applicable spreadsheet technique. Analyst 119:2161–2165
27. Chace DH, Luo Z, Jesus V, Haynes CA, Hannon WH (2010) Potential loss of methionine following extended storage of newborn screening sample prepared for tandem mass spectrometry analysis. Clin Chim Acta 441:1284–1286
28. Hustad S, Eussen S, Midttun O, Ulvik A, Kant PM, Morkrid L, Gislefoss R, Ueland PM (2012) Kinetic modeling of storage effects on biomarkers related to B vitamin status and one-carbon metabolism. Clin Chem 58:402–410
29. Tian J, Shi C, Gao P, Yuan K, Yang D, Lu X, Xu G (2008) Phenotype differentiation of three *E. coli* strains by GC-FID and GC-MS based metabolomics. J Chromatogr B 871:220–226
30. Plassmeier J, Barsch A, Persicke M, Niehaus K, Kalinowski J (2007) Investigation of central carbon metabolism and the 2-methylcitrate cycle in *Corynebacterium glutamicum* by metabolic profiling using gas chromatography–mass spectrometry. J Biotechnol 130:354–363
31. Yang S, Sadilek M, Synovec RE, Lidstrom ME (2009) Liquid chromatography–tandem quadrupole mass spectrometry and comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry measurement of targeted metabolites of *Methylobacterium extorquens* AM1 grown on two different carbon sources. J Chromatogr A 1216:3280–3289

Metabolic Engineering 24 (2014) 129–138

Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/ymben

Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production

Justyna Nocon [a], Matthias G. Steiger [a,d], Martin Pfeffer [a,1], Seung Bum Sohn [b], Tae Yong Kim [b], Michael Maurer [c,d], Hannes Rußmayer [a,d], Stefan Pflügl [a,d], Magnus Ask [a,d], Christina Haberhauer-Troyer [d,e], Karin Ortmayr [e], Stephan Hann [d,e], Gunda Koellensperger [d,e], Brigitte Gasser [a,d], Sang Yup Lee [b,f], Diethard Mattanovich [a,d,*]

[a] Department of Biotechnology, University of Natural Resources and Life Sciences Vienna, Muthgasse 18, Vienna, Austria
[b] Bioinformatics Research Center, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea
[c] School of Bioengineering, University of Applied Sciences FH Campus Wien, Vienna, Austria
[d] Austrian Centre of Industrial Biotechnology, Vienna, Austria
[e] Department of Chemistry, University of Natural Resources and Life Sciences, Vienna, Austria
[f] Metabolic Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering (BK21 plus program), BioProcess Engineering Research Center, Center for Systems and Synthetic Biotechnology, KAIST Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea

ARTICLE INFO

Article history:
Received 20 January 2014
Received in revised form
9 May 2014
Accepted 12 May 2014
Available online 20 May 2014

Keywords:
Metabolic modeling
GEM
Recombinant protein
Yeast
Flux analysis
NADPH

ABSTRACT

The production of recombinant proteins is frequently enhanced at the levels of transcription, codon usage, protein folding and secretion. Overproduction of heterologous proteins, however, also directly affects the primary metabolism of the producing cells. By incorporation of the production of a heterologous protein into a genome scale metabolic model of the yeast *Pichia pastoris*, the effects of overproduction were simulated and gene targets for deletion or overexpression for enhanced productivity were predicted. Overexpression targets were localized in the pentose phosphate pathway and the TCA cycle, while knockout targets were found in several branch points of glycolysis. Five out of 9 tested targets led to an enhanced production of cytosolic human superoxide dismutase (hSOD). Expression of bacterial β-glucuronidase could be enhanced as well by most of the same genetic modifications. Beneficial mutations were mainly related to reduction of the NADP/H pool and the deletion of fermentative pathways. Overexpression of the hSOD gene itself had a strong impact on intracellular fluxes, most of which changed in the same direction as predicted by the model. *In vivo* fluxes changed in the same direction as predicted to improve hSOD production. Genome scale metabolic modeling is shown to predict overexpression and deletion mutants which enhance recombinant protein production with high accuracy.

1. Introduction

Heterologous protein production is a multi-billion dollar market, mainly covering biopharmaceuticals and industrial enzymes. Microbial production systems, like bacteria and yeasts, are in many cases characterized by high synthesis rates and product titers (Porro et al., 2011). Nevertheless, there are still limits in productivity observed. In particular, a negative impact of recombinant protein production on cell growth is observed which has to be attributed to more than just a stoichiometric drain of energy from biomass formation towards product formation. In yeast systems, these limitations are ascribed to bottlenecks in protein folding and secretion in case of secretory proteins (recently reviewed by Idiris et al., (2010) and Damasceno et al., (2012), limited transcriptional efficiency due to weak promoters or low gene copy numbers (reviewed in Gasser et al., 2013) or to intracellular proteolytic degradation (Pfeffer et al., 2011). Additionally, metabolic limitations cannot be ruled out as a cause for suboptimal productivity of heterologous proteins (Heyland et al., 2011a; Heyland et al., 2011b; Kaleta et al., 2013; Klein et al., 2014; Kazemi Seresht et al., 2013). To address this problem, metabolic modeling can be applied for

* Corresponding author at: Department of Biotechnology, University of Natural Resources and Life Sciences, Muthgasse 18, Vienna, Austria.
E-mail address: diethard.mattanovich@boku.ac.at (D. Mattanovich).
[1] Current address: ERBER AG, Tulln, Austria.

http://dx.doi.org/10.1016/j.ymben.2014.05.011

both computational simulation of metabolic flux networks and experimental flux analyses. These analyses are then employed to predict targets for metabolic engineering to enhance productivity (Xu et al., 2013).

Vijayasankaran et al., (2005) showed that the amino acid composition of heterologous proteins has a profound impact on the predicted elementary flux modes leading to high protein production in *Escherichia coli*. Oddone et al., (2009) have used a dynamic flux balance analysis model to predict targets for gene downregulation to enhance production of recombinant green fluorescent protein (GFP) in *Lactococcus lactis*. Knock down of two of the predicted genes led to a 15% increase of GFP per cell mass. Furthermore, the process of protein synthesis is highly energy demanding. Consequently higher intracellular ATP levels should enhance the cellular capacity to produce recombinant proteins, as it was shown for *E. coli* (Kim et al., 2012).

Overexpression of a recombinant (in this case homologous) enzyme in *Aspergillus niger* led to significant changes in metabolic flux distribution, mainly enhancing reactions towards NADPH production and reducing the tricarboxylic acid (TCA) cycle flux (Driouch et al., 2012). Comparing the *in vivo* flux distribution with theoretical flux distributions calculated by elementary mode analysis revealed that the metabolism was already shifted to a large extent towards optimum flux distribution just by overproduction of the recombinant protein. The same pattern of flux changes was also observed for *Aspergillus oryzae* overproducing α-amylase (Pedersen et al., 1999).

The yeast *Pichia pastoris* is well established as a host for the production of heterologous proteins (Gasser et al., 2013). Recombinant overexpression of heterologous proteins in *P. pastoris* was shown to increase TCA cycle flux and ATP production slightly (Dragosits et al., 2009; Heyland et al., 2010). *De novo* synthesis of amino acids, in particular the synthesis of energetically costly ones, was identified as a limiting metabolic process for protein overproduction (Heyland et al., 2011b).

Genome sequencing of *P. pastoris* (Mattanovich et al., 2009; De Schutter et al., 2009) laid the basis for establishing genome-scale metabolic models (GEM) of this yeast (Sohn et al., 2010; Chung et al., 2010; Caspeta et al., 2012). The incorporation of heterologous protein production into the metabolic model allowed the investigation of the interplay among protein production, energy demand and biomass formation (Sohn et al., 2010). In the present work we aimed to verify predicted flux changes caused by protein overproduction by ^{13}C labeling based flux analysis, and to employ the GEM to predict beneficial mutations in the *P. pastoris* central carbon metabolism. The model was used to predict gene knock out targets by Minimization of Metabolic Adjustment (MOMA) and overexpression targets by Flux Scanning based on Enforced Objective Function (FSEOF) for enhanced production of intracellular human copper/zinc superoxide dismutase (hSOD). We present the effects of manipulating central metabolic fluxes on the production of hSOD in *P. pastoris*.

