

Genome-Scale Metabolic Model of *Pichia pastoris* With Native and Humanized Glycosylation of Recombinant Proteins

Zahra Azimzadeh Irani,¹ Eduard J. Kerkhoven,² Seyed Abbas Shojaosadati,¹ Jens Nielsen^{2,3}

¹Biotechnology Group, Faculty of Chemical Engineering, Tarbiat Modares University, Tehran, Iran; telephone: +98 21 82883341; fax: +98 21 82884931; e-mail: shoja_sa@modares.ac.ir

²Systems and Synthetic Biology, Department of Biology and Biological Engineering, Chalmers University of Technology, Göteborg, Sweden

³Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm, Denmark; telephone: +46 31 772 3804; fax: +46 31 772 3801; e-mail: nielsenj@chalmers.se

ABSTRACT: *Pichia pastoris* is used for commercial production of human therapeutic proteins, and genome-scale models of *P. pastoris* metabolism have been generated in the past to study the metabolism and associated protein production by this yeast. A major challenge with clinical usage of recombinant proteins produced by *P. pastoris* is the difference in N-glycosylation of proteins produced by humans and this yeast. However, through metabolic engineering, a *P. pastoris* strain capable of producing humanized N-glycosylated proteins was constructed. The current genome-scale models of *P. pastoris* do not address native nor humanized N-glycosylation, and we therefore developed *ihGlyco-pastoris*, an extension to the iLC915 model with both native and humanized N-glycosylation for recombinant protein production, but also an estimation of N-glycosylation of *P. pastoris* native proteins. This new model gives a better prediction of protein yield, demonstrates the effect of the different types of N-glycosylation of protein yield, and can be used to predict potential targets for strain improvement. The model represents a step towards a more complete description of protein production in *P. pastoris*, which is required for using these models to understand and optimize protein production processes.

Biotechnol. Bioeng. 2016;113: 961–969.

© 2015 Wiley Periodicals, Inc.

KEYWORDS: genome scale metabolic model; humanized glycosylation; *Pichia pastoris*; recombinant protein production

Introduction

Pichia pastoris has been used for the expression of over 200 different heterologous proteins (Cregg et al., 2000), including the production of numerous therapeutic recombinant proteins such as human serum albumin (Kobayashi et al., 1998), human superoxide dismutase (rhSOD) (Marx et al., 2009), human erythropoietin (EPO) (Zhan et al., 1999), and human monoclonal antibody 3H6 Fab fragment (FAB) (Gach et al., 2007). The main reasons for using *P. pastoris* as a host for heterologous protein production include the availability of the alcohol oxidase I gene (*AOX1*) promoter for controlled gene expression (Cregg et al., 2000), the ability of growing to high cell densities in bioreactors (Cereghino, 2000), and the ability of yeasts to perform eukaryotic post-translational modifications such as glycosylation, but also ubiquitination, sumoylation, and myristoylation.

Many human therapeutic proteins undergo a post-translational modification known as N-glycosylation, which can affect the protein's secretion, folding, and bioactivity (Helenius and Aebi, 2001). Over 70% of the therapeutic proteins under preclinical and clinical development are glycosylated (Chung et al., 2010), and it is therefore, important to include N-glycosylation in studies of therapeutic recombinant protein expression.

During N-linked glycosylation, which is conserved across yeast and other eukaryotes (Gemmil and Trimble, 1999), a 14-saccharide “core” unit is assembled as a membrane-bound dolichyl pyrophosphate precursor by enzymes located on both sides of the endoplasmic reticulum (ER), and this core is subsequently transferred to targeted asparagine residues of proteins (Gemmil and Trimble, 1999).

Besides N-linked glycosylation, *P. pastoris* is also capable of O-linked glycosylation, where the glycan is attached onto the hydroxyl groups of serine or threonine residue (Puxbaum et al., 2015). In human cells, there are different types of O-linked glycan structures, with separate pathways, and their large

Correspondence to: S.A. Shojaosadati and J. Nielsen

Contract grant sponsor: Ministry of Science, Research and Technology of the Islamic Republic of Iran

Contract grant sponsor: Novo Nordisk Foundation

Contract grant sponsor: U.S. Department of Energy, Genomic Science Program

Contract grant number: DE-SC0008744

Received 18 June 2015; Revision received 21 September 2015; Accepted 12 October 2015

Accepted manuscript online 19 October 2015;

Article first published online 2 November 2015 in Wiley Online Library

(<http://onlinelibrary.wiley.com/doi/10.1002/bit.25863/abstract>).

DOI 10.1002/bit.25863

difference with *P. pastoris* N-linked glycans has limited glycoengineering of yeast to reproduce humanized O-glycosylation (Puxbaum et al., 2015). However, the percentage of O-linked glycosylation in *Pichia pastoris* is small (Gemmill and Trimble, 1999) and the biological role of O-linked glycosylation has not been elucidated, although it seems to play some role in protein folding quality control and is essential for survival (Delic et al., 2013).

