FEMS Yeast Research, 17, 2017, fox068

doi: 10.1093/femsyr/fox068
Advance Access Publication Date: 23 August 2017
Minireview

MINIREVIEW

Systems biotechnology for protein production in *Pichia pastoris*

Richard J. Zahrl[1,2], David A. Peña[1], Diethard Mattanovich[1,2,*,†] and Brigitte Gasser[1,2]

[1]Department of Biotechnology, BOKU – University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria and [2]Austrian Centre of Industrial Biotechnology, Muthgasse 11, 1190 Vienna, Austria

*Corresponding author: Department of Biotechnology, University of Natural Resources and Life Sciences (BOKU), Muthgasse 18, 1190 Vienna, Austria. Tel: +43 1 47654 79042; Fax: +43 1 47654 79009; E-mail: diethard.mattanovich@boku.ac.at

One sentence summary: Systems biotechnology enables to understand and improve recombinant protein production in the methylotrophic yeast *Pichia pastoris*.

Editor: Irina Borodina

†Diethard Mattanovich, http://orcid.org/0000-0002-0907-4167

ABSTRACT

The methylotrophic yeast *Pichia pastoris* (syn. *Komagataella* spp.) is one of the most important production systems for heterologous proteins. After the first genome sequences were published in 2009, tremendous effort was made to establish systems-level analytical methods. Methylotrophic lifestyle was one of the most thoroughly investigated topics, studied at the levels of transcriptome, proteome and metabolic flux. Also the responses of *P. pastoris* to environmental stress conditions experienced during high cell density production processes were studied. Metabolomics and flux analysis revealed the plasticity of the cellular metabolism in its adaption to the production of foreign proteins and served as blueprints for subsequent cell engineering and/or process design. The transcriptional response elicited by overexpression of heterologous proteins seems to depend on the nature and complexity of the recombinant product. Based on these data, novel targets for strain engineering could be deduced from transcriptomics and proteomics data mining and effectively enhanced protein secretion. Transcriptional regulation data also served as a valuable resource to identify novel promoters with the desired regulatory characteristics. This review aims to provide a comprehensive overview of systems biology applications in *P. pastoris* ranging from increased understanding of cell physiology to improving recombinant protein production in this cell factory.

Keywords: systems biology; yeast; cell engineering; recombinant protein production; bioprocess engineering; synthetic biology

INTRODUCTION

The methylotrophic yeast *Pichia pastoris* was first isolated from a chestnut tree in France and described as *Zygosaccharomyces pastori* by Alexandre Guilliermond (Guilliermond 1920). In the 1950s, Herman Phaff isolated further strains from black oak trees in California, USA, and renamed the species as *Pichia pastoris* (Phaff, Miller and Shifrine 1956). In 1995, the *P. pastoris* strains were moved to a new genus, *Komagataella* (Yamada *et al.* 1995), and later separated into two species (Kurtzman 2005): *Komagataella pastoris* (including the French type strain) and *K. phaffii* (including the American isolates). Today, the genus *Komagataella* consists of six species (Naumov *et al.* 2013). Both *K. pastoris* and *K. phaffii* strains are in use for recombinant protein production under the name *P. pastoris*. For ease of understanding and for consistency with the literature, we refer here to all strains as *P. pastoris*.

Received: 28 May 2017; **Accepted:** 22 August 2017

As a methylotrophic yeast, *P. pastoris* is able to grow on methanol as the sole carbon and energy source. The extraordinarily high and methanol-inducible expression level of methanol utilization-related genes, namely the alcohol oxidase gene *AOX1*, was utilized to develop an efficient protein production system (Cregg *et al.* 1985) and has proven valuable for a growing list of therapeutic proteins and industrial enzymes (Weinacker *et al.* 2013; Spadiut *et al.* 2014). The advantages of *P. pastoris* mainly relate to a high secretory capacity (Puxbaum, Mattanovich and Gasser 2015) and the strong methanol inducible promoters (Weinhandl *et al.* 2014; Gasser, Steiger and Mattanovich 2015). Growth on methanol, however, adds technological challenges for industrial-scale production, such as high oxygen demand, high heat production and the demand for safety precautions for the use of a flammable substrate (Mattanovich *et al.* 2014). To circumvent the disadvantages of using methanol, variants of the *AOX1* promoter were developed which are active without the addition of methanol (Hartner *et al.* 2008; Shen *et al.* 2016; Looser *et al.* 2017; Wang *et al.* 2017). Alternatively, co-feeding of methanol with glycerol or sorbitol reduces both biological heat production and oxygen demand (Niu *et al.* 2013; Berrios *et al.* 2017). To avoid the use of methanol completely, constitutive promoters can be applied (Vogl *et al.* 2014a; Weinhandl *et al.* 2014), and methanol independent regulated promoters were identified (Prielhofer *et al.* 2013; Landes *et al.* 2016).

Besides strong promoters, the high efficiency of protein production by *P. pastoris* is attributed to an efficient secretory pathway. On a genomic level, the core secretory pathway functions are rather conserved among yeasts (Delic *et al.* 2013b), while Papanikou and Glick (2009) described distinct morphological differences between the Golgi apparatus of *P. pastoris* and *Saccharomyces cerevisiae*. These morphological differences may be a reason why *P. pastoris* generally secretes heterologous proteins more efficiently.

Systems biotechnology has been defined as the integration of high-throughput genome-scale data of biotechnologically relevant strains with *in silico* modeling and simulation to facilitate strain development for efficient industrial production (Lee, Lee and Kim 2005). Among others, this approach is applied for the optimization of yeast cells for heterologous protein production (Graf *et al.* 2009; Gutierrez and Lewis 2015). Here, we discuss the developments and the current state of the art of systems biotechnology of *P. pastoris* to study and enhance heterologous protein production. Figure 1 depicts a timeline, which highlights major milestones in the history of *P. pastoris* systems biology.

SYSTEMS BIOLOGY ANALYSIS OF *PICHIA PASTORIS*

Genomics

Genomic information is a prerequisite for genome-wide systems-level analysis. Since the first genome sequences of different *P. pastoris* strains were published in 2009 (*Komagataella pastoris* type strain DSMZ70382 and *K. phaffii* strain GS115) (De Schutter *et al.* 2009; Mattanovich *et al.* 2009), they have been a valuable resource to investigate the genetic repertoire of the methylotrophic yeast. Important cellular functions that are different to the model yeast *S. cerevisiae* were found, such as the presence of mitochondrial complex I (Bridges, Fearnley and Hirst 2010), specific enzymes of sphingolipid biosynthesis and fatty acid metabolism (Ternes *et al.* 2011; Yu *et al.* 2012; Tomas-Gamisans, Ferrer and Albiol 2016) or the pathway needed for rhamnose utilization (Prielhofer *et al.* 2015; Liu *et al.* 2013). Genes encoding isoenzymes of core metabolic enzymes with an alternative localization to the peroxisomes were found to be located adjacent to their cytosolic or mitochondrial isoforms, indicating tandem gene duplication and positive evolutionary selection for their newly acquired function in methanol utilization (Russmayer *et al.* 2015a). Furthermore, for some proteins involved in amino acid metabolism different cellular localizations compared to *S. cerevisiae* were predicted and confirmed (Forster *et al.* 2014). Also, the genomic setup of the mating type locus was elucidated (Hanson, Byrne and Wolfe 2014). Detailed analysis of the gene predictions also determined the endowment of the secretory pathway with chaperones and folding-associated enzymes (Delic *et al.* 2013b). In combination with other genome-scale high-throughput techniques, such as nucleosome mapping and ChIP-seq, the genomic information was also used to identify a novel type of G/C-rich autonomous replicating sequences (Liachko *et al.* 2014) and it revealed that *P. pastoris* contains inverted repeat-structured centromer

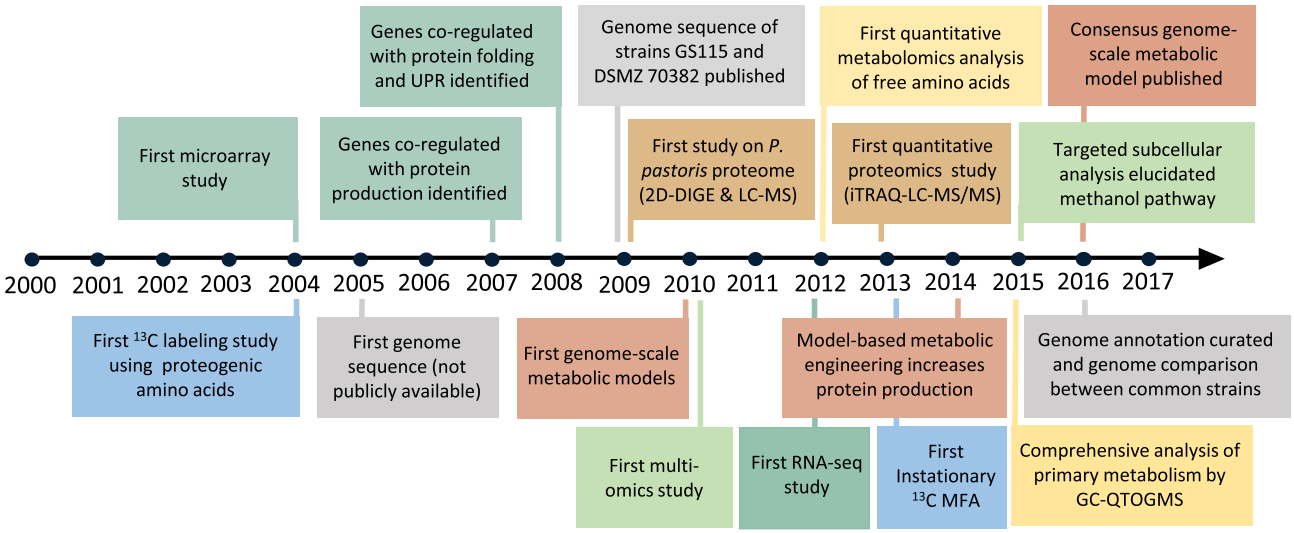

Figure 1. Hallmarks of systems biology development in *Pichia pastoris*. Genomics (gray), transcriptomics (dark green), proteomics (brown), metabolomics (yellow), fluxomics (blue), genome-scale metabolic model and model-based engineering (red), and combined omics studies (light green).

sequences (Coughlan *et al.* 2016; Sturmberger *et al.* 2016). In 2016, the genome sequences of the two previously sequenced strains and *K. phaffii* CBS7435 (the parental strain of GS115) were refined by using PacBio sequencing (Love *et al.* 2016; Sturmberger *et al.* 2016), and comparative genomics were performed, which revealed syntenic rearrangements between the species and 35 single nucleotide polymorphisms between GS115 and CBS7435 (Love *et al.* 2016). Furthermore, the gene predictions and their functional annotations were manually curated and are available through pichiagenome.org (Valli *et al.* 2016). Schwarzhans *et al.* (2016a) analyzed the genomic sequences of 31 individual recombinant *P. pastoris* transformants expressing different levels of green fluorescent protein and found that productivity correlated with the head-to-tail orientation of cassette integration, while low productivity strains with sometimes peculiar colony morphology were found to occur due to non-canonical integration events of the expression cassette or other vector-derived sequences (Schwarzhans *et al.* 2016b).

Transcriptomics

Transcription is the first possible level of regulation in a biological system concerning the flow of information in the central dogma of molecular biology. Understanding what is transcribed when is therefore essential for controlling, manipulating and designing a yeast cell factory for recombinant protein production. While initial studies focused just on some genes or used heterologous microarrays (Sauer *et al.* 2004), publication of the genome sequences enabled the study of transcriptional regulation patterns of the whole genome. To this end, specific microarrays designed for *P. pastoris* were used, and have recently been complemented by the next generation sequencing method of RNA sequencing (Graf *et al.* 2008; Liang *et al.* 2012).