2. Materials and methods

2.1. Strains and vectors

The *P. pastoris* strains used in this study were based on the wild type strain X-33 (Invitrogen). The strain producing intracellular human superoxide dismutase under control of the strong glycolytic glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter (X-33_hSOD_NTS), carrying 15 copies of the hSOD gene was described by Marx et al. (2009). For production of intracellular β-glucuronidase A (GusA) from *E. coli* (Jefferson et al., 1986), the respective gene was cloned into the *P. pastoris* Puzzle vector (Stadlmayr et al., 2010) and integrated into the *P. pastoris* genome into the *AOX1* terminator locus.

E. coli strain top 10 (invitrogen) was used as a DNA manipulation host.

2.2. Cultivation of P. pastoris

For metabolic flux analysis cells were grown at 25 °C in 50 mL of YNB medium (3.4 g L^{-1} YNB w/o amino acids and ammonia sulfate, 10 g L^{-1} $(NH_4)_2SO_4$, 400 mg L^{-1} biotin, 20 g L^{-1} glucose) in 250 mL wide neck shake flasks without baffles. For ^{13}C labeling glucose was a mixture of 17% uniformly labeled ^{13}C glucose and 83% naturally labeled glucose. The cells were inoculated at an OD = 0.03 and grown until exponential phase (OD around 1). Glucose uptake and extracellular metabolites were determined in cultures grown on YNB with unlabeled glucose.

For determination of hSOD production, strains were cultivated at 25 °C in 10 mL YPD medium supplemented with 2 mM $CuCl_2$ and 0.02 mM $ZnSO_4$ in 100 mL wide neck shake flasks without baffles. The cultures were grown for 48 h and fed three times in 12 h intervals with 100 μL of 50 g L^{-1} glucose solution. GusA production was determined in cultures grown in minimal medium as described by Gasser et al. (2013), cultivated as above.

Optical density was measured at 600 nm wavelength in 1 mL of culture broth using a WPA CO8000 Cell Density Meter. For determination of yeast dry mass (YDM) 9 mL of culture broth was centrifuged, washed with 10 mL of ddH_2O and dried for 24 h at 105° in preweighed tubes. YDM was then related to OD to calculate biomass concentrations. Unless described differently all cultures were repeated in triplicate.

2.3. Recombinant protein quantification

At the end of the cultivation, 1 mL aliquots of cells were harvested and wet cell weight was determined. For further analysis the cell pellets were kept at −20 °C. The harvested cell pellets were re-suspended in 500 μL extraction buffer (20 mM Tris–HCl pH 8.2, 5 mM EDTA, 0.1% Triton X-100, 7 mM β-mercaptoethanol, 1 mM $CuCl_2$, 0.1 mM $ZnSO_4$ and protease inhibitor cocktail (Sigma)) and mechanically disrupted with 500 μL glass beads (diameter 0.5 mm) on a FastPrep® in 3 cycles of 20 s at 6.5 m s^{-1} and 5 min rest on ice. The hSOD concentration in the cell extracts was determined by ELISA (Marx et al., 2009) and correlated with total protein content (determined by Coomassie Protein Assay, Thermo Scientific).

For GusA determination the cell pellets were re-suspended in an extraction buffer (100 mM sodium phosphate buffer pH 7.0, 2 mM EDTA, 0.02% Triton X-100, 10 M DTT) and mechanically disrupted with glass beads as described above. The GusA concentration in the cell extracts was determined by 4-methylumbelliferyl-β-d-glucuronide (MUG) activity assay as described by Blumhoff et al. (2013).

2.4. Metabolic flux analysis

^{13}C labeling patterns of protein-bound amino acids were analyzed with GC–MS according to Zamboni et al. (2009) using an Agilent 6890N gas chromatograph coupled to a 5975 mass spectrometer equipped with a Phenomenex Zebron ZB-5MS column (30 m × 0.25 mm i.d. × 0.25 μm film thickness). In order to improve the selectivity of histidine detection the original temperature program was modified as follows: 160 °C for 1 min, with 20 °C/min to 250 °C, with 5 °C/min to 270 °C, with 30 °C/min to 310 °C, hold for 0.5 min.

Specific growth rates, glucose uptake rates and metabolite secretion rates were determined during the exponential growth phase between 17 and 23 h after inoculation.

Flux calculations were performed with OpenFLUX using standard settings and applying the gradient based search algorithm for sensitivity analysis (Quek et al., 2009). The stoichiometric model (provided in the supplementary file) was based on a previously published *P. pastoris* model of the central carbon metabolism (Baumann et al., 2010). Uptake and secretion rates of extracellular metabolites were measured by HPLC. The biomass composition of *P. pastoris* wild type strain X-33 and recombinant protein expressing strains were determined previously. For the hSOD-expressing strain the composition as for the Fab-expressing strain was assumed (Carnicer et al., 2009). Mass distribution vectors of amino acids were extracted from GC–MS raw data using FiatFlux (Zamboni et al., 2005). Mass distribution values are provided in the supplementary file.

2.5. Analysis of intracellular $NADP^+$ and NADPH

The protocol for extraction of $NADP^+$ and NADPH from cell pellets was adapted from the works of Canelas et al., (2009). One milliliter samples of 24 h grown cultures were centrifuged and the cell pellets were extracted in 1 mL of 5 mM ammonium acetate (pH 8.0) for 3 min at 85 °C with intermediate mixing. After centrifugation, supernatants were separated on a silica-based C18 column (Waters Atlantis T3, 2.1 × 150 mm^2, 3 μm). The HPLC system (Thermo Accela 1250) was coupled to a triple quadrupole MS system (Thermo TSQ Vantage) via a heated ESI ion source for quantification. As described by Ortmayr et al., (2014) an absolute quantification of NADPH is limited by its rapid oxidation and the lack of an isotopically labeled standard to compensate for losses during sample preparation. However relative quantification can be achieved reproducibly. NADPH was also measured in the same samples with the enzymatic cycling assay as described by Ask et al. (2013) showing the same relative NADPH levels.

2.6. Prediction of knockout and overexpression targets

Identification of gene overexpression targets for increasing recombinant protein production was performed using Flux Scanning based on Enforced Objective Function (FSEOF) (Choi et al., 2010) using the GEM of Sohn et al. (2010). Gene targets that were selected showed increased flux through their respective metabolic reactions when an increased flux towards protein production was enforced as an additional constraint to the metabolic model.

Identification of gene knockouts was performed with Minimization of Metabolic Adjustment (MOMA) (Segrè et al., 2002). Six gene knockout targets were selected that showed minor decrease in biomass production and an increase in specific protein synthesis.

2.7. Overexpression of target genes

The target genes for overexpression were amplified from X-33 genomic DNA and cloned under control of the GAP promoter into the Puzzle vector (Stadlmayr et al., 2010) containing a HphMX (hygromycin) resistance cassette. Primers are given in the supplementary file. The vectors were integrated into the *AOX1* terminator locus of the *P. pastoris* genome.

2.8. Knock out of target genes

The genes predicted as knock out targets were disrupted using the split marker cassette method established recently for *P. pastoris* (Heiss et al., 2013). Thereby, two fragments of 700 bp, one located around 200 bp upstream of the start codon and one located around 200 bp downstream of the start codon of the respective genes, were used to flank the HphMX resistance cassette. Upon transformation, the homologous recombination event replaces a 5′ fragment of the gene and its promoter with the antibiotic resistance. To increase the gene-targeting efficiency, the resistance gene is split into two overlapping fragments, which can only integrate into the host genome upon a successful homologous recombination event. For the verification of gene knockouts, genomic DNA of the modified strains was isolated on FTA™ cards (Whatmann) and used for PCR with primers located outside of the split marker cassette, where positive transformants give an approximately 1000 bp larger fragment compared to the wild type. All primers are given in the supplementary file.

2.9. Enzyme activity assays

Strains overexpressing glucose-6-phosphate dehydrogenase (*ZWF1*) and malate dehydrogenase (*MDH1*) were screened for the respective enzyme activity. The glucose-6-phosphate dehydrogenase assay was performed according to Souza et al. (2002). Briefly, cells were mechanically disrupted with glass beads, and the cell extracts were added to the reaction buffer (100 mM Tris–HCl pH = 7.5) containing 3 mM $MgCl_2$, 0.2 mM $NADP^+$ and 4 mM glucose-6-phosphate. Reduction of $NADP^+$ was followed by the increase of absorbance at 340 nm in 30 s intervals over 6 min using a spectrophotometer.