During N-linked glycosylation, after coupling of the core unit to the protein, terminal glucose and mannose residues are removed by ER glucosidases and mannosidases prior to entry of the glycoprotein into the Golgi (Gemmill and Trimble, 1999). The difference between yeast and mammalian cells occurs once the glycoprotein exits the ER and enters the Golgi apparatus (Hamilton and Gerngross, 2007). In contrast to mammals, yeast does not further trim the N-glycans in the Golgi, but rather extends them with additional mannose sugars, to produce hypermannosylated glycans. Recent advances have allowed for the generation of yeast strains capable of replicating the most essential glycosylation pathways found in mammals (Hamilton et al., 2006). Glycoengineering has been accomplished to eliminate the hyper-mannosylated yeast glycans, and to introduce the required elements for producing human-like sialylated complex glycans (Bretthauer, 2003; Chiba and Akeboshi, 2009). These humanized glycoproteins can be attained using metabolic engineering of *P. pastoris*.

One of the tools of metabolic engineering is the use of computational models that describe the metabolic pathways in cells and can aid in the identification of new targets or strategies for further metabolic engineering (Kerkhoven et al., 2015). A metabolic model describing N-glycosylation in *P. pastoris* has previously been constructed (Eskitoros et al., 2014), however, this model only consisted of a subset of the metabolic reaction present in *P. pastoris*. Metabolic networks can alternatively be studied in silico with the help of genome-scale metabolic models (GEMs), which contain all known metabolic reactions within an organism. GEMs have become instrumental for system-level understanding of metabolism and its applications in metabolic engineering (Kim et al., 2014). Computational algorithms such as constraint-based flux balance analysis (FBA) can then be applied to investigate GEMs and these methods have been used to study the objectives and functions of metabolic networks (Kerkhoven et al., 2015).

To date, three GEMs for *P. pastoris* have been reported. The iPP668 model constructed by Chung et al. (2010), the PpaMBEL1254 model constructed by Sohn et al. (2010) and the latest GEM, iLC915 model constructed by Caspeta et al. (2012). While these models contain various reactions involved in glycoprotein metabolism, they neither include functional N-glycosylation of recombinant protein nor include reactions describing the aforementioned humanized glycosylation.

In this study, we developed a functional GEM that describes both native *P. pastoris* N-glycosylation and humanized N-glycosylation for a number of recombinant proteins. An average N-glycosylation is assumed such that micro- and macroheterogeneity do not have to be considered. Additionally, an estimation of N-glycosylation of *P. pastoris* native proteins is

included. This model facilitates further investigation of the effect of N-glycosylation on protein production yields.

Materials and Methods

The iLC915 model (Caspeta et al., 2012) was extended as detailed in this paper to yield the new model, which we named *ihGlycopastoris*, for humanized glycosylation in *P. pastoris*. This model is provided in SBML format. Model curation and simulation was performed using the SBML (Keating et al., 2006) and RAVEN Toolbox (Agren et al., 2013) for MATLAB (Mathworks Inc., MA). MOSEK (MOSEK ApS, Denmark) was used to solve linear programming problems. The model was constrained with relative in and outflow fluxes as mentioned in the Results section. Theoretical yields were normalized to the uptake rate in Cmol.

N-glycosylation of *P. pastoris* native proteins was estimated based on experimental data from *Saccharomyces cerevisiae*. By proteomics, 645 N-glycosylations were detected in the whole *S. cerevisiae* proteome (Chen et al., 2014). The length of all *S. cerevisiae* protein sequences in the genome is roughly 3 million amino acids, which translates to an average N-glycosylation on every 4691st amino acid. With iLC915 biomass containing 3.2956 mmol gDCW⁻¹ amino acids, 7.03 μmol gDCW⁻¹ of native, mannose-rich, N-glycan was added to the biomass equation.

Results and Discussion

The most recent GEM of *P. pastoris*, iLC915, has been showed to provide good predictions of both in vivo growth and antibody-production yields (Caspeta et al., 2012). This model was selected for further extension with native and humanized N-glycosylation pathways. Before any additional pathways were added, however, additional curation of this model was performed, primarily to render the model more robust. The curations are detailed in Supplementary Information 1. After this manual curation of iLC915, we extended the updated model to incorporate both *P. pastoris* native N-glycosylation and humanized N-glycosylation, where the N-glycan group is modified to resemble human N-glycans.

Reconstructing the Dolichol Pathway of Native N-Linked Glycosylation

P. pastoris native N-glycosylation begins in the ER with a 14-residue sugar that contains three glucose (Glc), nine mannose (Man), and two N-acetylglucosamine (GlcNAc) residues (Fig. 1), which are attached to a dolichol carrier (Samuelson et al., 2005). This core lipid-linked oligosaccharide, (glucosyl)₃(mannosyl)₉-(N-acetylglucosaminyl)₂-diphosphodolichol is subsequently transferred by an oligosaccharyltransferase (OST) to asparagine residues in a protein (Karaoglu et al., 1997). Following transfer to the nascent polypeptide, forming a glycopeptide, three glucose residues and one mannose residue are removed to produce Man₈GlcNAc₂, at which stage the glycopeptide is transported to the Golgi.