Recombinant protein production usually depends on high transcription levels of the protein of interest. Therefore, the genes of interest are mostly expressed under control of strong constitutive or methanol-inducible promoters (Weinhandl *et al.* 2014). Additionally, productivity can easily be improved by increasing the gene copy number of the protein of interest independent of the promoter system used (Vassileva *et al.* 2001; Mansur *et al.* 2005). However, an optimal gene copy number seems to exist for each recombinant protein (Marx *et al.* 2009). The reason for a decline in productivity upon increasing the copy number might be a limitation in protein folding and secretion (Hohenblum *et al.* 2004). The transcriptional responses elicited by production of recombinant proteins were studied for a selection of different proteins. In most cases, these were difficult to produce proteins, where the aim was to understand the bottleneck or limitations of a specific strain or production process (Sauer *et al.* 2004; Baumann *et al.* 2010; Dragosits *et al.* 2010; Hesketh *et al.* 2013; Vogl *et al.* 2014b; Edward-Jones *et al.* 2015; Aw, Barton and Leak 2017; Camara *et al.* 2017). The impact of protein production in *P. pastoris* was investigated for the first time by microarray analysis of a recombinant trypsinogen secreting *P. pastoris* strain as compared to a non-expressing strain, both cultivated in methanol-induced fed-batch cultivations (Sauer *et al.* 2004; Gasser *et al.* 2007b). The study allowed identification of 524 significantly regulated genes and provided an important picture of how the cells natively adapt to the additional stress of recombinant protein synthesis and secretion. Among others, induction of the unfolded protein response (UPR) was observed in the recombinant strain, a stress response pathway associated with folding and secretion overload. A recent study compared nine different trypsinogen-producing transformants grown in shake flasks. No specific traits of high and low production strains could be identified on the transcriptome level (Aw, Barton and Leak 2017). Interestingly, a limitation may also occur at the translational (Edwards-Jones *et al.* 2015) or transcriptional level (Camara *et al.* 2017). In the latter study, the transcriptomic analyses of strains producing *Rhizopus oryzae* lipase (ROL) under control of P_{AOX1} with different gene copy numbers revealed a possible titration effect of a transcription factor regulating methanol utilization genes including *AOX1*, alongside an attenuation of the methanol metabolism and peroxisome biogenesis. A benefit of methanol-grown cells, however, is that, despite their lower growth rates, translation is higher than when glycerol or glucose is utilized as the carbon source (Prielhofer *et al.* 2015). This is also supported by the higher total protein concentration per biomass (Russmayer *et al.* 2015a). This profile of lower growth in the production phase yet increased protein synthesis matches the required profile of a cell factory.

An ideal production process involves a phase of rapid growth followed by a production phase of minimal growth. Thus, it is interesting to investigate the interrelation between growth rate and specific productivity at the gene expression level. Using transcriptomics data, Rebnegger *et al.* (2014) observed an increase of translation-related genes and a decrease in proteolytic processes with increasing growth rate in glucose-limited chemostats. As a consequence, specific protein secretion rate was observed to decrease with decreasing specific growth rate. In a study exploring the role of the cell cycle on the link between protein production and growth rate, a strain overexpressing *CLB2*, encoding a B-type cyclin, was constructed (Buchetics *et al.* 2011). Cell cycle distribution in the engineered strain was shifted towards G2 and mitosis phases, which enabled a 1.3-fold increase in antibody fragment secretion at low growth rates.

Membrane proteins represent an especially difficult product as they are not secreted and perturb membrane homeostasis in the cell (Guerfal *et al.* 2010). To understand the cellular response and to identify engineering targets, transcriptomes of *P. pastoris* producing different classes of membrane proteins were analyzed (Vogl *et al.* 2014b). The effects on cellular regulation mainly depended on the localization of the overexpressed membrane protein. If the membrane protein was targeted to the inner mitochondrial membrane, energy metabolism was altered, whereas the UPR was triggered when the membrane protein was targeted to the endoplasmic reticulum (ER). Notably, coexpression of $HAC1^{i}$ (activated, spliced variant of the *HAC1* gene encoding the UPR transcriptional activator) (Guerfal *et al.* 2010) could improve the yields of ER resident membrane proteins 1.5-fold to 2.1-fold. Furthermore, transcriptomics has been used to characterize already engineered strains created by overexpression of transcription factors, namely Aft1 (Ruth *et al.* 2014), Yap1 (Delic *et al.* 2014b) and Hac1 (Graf *et al.* 2008; Vogl *et al.* 2014b).

The great majority of currently available transcriptomic data has been obtained from microarrays. However, some studies have already taken advantage of the next generation RNA sequencing (RNA-Seq) technology to characterize global gene regulation in *P. pastoris*. The first transcriptomic studies using RNA-Seq revealed the first internal ribosome entry sites in *P. pastoris* (Liang *et al.* 2012) and suggested a possible role for untranslated RNA molecules in regulating gene action in response to the production of misfolded proteins (Hesketh *et al.* 2013). The latter report also provided evidence of constitutive *HAC1* mRNA splicing under different growth conditions and detrimental stress responses during repeated methanol feeds. Finally, RNA-Seq data was used to significantly improve the percentage of annotated open reading frames (ORFs) from 48% to 73% (Valli *et al.* 2016).

Proteomics

The proteome portrays the functional picture of information conveyed by the transcriptome. Capturing this picture of the proteome by means of proteomics methods is therefore essential for complementing transcriptomic data and understanding the system as a whole.

The first proteomes of *P. pastoris* were assessed from cells grown in glucose-limited chemostats at a fixed specific growth rate at three different cultivation temperatures (Dragosits *et al.* 2009). The specific productivity of an antibody fragment was 3-fold higher at 20°C compared to 30°C. Interestingly, the expression of folding chaperones was decreased at lower temperatures and proteolytic activity did not change, thus challenging previous assumptions that the increase in productivity was related to proteolytic activity. The secretome also seems to be influenced by the cultivation temperature, as more host proteins were found in the supernatant at 30°C cultivations than at 25°C of methanol-induced cultures (Huang *et al.* 2011). Furthermore, methanol-grown cultures exhibited higher concentrations of native proteins in the supernatant than glucose-grown cultures (Mattanovich *et al.* 2009). However, in general the secretion of native proteins was very low, which is another major advantage of recombinant protein production in *P. pastoris*.

Proteomic investigations not related to protein production also provided vital knowledge about the biological system. Of utmost interest is the secretory pathway, which was approached by investigating the proteomic composition of the secretory organelles (Klug *et al.* 2014). Verification of the protein localizations is crucial as they were previously merely based on sequence homology to *S. cerevisiae* proteins. Moreover, even proteins which are not present in *S. cerevisiae*, such as the subunits of respiratory complex I (NADH:quinone oxidoreductase), could be identified and characterized by proteomics approaches (Bridges *et al.* 2009; Bridges, Fearnley and Hirst 2010).

Another investigated hotspot is the methanol utilization (MUT) pathway, where proteomics helped to better understand which processes are truly active in methanol-grown cells. The first quantitative proteomic analysis of methanol metabolism revealed evidence for the action of the malate-aspartate NADH shuttle (Austin *et al.* 2011). This transports cytosolically produced electrons of NADH from the methanol dissimilation pathway into the mitochondria to generate ATP.

Methanol cultivations usually lead to the highest productivities. Therefore, a proteome study was conducted on the time-dependent physiological response to methanol-induced high-level production of the hepatitis B virus surface antigen (Vanz *et al.* 2012). Processes upregulated in the methanol fed-batch phase included the oxidative stress response, induction of the UPR (with Pdi1 as most prominent indicator), a cytosolic chaperone (Hsp104) and an ER-associated degradation (ERAD)-related AAA ATPase (Cdc48). Autophagic processes, on the other hand, were mainly related to peroxisome generation and not to hepatitis B virus surface antigen deposits.

Also, xylanase A triggered the UPR (Lin *et al.* 2013). Therefore, *HAC1*^{i} was co-overexpressed, which decreased protein synthesis and increased ER protein folding. The quantitative proteomic data was further verified by quantitative PCR analysis, suggesting that less abundant proteins also had lower transcript levels. This further supports the notion that regulation occurs at the transcriptional rather than at a translational level (Prielhofer *et al.* 2015). Interestingly, a downregulation of the UPR was also observed after the shift from glycerol to methanol in fed-batch processes. A proteomic study observed a high constitutive induction of the UPR and ERAD pathways in the pre-induction glycerol batch, independent of the production of recombinant protein (Vanz, Nimtz and Rinas 2014).

Furthermore, the interactome of an antibody fragment was studied (Pfeffer *et al.* 2012). Many proteasome-associated proteins were identified and partial inhibition of the proteasome resulted in a significant increase in antibody fragment secretion. In a complementary study, it could be shown that the majority of the recombinant protein (58%) was actually degraded within the cell instead of being secreted (Pfeffer *et al.* 2011), thus suggesting that proteasomal- and ER-associated degradation could be a promising target for strain engineering.

Metabolomics and metabolic flux analysis

Metabolomics aims at the comprehensive quantitative analysis of the metabolites synthesized by a biological system (Fiehn 2002). In the context of *P. pastoris* biotechnology, most analyses have focused on metabolites associated with specific pathways of interest, an approach known as metabolite profiling (Goodacre *et al.* 2004). Major techniques developed for this purpose include gas chromatography with mass spectrometry (GC-MS) of derivatized metabolites, direct nuclear magnetic resonance (NMR) of complex mixtures and liquid chromatography coupled with mass spectrometry (LC-MS) (Goodacre *et al.* 2004). In particular, intracellular metabolite quantification for *P. pastoris* has benefited from advances in internal standards (Neubauer *et al.* 2012; Haberhauer-Troyer *et al.* 2013), improved protocols on cell quenching and filtration (Tredwell *et al.* 2011; Carnicer *et al.* 2012a; Russmayer *et al.* 2015b; Mattanovich *et al.* 2017) and specialized analytical methods (Klavins *et al.* 2014; Mairinger *et al.* 2015; Guo *et al.* 2016).

Quantitative analysis of intracellular free amino acids of *P. pastoris* has been used to study the metabolic impact of assimilation of methanol as a carbon source (Russmayer *et al.* 2015a) and of oxygen availability in glucose-limited chemostat cultivations (Carnicer *et al.* 2012b). Measured variations of free amino acid concentrations were found to correlate with changes in the cellular proteome. Also, recombinant protein production resulted in a reduced concentration of energy costly amino acids in an oxygen-dependent manner. A further application of metabolite profiling identified isoleucine, aspartate and arabitol as potential markers for UPR using NMR (Tredwell *et al.* 2017).

Expanding on the static picture provided by the metabolome, ^{13}C- metabolic flux analysis (MFA) aims to quantify *in vivo* reactions rates (i.e. metabolic fluxes) within a metabolic network (Wiechert 2001). The metabolic flux distribution represents the integrated output of protein-metabolite networks (Sauer 2006), thus providing a valuable source of information for systems biology and metabolic engineering. In order to quantify pathway activity, cells are grown on ^{13}C labeled substrates and isotope patterns are measured by MS or NMR. Finally, information on extracellular rates and biomass composition are combined in a metabolic model to computationally infer intracellular fluxes (Sauer 2006).