Malate dehydrogenase activity was measured according to (McAlister-Henn and Thompson, 1987) by addition of the cell extracts to reaction buffer (100 mM imidazol, 200 mM KCl, 2 mM EDTA, 10 mM $MgSO_4 \cdot 7H_2O$) containing 0.3 mM NADH and 100 mM oxaloacetate. The oxidation of NADH corresponding to decrease in the absorbance at 340 nm was followed in 30 s intervals for 6 min.

2.10. Analysis of external metabolites

To measure external metabolites the strains were cultivated as for ^{13}C labeling in liquid cultures with a starting OD = 0.1 until a stationary phase was reached. Sampling was performed every 12 h. At each time point optical density of the culture was measured and supernatant was collected for measurement of external metabolites. Glucose, ethanol, arabitol, acetate, acetaldehyde and pyruvate were quantified by HPLC analysis, as described by Pflügl et al. (2012) using a Rezex ROA-Organic Acid H^+ 300 mm × 7.8 mm column (Phenomenex, USA).

3. Results and discussion

3.1. Effect of hSOD overexpression on predicted and measured metabolic fluxes

In a previous study we correlated enhanced hSOD production with high gene copy number of hSOD expression cassettes. The best hSOD strain produced 4 mg g^{-1} hSOD per dry biomass, reaching 272 mg L^{-1} hSOD in fed batch (Marx et al., 2009). However, overproduction of hSOD led to a decrease of biomass yield to 75% of the wild type strain, yielding lower biomass formation in strains with higher hSOD productivity, as observed for other heterologous proteins in different hosts too. Significant intracellular flux changes between wild-type and hSOD strain were predicted by the GEM. To verify these flux changes experimentally, we performed flux analysis of the hSOD strain and the X-33 control strain based on ^{13}C labeling in unlimited batch cultures (which resemble the simulations of the GEM). Metabolic fluxes of both strains are shown in Fig. 1. In the wild type strain, the split ratio between glycolysis and PPP was 0.72 to 0.28. From the pyruvate node, 5% are directed to the fermentative pathway,

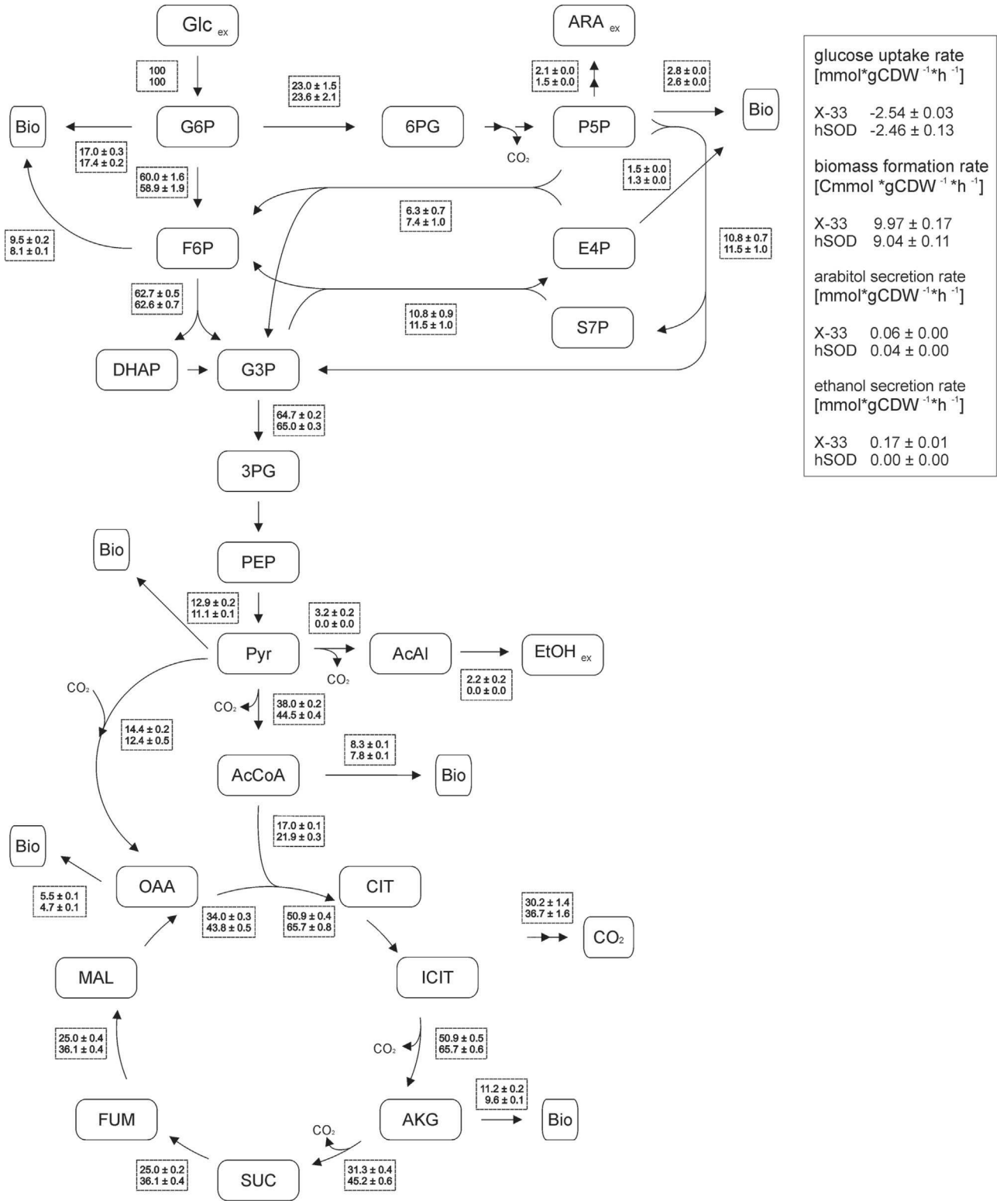

Fig. 1. Flux distribution in strains X-33 and hSOD. The flux values are normalized to glucose uptake and presented in [% Cmol]. The upper value in the rectangular boxes represents the flux distribution of strain X-33 and the lower value the flux distribution of hSOD strain. For reversible reactions only the net fluxes are presented.

75% to the TCA cycle and 20% to biomass. The PPP flux in the hSOD strain was slightly higher than in the wild type strain (although statistically not significant). TCA flux increased by 29%, and the fermentative branch through pyruvate decarboxylase decreased below the detection limit in the hSOD strain.

The production of hSOD leads to an increase of the total NADP/H pool as well as the NADPH concentration, so that the anabolic reduction charge is increased almost threefold (Table 1). Protein production requires anabolic reductive power for the synthesis of amino acids. Apparently the extra demand is compensated in the hSOD strain by the observed changes in the NADP/H pool.

The predicted flux changes of hSOD *vs.* X-33 strain showed mainly the same trends as the actually measured flux changes.

Table 1
Total NADP/H pools and relative anabolic reduction charge of strains X-33, hSOD, ZWF1 and SOL3. SEM=standard error of the mean. FC=fold change.