In the Golgi, this Man₈GlcNAc₂ glycan can be further extended with additional mannose sugars. The number of mannoses is not set but is affected by various factors such as expression level, carbon

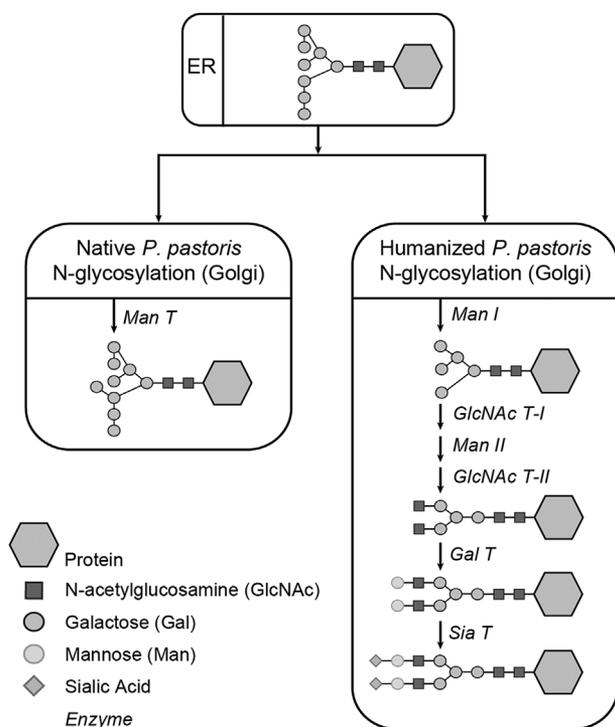


Figure 1. Schematic overview of N-linked glycosylation pathways in *P. pastoris*. Both native (left) and glycoengineered humanized glycosylation (right) are shown. On top is the universal dolichol core complex that is made in the ER, while subsequent steps take part in the Golgi. Man I: mannosidase I; Man II: mannosidase II; GlcNAcT-I: N-acetylglucosaminyltransferase I; GlcNAcT-II: N-acetylglucosaminyltransferase II; Gal T: galactosyltransferase; Sia T: sialyltransferase; ManT: mannosyltransferase.

source, and site accessibility. This is referred to as microheterogeneity, and in *P. pastoris*, the number of mannoses ranges between 8 and 11 (Trimble et al., 1991). In *ihGlycopastoris*, microheterogeneity is not taken into account, and a weighted median of 9 mannoses is used in native N-glycosylation. The reactions of the dolichol and native N-glycosylation pathways were added to the iLC915 model using a recent model of *S. cerevisiae* (Feizi, Marras, Nielsen, manuscript in preparation) as a template, while the reactions were annotated with *P. pastoris* genes identified using literature data and homology with annotated *S. cerevisiae* genes.

The model was further extended by incorporating reactions for N-glycosylation of *P. pastoris*' native proteome. Experimental determination of N-glycosylation on a proteome scale is technologically challenging and while such data is currently not available for *P. pastoris*, these measurements have been performed for *S. cerevisiae* (Chen et al., 2014). We therefore incorporated a rough estimation of N-glycosylation of the *P. pastoris* proteins based on the *S. cerevisiae* data, as described in Materials and Methods, without taking micro- and macro-heterogeneity into account.

Addition of Humanized N-Linked Glycosylation

In contrast to yeast, mammals do not extend the $\text{Man}_8\text{GlcNAc}_2$ glycan in the Golgi, but rather trim the existing high mannose

structure in three steps to $\text{Man}_5\text{GlcNAc}_2$ by α -1,2-mannosidase, followed by the addition of GlcNAc by N-acetylglucosaminyltransferase I (GnTI). A number of subsequent trimming and extension reactions lead to the production of a sialylated structure (Fig. 1), which represents the final N-glycosylated protein. *P. pastoris* strains have been constructed that allow the construction of these human-like N-glycans.

Both native and humanized N-glycosylation pathways for *P. pastoris* (Fig. 1) were added to the iLC915 model in order to analyze the effect of these changes on the yeast metabolism, and we named this extended model *ihGlycopastoris*. To allow the investigation of various recombinant proteins, we added the polymerization reactions of several proteins (Table 1), differing in both size and number of glycosylation sites, and generated native and humanized glycosylation reactions for each of these proteins. The model can easily be adapted to facilitate the study of N-glycosylation of other proteins. A detailed overview and description of the reactions describing N-glycosylation in *ihGlycopastoris* are given in Supplementary Information 2–3.

In comparison with other models of N-glycosylation in *P. pastoris* (Barrigon et al., 2015; Eskitoros et al., 2014), our model is in the context of the whole of *P. pastoris* metabolism, allows both native and humanized N-glycosylation of multiple recombinant proteins and estimates N-glycosylation of the native proteome.