The metabolic response of *P. pastoris* to protein production on different media and growth conditions has been extensively studied using ^{13}C-MFA (Ferrer and Albiol 2014). First systematic studies estimated intracellular flux distributions using ^{13}C labeling of proteinogenic amino acids, which were measured by NMR or GC-MS/MS. Using this technique, significant impacts of growth rate and different carbon sources (glycerol, glucose and methanol) on the central carbon metabolism were observed (Sola *et al.* 2004, 2007; Jordà *et al.* 2012, 2014a). Further studies

on the metabolic response of *P. pastoris* to protein production revealed increased activity of the tricarboxylic acid (TCA) cycle (Heyland *et al.* 2011), increased glycolysis and methanol dissimilation (Jordà *et al.* 2012), as well as reduced growth-rate and by-product formation (Heyland *et al.* 2011).

The metabolic effect of feeding a mixture of sorbitol and methanol in fed-batch cultivations was investigated by dynamic flux balance analysis (Celik, Calik and Oliver 2010; Calik *et al.* 2011). In this approach, extracellular fluxes were measured at different time points and used as constraints in a metabolic model. Finally, the model was solved by minimization of the protein secretion rate. In this case, the analysis resulted in an improved feeding strategy for higher methanol assimilation (Calik *et al.* 2011).

Jorda *et al.* (2013) introduced the use of instationary ^{13}C-MFA and metabolomics to obtain a complete flux distribution from a medium-scale (79 reactions) metabolic model of *P. pastoris*. This experiment provided deep insights into cofactor generation and consumption, thermodynamic efficiency and reversibility of central metabolic reactions, as well as the buffering capacity of key metabolites by converting amino acids pools. A subsequent study using this approach validated previous findings on the higher energy and redox demand of protein producing strains. It also revealed a possible role of trehalose on stress response (Jorda *et al.* 2014b), which has been also observed in retentostat experiments (Rebnegger *et al.* 2016). Compared with previous studies, the use of intracellular amino acids and sugar pools instead of proteinogenic amino acids enabled shorter experiments and estimation of metabolite turnover rates.

Nie *et al.* (2014) applied this technique to compare two strains with a 4-fold difference in productivity of recombinant ß-galactosidase. Owing to protein production, maximum specific growth rate and glucose uptake decreased by 10% and 22%, respectively. Fluxes through pentose phosphate pathway and glycolysis increased whereas by-product formation decreased, resulting in higher energy and NADPH generation. A second study on the same strains observed a 32% increase in protein productivity after the addition of glutamate to the medium (Liu *et al.* 2016). Intracellular pools of amino acids that were reduced in the overproducing strain were restored to levels even higher than the control strain, thus providing evidence for the limitation of amino acid synthesis on protein production. Moreover, the improved amino acid supply freed up resources for higher NADH and ATP regeneration. On the contrary, NADPH supply was reduced, supporting the idea that the pentose phosphate pathway might be upregulated in response to the high demand of reducing power and precursors for amino acids biosynthesis.

Metabolic modeling

Genome-scale metabolic models (GEMs) provide a tool for predicting metabolic phenotypes from the genomic information. After initial genome sequencing of *P. pastoris*, multiple constraint-based metabolic reconstructions became available. The first GEM, PpaMBEL1254, was published in 2010 (Sohn *et al.* 2010). It contained information on 540 metabolic genes from the strain DSMZ70382, 1254 reactions and eight subcellular compartments: cytoplasm, ER, Golgi apparatus, mitochondria, nucleus, peroxisome and vacuole (Sohn *et al.* 2010). Around the same time, Chung *et al.* (2010) published the model iPP668, with information on 668 genes from the strain GS115, 1362 metabolic reactions and the same compartments as PpaMBEL1254. PpaMBEL1254 was found to predict 9.4% of the metabolic genes as essential and correctly reproduced the effect of oxygen limitation on growth rate for two model proteins. Similarly, iPP668 was found consistent with chemostat data using glucose and glycerol-methanol mixtures as carbon sources. Moreover, it correctly predicted the positive effect of sorbitol for amino acid synthesis and the high NADH turnover when using methanol as a substrate.

In 2012, the GEM iLC915 containing 1423 reactions, 899 metabolites and seven compartments was published (Caspeta *et al.* 2012). Evaluation of model predictions found good agreement with experimental data. Given the improved gene coverage of iLC915 (17.2% of annotated genes), more genes (12%) were found to be essential *in silico* than in previous models. Irani *et al.* (2016) curated the model iLC915 and included native and human N-glycosylation capabilities to create the GEM ihGlycopastoris. By accounting for N-glycosylation of different model proteins, it was possible to reduce the overestimation of protein synthesis, especially for small and highly glycosylated proteins. Eskitoros, Ata and Çalık (2017) also studied the impact of glycosylation on the prediction of flux distributions using a medium-scale model.

An integration of the GEMs iPP668, PpaMBEL1254 and iLC915 into the model iMT1026 was published in 2016 (Tomas-Gamisans, Ferrer and Albiol 2016). The model included 1026 genes, 2035 reactions and the mitochondrial intermembrane space as a new compartment. Pathways in the fatty acid metabolism, sphingolipid synthesis, GPI-anchor biosynthesis, N-glycosylation as well as the proton stoichiometry in oxidative phosphorylation were curated. Also, missing complexes in the electron transport chain were integrated. The resulting model predicted more accurately growth rate and by-product formation than previous models.

Recently, a dynamic GEM based on a curated version of the iPP668 model was also published (Saitua *et al.* 2017). The model enabled investigation of fluxes during batch and fed-batch cultivations by dynamic flux balance analysis. An optimal feeding strategy, as well as a gene deletion improving human serum albumin production, was predicted to test model capabilities. Similarly, stoichiometric models of *P. pastoris* have been applied for process control under scarce or imprecise data (Tortajada *et al.* 2010), as well as for process optimization (Barrigon, Valero and Montesinos 2015).

Integrated omics analyses combining several layers

Systems biology is defined as the quantitative analysis of biological systems on different cellular levels, following a holistic approach (Graf *et al.* 2009). By using high-throughput technologies such as genome, transcriptome, proteome, metabolome and fluxome analysis, in combination with advanced mathematical modeling tools, it aims to study the interactions between the components of biological systems. However, when looking at the studies conducted in *P. pastoris*, mostly just one layer of regulation was investigated, and there are only a few reports which were truly 'multi-omics' analyses or combinations of different omics approaches (simultaneous or in a stepwise manner). Interestingly, most of the multi-omics studies focused on the physiological reactions of *P. pastoris* to different environmental conditions rather than on elucidating the response triggered by recombinant protein production.

The responses of recombinant Fab producing and control *P. pastoris* strains to oxygen limitations and hypoxic conditions in glucose-limited chemostat cultivations were studied on several cellular layers over the last 10 years. These studies are a good

example of how stepwise integration of omics technologies delivered mechanistic insights and advanced our understanding of why cultivation of *P. pastoris* at lower oxygen concentrations resulted in increased production of secretory proteins under control of P_{GAP} in glucose-based chemostats and fed-batch cultures (Baumann *et al.* 2008). Initial transcriptome and proteome studies revealed that *P. pastoris* responds to oxygen limitation by upregulation of glycolysis, amino acid metabolism and general stress response, while TCA cycle activities were downregulated despite the fixed growth rate (Baumann *et al.* 2010). Upregulation of glycolytic genes and consequently higher mRNA levels of the recombinant Fab, expressed under control of the glycolytic P_{GAP} promoter, was identified as one factor explaining higher productivity in hypoxic conditions. In addition, alterations in the lipid metabolism were detected, which may lead to changed fluidity of the plasma membrane. These changes in the sterol–sphingolipid balance were thought to be associated with the higher secretory potential of the cells. Genes and proteins involved in oxidative stress response were upregulated at conditions of higher Fab secretion, which was also observed for chaperones. However, induction of the latter can be explained by the altered lipid biosynthesis which, in turn, leads to activation of the UPR and thus to increased transcription of chaperones (Baumann *et al.* 2010). The anticipated changes in lipid composition upon oxygen limitation in glucose-limited chemostats were recently confirmed by a comprehensive lipidomics study (Adelantado *et al.* 2017).

Even though many of the observed changes were triggered by lowered oxygen availability rather than being a consequence of heterologous protein production, they had a major impact on productivity and secretion. For example, the effect of hypoxia on ergosterol biosynthesis (Carnicer *et al.* 2009; Baumann *et al.* 2010) was mimicked by addition of the inhibitory substance fluconazole, which also led to a reduction in sterol content and a correlating increase in Fab secretion (Baumann *et al.* 2011).

Interestingly, analysis of the oxygen data revealed a high correlation of transcript levels with proteome and flux data for glycolysis (upregulated) and TCA cycle (downregulated) components (Baumann *et al.* 2010). This correlation was subsequently also confirmed at the level of free intracellular metabolites (Carnicer *et al.* 2012b). A high degree of correlation between transcriptome and proteome data was also observed by Russmayer *et al.* (2015a), suggesting that transcriptional regulation of metabolic enzymes is prevalent in *P. pastoris*. However, looking at the transcriptome alone may not be enough to understand all the changes occurring in cell metabolism.

Quantitative measurement of the free intracellular metabolite pools clarified that the observed gene regulation patterns of amino acid metabolism in oxygen-limited conditions negatively correlated with the response of free amino acid levels and fluxes through those reactions (Carnicer *et al.* 2012b). Both free amino acid pools and total cellular (proteinogenic) amino acid abundances were more responsive to different oxygenation levels than to Fab production, while recombinant protein production mainly impacted amino acids based on their energy costs, a regulatory logic already reported previously by Heyland *et al.* (2011).

This comprehensive examination also led to the first determination of the macromolecular and elemental composition of *P. pastoris* (Carnicer *et al.* 2009), as well as to an update of the GEM with reactions of the lipid and energy metabolism (Tomas-Gamisans, Ferrer and Albiol 2016). Taken together, all these data highlight that oxygen availability (and also other environmental conditions such as temperature or osmolarity) had a major impact on cell physiology and cellular regulation patterns. In contrast, there was not much regulation triggered by the recombinant product itself under the conditions studied (Carnicer *et al.* 2009, 2012b; Dragosits *et al.* 2009, 2010; Baumann *et al.* 2010; Adelantado *et al.* 2017), even though the antibody Fab fragment is a rather complex protein.

Apart from strain producing the antibody Fab fragment under control of P_{GAP}, only the production of ROL was studied using multi-omics approaches. ROL was produced under control of P_{AOX1}, and the strains were cultivated in chemostats with methanol/glycerol or methanol/glucose co-feed. The impact of ROL on the *P. pastoris* central carbon metabolism included a redistribution of carbon fluxes towards increased glycolysis, TCA cycle and NADH regeneration fluxes, as well as higher methanol dissimilation rates (Jordà *et al.* 2012, 2014a). This and the lower biomass yield were considered to be an indication of metabolic burden by ROL production. Furthermore, accumulation of trehalose was observed in the ROL-producing strain (Jorda *et al.* 2014b). Especially in multicopy ROL-producing strains, a strong downregulation of methanol metabolism and peroxisome biogenesis was apparent, which also correlated to decreased methanol uptake rates. Again, an impact on the fatty acid metabolism was observed by transcriptomics analysis, in particular in the high copy number strains (Camara *et al.* 2017). Although ROL was previously shown to induce the UPR in fed-batch cultivations (Resina *et al.* 2007), only limited reactions related to folding and secretion stress were observed in chemostats with glycerol/methanol co-feeding (Camara *et al.* 2017). The same holds true for the Fab-producing strains analyzed in the environmental stress studies, which showed clear UPR induction in glucose-limited fed-batch but not in chemostat cultivations, indicating that the cultivation conditions have a severe impact on how the cells react to additional burden of heterologous protein production.