Strain	NADP/H pool (μmol/g YDM)	SEM	Relative reduction charge (FC to X-33)	SEM
X-33	0.38	0.025	1.0	0.52
SOD	0.55	0.040	2.9	0.64
ZWF1	0.49	0.002	3.1	0.03
SOL3	0.58	0.029	1.6	0.30

In the hSOD strain, mainly the TCA flux increases, while PPP and glycolytic fluxes remain almost unchanged, and the fermentative branches to ethanol and arabitol decrease in the production strain. While directions of flux changes are well predicted the model underestimates the change in TCA flux. The GEM predicted a higher PPP flux in the hSOD strain indicating a higher demand of NADPH in this strain. The PPP flux increased only slightly in the hSOD strain, but the total NADP/H concentration and the fraction of the reduced form increased in this strain. Driouch et al. (2012) have observed a similar overlap of predicted flux changes upon overproduction of recombinant fructofuranosidase in *A. niger*. They employed elementary flux mode calculation to predict fluxes, and found a 63% overlap of measured flux changes with the prediction. An increase of PPP flux was observed, while TCA cycle flux decreased in the *A. niger* overproduction strain. A very similar pattern was also observed with *A. oryzae* producing recombinant α-amylase (Pedersen et al., 1999), with *Schizosaccharomyces pombe* overproducing maltase (Klein et al., 2014) and with *S. cerevisiae* producing human SOD (Gonzalez et al., 2003). An increase of TCA cycle flux was observed here and consistently before in *P. pastoris* strains producing different recombinant proteins (Dragosits et al., 2009; Heyland et al., 2011b; Jordà et al., 2012). A literature review reveals an increased TCA flux attributed to higher energy demand also in *E. coli* (Weber et al., 2002) and CHO cells (Sheikholeslami et al., 2013) during recombinant protein production. In contrast, a decreased TCA flux was also observed in *E. coli* (Wittmann et al., 2007). Especially the latter case shows that this effect is not only host species dependent but also related to the produced protein. While Klein et al. (2014) describe a downregulation of TCA cycle flux in recombinant *S. pombe*, they observed an increase of recombinant maltase production when TCA cycle was enforced by a substrate change. We conclude that increased energy production by enhanced TCA cycle flux is an important prerequisite for recombinant protein production. However, the ability to shift the metabolism towards enhanced energy production depends obviously both on the host organism and specific features of the produced protein and the production strain.

3.2. Prediction of overexpression and knockout targets

The GEM by Sohn et al. (2010) was used for prediction of metabolic engineering targets beneficial to hSOD production. Nine metabolic reactions were predicted by FSEOF to enhance hSOD productivity upon their overexpression (Table 1). They include 6 consecutive steps of the pentose phosphate pathway (PPP) (*ZWF1*, *SOL3*, *GND2*, *RPE1*, *TKL1* and *TAL1*), two genes related to the TCA cycle (*MDH1* and *GDH3*), and *GPD1* of the glycerol pathway. To visualize the affected pathways all these reactions are indicated in Fig. 2.

Prediction of beneficial single gene knockouts with MOMA (Fig. 3) led to 6 targets (Table 2), mainly related to the fermentative pathway downstream of the pyruvate nodes (*PDC1*, *ADH2*, *ALD4* and *PDA1*), to glycolysis (*TPI1*) and to the glycerol pathway (*GUT2*). Most of these knockouts may increase the pyruvate pool and will thus lead to an increase of the TCA cycle flux (with the exception of *PDA1*). *GUT2* knockout may increase the glycerol pool by reducing its degradation. These proposed knockouts are visually presented in Fig. 2.

The predicted overexpressions in the upper part of the PPP and of *GDH3* are directly related to the reduction of $NADP^+$. Also a potential deletion of *TPI1* should enhance the oxidative part of the PPP and thus NADPH levels (Krüger et al., 2011). *De novo* synthesis of proteins costs both energy and reduction equivalents

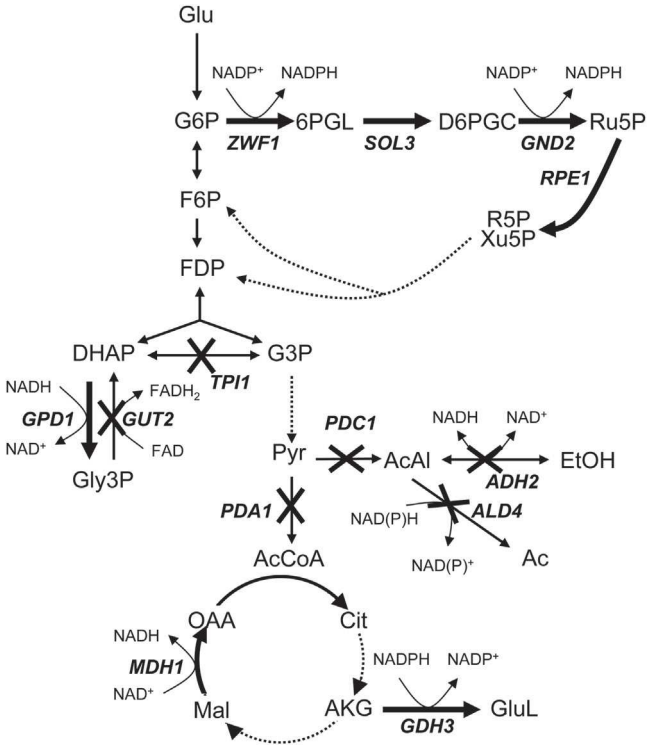

Fig. 2. Overexpression (OE) and knockout (KO) targets illustrated in the metabolic map of the central metabolism. The targets for single gene KO and OE are indicated by their gene names. OE targets are illustrated by a thick arrow, and KO targets by a crossed reaction. Turnover of redox cofactors is shown for the selected reactions.

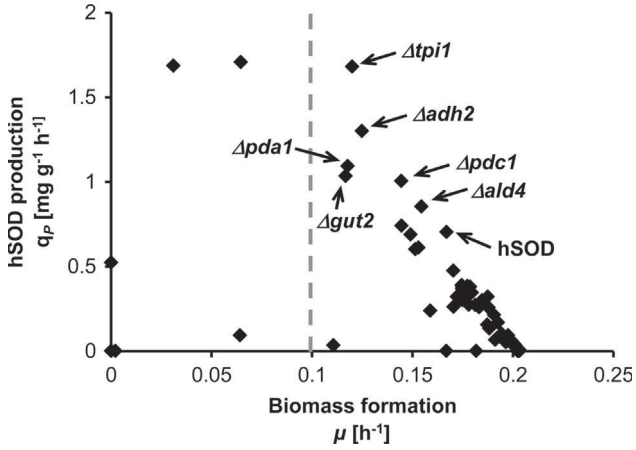

Fig. 3. Simulation results for single gene knockout targets using MOMA to increase the production of hSOD. A cut-off value of 0.1 h^{-1} was used for the biomass formation (dashed line), to select mutants that are capable of achieving increased hSOD production while at the same time, do not greatly inhibit growth. The base strain expressing hSOD production and the top knockout candidates are pointed out with their respective gene names. (Δ*adh2*: alcohol dehydrogenase, Δ*ald4*: aldehyde dehydrogenase, Δ*gut1*: glycerol-3-phosphate dehydrogenase, Δ*pda1*: pyruvate dehydrogenase, Δ*pdc1*: pyruvate decarboxylase, Δ*tpi1*: triose-phosphate isomerase).

Table 2
Predicted knockout (MOMA) and overexpression (FSEOF) targets for increased production of hSOD.

Genes	Enzyme names	Gene name	$\log_2$ flux change hSOD/X-33
MOMA target			
PIPA03164 PIPA01726	Pyruvate decarboxylase	*PDC1*	$\ll -5$
PIPA03313 PIPA02544	Alcohol dehydrogenase (ethanol)	*ADH2*	$\ll -5$
PIPA00390	Aldehyde dehydrogenase (acetylaldehyde, NAD)	*ALD4*	n.d.*
PIPA02794 PIPA03785 PIPA04299 PIPA03623	Pyruvate dehydrogenase	*PDA1*	0.23 ± 0.02
PIPA03441	Triose-phosphate isomerase	*TPI1*	0.00 ± 0.03
PIPA02567	Glycerol-3-phosphate dehydrogenase (FAD)	*GUT2*	0.00 ± 0.03
FSEOF target			
PIPA08178	Glucose 6-phosphate dehydrogenase	*ZWF1*	0.04 ± 0.23
PIPA04435	6-phosphogluconolactonase	*SOL3*	0.04 ± 0.23
PIPA03124	Phosphogluconate Dehydrogenase	*GND2*	0.04 ± 0.23
PIPA03251	Ribulose 5-phosphate 3-epimerase	*RPE1*	-0.15 ± 0.02
PIPA02093	Transketolase	*TKL1*	0.09 ± 0.22
PIPA03744	Transaldolase	*TAL1*	0.09 ± 0.22
PIPA06084	Glycerol-3-phosphate dehydrogenase	*GPD1*	0.01 ± 0.02
PIPA02244	Malate dehydrogenase	*MDH1*	0.56 ± 0.01
PIPA03564	Glutamate dehydrogenase (NADP)	*GDH3*	n.d.*

* Not determined.