Validation of ihGlycopastoris

The updated model (*ihGlycopastoris*) was validated for its ability to predict growth rates. Experimental data on exchanges fluxes were set as constraints in *ihGlycopastoris* while biomass production was set as the objective function, and the resulting growth predictions were compared to reported experimentally measured growth rate (Dragosits et al., 2009; Jordà et al., 2014; Jungo et al., 2006, 2007a; Solà et al., 2004, 2007; Tortajada Serra, 2012; Zhang et al., 2004). Additional to growth data, the model's ability to predict recombinant protein production rates in the reported experimental conditions were also compared to experimental yields (Dragosits et al., 2009; Jungo et al., 2007a, 2007b; Solà et al., 2007; Zhang et al., 2004). For each source, several different operating conditions are compared, e.g., different temperatures using glucose as the carbon source to produce FAB (Dragosits et al., 2009); high and low dilution rates using glycerol/methanol mixtures (Solà et al., 2007); glycerol/methanol mixture with linear increase of the methanol fraction in

Table 1. Different proteins with different length and glycosylation sites.

Protein	Amino acids	N-glycosylation sites	$\frac{\text{N-glycosylation sites}}{\text{Amino acids}}$
Human serum albumin Redhill variant (Alb-Redhill) ^a	609	1	0.0016
Interferon- γ (INF- γ)	195	1	0.0051
Human Interleukin 4 (hIL4)	153	1	0.0065
Human mast cell Chymase (rhChymase)	247	2	0.0081
Erythropoietin (EPO)	193	3	0.0155
Prostaglandin H synthases (PGHS)	604	4	0.0066

^aWhere normal human serum albumin has no N-glycosylation sites, Albumin-Redhill is mutated A320T, which generates an N-glycosylation site (Brennan et al., 1990).

the feed to produce avidin (Jungo et al., 2007a). Both measured exchange fluxes and growth rates were used to constraint *ihGlycopastoris*, while the objective function was set as the excretion of the relevant protein. As no data was available on protein yields on humanized N-glycosylated protein production, only native N-glycosylated protein production predictions could be compared to experimental data.

The model predictions for growth were in good agreement ($r^2=0.92$) with experimental reported growth (Fig. 2A) and as such, *ihGlycopastoris* is validated by more experimental growth data than the iLC915 model (Caspeta et al., 2012). For a number of experimental conditions, the model was unable to predict protein production (Fig. 2B). These cases corresponded to conditions where the model was also under-predicting the growth rate, from which leads that there are no capacity to produce recombinant protein. For the cases where the model allowed for protein production, the predicted rates were substantially higher than the experimental rates (Fig. 2B). In these cases, other cellular processes, which are not taken into account in *ihGlycopastoris*, limit the recombinant protein production. These processes can be reduced activation of the protein expression system, or limitations in the protein secretion capacity (Hohenblum et al., 2004). In particular, the very high affinity of avidin to biotin, an essential vitamin, might have detrimental effects when its production increases (Jungo et al., 2007c). While the model is unable to capture this level of regulation, it does provide an upper boundary of protein yield.

The Effect of N-Glycosylation on the Protein Yield

To investigate the effect of different forms of N-glycosylation on recombinant protein yields, theoretical yields were calculated for a variety of recombinant proteins undergoing native N-glycosylation, humanized N-glycosylation, or no N-glycosylation at all, where the latter represents the theoretical yield when N-glycosylation is not taken into account. Yields were calculated using FBA by setting the objective function to the excretion of a particular glycosylated form

of a recombinant protein. From this comparison, it becomes clear that taking N-glycosylation into account is of importance, as the theoretical yields of native N-glycosylated protein is substantial lower than non-glycosylated yields (Fig. 3A), i.e., not taking N-glycosylation into account results in overestimation of the protein yield. For Alb-Redhill, this indicates the difference between the theoretical yield of predominant human serum albumin, which has no N-glycosylation site, and the Redhill variant, that has an N-glycosylation site introduced due to an A334T mutation (Brennan et al., 1990).

Glycoengineering of *P. pastoris* to produce humanized glycosylated protein results in a further reduction in theoretical protein yields (Fig. 3A). This is in contrast to previous computational modelling that was performed of N-glycosylation in *P. pastoris* (Eskitoros et al., 2014), where it was concluded that yields of humanized N-glycosylated proteins are higher than of native N-glycosylated proteins. The reduced theoretical yield of humanized in comparison to native N-glycosylation can be explained by the size of the N-glycan complex, which contains 70 and 84 carbons in native and humanized forms respectively. In comparison, the model of (Eskitoros et al., 2014) produces partially humanized N-glycosylation, which requires lower amount of carbons.

The reduction in yield due to native and humanized N-glycosylation is not the same for each different recombinant protein, for example, the theoretical yield of Alb-Redhill is only nominally affected by N-glycosylation while the reduction in theoretical yield in EPO is substantial. These differences are caused by the size of the recombinant protein and their number of N-glycosylation sites; smaller proteins and more glycosylation sites result in a larger reduction in yield (cf. Table I and Fig. 3A).