Finally, the systems-level response of *P. pastoris* to the use of methanol as a carbon source was studied by a combination of several 'omics' technologies in a single study (Russmayer *et al.* 2015a). An overview is given in Fig. 2. Cell physiology was generally characterized by a high protein production state, driven by the massive synthesis of enzymes for methanol catabolism. As a consequence, the protein translation machinery was upregulated on the proteome level, leading to the depletion of free amino acids, despite the upregulation of their synthesis pathways. Interestingly, the resulting flux distribution in the central carbon metabolism allowed a high flux through the pentose phosphate pathway, thus not only increasing the supply of reduced NADPH but also facilitating the synthesis of vitamins. In particular, pathways of riboflavin and thiamine synthesis were upregulated to produce the coenzymes required for methanol assimilation. Remarkably, the combination of regulatory changes observed by proteomics, metabolomics and transcriptomics provided strong evidence that methanol assimilation is performed entirely in the peroxisomes by a set of paralogous enzymes. Targeted application of subcellular proteomics and metabolomics then confirmed the peroxisomal localization of these enzymes (Fba1–2, Rpe1–2, Rki1–2 and Shb17), as well as the occurrence of the intermediate sedoheptulose-1,7-bisphosphate in methanol-grown cells, thereby verifying the presence of a dedicated and compartmentalized xylulose-monophosphate (XuMP) cycle (Fig. 3). The reactions and compartmentalization of the XuMP cycle were also integrated in the consolidated genome-scale model iMT1026 (Tomas-Gamisans, Ferrer and Albiol 2016), showing that the findings of omics studies advance the accuracy of GEMs.

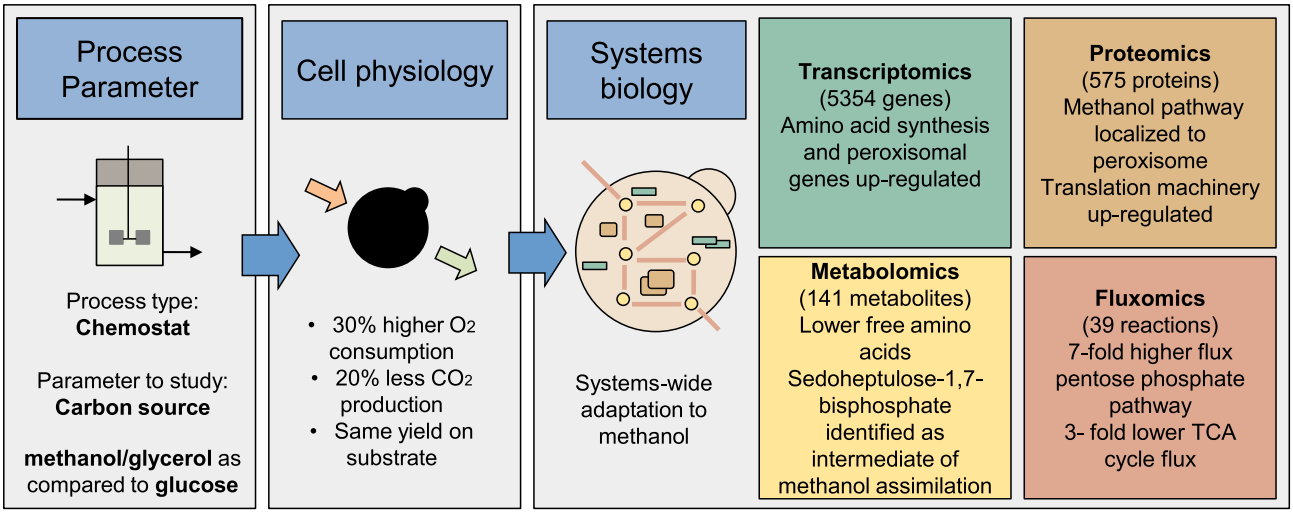

Figure 2. Systems-level response of *Pichia pastoris* to methanol as a carbon source. Integrated systems biology enabled an understanding beyond a black box model (global cell physiology parameters) and revealed major systemic adaptations of *P. pastoris* to methylotrophic lifestyle. Data are taken from Russmayer *et al.* (2015b). *P. pastoris* CBS7435 was grown in chemostat cultivations at $\mu = 0.1\ h^{-1}$, using either glucose or methanol/glycerol as carbon source. Steady-state samples were analyzed with transcriptomics (DNA microarrays), proteomics (whole cell and organelle-specific LC-MS/MS), metabolomics (LC-MS/MS and GC-MS/MS) and metabolic flux analysis (^{13}C-labelled cell extracts, GC-MS of proteinogenic amino acids). The number of quantified analytes (genes, proteins, metabolites or fluxes) is given in brackets for each type of omics-analysis.

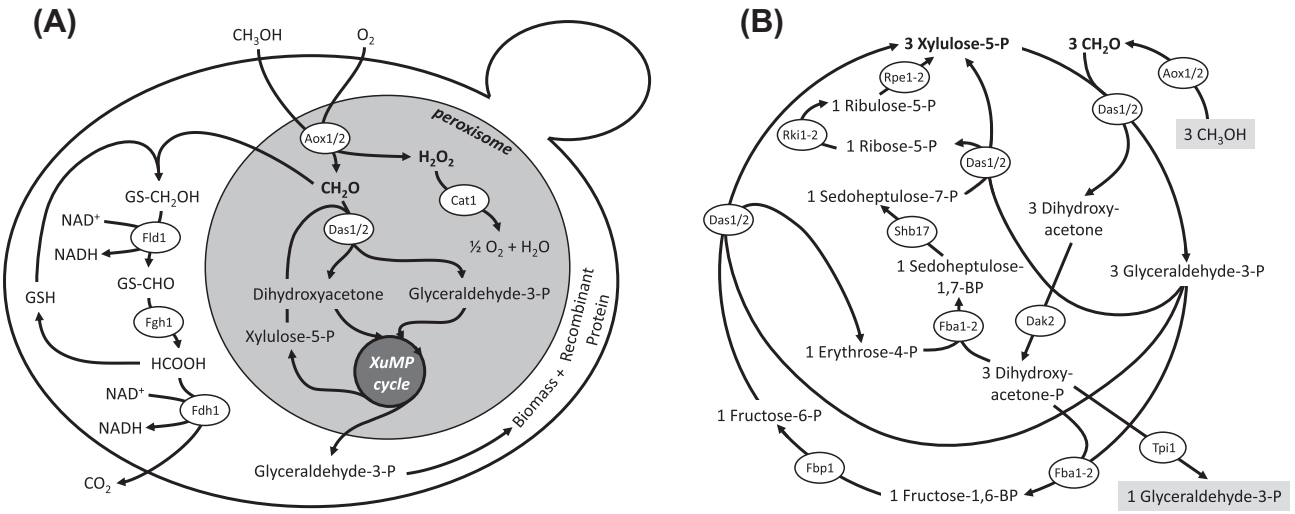

Figure 3. Metabolic pathways of methanol utilization in *Pichia pastoris*. (**A**) Overview of assimilation and dissimilation pathways. (**B**) Xylulose monophosphate (XuMP) cycle for methanol assimilation. Metabolites entering and leaving the peroxisomes are marked in gray boxes. GSH: reduced glutathione; GS-CH2OH: S-(hydroxymethyl)glutathione; GS-CHO: S-formylglutathione; Cat1: catalase; Fld1: formaldehyde dehydrogenase; Fgh1: S-formylglutathione hydrolase; Fdh1: formate dehydrogenase; Aox1/2: alcohol oxidase 1 and 2; Das 1/2: dihydroxyacetone synthase 1 and 2; Dak2: dihydroxyacetone kinase; Tpi1: triosephosphate isomerase; Fba1–2: fructose 1,6-bisphosphate aldolase; Fbp1: fructose 1,6-bisphosphatase; Shb17:sedoheptulose 1,7-bisphosphatase; Rki1–2: Ribose 5-phosphate ketol-isomerase; Rpe1–2: D-ribulose 5-phosphate 3-epimerase. (from Mattanovich, Sauer and Gasser 2017, with permission)

SYSTEMS BIOLOGY BASED STRAIN ENGINEERING FOR ENHANCED PROTEIN PRODUCTION

Synthetic biology toolbox for rapid strain generation and improvement of *Pichia pastoris*

Transcriptomics, proteomics, metabolomics and, ideally, the combination of all these information levels allow us to gain an all-encompassing picture of the biological system *P. pastoris*. This picture assists in identifying engineering targets to increase recombinant protein production. A major advantage of *P. pastoris* and yeast in general is the great variety of applicable molecular biological tools, which have been developed during the last decades. Thus, extensive cell engineering to create a designer cell factory is theoretically already possible.

Promoters are a key factor in production hosts and in strain engineering as they define gene expression levels (Mattanovich *et al.* 2012; Weinhandl *et al.* 2014; Vogl *et al.* 2016). A special type are bidirectional promoters, which are short DNA sequences driving gene expression in both directions (Geier *et al.* 2015). Promoter sequences can also be used to change the expression levels of native genes, i.e. create conditional mutants by replacing the native promoter (Delic, Mattanovich and Gasser 2013a). Transcriptomics data can be used to deduce promoters with the desired regulatory properties (e.g. strong expression,

easily regulatable). For instance, transcriptional regulation patterns obtained from cells grown under different conditions have enabled the identification of promoters with the desired characteristics for recombinant protein production: repression during the growth phase and activation during the production phase (Hartner *et al.* 2008; Stadlmayr *et al.* 2010; Gasser, Steiger and Mattanovich 2015; Redden, Morse and Alper 2015; Vogl *et al.* 2016). Similarly, P_{THI11}, the first regulated promoter independent of the main nutritional components such as carbon and nitrogen, has been identified, alongside other repressible promoters (Landes *et al.* 2016). This extended set of promoters could be used, for example, to fine-tune multigene coexpressions or synthetic biology applications (Vogl *et al.* 2016).

Recently, advanced cloning techniques such as Golden Gate assembly (Engler *et al.* 2009; Werner *et al.* 2012) or Gibson assembly (Gibson *et al.* 2009) have been adapted for *P. pastoris* and applied to facilitate generation of production strains. They enable rapid combination of different promoters, secretion signals and transcription terminators (Weninger *et al.* 2016; Obst, Lu and Sieber 2017; Prielhofer *et al.* submitted; Schreiber *et al.* 2017). Furthermore, these techniques provide an attractive solution for rapid strain engineering approaches, and were used to determine the optimal setup for CRISPR/Cas9 in *P. pastoris* (Prielhofer *et al.* submitted; Weninger *et al.* 2016).

In many cases, the chosen strategy for improving a production strain involves overexpressing a set of chosen genes. Generally, overexpression of a single helper gene already leads to an increase in the secretion of the recombinant protein (Delic *et al.* 2014b). This is routinely done by integrating an additional gene copy under the control of a strong promoter. For instance, the secretion of an antibody fragment could be improved by overexpressing genes, which were already upregulated in the protein-producing strains (Gasser *et al.* 2007b).

If more than one gene is intended to be overexpressed, this usually involves subsequent transformations with vectors containing different selection markers, possibly requiring marker recycling steps in between. As the procedure is time consuming and laborious, usually not more than two factors were combined in one strain. Alternatively, multiple expression cassettes can be assembled in a modular way. The Golden Gate assembly was recently adapted for strain engineering of *P. pastoris*, allowing combination of at least five expression cassettes on one plasmid. This work resulted in establishing the GoldenPiCS toolbox containing suitable promoters, terminators and integration loci (Prielhofer *et al.* submitted). Gibson assembly is another approach, allowing rapid and scarless cloning (Gibson *et al.* 2009), but with the drawback that for each assembly the fragments need to be PCR-amplified with customized overhangs. The modular assembly of expression cassettes allows a combinatorial cell engineering approach. Vogl *et al.* (2015) provided early examples for the application of this modular approach. They observed an up to 31-fold difference in volumetric activity in the culture supernatant, when assessing secretory expression of HRP fused to different epitope tags and fusion proteins. Schreiber *et al.* (2017) and Obst, Lu and Sieber (2017) used the assembly to optimize their genetic construct for protein secretion mainly by varying the promoter and secretion signal. Ata *et al.* (2017) used the GoldenPiCS system to assemble and investigate promoter variants.