(Kaleta et al., 2013). Heyland et al. (2011b) have shown that metabolically costly amino acids constitute a bottleneck in the production of recombinant aminopeptidase in *P. pastoris*. As the synthesis of these amino acids requires reduced NADPH a benefit of the enhanced reduction of the NADP/H pool on recombinant protein production appears plausible.

The benefit of other predicted mutations requires more in-depth analysis to be fully understood. Deletion of genes encoding reactions from pyruvate towards ethanol or acetate may lead to an increased TCA cycle flux and hence increased energy production, which may counteract the increased energy demand for synthesis of amino acids and recombinant protein. Actually it has been shown that recombinant protein production deviates the cellular metabolism of *P. pastoris* towards the TCA cycle (Dragosits et al., 2009; Heyland et al., 2011b; Jordà et al., 2012). The appearance of pyruvate dehydrogenase (*PDA1*) as a knockout target seems contradictory to overexpression targets of TCA cycle enzymes, as it is the entry point of carbon flux into the TCA cycle. To determine how *PDA1* can be a knockout target, we re-examined the flux in the GEM under the condition of *PDA1* deletion. It was found that in the *PDA1* mutant, the flux from pyruvate to acetyl-CoA can be re-routed through acetaldehyde to acetate, which is then converted to acetyl-CoA. This physiological state, however, was not reproduced experimentally, as the *PDA1* mutant strain was determined to be not viable (see Section 3.3).

The benefits of a decrease of fermentation ($\Delta pdc1$, $\Delta adh2$) and an increased TCA cycle flux (*MDH1*) on hSOD production were anticipated already in the hSOD strain. While the predicted gene overexpressions correlated to an increased flux through the respective reaction, the predicted gene deletions were mirrored by down regulation of the respective flux compared to the wild type (Table 2). The PPP flux increase in the hSOD strain however was not significant indicating that a further benefit on recombinant protein production should be achievable by overexpressing genes encoding PPP enzymes as proposed by the model.

3.3. Gene knockout and overexpression in the hSOD strain

To study the impact of the model-predicted engineering targets on recombinant protein production, most of the aforementioned genes (Table 2) were overexpressed or knocked out, in the hSOD producing parental strain. Out of 6 predicted knockout targets, the deletion of pyruvate dehydrogenase subunit α (*PDA1*) and triose-phosphate isomerase (*TPI1*) were not achieved and thus assumed to be inviable under the tested conditions. Overexpression of 7 genes was performed, including the four consecutive steps of the upper oxidative part of the PPP, the two genes related to the TCA cycle, and *GPD1*. *TKL1* and *TAL1* were not selected due to uncertainties of their annotation.

Knockouts were verified by PCR, while enhanced activities of selected overexpressions (*ZWF1* and *MDH1*) were proven by enzymatic assays (data not shown).

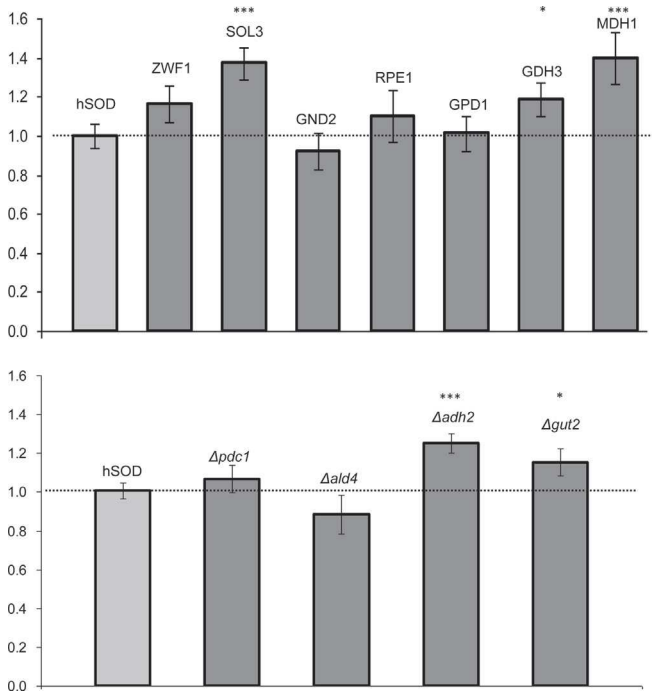

Fig. 4. Impact of mutations on hSOD expression levels. Relative changes of hSOD yield (μg hSOD per mg of total extracted proteins, relative to the hSOD control strain) are indicated. Data are means of 10 independent overexpression clones (upper panel) or 8 independent cultures of a knockout clone (lower panel), respectively. Error bars indicate the standard errors of the mean. Significance of differences to the control was calculated with Student's *t*-test. ***p-value < 0.01; *p-value < 0.1.

3.4. Comparison of engineered strains and the parental hSOD strain

Overexpression targets were evaluated in 10 clones per strain, while each knockout was tested in 8 replicates per strain. The effect of cell engineering on recombinant protein production varied among the different mutant strains (Fig. 4). Among the 4 overexpressed PPP genes, *SOL3* had the strongest impact on hSOD production, leading to 40% increase in the produced protein. A beneficial effect was also seen by overexpression of *ZWF1*, the initial enzyme of the PPP, while no significant effect was seen by overexpression of the two latter PPP enzymes. Increased hSOD production was also observed upon *MDH1* overexpression (plus 40%) and to a lesser extent by *GDH3* overexpression. Out of the 4 knockout strains, an average of 1.2 fold higher hSOD levels could be detected in the Δ*adh2* strain. Most of the other strains showed minor positive effects.

It has to be noted that overexpression of single genes of a linear pathway may be suited to enhance the flux if the respective step has a major role in controlling this flux (Fell, 1998). Glucose-6-phosphate dehydrogenase (*Zwf1*) as the entry point of PPP should be the most plausible controlling step for the competition with the upper glycolytic pathway. In *S. cerevisiae* Zwf1 is regulated by the NADPH/$NADP^+$ ratio (Zubay, 1988) which would explain that overexpression of this enzyme has only a limited capacity to enhance the PPP flux. It was shown recently that the second step of the PPP, 6-phosphogluconolactonase (Sol3 and Sol4 in *S. cerevisiae*) is controlled both at the transcriptional and translational level in this yeast (Castelli et al., 2011; Zampar et al., 2013). Flux control analysis supports the claim that 6-phosphogluconolactonase is the major PPP flux controlling step in yeast (Messiha et al., 2014). A similar regulation of PPP enzymes can also be assumed in *P. pastoris*, as reflected by the increases in hSOD production by overexpression of Sol3 and Zwf1. Regarding TCA cycle, Driouch et al. (2012) propose an attenuation or knockout of malate dehydrogenase in *A. niger* to further improve protein production (awaiting experimental verification), which stands in contrast to the actual benefit of an overexpression of this gene in *P. pastoris*, as shown here. There is no clear connection identified between Mdh1 activity and protein synthesis, and the difference between *A. niger* and *P. pastoris* may be associated with the opposite changes of the TCA cycle flux in these two species.