Assessments of theoretical yields as described above were performed using glucose as a carbon source. Also, the use of different carbon sources was investigated by changing the carbon source while keeping the Cmol uptake rate constant. These simulations indicated that theoretical protein yields vary depending on carbon source, while this effect is almost identical for all

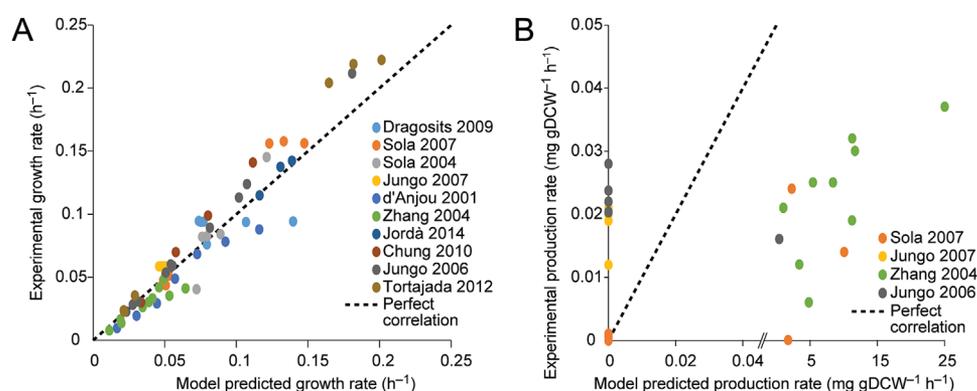


Figure 2. Comparison of experimental growth rates and proteins productions with model predicted values. Experimentally measured exchange rates of carbon sources, oxygen etc. from various sources were used to constrain *ihGlycopastoris*. **(A)** Specific growth rates were predicted by setting the biomass growth as objective function. **(B)** Protein production rates were predicted by constraining the growth, while setting the protein excretion as objective function. Experimental data for protein production rates were obtained from: FAB (Dragosits et al., 2009), ROL (Solà et al., 2007), avidin (Jungo et al., 2006, 2007a), INF-t (Zhang et al., 2004).

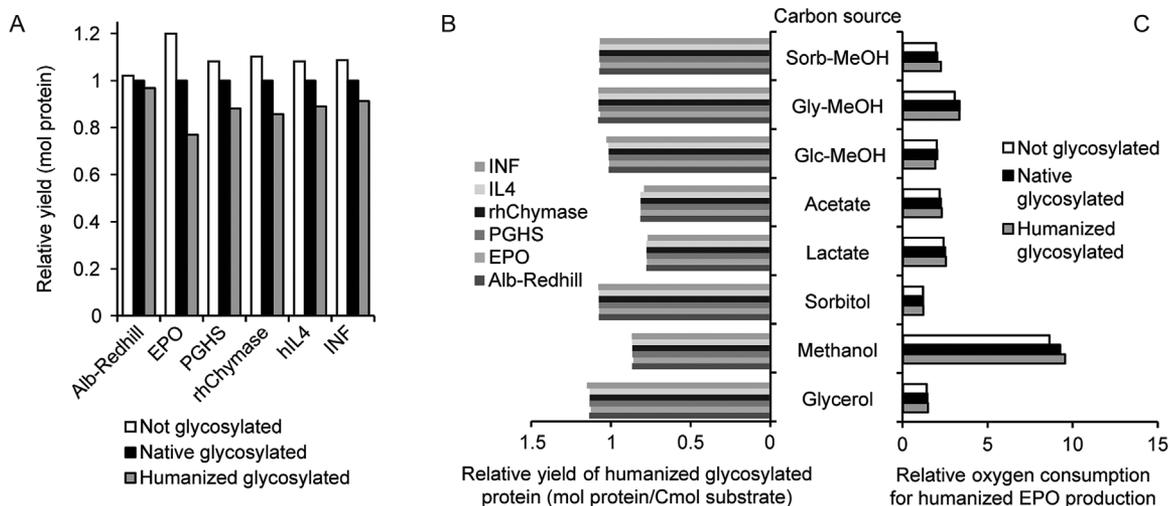


Figure 3. The effect of glycosylation on protein production. **(A)** Protein yields when no N-glycosylation is taken into account (white) and when undergoing humanized N-glycosylation (grey), relative to protein yield during native N-glycosylation (which would occur in non-glycoengineered strains). Protein yields are given as relative moles. **(B)** Humanized protein yields on different carbon sources, relative to the yield on glucose. Units are moles of proteins per Cmol of the substrate. **(C)** Oxygen requirement of EPO production for different carbon sources, normalized to the oxygen requirements on glucose (mol/Cmol).

recombinant proteins and for the different modes of N-glycosylation (Fig. 3B). In particular, the theoretical protein yield is primarily decreased in methanol, due to the increased oxidative phosphorylation. This is also evidenced by the increased oxygen requirements (Jungo et al., 2007a) (Fig. 3C), while it has been described previously that the protein yield on methanol increases significantly when a co-substrate is used (Çalik et al., 2010; Eskitoros and Çalik, 2014; Jungo et al., 2007a).

Prediction of Potential Targets for Protein Yield Improvement

ihGlycopastoris was subsequently used to predict potential amplification targets that would increase the protein yield. These predictions were performed for a large protein with few (one) N-glycosylation sites (Alb-Redhill) and a small protein with multiple (three) N-glycosylation sites (EPO). Target identification was performed using Flux Scanning based on Enforced Objective Function (FSEOF), where the model is constrained with an experimentally measured carbon uptake rate while the experimentally observed protein yield is stepwise increased to the theoretical yield (Choi et al., 2010). Simultaneously, the biomass production, which has been set as objective function, decreases due to competition for nutrients with the protein production. Potential targets are those reactions whose flux is systematically increasing at each iterative increase of protein yield. It is worth to note that while this strategy may increase the yield of protein, it does not address the quality and activity of the protein and in particular it might affect the heterogeneity of N-glycosylation.