In addition to the commonly used genomic integration (Cregg, Vedvick and Raschke 1993), an episomal vector was constructed using an autonomously replicating sequence from *Kluyveromyces lactis* (Camattari *et al.* 2016). The vector enabled the overexpression of genetic constructs with low interclonal variability. Moreover, *in vivo* assembly of DNA fragments using overlapping regions has been demonstrated and might facilitate high-throughput construction of an expression system (Mizutani *et al.* 2011).

Finally, the creation of gene knockouts or replacements can be performed using split-marker cassettes or the CRISPR/Cas9 system (Gasser *et al.* 2013; Weninger *et al.* 2016). The CRISPR/Cas9 system allows marker-less genome engineering and the creation of multiple simultaneous knockouts in *P. pastoris*.

Enhanced protein folding and secretion

Components of the secretory pathway are often identified as the target for strain engineering after gaining knowledge from omics analyses. An example for this workflow is a transcriptomic study of *P. pastoris*, which identified significantly regulated genes in a strain overexpressing human trypsinogen versus that of a non-expressing strain in a methanol fed batch (Gasser *et al.* 2007b). The identified factors were chosen for an overexpression construct and were also able to increase secretion of a Fab fragment in glucose-based fed-batch cultivations, indicating general bottlenecks in the secretory pathway. The majority of work focused on overexpressing factors involved in protein folding, such as chaperones or disulfide isomerases (reviewed by Delic *et al.* 2014a; Puxbaum, Mattanovich and Gasser 2015). The central transcription factor for chaperones in the ER is Hac1, which induces the UPR. Overexpression of the induced variant of this transcription factor increased the yields of recombinant proteins significantly (Gasser *et al.* 2006; Guerfal *et al.* 2010). Interestingly, the overexpression of the ER resident disulfide bond isomerase Pdi1 improved production of antibody fragments even more than Hac1 overexpression (Gasser *et al.* 2006). It also increased secretion of a recombinant potential vaccine candidate in a dose-dependent manner (Inan *et al.* 2006). Extensive disulfide bond formation in the ER causes oxidative stress (Delic *et al.* 2012). Overexpression of the transcription factor Yap1, regulating the transcriptional response to oxidative stress, reduced oxidative stress levels and significantly increased trypsinogen secretion (Delic *et al.* 2014b).

Secretion signals are needed for targeting the recombinant proteins into the secretory pathway. The *S. cerevisiae* α-mating factor (MFα) prepro-peptide is still the most commonly used signal sequence (Lin-Cereghino *et al.* 2013). Proteomics analysis of the *P. pastoris* secretome revealed Epx1 as the most abundant secretory protein (Heiss *et al.* 2013). From this Epx1 secretion signal, variants were constructed and they proved to be even more efficient than the commonly used secretion leaders (Heiss *et al.* 2015). Other signal sequences were determined *in silico* from analyzing *P. pastoris*' secretome (Massahi and Calik 2015). A selection thereof was tested and successfully secreted the recombinant protein, but slightly less efficiently than the MFα signal sequence (Massahi and Calik 2016).

Another possible major branch-off on the secretory pathway towards degradation is the ERAD system. ERAD is thought to be linked to the UPR (Friedlander *et al.* 2000) which, in turn, is triggered by the high expression of secretory recombinant protein (Mattanovich *et al.* 2004; Sha *et al.* 2013). Both ERAD components and vacuolar proteins were identified as intracellular interaction partners of a recombinant Fab fragment in the interactome study (Pfeffer *et al.* 2012). In ERAD, the final step is proteasomal degradation; its significance was shown and the amount of antibody fragment lost in this way was determined to be 58% (Pfeffer *et al.* 2011). On the other hand, the inhibition of vacuolar protein targeting (using either a selective mutation of the sorting

receptor Vps10 or cells disrupted in the CORVET vacuole tethering pathway) was able to suppress mistargeting of the recombinant protein to the vacuole and boosted secretion efficiency (Fitzgerald and Glick 2014; Marsalek *et al.* 2017).

The final step on the secretory pathway is the exocytotic fusion of secretory vesicles with the plasma membrane and the diffusion of the recombinant protein through the cell wall. The cell wall may or may not be a barrier (De Nobel and Barnett 1991). The multi-omics study investigating hypoxic cultivation conditions on glucose identified ergosterol depletion, which in turn hampers Gas1 incorporation into the cell wall (Baumann *et al.* 2011). This may affect the cell wall in a similar way as *GAS1* knockout strains (Marx *et al.* 2006). Gas1 highly affects the structure and permeability of the yeast cell wall. The knockout had a positive effect on lipase secretion (Marx *et al.* 2006).

Engineering of metabolism

Recombinant protein production is an energy and redox-intensive process. To cope with the metabolic burden of protein overexpression, significant metabolic rearrangements have been observed in protein-producing strains of *P. pastoris*. In particular, cell metabolism adapts to increase the supply of ATP, NADH and NADPH while reducing by-product formation (Heyland *et al.* 2011; Nie *et al.* 2014; Nocon *et al.* 2014). However, this metabolic plasticity is limited. Thus, the available cofactors and precursors can become insufficient, leading to depletion of the intracellular pools of nucleotides, tRNAs and amino acids (Klein, Niklas and Heinzle 2015; Liu *et al.* 2016). Protein secretion further drains precursors and energy from the central carbon metabolism (Klein, Niklas and Heinzle 2015). Proper folding and disulfide bond formation in the ER consumes ATP and NADPH molecules. Excessive protein misfolding can activate the UPR, which demands further energy for chaperone activity. Also, protein glycosylation requires the synthesis of expensive nucleotide-activated sugars. Finally, synthesis of the secretory machinery drains acetyl-CoA and NADPH from the cytoplasm, and protein transportation itself consumes energy in the form of GTP.

In order to meet these demands, the central carbon metabolism can be modified towards an optimal metabolic flux distribution. Although this approach is still an emerging field within the systems biotechnology of *P. pastoris* and yeasts in general, some interesting examples can already be mentioned. Using a GEM of *P. pastoris*, genetic engineering targets increasing the synthesis of human superoxide dismutase were identified (Nocon *et al.* 2014). Genes to be overexpressed encoded enzymes in the pentose phosphate pathway, TCA cycle and glutamate metabolism, whereas gene deletions were found in by-product pathways. Most of the tested predictions improved intracellular protein production in shake flask cultivations, using glucose as a carbon source. In particular, combinatorial overexpression of enzymes in the pentose phosphate pathway enabled a higher flux through the oxidative part of the pathway, which resulted in a 3.8-fold increase in protein production (Nocon *et al.* 2016). Overexpression of mitochondrial malate dehydrogenase also resulted in a significant increase in protein production as well as reduced fermentation products. In this case, a higher metabolic flux was diverted through the TCA cycle, possibly increasing the supply of ATP. Gene deletions on fermentation pathways resulted in improved protein production except for the deletion of *ALD4*, which generated a metabolic overflow towards ethanol production (Nocon *et al.* 2014).

Further metabolic engineering strategies beyond the central carbon metabolism could also benefit protein production in *P. pastoris*. By using a kinetic model of the methanol dissimilation pathway, formaldehyde dehydrogenase (FLD) was identified as the flux controlling step. As expected, FLD overexpression resulted in a higher NADH regeneration capacity, which could be exploited for protein production under the *AOX1* promoter (Schroer *et al.* 2010). Similarly, a biotin-prototrophic strain of *P. pastoris* was constructed by expression of enzymes missing in the biosynthesis pathway of biotin (Gasser, Dragosits and Mattanovich 2010). The strain was able to grow on a minimal medium without vitamins, which could reduce the batch to batch variability and costs of the medium.

Protein production can be limited by factors beyond the metabolic burden. Depending on the growth conditions, protein overexpression reduces growth rate via a transcriptional or translational burden (Kafri *et al.* 2016). For *S. cerevisiae* growing in rich medium, the reduced growth was explained by a limitation in translation initiation. Possibly, the highly abundant transcripts of the recombinant protein compete for limited ribosomes that would otherwise translate endogenous proteins (Kafri *et al.* 2016). Thus, even in the favorable metabolic state provided by cultivation in a rich medium, a limited translational machinery guarantees a trade-off between growth rate and heterologous protein production.

Process engineering based on systems biology

So far, systems biology has mainly supported process engineering by providing mechanistic explanations of the effect of process variables on cell factory performance. Several process parameters affecting protein production in *P. pastoris* have been studied. Among the environmental factors, pH, osmolarity, oxygen availability and temperature appear to be particularly important (Mattanovich *et al.* 2004). Recently, Burgard *et al.* (2017) applied transcriptomics of *P. pastoris* cells grown in glucose- or methanol-limited fed-batch cultivations to identify marker genes for macronutrient limitation, and subsequent media optimization.

Regarding the optimal temperature, lowering the cultivation temperature from 30°C or 25°C to 20°C led to a 3-fold increase in specific productivity of two different secreted Fab antibody fragments in glucose-based bioreactor cultivations (Gasser *et al.* 2007a; Dragosits *et al.* 2009, 2011). Even though there was a remarkable rearrangement of several metabolic enzymes in the central carbon and amino acid metabolism, the benefit of the reduced temperature was mainly attributed to reduced cell and folding stress at the low temperature. The positive effect on Fab secretion was thus explained as a side effect of the lower cellular stress levels. On the other hand, genes involved in vesicular transport were shown to be upregulated in this condition, and may also contribute to higher secretion of heterologous proteins at the lower temperature (Gasser *et al.* 2007a). Decreasing temperature was also shown to be beneficial for methanol-grown cells but, in this case, low temperature was reported to decrease cell lysis and thus the release of proteases (Jahic *et al.* 2003).

Oxygen availability is another critical parameter for high cell density fermentations of *P. pastoris*. It typically limits fermentation scale-up because uneconomical oxygen transfer rates are required. Although usually normoxic conditions are applied during bioreactor cultivations, increased secretion capacity and preferred product quality are observed when cells are cultivated in hypoxic or oxygen-limited conditions (Baumann *et al.* 2008;

Berdichevsky *et al.* 2011; Gunes and Calik 2016). It should be noted that the biological consequences of oxygen availability are fundamentally different if cells are grown on methanol compared to glucose. Potential cellular mechanisms that correlate with increased productivity in glucose-based processes were already discussed above. Oxygen supply is even more critical during methanol-based production, as oxygen is a co-substrate of methanol oxidation and stoichiometrically linked to MUT by alcohol oxidase. No studies have yet been made to explain why oxygen limitation positively influences heterologous protein production in methanol-grown cells.