A second intracellular model protein, GusA, was chosen to verify the effect of the model predictions on protein production. The majority of beneficial effects of mutations in the central carbon

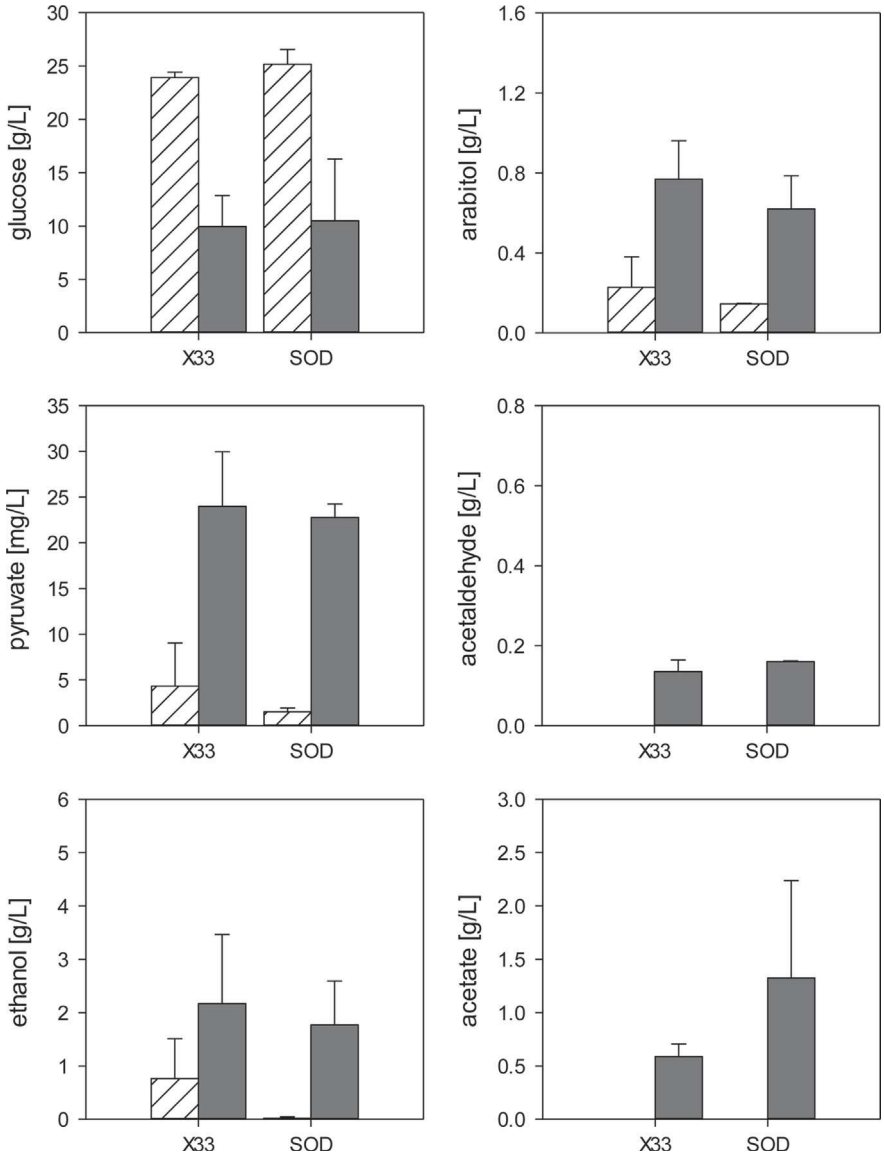

Fig. 5. Extracellular metabolites of strains X-33 and hSOD. Concentrations of residual glucose, arabitol, pyruvate, acetaldehyde, ethanol and acetate were measured in shake flasks at two time points: at exponential growth (striped bars) and at a late growth phase (filled bars). Error bars indicate standard errors of the mean of two parallel cultures.

metabolism on protein production could be reproduced, indicating a broader validity of the approach discussed here (data not shown).

3.5. Intracellular fluxes and extracellular metabolites of mutant strains

Residual glucose and extracellular metabolites (pyruvate, acetaldehyde, ethanol, acetate and arabitol) were measured in batch cultures of the wild-type and the hSOD strain, and all mutants during exponential growth (same as for ^{13}C labeling) and at a late growth phase. Only minor differences were observed between the wild-type and the hSOD strain (Fig. 5). In line with the predicted flux changes, ethanol and arabitol concentrations were lower in supernatants of the production strain while secreted acetate was increased at the late growth phase. It should be noted that extracellular acetate and also acetaldehyde concentrations were below the detection limit during exponential growth in all strains. Overexpression of PPP genes had minor impact on extracellular metabolites (Fig. 6), except for *GND2*, reducing acetaldehyde, ethanol and acetate in the supernatant. *GDH3* had no significant impact on extracellular metabolites, while *MDH1* overexpression reduced ethanol and acetate secretion. Deletions in the fermentative pathway led to the accumulation of upstream metabolites, and in case of Δ*pdc1* acetaldehyde and ethanol were not produced at all (Fig. 6). Both Δ*adh2* and Δ*ald4* did not abolish the respective production of ethanol and acetate. The alcohol dehydrogenase genes of *P. pastoris* have not yet been clearly annotated according to their biochemical function, indicating that more than one knockout is necessary to eliminate ethanol formation. Similarly, *Ald4* may not be the only enzyme responsible for acetaldehyde oxidation. Moreover, Δ*ald4* led to a 2-fold increase in ethanol production, probably by channeling the acetaldehyde overflow. This strong increase in ethanol production may explain that Δ*ald4* did not improve hSOD production as predicted. It needs to be evaluated in future if an attenuation or knockout of both reactions is needed to increase recombinant protein production. The intracellular flux distribution of the mutant strains (data not shown) was measured, but no statistically significant changes compared to the flux distribution of the SOD strain were obtained. As fluxes in the hSOD strain already changed toward the direction in which the mutants would be attained we assume that further significant flux changes need additional metabolic engineering of the production

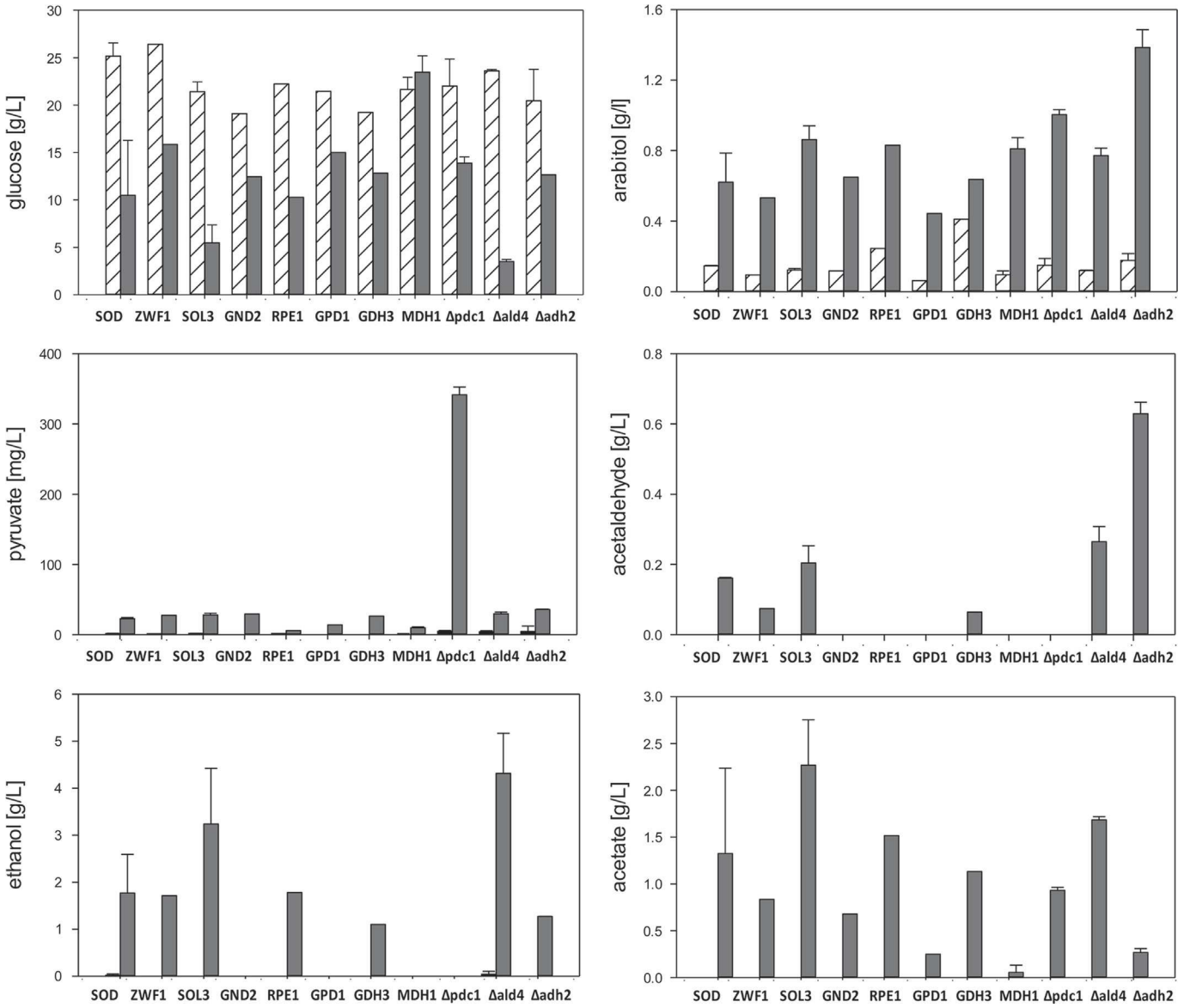

Fig. 6. Extracellular metabolites of mutant strains, compared to the parental control strain hSOD. Data are derived and displayed as described for fig. 6.

host. The total NADP/H pool and the anabolic reduction charge were unchanged in the *ZWF1* strain compared to the hSOD strain (but still threefold more reduced than in the wild type strain). Surprisingly the NADP/H system was more oxidized in the *SOL3* strain than in the hSOD parent strain (Table 1) while *SOL3* overexpression had a strong positive impact on hSOD production. This has been observed consistently in several measurements and may lead to the conclusion that the impact of PPP may go beyond the provision of reduced equivalents.