FSEOF was performed using both native and humanized N-glycosylation. Experimental protein production rates were set at $0.064 \mu\text{mol gDCW}^{-1} \text{h}^{-1}$ for EPO (Eskitoros and Çalik, 2014) and

$0.00169 \mu\text{mol gDCW}^{-1} \text{h}^{-1}$ human serum albumin (Maccani et al., 2014) (used here as a proxy for Alb-Redhill). Given the reported carbon uptake rates, the theoretical production rates of humanized N-glycosylated proteins were calculated as $6.1 \mu\text{mol gDCW}^{-1} \text{h}^{-1}$ for EPO and $0.84 \mu\text{mol gDCW}^{-1} \text{h}^{-1}$ for Alb-Redhill (Supplementary Information 4) representing a 95 and 501 fold-change increase over the experimental production rates. A large number of reactions that were identified as consistently increasing at improved protein yields, but the majority of these reactions could be contributed to general amino acid biosynthesis, and these reactions could have been identified as targets without the use of computational models. However, it is re-assuring that the model captures these targets. When the fate of the carbons entering the model shifts from biomass production to protein production, the amino acid requirements increase.

Interestingly, N-glycan metabolism was identified as a potential target (Table II), indicating that the increased N-glycan requirements due to increased recombinant protein production is not compensated by the decrease in N-glycosylation of native protein. Also biosynthetic pathways for precursors of N-glycosylation are suggested targets, such as UDP-GlcNAc and GDP-Man biosynthesis.

Four targets (*GNA1*, *GFA1*, *UAP1*, and *PCMI*) are involved in the biosynthesis of UDP-GlcNAc, where fructose 6-phosphate is derived from glucose and converted to glucosamine-6-phosphate (GlcN-6P) to finally form UDP-GlcNAc (Table II and Fig. 4). Another 4 targets (*GLK1*, *PSA1*, *PMI40* and *PMM1*) are in the GDP-Man biosynthetic pathway. UDP-GlcNAc and GDP-Man are nucleotide sugar donors that reside in the cytoplasmic side of the ER membrane, where GlcNAc and mannose molecules are added in the N-linked glycan. Increased production of glycosylated protein would therefore also increase the requirements of these precursors.

Table II. Targets predicted by FSEOF for increased production of EPO or Alb-Redhill using FSEOF.