The use of methanol as a carbon source and inducer of protein production leads to high translational capacity, thus partially explaining the success of *P. pastoris* as an expression host (Prielhofer *et al.* 2015; Russmayer *et al.* 2015a). In this respect, metabolomics studies and flux analyses suggested cellular mechanisms by which the use of mixed feeds of methanol and glycerol, glucose or sorbitol alleviates the metabolic burden derived from heterologous protein production (Celik, Calik and Oliver 2010) and increase productivity (Jungoand Marison 2007; Jungo *et al.* 2007; Celik, Calik and Oliver 2009; Paulova *et al.* 2012). However, methanol concentrations and feeding rates used for fed-batch cultivations must be carefully selected, particularly when using pure methanol as it has been shown to impact protein synthesis capabilities. High productivity of scFv productivity was observed at high methanol concentrations (up to 25 g/L), but autolysis and degradation also increased (Fujiki, Kumada and Kishimoto 2015). Furthermore, a glycerol batch phase is typically used prior to the production phase with methanol as it has been found to be beneficial for recombinant protein production. The higher levels of UPR-related proteins in the glycerol batch seem to prepare the cells for effective secretion in the methanol induction phase (Vanz, Nimtz and Rinas 2014). Similarly, in the usually applied batch media, cells face a hyperosmotic environment due to the initial high salt concentrations in the medium. Using systems-level analyses, it was shown that the UPR may also pre-condition *P. pastoris* to withstand high levels of extracellular osmolytes (Dragosits *et al.* 2010). Thus, certain environmental stimuli may elicit a cross-protective effect to a second stress situation in *P. pastoris*, a phenomena already known for other yeasts (Jamieson 1998).

CONCLUSIONS AND OUTLOOK

Systems biotechnology of *P. pastoris* has seen significant progress over the last two decades. Driven by the advances in omics technologies, it has improved our understanding of *P. pastoris* as a host for protein production, and its genomic information and regulatory responses at the transcriptomic, proteomic, metabolomic and fluxomic levels. Transcriptomics studies, supported by whole cell proteomics analyses, showed that a high degree of gene regulation occurs already at the transcript level. Targeted proteomics proved very useful in elucidating the correct localization of several metabolic enzymes, including the peroxisomal localization of the methanol assimilation pathway. The development of technologies for metabolome quantification and fluxomics revealed metabolic adjustments to different media and the stress of protein production. Benefiting from the constantly improving genome annotation, several metabolic models have now been published, enabling the prediction of rational cell engineering targets. In particular, multiple successful engineering targets improving protein production have been identified in the central carbon metabolism, confirming the relevance of the metabolic responses observed in many of the reviewed studies.

Most omics studies so far have been conducted in chemostat cultures, which is not the preferred way of production process to date. Hence, there is a need to understand cellular responses occurring during fed-batch cultivation, moving beyond the phases of methanol induction and adaptation in early fed batch to truly understand the limitations occurring during a production process. On the other hand, continuous processing might be a promising alternative based on the observations made during the omics studies. Interestingly, several recombinant strains did not show strong activation of the UPR and other stress responses when cultivated in chemostat, while a clear UPR induction and related reactions were observed when they were cultivated in fed-batch mode. This points towards lower stress levels in continuously cultivated cells. Elucidating the cellular processes underlying this observation, as well as the genomic stability and adaptation of *P. pastoris* in long continuous cultures, would require further systems biology studies. Ideally, interpretation of data resulting from these studies should be followed up by confirmative experiments of the observed changes and, if possible, by experimental testing of the hypotheses developed.

In addition, process engineering could benefit from a systems biotechnology perspective. As exemplified by studies on the impact of oxygen, temperature or osmolarity on cell physiology, similar stress responses are triggered by different bioreactor parameter, which suggests the possibility of exploiting shared regulatory mechanisms to obtain strains that are robust against multiple environmental changes.

Furthermore, the design of optimal strains would require expanding our biological models to include kinetic parameters, compartment-specific measurements, regulatory networks and the secretory machinery. Integrating omics-derived large datasets into such comprehensive models would allow us to understand non-linear interactions between cellular components, as well as to identify complex combinations leading to improved production and reduced degradation of recombinant proteins. In the long term, in-depth understanding of the systems biology of *P. pastoris* could thus enable major cell reprogramming into a more robust and efficient secretory cell factory.

AUTHORS' CONTRIBUTIONS

RJZ and DAP contributed equally to this manuscript. RJZ focused on transcriptomics and proteomics, DAP on metabolomics and process engineering, and BG on protein folding and secretion. DM drafted the introduction. All authors contributed to the structure of this review, and to the outlook.

FUNDING

Work on *Pichia pastoris* in our laboratory was 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 Business Agency Vienna through the COMET Funding Program managed by the Austrian Research Promotion Agency FFG. DAP is funded by the Austrian Science Fund (FWF): Doctoral Program BioToP—Biomolecular Technology of Proteins (FWF W1224).

Conflict of Interest. None declared.

REFERENCES

Adelantado N, Tarazona P, Grillitsch K *et al.* The effect of hypoxia on the lipidome of recombinant *Pichia pastoris*. *Microb Cell Fact* 2017;**16**:86.

Ata Ö, Prielhofer R, Gasser B *et al.* Transcriptional engineering of the glyceraldehyde-3-phosphate dehydrogenase promoter for improved heterologous protein production in *Pichia pastoris*. *Biotechnol Bioeng* 2017;**114**:2319–27.

Austin RJ, Kuestner RE, Chang DK *et al.* SILAC compatible strain of *Pichia pastoris* for expression of isotopically labeled protein standards and quantitative proteomics. *J Proteome Res* 2011;**10**:5251–9.

Aw R, Barton GR, Leak DJ. Insights into the prevalence and underlying causes of clonal variation through transcriptomic analysis in *Pichia pastoris*. *Appl Microbiol Biot* 2017;**101**:5045–58.

Barrigon JM, Valero F, Montesinos JL. A macrokinetic model-based comparative meta-analysis of recombinant protein production by *Pichia pastoris* under *AOX1* promoter. *Biotechnol Bioeng* 2015;**112**:1132–45.

Baumann K, Adelantado N, Lang C *et al.* Protein trafficking, ergosterol biosynthesis and membrane physics impact recombinant protein secretion in *Pichia pastoris*. *Microb Cell Fact* 2011;**10**:93.

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

Baumann K, Maurer M, Dragosits M *et al.* Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins. *Biotechnol Bioeng* 2008;**100**:177–83.

Berdichevsky M, d'Anjou M, Mallem MR *et al.* Improved production of monoclonal antibodies through oxygen-limited cultivation of glycoengineered yeast. *J Biotechnol* 2011;**155**:217–24.

Berrios J, Flores MO, Diaz-Barrera A *et al.* A comparative study of glycerol and sorbitol as co-substrates in methanol-induced cultures of *Pichia pastoris*: temperature effect and scale-up simulation. *J Ind Microbiol Biot* 2017;**44**:407–11.

Bridges HR, Fearnley IM, Hirst J. The subunit composition of mitochondrial NADH:ubiquinone oxidoreductase (complex I) from *Pichia pastoris*. *Mol Cell Proteomics* 2010;**9**:2318–26.

Bridges HR, Grgic L, Harbour ME *et al.* The respiratory complexes I from the mitochondria of two *Pichia* species. *Biochem J* 2009;**422**:151–9.

Buchetics M, Dragosits M, Maurer M *et al.* Reverse engineering of protein secretion by uncoupling of cell cycle phases from growth. *Biotechnol Bioeng* 2011;**108**:2403–12.

Burgard J, Valli M, Graf AB *et al.* Biomarkers allow detection of nutrient limitations and respective supplementation for elimination in *Pichia pastoris* fed-batch cultures. *Microb Cell Fact* 2017;**16**:117.

Calik P, Sahin M, Taspinar H *et al.* Dynamic flux balance analysis for pharmaceutical protein production by *Pichia pastoris*: human growth hormone. *Enzyme Microb Technol* 2011;**48**:209–16.

Camara E, Landes N, Albiol J *et al.* Increased dosage of *AOX1* promoter-regulated expression cassettes leads to transcription attenuation of the methanol metabolism in *Pichia pastoris*. *Sci Rep* 2017;**7**:44302.

Camattari A, Goh A, Yip LY *et al.* Characterization of a panARS-based episomal vector in the methylotrophic yeast *Pichia pastoris* for recombinant protein production and synthetic biology applications. *Microb Cell Fact* 2016;**15**:139.

Carnicer M, Baumann K, Toplitz I *et al.* Macromolecular and elemental composition analysis and extracellular metabolite balances of *Pichia pastoris* growing at different oxygen levels. *Microb Cell Fact* 2009;**8**:65.

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.

Caspeta L, Shoaie S, Agren R *et al.* Genome-scale metabolic reconstructions of *Pichia stipitis* and *Pichia pastoris* and *in silico* evaluation of their potentials. *BMC Syst Biol* 2012;**6**:24.

Celik E, Calik P, Oliver SG. Fed-batch methanol feeding strategy for recombinant protein production by *Pichia pastoris* in the presence of co-substrate sorbitol. *Yeast* 2009;**26**:473–84.

Celik E, Calik P, Oliver SG. Metabolic flux analysis for recombinant protein production by *Pichia pastoris* using dual carbon sources: Effects of methanol feeding rate. *Biotechnol Bioeng* 2010;**105**:317–29.

Chung BK, Selvarasu S, Andrea C *et al.* Genome-scale metabolic reconstruction and *in silico* analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. *Microb Cell Fact* 2010;**9**:50.

Coughlan AY, Hanson SJ, Byrne KP *et al.* Centromeres of the yeast *Komagataella phaffii* (*Pichia pastoris*) have a simple inverted-repeat structure. *Genome Biol Evol* 2016;**8**:2482–92.

Cregg JM, Barringer KJ, Hessler AY *et al.* *Pichia pastoris* as a host system for transformations. *Mol Cell Biol* 1985;**5**:3376–85.

Cregg JM, Vedvick TS, Raschke WC. Recent advances in the expression of foreign genes in *Pichia pastoris*. *Biotechnology* 1993;**11**:905–10.

De Nobel JG, Barnett JA. Passage of molecules through yeast cell walls: a brief essay-review. *Yeast* 1991;**7**:313–23.

De Schutter K, Lin YC, Tiels P *et al.* Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nat Biotechnol* 2009;**27**:561–6.

Delic M, Gongrich R, Mattanovich D *et al.* Engineering of protein folding and secretion-strategies to overcome bottlenecks for efficient production of recombinant proteins. *Antioxid Redox Signal* 2014a;**21**:414–37.

Delic M, Graf AB, Koellensperger G *et al.* Overexpression of the transcription factor Yap1 modifies intracellular redox conditions and enhances recombinant protein secretion. *Microb Cell* 2014b;**1**:376–86.

Delic M, Mattanovich D, Gasser B. Repressible promoters - a novel tool to generate conditional mutants in *Pichia pastoris*. *Microb Cell Fact* 2013a;**12**:6.

Delic M, Rebnegger C, Wanka F *et al.* Oxidative protein folding and unfolded protein response elicit differing redox regulation in endoplasmic reticulum and cytosol of yeast. *Free Radic Biol Med* 2012;**52**:2000–12.

Delic M, Valli M, Graf AB *et al.* The secretory pathway: exploring yeast diversity. *FEMS Microbiol Rev* 2013b;**37**:872–914.

Dragosits M, Frascotti G, Bernard-Granger L *et al.* Influence of growth temperature on the production of antibody Fab fragments in different microbes: a host comparative analysis. *Biotechnol Prog* 2011;**27**:38–46.

Dragosits M, Stadlmann J, Albiol J *et al.* The effect of temperature on the proteome of recombinant *Pichia pastoris*. *J Proteome Res* 2009;**8**:1380–92.

Dragosits M, Stadlmann J, Graf A *et al*. The response to unfolded protein is involved in osmotolerance of *Pichia pastoris*. *BMC Genomics* 2010;**11**:207.

Edwards-Jones B, Aw R, Barton GR *et al*. Translational arrest due to cytoplasmic redox stress delays adaptation to growth on methanol and heterologous protein expression in a typical fed-batch culture of *Pichia pastoris*. *PLoS One* 2015;**10**:e0119637.