4. Conclusions

We could demonstrate that recombinant protein production in *P. pastoris* causes significant flux changes which could be reconstructed by a GEM with high accuracy. MOMA and FSEOF turned out as efficient tools to predict which gene knockouts or overexpressions may improve production of a recombinant protein. About 70% of the predicted beneficial mutations were already anticipated by flux changes in the same direction the non-mutated hSOD overproduction strain compared to the wild type. Obviously regulatory plasticity of the yeast metabolism directs fluxes towards what is needed for recombinant protein production. Five out of 9 tested single gene modifications (overexpression or knockout) led to a significant improvement of recombinant protein production, indicating the benefit of a further shift of these pathways. A further positive effect of combined modifications may be anticipated, especially of consecutive steps of the PPP. This work is currently in progress in our laboratory.

Metabolic models provide the power to describe catabolic and anabolic processes of cellular growth and product formation. Prediction of metabolic engineering targets was applied successfully to enhance production of metabolites, such as ethanol (Bro et al., 2006), succinate (Kim et al., 2007), lysine (Becker et al., 2011), sesquiterpenes (Asadollahi et al., 2009), or polyhydroxyalkanoates (Poblete-Castro et al., 2013). Obviously model based metabolic engineering works successfully not only for products of the primary metabolism, but also for secondary metabolites as well as for biopolymers including proteins.

Metabolic models also consider the demand of energy and reductive power for the synthesis of biomass and even complex products like recombinant proteins. The protein secretion process, however, adds further complexity. Chaperone activity consumes ATP in multiple cycles of binding and release of a folding polypeptide (Walter and Buchner, 2002) which cannot be described stoichiometrically. Similarly, the demand of NADPH for redox balancing of oxidative protein folding is not stoichiometrically coupled to the number of disulfide bonds to be closed (Margittai and Sitia, 2011). To exclude this complexity in this work we have focused on non-secreted proteins. First attempts to model the secretory pathway of yeast have been described by Feizi et al. (2013) and will pave the way towards integration of secretion into the reconstruction of protein production.

Conflict of interest

The authors declare no financial or commercial conflict of interest.

Acknowledgments

This work was funded by the Austrian Science Fund (FWF): Doctoral Program BioToP—Biomolecular Technology of Proteins (FWF W1224). Further support 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 is acknowledged. EQ BOKU VIBT GmbH is acknowledged for providing mass spectrometry instrumentation. This work was also supported by the Technology Development Program to Solve Climate Changes on Systems Metabolic Engineering for Biorefineries (NRF-2012-C1AAA001-2012M1A2A2026556) from the Ministry of Education, Science and Technology through the National Research Foundation of Korea. The authors thank Stefanie Müller for her excellent technical assistance.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ymben.2014.05.011.

References

Asadollahi, M.A., Maury, J., Patil, K.R., Schalk, M., Clark, A., Nielsen, J., 2009. Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through in silico driven metabolic engineering. Metab. Eng. 11, 328–334.

Ask, M., Bettiga, M., Duraiswamy, V.R., Olsson, L., 2013. Pulsed addition of HMF and furfural to batch-grown xylose-utilizing *Saccharomyces cerevisiae* results in different physiological responses in glucose and xylose consumption phase. Biotechnol. Biofuels 6, 181.

Baumann, K., Carnicer, M., Dragosits, M., Graf, A.B., Stadlmann, J., Jouhten, P., Maaheimo, H., Gasser, B., Albiol, J., Mattanovich, D., Ferrer, P., 2010. A multilevel study of recombinant *Pichia pastoris* in different oxygen conditions. BMC Syst. Biol. 4, 141.

Becker, J., Zelder, O., Häfner, S., Schroder, H., Wittmann, C., 2011. From zero to hero: design-based systems metabolic engineering of *Corynebacterium glutamicum* for L-lysine production. Metab. Eng. 13, 159–168.

Blumhoff, M., Steiger, M.G., Marx, H., Mattanovich, D., Sauer, M., 2013. Six novel constitutive promoters for metabolic engineering of *Aspergillus niger*. Appl. Microbiol. Biotechnol. 97, 259–267.

Bro, C., Regenberg, B., Forster, J., Nielsen, J., 2006. *In silico* aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. Metab. Eng. 8, 102–111.

Canelas, A.B., ten Pierick, A., Ras, C., Seifar, R.M., van Dam, J.C., van Gulik, W.M., Heijnen, J.J., 2009. Quantitative evaluation of intracellular metabolite extraction techniques for yeast metabolomics. Anal. Chem. 81, 7379–7389.

Carnicer, M., Baumann, K., Toplitz, I., Sánchez-Ferrando, F., Mattanovich, D., Ferrer, P., Albiol, J., 2009. Macromolecular and elemental composition analysis and extracellular metabolite balances of *Pichia pastoris* growing at different oxygen levels. Microb. Cell Fact. 8, 65.

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.

Castelli, L.M., Lui, J., Campbell, S.G., Rowe, W., Zeef, L.A., Holmes, L.E., Hoyle, N.P., Bone, J., Selley, J.N., Sims, P.F., Ashe, M.P., 2011. Glucose depletion inhibits translation initiation via eIF4A loss and subsequent 48S preinitiation complex accumulation, while the pentose phosphate pathway is coordinately upregulated. Mol. Biol. Cell. 22, 3379–3393.

Choi, H.S., Lee, S.Y., Kim, T.Y., Woo, H.M., 2010. *In silico* identification of gene amplification targets for improvement of lycopene production. Appl. Environ. Microbiol. 76, 3097–3105.

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

Damasceno, L.M., Huang, C.J., Batt, C.A., 2012. Protein secretion in *Pichia pastoris* and advances in protein production. Appl. Microbiol. Biotechnol. 93, 31–39.

De Schutter, K., Lin, Y.C., 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, 561–566.

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, 1380–1392.

Driouch, H., Melzer, G., Wittmann, C., 2012. Integration of *in vivo* and *in silico* metabolic fluxes for improvement of recombinant protein production. Metab. Eng. 14, 47–58.

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

Fell, D.A., 1998. Increasing the flux in metabolic pathways: a metabolic control analysis perspective. Biotechnol. Bioeng. 58, 121–124.

Gasser, B., Prielhofer, R., Marx, H., Maurer, M., Nocon, J., Steiger, M., Puxbaum, V., Sauer, M., Mattanovich, D., 2013. *Pichia pastoris*: protein production host and model organism for biomedical research. Future Microbiol. 8, 191–208.

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

Heiss, S., Maurer, M., Hahn, R., Mattanovich, D., Gasser, B., 2013. Identification and deletion of the major secreted protein of *Pichia pastoris*. Appl. Microbiol. Biotechnol. 97, 1241–1249.

Heyland, J., Blank, L.M., Schmid, A., 2011a. Quantification of metabolic limitations during recombinant protein production in *Escherichia coli*. J. Biotechnol. 155, 178–184.

Heyland, J., Fu, J., Blank, L.M., Schmid, A., 2010. Quantitative physiology of *Pichia pastoris* during glucose-limited high-cell density fed-batch cultivation for recombinant protein production. Biotechnol. Bioeng. 107, 357–368.