Reaction	Gene	Abbreviation	Reaction	Fold change Alb-Redhill	Fold change EPO	Group
r550	PAS_chr3_1192	GLK1	ATP + mannose → ADP + mannose 6-P	111	12	GDP-Man
r635	PAS_chr2-1_0093	PSA1	GDP + mannose 1-P ↔ phosphate + GDP-mannose	21	12	GDP-Man
r992	PAS_chr2-2_0053	PMM1	Mannose 6-P ↔ mannose 1-P	21	12	GDP-Man
r983	PAS_chr3_1115	PMI40	Mannose 6-P ↔ fructose 6-P	6.2	10	GDP-Man
r526	PAS_chr2-1_0626	GFA1	Fructose 6-P + glutamine → glutamate + GlcN 6-P	63	12	UDP-GlcNAc
r390	PAS_chr4_0060	GNA1	Acetyl-CoA + GlcN 6-P ↔ CoA + GlcNAc 6-P	63	12	UDP-GlcNAc
r636	PAS_chr3_0676	UAP1	UTP + GlcNAc 1-P ↔ diphosphate + UDP-GlcNAc	63	12	UDP-GlcNAc
r991	PAS_chr1-1_0067	PCM1	GlcNAc 6-P ↔ GlcNAc 1-P	63	12	UDP-GlcNAc
r448	PAS_chr1-4_0417	ALG11	GDP-mannose + (GlcNAc) ₂ (Man) ₃ (PP-Dol) → GDP + (GlcNAc) ₂ (Man) 4 (PP-Dol)	23	12	N-glycan
r452	PAS_c121_0002	ALG2	GDP-mannose + (GlcNAc) ₂ (Man) ₁ (PP-Dol) → GDP + (GlcNAc) ₂ (Man) 2 (PP-Dol)	23	12	N-glycan
r455	PAS_chr2-1_0759	ALG1	GDP-mannose + Chitobiosyl-PP-dolichol → GDP + (GlcNAc) ₂ (Man) 1(PP-Dol)	23	12	N-glycan
r454	PAS_chr3_0944	ALG13	UDP-GlcNAc + GlcNAc-PP-dolichol → UDP + Chitobiosyl-PP-dolichol	23	12	N-glycan
r661	PAS_chr2-1_0727	ALG7	UDP-GlcNAc + P-dolichol → UMP + GlcNAc-PP-dolichol	23	12	N-glycan
r465	PAS_chr1-1_0459	DPM1	GDP-mannose + P-dolichol → GDP + P-dolichol mannose	23	12	N-glycan
r2025	PAS_chr4_0712	ALG3	P-dolichol mannose + (GlcNAc) ₂ (Man) ₄ (PP-Dol) → P-dolichol + (GlcNAc) ₂ (Man) ₆ (PP-Dol)	23	12	N-glycan
r2026	PAS_chr2-2_0036	ALG9	P-dolichol mannose + (GlcNAc) ₂ (Man) ₆ (PP-Dol) → P-dolichol + (GlcNAc) ₂ (Man) ₇ (PP-Dol)	23	12	N-glycan
r2027	PAS_chr4_0544	ALG12	P-dolichol mannose + (GlcNAc) ₂ (Man) ₇ (PP-Dol) → P-dolichol + (GlcNAc) ₂ (Man) ₈ (PP-Dol)	23	12	N-glycan
r2031	PAS_chr2-1_0549	ALG6	P-dolichol glucosyl + (GlcNAc) ₂ (Man) ₉ (PP-Dol) → P-dolichol + (Glc) ₁ (GlcNAc) ₂ (Man) ₉ (PP-Dol)	23	12	N-glycan
r2032	PAS_chr3_0999	ALG8	P-dolichol glucosyl + (Glc) ₁ (GlcNAc) ₂ (Man) ₉ (PP-Dol) → P-dolichol + (Glc) ₂ (GlcNAc) ₂ (Man) ₉ (PP-Dol)	23	12	N-glycan
r2033	PAS_chr1-4_0475	ALG10	P-dolichol glucosyl + (Glc) ₂ (GlcNAc) ₂ (Man) ₉ (PP-Dol) → P-dolichol + (Glc) ₃ (GlcNAc) ₂ (Man) ₉ (PP-Dol)	23	12	N-glycan
r2029	PAS_chr2-2_0552	ALG5	UDP-glucose + P-dolichol → UDP + P-dolichol glucosyl	23	12	N-glycan
r72	PAS_chr3_0277	GND2	NADP ⁺ + 6-P-gluconate → CO ₂ + NADPH + Ribulose 5-P	23	12	PPP
r558	PAS_chr1-4_0669	GNK1	ATP + gluconic acid → ADP + 6-P-gluconate	23	12	PPP
r915	PAS_chr2-2_0137	CYS4	Serine + H ₂ S → Cysteine	102	1.7	Sulfur
r567	PAS_chr3_0667	MET14	ATP + adenylyl sulfate → ADP + 3'-P-adenylyl sulfate	23	1.7	Sulfur
r144	PAS_chr3_0058	ARH1	NADP ⁺ + 2 reduced ferredoxin ↔ NADPH + 2 oxidized ferredoxin	23	—	Sulfur
r598	PAS_chr1-4_0253	MET3	ATP + sulfate → diphosphate + adenylyl sulfate	23	1.7	Sulfur
r733	PAS_chr2-1_0547	MET22	adenosine 3',5'-PP → phosphate + AMP	23	1.7	Sulfur
r281	PAS_chr2-2_0480	TRR1	NADP ⁺ + thioredoxin ↔ NADPH + thioredoxin-S ₂	20	1.7	Sulfur
r228	PAS_chr3_0225	SDH1	Succinate + FAD ↔ fumarate + FADH ₂	13	17	Glyoxylate
r878	PAS_chr1-4_0338	ICL1	Isocitrate → glyoxylate + succinate	13	17	Glyoxylate
r534	PAS_chr4_0416	AGX1	Glyoxylate + alanine → pyruvate + glycine	13	17	Glyoxylate

Abbreviations for gene names were taken from their *S. cerevisiae* homologs when they have not been defined for *P. pastoris*. Fold changes indicate the increase that is required to go from the experimental to the theoretical protein production yield.

Two targets are related to the pentose phosphate pathway (*GND2* and *GNK1*). The pentose phosphate pathway can be argued to have two biological functions: providing pentose sugars for the biosynthesis of ribonucleotides, which are used in DNA and RNA, and the regeneration of NADP⁺ to NADPH, providing reducing potential for reductive processes in the cell. While an increased protein production yield would not increase the ribonucleotide requirements, it is plausible that the NADPH requirements of the cell increase. The potential of PPP as an overexpression target for increased protein production has previously been identified and validated for hSOD production in *P. pastoris* (Nocon et al., 2014).

Another group of targets are related to sulfur metabolism (Table II). When comparing the amino acid composition of the

whole *P. pastoris* proteome (Carnicer et al., 2009) with EPO, most of the amino acids are present in comparable amounts, mostly within two-fold differences (Table III). Cysteine, however, is present at a 15-fold higher amount in EPO compared to the proteins that make up *P. pastoris* biomass, while the cysteine content of *P. pastoris* is less than 50% the content in *S. cerevisiae* (Sohn et al., 2010). As cysteine is a sulfur containing protein, an increased requirement of cysteine will demand a similar increase in sulfur. Interestingly, the potential of the MET genes (Table II) as potential targets is corroborated with the findings that genes of methionine pathway are upregulated after a shift from glycerol to methanol as a carbon source, which induced the expression of recombinant protein (Sauer et al., 2004).