Engler C, Gruetzner R, Kandzia R *et al*. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One* 2009;**4**:e5553.

Eskitoros MŞ, Ata Ö, Çalık P. Metabolic reaction network of *Pichia pastoris* with glycosylation reactions: Flux analysis for erythropoietin production. *J Chem Technol Biot* 2017;**89**:1675–85.

Ferrer P, Albiol J. 13C-based metabolic flux analysis of recombinant *Pichia pastoris*. *Methods Mol Biol* 2014;**1191**:291–313.

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

Fitzgerald I, Glick BS. Secretion of a foreign protein from budding yeasts is enhanced by cotranslational translocation and by suppression of vacuolar targeting. *Microb Cell Fact* 2014;**13**:125.

Forster J, Halbfeld C, Zimmermann M *et al*. A blueprint of the amino acid biosynthesis network of hemiascomycetes. *FEMS Yeast Res* 2014;**14**:1090–100.

Friedlander R, Jarosch E, Urban J *et al*. A regulatory link between ER-associated protein degradation and the unfolded-protein response. *Nat Cell Biol* 2000;**2**379–84.

Fujiki Y, Kumada Y, Kishimoto M. Phase analysis in single-chain variable fragment production by recombinant *Pichia pastoris* based on proteomics combined with multivariate statistics. *J Biosci Bioeng* 2015;**120**:187–92.

Gasser B, Dragosits M, Mattanovich D. Engineering of biotin-prototrophy in *Pichia pastoris* for robust production processes. *Metab Eng* 2010;**12**:573–80.

Gasser B, Maurer M, Gach J *et al*. Engineering of *Pichia pastoris* for improved production of antibody fragments. *Biotechnol Bioeng* 2006;**94**:353–61.

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

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.

Gasser B, Sauer M, Maurer M *et al*. Transcriptomics-based identification of novel factors enhancing heterologous protein secretion in yeasts. *Appl Environ Microb* 2007b;**73**:6499–507.

Gasser B, Steiger MG, Mattanovich D. Methanol regulated yeast promoters: production vehicles and toolbox for synthetic biology. *Microb Cell Fact* 2015;**14**:196.

Geier M, Fauland P, Vogl T *et al*. Compact multi-enzyme pathways in *P. pastoris*. *Chem Commun* 2015;**51**:1643–6.

Gibson DG, Young L, Chuang RY *et al*. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 2009;**6**:343–5.

Goodacre R, Vaidyanathan S, Dunn WB *et al*. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol* 2004;**22**:245–52.

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

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

Guerfal M, Ryckaert S, Jacobs PP *et al*. The HAC1 gene from *Pichia pastoris*: characterization and effect of its overexpression on the production of secreted, surface displayed and membrane proteins. *Microb Cell Fact* 2010;**9**:49.

Guilliermond A. *Zygosaccharomyces pastori*, nouvelle espèce de levures copulation hétérogamique. *Bull Soc Mycol France* 1920;**36**:203–11.

Gunes H, Calik P. Oxygen transfer as a tool for fine-tuning recombinant protein production by *Pichia pastoris* under glyceraldehyde-3-phosphate dehydrogenase promoter. *Bioprocess Biosyst Eng* 2016;**39**:1061–72.

Guo M, Liu X, Huang M *et al*. ^{13}C-assisted Ultra-high performance liquid chromatography-triple quadrupole mass spectrometry method for precise determination of intracellular metabolites in *Pichia pastoris*. *Chinese J Anal Chem* 2016;**44**:232–40.

Gutierrez JM, Lewis NE. Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. *Biotechnol J* 2015;**10**:939–49.

Haberhauer-Troyer C, Delic M, Gasser B *et al*. Accurate quantification of the redox-sensitive GSH/GSSG ratios in the yeast *Pichia pastoris* by HILIC-MS/MS. *Anal Bioanal Chem* 2013;**405**:2031–9.

Hanson SJ, Byrne KP, Wolfe KH. Mating-type switching by chromosomal inversion in methylotrophic yeasts suggests an origin for the three-locus *Saccharomyces cerevisiae* system. *P Natl Acad Sci USA* 2014;**111**:E4851–8.

Hartner FS, Ruth C, Langenegger D *et al*. Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. *Nucleic Acids Res* 2008;**36**:e76.

Heiss S, Maurer M, Hahn R *et al*. Identification and deletion of the major secreted protein of *Pichia pastoris*. *Appl Microbiol Biot* 2013;**97**:1241–9.

Heiss S, Puxbaum V, Gruber C *et al*. Multistep processing of the secretion leader of the extracellular protein Epx1 in *Pichia pastoris* and implications for protein localization. *Microbiology* 2015;**161**:1356–68.

Hesketh AR, Castrillo JI, Sawyer T *et al*. Investigating the physiological response of *Pichia* (*Komagataella*) pastoris GS115 to the heterologous expression of misfolded proteins using chemostat cultures. *Appl Microbiol Biot* 2013;**97**:9747–62.

Heyland J, Fu J, Blank LM *et al*. Carbon metabolism limits recombinant protein production in *Pichia pastoris*. *Biotechnol Bioeng* 2011;**108**:1942–53.

Hohenblum H, Gasser B, Maurer M *et al*. Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*. *Biotechnol Bioeng* 2004;**85**:367–75.

Huang CJ, Damasceno LM, Anderson KA *et al*. A proteomic analysis of the *Pichia pastoris* secretome in methanol-induced cultures. *Appl Microbiol Biot* 2011;**90**:235–47.

Inan M, Aryasomayajula D, Sinha J *et al*. Enhancement of protein secretion in *Pichia pastoris* by overexpression of protein disulfide isomerase. *Biotechnol Bioeng* 2006;**93**:771–8.

Irani ZA, Kerkhoven EJ, Shojaosadati SA *et al*. Genome-scale metabolic model of *Pichia pastoris* with native and humanized glycosylation of recombinant proteins. *Biotechnol Bioeng* 2016;**113**:961–9.

Jahic M, Wallberg F, Bollok M *et al*. Temperature limited fed-batch technique for control of proteolysis in *Pichia pastoris* bioreactor cultures. *Microb Cell Fact* 2003;**2**:6.

Jamieson DJ () Oxidative stress responses of the yeast *Saccharomyces cerevisiae*. *Yeast* 1998;**14**:1511–27.

Jorda J, de Jesus SS, Peltier S *et al*. 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* 2014a;**31**:120–32.

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

Jorda J, Rojas HC, Carnicer M *et al*. Quantitative metabolomics and instationary ^{13}C-metabolic flux analysis reveals impact of recombinant protein production on trehalose and energy metabolism in *Pichia pastoris*. *Metabolites* 2014b;**4**:281–99.

Jorda J, Suarez C, Carnicer M *et al*. Glucose-methanol co-utilization in *Pichia pastoris* studied by metabolomics and instationary 13C flux analysis. *BMC Syst Biol* 2013;**7**:17.

Jungo C, Marison I, von Stockar U Mixed feeds of glycerol and methanol can improve the performance of *Pichia pastoris* cultures: A quantitative study based on concentration gradients in transient continuous cultures. *J Biotechnol* 2007;**128**:824–37.

Jungo C, Schenk J, Pasquier M *et al*. A quantitative analysis of the benefits of mixed feeds of sorbitol and methanol for the production of recombinant avidin with *Pichia pastoris*. *J Biotechnol* 2007;**131**:57–66.

Kafri M, Metzl-Raz E, Jona G *et al*. The cost of protein production. *Cell Rep* 2016;**14**:22–31.

Klavins K, Chu DB, Hann S *et al*. Fully automated on-line two-dimensional liquid chromatography in combination with ESI MS/MS detection for quantification of sugar phosphates in yeast cell extracts. *Analyst* 2014;**139**:1512–20.

Klein T, Niklas J, Heinzle E. Engineering the supply chain for protein production/secretion in yeasts and mammalian cells. *J Ind Microbiol Biotechnol* 2015;**42**:453–64.

Klug L, Tarazona P, Gruber C *et al*. The lipidome and proteome of microsomes from the methylotrophic yeast *Pichia pastoris*. *Biochim Biophys Acta* 2014;**1841**:215–26.

Kurtzman CP. Description of *Komagataella phaffii* sp. nov. and the transfer of *Pichia pseudopastoris* to the methylotrophic yeast genus *Komagataella*. *Int J Syst Evol Micr* 2005;**55**:973–6.

Landes N, Gasser B, Vorauer-Uhl K *et al*. The vitamin-sensitive promoter P_{THI11} enables pre-defined autonomous induction of recombinant protein production in *Pichia pastoris*. *Biotechnol Bioeng* 2016;**113**:2633–43.

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

Liachko I, Youngblood RA, Tsui K *et al*. GC-rich DNA elements enable replication origin activity in the methylotrophic yeast *Pichia pastoris*. *PLoS Genet* 2014;**10**:e1004169.

Liang S, Wang B, Pan L *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.

Lin XQ, Liang SL, Han SY *et al*. 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.

Lin-Cereghino GP, Stark CM, Kim D *et al*. The effect of alpha-mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*. *Gene* 2013;**519**:311–7.

Liu B, Zhang Y, Zhang X *et al*. Discovery of a rhamnose utilization pathway and rhamnose-inducible promoters in *Pichia pastoris*. *Sci Rep* 2013;**6**:27352.

Liu P, Huang M, Guo M *et al*. Combined 13C-assisted metabolomics and metabolic flux analysis reveals the impacts of glutamate on the central metabolism of high β-galactosidase-producing *Pichia pastoris*. *Bioresource Bioprocess* 2016;**3**:47.

Looser V, Luthy D, Straumann M *et al*. Effects of glycerol supply and specific growth rate on methanol-free production of CALB by *P. pastoris*: functional characterisation of a novel promoter. *Appl Microbiol Biot* 2017;**101**:3163–76.

Love KR, Shah KA, Whittaker CA *et al*. Comparative genomics and transcriptomics of *Pichia pastoris*. *BMC Genomics* 2016;**17**:550.

Mairinger T, Steiger M, Nocon J *et al*. Gas chromatography-quadrupole time-of-flight mass spectrometry-based determination of isotopologue and tandem mass isotopomer fractions of primary metabolites for ^{13}C-metabolic flux analysis. *Anal Chem* 2015;**87**:11792–802.

Mansur M, Cabello C, Hernandez L *et al*. Multiple gene copy number enhances insulin precursor secretion in the yeast *Pichia pastoris*. *Biotechnol Lett* 2005;**27**:339–45.

Marsalek L, Gruber C, Altmann F *et al*. Disruption of genes involved in CORVET complex leads to enhanced secretion of heterologous carboxylesterase only in protease deficient *Pichia pastoris*. *Biotechnol J* 2017;**12**:1600584.

Marx H, Mecklenbrauker A, Gasser B *et al*. Directed gene copy number amplification in *Pichia pastoris* by vector integration into the ribosomal DNA locus. *FEMS Yeast Res* 2009;**9**:1260–70.

Marx H, Sauer M, Resina D *et al*. Cloning, disruption and protein secretory phenotype of the *GAS1* homologue of *Pichia pastoris*. *FEMS Microbiol Lett* 2006;**264**:40–7.

Massahi A, Calik P. *In-silico* determination of *Pichia pastoris* signal peptides for extracellular recombinant protein production. *J Theor Biol* 2015;**364**:179–88.

Massahi A, Calik P. Endogenous signal peptides in recombinant protein production by *Pichia pastoris*: From *in-silico* analysis to fermentation. *J Theor Biol* 2016;**408**:22–33.

Mattanovich D, Branduardi P, Dato L *et al*. Recombinant protein production in yeasts. *Methods Mol Biol* 2012;**824**:329–58.