Heyland, J., Fu, J., Blank, L.M., Schmid, A., 2011b. Carbon metabolism limits recombinant protein production in *Pichia pastoris*. Biotechnol. Bioeng. 108, 1942–1953.

Idiris, A., Tohda, H., Kumagai, H., Takegawa, K., 2010. Engineering of protein secretion in yeast: strategies and impact on protein production. Appl. Microbiol. Biotechnol. 86, 403–417.

Jefferson, R.A., Burgess, S.M., Hirsh, D., 1986. beta-glucuronidase from *Escherichia coli* as a gene-fusion marker. Proc. Natl. Acad. Sci. USA 83, 8447–8451.

Jordà, J., Jouhten, P., Cámara, E., Maaheimo, H., Albiol, J., Ferrer, P., 2012. Metabolic flux profiling of recombinant protein secreting *Pichia pastoris* growing on glucose:methanol mixtures. Microb. Cell Fact. 11, 57.

Kaleta, C., Schäuble, S., Rinas, U., Schuster, S., 2013. Metabolic costs of amino acid and protein production in *Escherichia coli*. Biotechnol. J. 8, 1105–1114.

Kazemi Seresht, A., Cruz, A.L., de Hulster, E., Hebly, M., Palmqvist, E.A., van Gulik, W., Daran, J.M., Pronk, J., Olsson, L., 2013. Long-term adaptation of *Saccharomyces cerevisiae* to the burden of recombinant insulin production. Biotechnol. Bioeng. 110, 2749–2763.

Kim, H.J., Kwon, Y.D., Lee, S.Y., Kim, P., 2012. An engineered *Escherichia coli* having a high intracellular level of ATP and enhanced recombinant protein production. Appl. Microbiol. Biotechnol. 94, 1079–1086.

Kim, T.Y., Kim, H.U., Park, J.M., Song, H., Kim, J.S., Lee, S.Y., 2007. Genome-scale analysis of *Mannheimia succiniciproducens* metabolism. Biotechnol. Bioeng. 97, 657–671.

Klein, T., Lange, S., Wilhelm, N., Bureik, M., Yang, T.H., Heinzle, E., Schneider, K., 2014. Overcoming the metabolic burden of protein secretion in *Schizosaccharomyces pombe*: a quantitative approach using ^{13}C-based metabolic flux analysis. Metab. Eng. 21, 34–45.

Krüger, A., Grüning, N.M., Wamelink, M.M., Kerick, M., Kirpy, A., Parkhomchuk, D., Bluemlein, K., Schweiger, M.R., Soldatov, A., Lehrach, H., Jakobs, C., Ralser, M., 2011. The pentose phosphate pathway is a metabolic redox sensor and regulates transcription during the antioxidant response. Antioxid. Redox Signal. 15, 311–324.

Margittai, E., Sitia, R., 2011. Oxidative protein folding in the secretory pathway and redox signaling across compartments and cells. Traffic 12, 1–8.

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, 1260–1270.

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

McAlister-Henn, L., Thompson, L.M., 1987. Isolation and expression of the gene encoding yeast mitochondrial malate dehydrogenase. J. Bacteriol. 169, 5157–5166.

Messiha, H., Kent, E., Malys, N., Carroll, K., Swainston, N., Mendes, P., Smallbone, K., 2009. Enzyme characterisation and kinetic modelling of the pentose phosphate pathway in yeast:PeerJ PrePrints vol 2,e146v4 2014.

Oddone, G.M., Mills, D.A., Block, D.E., 2009. A dynamic, genome-scale flux model of *Lactococcus lactis* to increase specific recombinant protein expression. Metab. Eng. 11, 367–381.

Ortmayr, K., Nocon, J., Gasser, B., Mattanovich, D., Hann, S., Kollensperger, G., 2014. Sample preparation workflow for the LC-MS based analysis of $NADP^+$ and NADPH in yeast. J. Sep. Sci., http://dx.doi.org/10.1002/jssc.201400290.

Pedersen, H., Carlsen, M., Nielsen, J., 1999. Identification of enzymes and quantification of metabolic fluxes in the wild type and in a recombinant *Aspergillus oryzae* strain. Appl. Environ. Microbiol. 65, 11–19.

Pfeffer, M., Maurer, M., Kollensperger, G., Hann, S., Graf, A.B., Mattanovich, D., 2011. Modeling and measuring intracellular fluxes of secreted recombinant protein in *Pichia pastoris* with a novel 34S labeling procedure. Microb. Cell Fact. 10, 47.

Pflügl, S., Marx, H., Mattanovich, D., Sauer, M., 2012. 1,3-Propanediol production from glycerol with *Lactobacillus diolivorans*. Bioresour. Technol. 119, 133–140.

Poblete-Castro, I., Binger, D., Rodrigues, A., Becker, J., Martins Dos Santos, V.A., Wittmann, C., 2013. *In-silico*-driven metabolic engineering of *Pseudomonas putida* for enhanced production of poly-hydroxyalkanoates. Metab. Eng. 15, 113–123.

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

Quek, L.E., Wittmann, C., Nielsen, L.K., Kromer, J.O., 2009. OpenFLUX: efficient modelling software for ^{13}C-based metabolic flux analysis. Microb. Cell Fact. 8, 25.

Segrè, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. USA 99, 15112–15117.

Sheikholeslami, Z., Jolicoeur, M., Henry, O., 2013. Probing the metabolism of an inducible mammalian expression system using extracellular isotopomer analysis. J. Biotechnol. 164, 469–478.

Sohn, S.B., Graf, A.B., Kim, T.Y., Gasser, B., Maurer, M., Ferrer, P., Mattanovich, D., Lee, S.Y., 2010. Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production. Biotechnol. J. 5, 705–715.

Souza, M.A., Ribeiro, M.Z., Silva, D.P., Pessoa, A., Vitolo, M., 2002. Effect of pH on the stability of hexokinase and glucose 6-phosphate dehydrogenase. Appl. Biochem. Biotechnol. 98-100, 265–272.

Stadlmayr, G., Mecklenbräuker, A., Rothmüller, M., Maurer, M., Sauer, M., Mattanovich, D., Gasser, B., 2010. Identification and characterisation of novel *Pichia pastoris* promoters for heterologous protein production. J. Biotechnol. 150, 519–529.

Vijayasankaran, N., Carlson, R., Srienc, F., 2005. Metabolic pathway structures for recombinant protein synthesis in *Escherichia coli*. Appl. Microbiol. Biotechnol. 68, 737–746.

Walter, S., Buchner, J., 2002. Molecular chaperones: cellular machines for protein folding. Angew. Chem. Int. Ed. Engl. 41, 1098–1113.

Weber, J., Hoffmann, F., Rinas, U., 2002. Metabolic adaptation of *Escherichia coli* during temperature-induced recombinant protein production: 2. Redirection of metabolic fluxes. Biotechnol. Bioeng. 80, 320–330.

Wittmann, C., Weber, J., Betiku, E., Kromer, J., Bohm, D., Rinas, U., 2007. Response of fluxome and metabolome to temperature-induced recombinant protein synthesis in *Escherichia coli*. J. Biotechnol. 132, 375–384.

Xu, C., Liu, L., Zhang, Z., Jin, D., Qiu, J., Chen, M., 2013. Genome-scale metabolic model in guiding metabolic engineering of microbial improvement. Appl. Microbiol. Biotechnol. 97, 519–539.

Zamboni, N., Fendt, S.M., Rühl, M., Sauer, U., 2009. (13)C-based metabolic flux analysis. Nat. Protoc. 4, 878–892.

Zamboni, N., Fischer, E., Sauer, U., 2005. FiatFlux—a software for metabolic flux analysis from ^{13}C-glucose experiments. BMC Bioinform. 6, 209.

Zampar, G.G., Kümmel, A., Ewald, J., Jol, S., Niebel, B., Picotti, P., Aebersold, R., Sauer, U., Zamboni, N., Heinemann, M., 2013. Temporal system-level organization of the switch from glycolytic to gluconeogenic operation in yeast. Mol. Syst. Biol., 9; p. 651.

Zubay, G., 1988. Biochemistry. Macmillan, New York.