Mattanovich D, Gasser B, Hohenblum H *et al*. Stress in recombinant protein producing yeasts. *J Biotechnol* 2004;**113**:121–35.

Mattanovich D, Graf A, Stadlmann J *et al*. Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*. *Microb Cell Fact* 2009;**8**:29.

Mattanovich D, Jungo C, Wenger J *et al*. *Yeast Suspension Culture*. Wiley-VCH Verlag GmbH and KGaA, Germany. 2014.

Mattanovich M, Russmayer H, Scharl-Hirsch T *et al*. Metabolomics of *Pichia pastoris*: Impact of buffering conditions on the kinetics and nature of metabolite loss during quenching. *FEMS Yeast Res* 2017;**17**:fox016.

Mattanovich D, Sauer M, Gasser B. *Industrial Microorganisms: Pichia pastoris*. Wiley-Verlag GmbH and KGaA, Germany. 2017.

Mizutani K, Yoshioka S, Mizutani Y *et al*. High-throughput construction of expression system using yeast *Pichia pastoris*, and its application to membrane proteins. *Protein Expr Purif* 2011;**77**:1–8.

Naumov GI, Naumova ES, Tyurin OV *et al*. *Komagataella kurtzmanii* sp. nov. a new sibling species of *Komagataella* (*Pichia*) *pastoris* based on multigene sequence analysis. *Anton Leeuw* 2013;**104**:339–47.

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.

Nie Y, Huang M, Lu J *et al*, Impacts of high beta-galactosidase expression on central metabolism of recombinant *Pichia pastoris* GS115 using glucose as sole carbon source via ^{13}C metabolic flux analysis. *J Biotechnol* 2014;**187**:124–34.

Niu H, Jost L, Pirlot N *et al.* A quantitative study of methanol/sorbitol co-feeding process of a *Pichia pastoris* Mut$^+$/pAOX1-lacZ strain. *Microbial Cell Fact* 2013;**12**:33.

Nocon J, Steiger M, Mairinger T *et al.* Increasing pentose phosphate pathway flux enhances recombinant protein production in *Pichia pastoris*. *Appl Microbiol Biot* 2016;**100**:5955–63.

Nocon J, Steiger MG, Pfeffer M *et al.* Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production. *Metab Eng* 2014;**24**:129–38.

Obst U, Lu TK, Sieber V. A modular toolkit for generating *Pichia pastoris* secretion libraries. *ACS Synth Biol* 2017;**6**:1016–25.

Papanikou E, Glick BS. The yeast Golgi apparatus: insights and mysteries. *FEBS Lett* 2009;**583**:3746–51.

Paulova L, Hyka P, Branska B *et al.* Use of a mixture of glucose and methanol as substrates for the production of recombinant trypsinogen in continuous cultures with *Pichia pastoris* Mut$^+$. *J Biotechnol* 2012;**157**:180–8.

Pfeffer M, Maurer M, Kollensperger G *et al.* Modeling and measuring intracellular fluxes of secreted recombinant protein in *Pichia pastoris* with a novel ^{34}S labeling procedure. *Microb Cell Fact* 2011;**10**:47.

Pfeffer M, Maurer M, Stadlmann J *et al.* Intracellular interactome of secreted antibody Fab fragment in *Pichia pastoris* reveals its routes of secretion and degradation. *Appl Microbiol Biot* 2012;**93**:2503–12.

Phaff H, Miller M, Shifrine M. The taxonomy of yeasts isolated from *Drosophila* in the Yosemite region of California. *Anton Leeuw* 1956;**22**:145–61.

Prielhofer R, Cartwright SP, Graf AB *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.

Prielhofer R, Maurer M, Klein J *et al.* Induction without methanol: novel regulated promoters enable high-level expression in *Pichia pastoris*. *Microb Cell Fact* 2013;**12**:5.

Puxbaum V, Mattanovich D, Gasser B. *Quo vadis?* The challenges of recombinant protein folding and secretion in *Pichia pastoris*. *Appl Microbiol Biot* 2015;**99**:2925–38.

Rebnegger C, Graf AB, Valli M *et al.* In *Pichia pastoris*, growth rate regulates protein synthesis and secretion, mating and stress response. *Biotechnol J* 2014;**9**:511–25.

Rebnegger C, Vos T, Graf AB *et al.* *Pichia pastoris* exhibits high viability and a low maintenance energy requirement at near-zero specific growth rates. *Appl Environ Microb* 2016;**82**:4570–83.

Redden H, Morse N, Alper HS. The synthetic biology toolbox for tuning gene expression in yeast. *FEMS Yeast Res* 2015;**15**:1–10.

Resina D, Bollok M, Khatri NK *et al.* Transcriptional response of *P. pastoris* in fed-batch cultivations to *Rhizopus oryzae* lipase production reveals UPR induction. *Microb Cell Fact* 2007;**6**:21.

Russmayer H, Buchetics M, Gruber C *et al.* Systems-level organization of yeast methylotrophic lifestyle. *BMC Biol* 2015a;**13**:80.

Russmayer H, Troyer C, Neubauer S *et al.* Metabolomics sampling of *Pichia pastoris* revisited: rapid filtration prevents metabolite loss during quenching. *FEMS Yeast Res* 2015b;**15**:fov049.

Ruth C, Buchetics M, Vidimce V *et al.* *Pichia pastoris* Aft1–a novel transcription factor, enhancing recombinant protein secretion. *Microb Cell Fact* 2014;**13**:120.

Saitua F, Torres P, Perez-Correa JR *et al.* Dynamic genome-scale metabolic modeling of the yeast *Pichia pastoris*. *BMC Syst Biol* 2017;**11**:27.

Sauer M, Branduardi P, Gasser B *et al.* Differential gene expression in recombinant *Pichia pastoris* analysed by heterologous DNA microarray hybridisation. *Microb Cell Fact* 2004;**3**:17.

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

Schreiber C, Muller H, Birrenbach O *et al.* A high-throughput expression screening platform to optimize the production of antimicrobial peptides. *Microb Cell Fact* 2017;**16**:29.

Schroer K, Luef KP, Hartner FS *et al.* Engineering the *Pichia pastoris* methanol oxidation pathway for improved NADH regeneration during whole-cell biotransformation. *Metab Eng* 2010;**12**:8–17.

Schwarzhans JP, Wibberg D, Winkler A *et al.* Integration event induced changes in recombinant protein productivity in *Pichia pastoris* discovered by whole genome sequencing and derived vector optimization. *Microb Cell Fact* 2016a;**15**:84.

Schwarzhans JP, Wibberg D, Winkler A *et al.* Non-canonical integration events in *Pichia pastoris* encountered during standard transformation analysed with genome sequencing. *Sci Rep* 2016b;**6**:38952.

Sha C, Yu XW, Li F *et al.* Impact of gene dosage on the production of lipase from *Rhizopus chinensis* CCTCC M201021 in *Pichia pastoris*. *Appl Biochem Biotech* 2013;**169**:1160–72.

Shen W, Xue Y, Liu Y *et al.* A novel methanol-free *Pichia pastoris* system for recombinant protein expression. *Microb Cell Fact* 2016;**15**:178.

Sohn SB, Graf AB, Kim TY *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.

Sola A, Jouhten P, Maaheimo H *et al.* 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.

Sola A, Maaheimo H, Ylonen K *et al.* Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*. *Eur J Biochem* 2004;**271**:2462–70.

Spadiut O, Capone S, Krainer F *et al.* Microbials for the production of monoclonal antibodies and antibody fragments. *Trends Biotechnol* 2014;**32**:54–60.

Stadlmayr G, Mecklenbrauker A, Rothmuller M *et al.* Identification and characterisation of novel *Pichia pastoris* promoters for heterologous protein production. *J Biotechnol* 2010;**150**:519–29.

Sturmberger L, Chappell T, Geier M *et al.* Refined *Pichia pastoris* reference genome sequence. *J Biotechnol* 2016;**235**:121–31.

Ternes P, Wobbe T, Schwarz M *et al.* Two pathways of sphingolipid biosynthesis are separated in the yeast *Pichia pastoris*. *J Biol Chem* 2011;**286**:11401–14.

Tomas-Gamisans M, Ferrer P, Albiol J. Integration and validation of the genome-scale metabolic models of *Pichia pastoris*: A comprehensive update of protein glycosylation pathways, lipid and energy metabolism. *PLoS One* 2016;**11**:e0148031.

Tortajada M, Llaneras F, Picó J *et al.* Validation of a constraint-based model of *Pichia pastoris* metabolism under data scarcity. *BMC Syst Biol* 2010;**4**:115.

Tredwell GD, Aw R, Edwards-Jones B *et al.* Rapid screening of cellular stress responses in recombinant *Pichia pastoris* strains using metabolite profiling. *J Ind Microbiol Biot* 2017;**44**:413–7.

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.

Valli M, Tatto NE, Peymann A *et al.* Curation of the genome annotation of *Pichia pastoris* (*Komagataella phaffii*) CBS7435 from gene level to protein function. *FEMS Yeast Res* 2016;**16**:fow051.

Vanz AL, Lunsdorf H, Adnan A *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.

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.

Vassileva A, Chugh DA, Swaminathan S *et al*. Effect of copy number on the expression levels of hepatitis B surface antigen in the methylotrophic yeast *Pichia pastoris*. *Protein Expr Purif* 2001;**21**:71–80.

Vogl T, Ahmad M, Krainer FW *et al*. Restriction site free cloning (RSFC) plasmid family for seamless, sequence independent cloning in *Pichia pastoris*. *Microb Cell Fact* 2015;**14**:103.

Vogl T, Ruth C, Pitzer J *et al*. Synthetic core promoters for *Pichia pastoris*. *ACS Synth Biol* 2014a;**3**:188–91.

Vogl T, Sturmberger L, Kickenweiz T *et al*. A toolbox of diverse promoters related to methanol utilization: Functionally verified parts for heterologous pathway expression in *Pichia pastoris*. *ACS Synth Biol* 2016;**5**:172–86.

Vogl T, Thallinger GG, Zellnig G *et al*. Towards improved membrane protein production in *Pichia pastoris*: general and specific transcriptional response to membrane protein overexpression. *N Biotechnol* 2014b;**31**:538–52.

Wang J, Wang X, Shi L *et al*. Methanol-independent protein expression by *AOX1* promoter with trans-acting elements engineering and glucose-glycerol-shift induction in *Pichia pastoris*. *Sci Rep* 2017;**7**:41850.

Weinacker D, Rabert C, Zepeda AB *et al*. Applications of recombinant *Pichia pastoris* in the healthcare industry. *Braz J Microbiol* 2013;**44**:1043–8.

Weinhandl K, Winkler M, Glieder A *et al*. Carbon source dependent promoters in yeasts. *Microb Cell Fact* 2014;**13**:5.

Weninger A, Hatzl AM, Schmid C *et al*. Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. *J Biotechnol* 2016;**235**:139–49.

Werner S, Engler C, Weber E *et al*. Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Bioeng Bugs* 2012;**3**:38–43.

Wiechert W. ^{13}C metabolic flux analysis. *Metab Eng* 2001;**3**:195–206.

Yamada Y, Matsuda M, Maeda K *et al*. The phylogenetic relationships of methanol-assimilating yeasts based on the partial sequences of 18S and 26S ribosomal RNAs: the proposal of *Komagataella gen. nov*. (Saccharomycetaceae). *Biosci Biotech Bioch* 1995;**59**:439–44.

Yu AQ, Zhu JC, Zhang B *et al*. Knockout of fatty acid desaturase genes in *Pichia pastoris* GS115 and its effect on the fatty acid biosynthesis and physiological consequences. *Arch Microbiol* 2012;**194**:1023–32.