In addition to EPO, FSEOF was also performed for Alb-Redhill, which contains less glycosylation sites per amino acid, has a

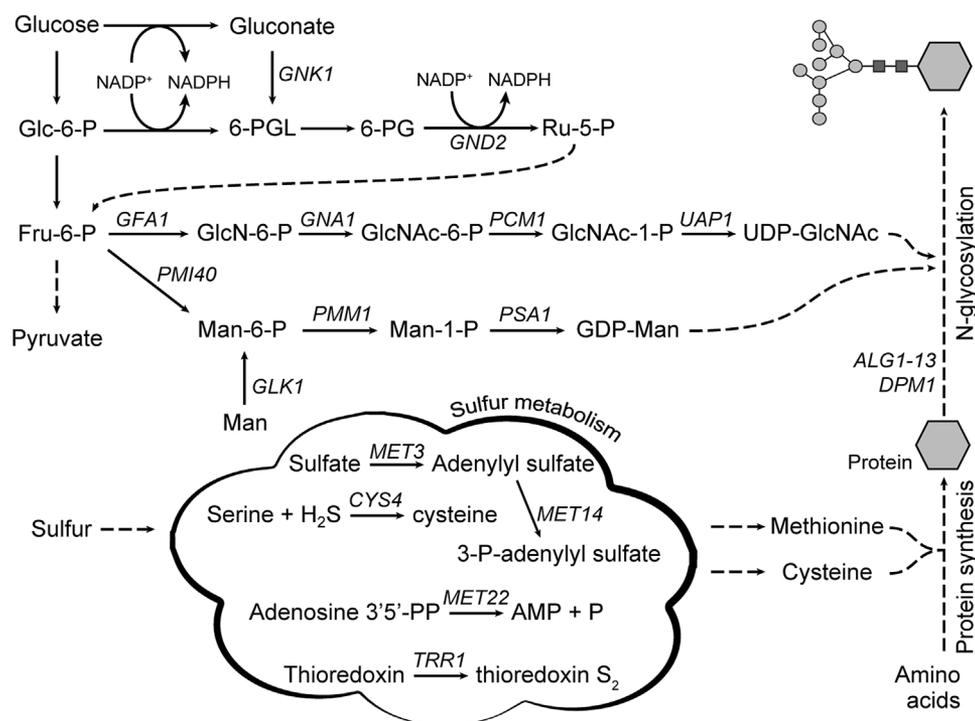


Figure 4. Schematic overview of amplification targets predicted by FSEOF. Names of potential amplification targets are indicated by their abbreviation, corresponding to Table II.

different amino acid composition, and the experimental data used for the experimental yield was based on growth on glucose instead of methanol and mannitol (Maccani et al., 2014). Regardless of these differences, amplification targets suggested by

FSEOF were similar for both proteins. While the fold-changes of the different groups are larger for Alb-Redhill, due to the larger difference in the experimental and the theoretical yield, they follow the same trend.

Table III. Amino acid composition of Alb-Redhill, EPO, and *P. pastoris* proteome.

Amino Acid	% in Alb-Redhill	% in EPO	% in proteome	Difference Alb-Redhill vs. proteome	Difference EPO vs. proteome
Alanine	10	10	10	1.0	1.0
Arginine	4.6	6.7	5.9	0.8	1.1
Asparagine	2.8	3.1	4.7	0.6	0.7
Aspartate	5.9	3.1	4.7	1.3	0.7
Cysteine	5.7	2.6	0.2	<u>32.8</u>	<u>14.8</u>
Glutamate	10	6.7	8.2	1.2	0.8
Glutamine	3.3	3.6	8.2	0.4	0.4
Glycine	2.1	6.2	7.3	0.3	0.9
Histidine	2.6	1.6	1.8	1.4	0.9
Isoleucine	1.5	2.6	4.4	0.3	0.6
Leucine	10	17	7.4	1.4	2.3
Lysine	9.8	4.1	6.7	1.5	0.6
Methionine	1.1	1.0	0.8	1.5	1.4
Phenylalanine	5.7	2.1	3.2	1.8	0.6
Proline	3.9	5.7	4.1	0.9	1.4
Serine	4.6	6.2	6.7	0.7	0.9
Threonine	4.9	5.7	5.9	0.8	1.0
Tryptophan	0.3	2.6	1.4	0.2	1.9
Tyrosine	3.1	2.1	2.3	1.4	0.9
Valine	7.0	6.7	6.2	1.1	1.1

Underlined is the largest difference between recombinant proteins and *P. pastoris* proteome. Values for *P. pastoris* proteome obtained from (Carnicer et al., 2009).

Conclusion

Here, we generated the first functional GEM of *P. pastoris*, *ihGlycopastoris* that includes N-glycosylation and is capable of simulating humanized glycosylation. Our results demonstrated that N-glycosylation should be taken into account when one uses GEMs to study recombinant protein expression in *P. pastoris*, as failing to do so, results in production yields that are too high. The genetic engineering of humanized N-glycosylation lowers the protein yield further, while the precise effect is dependent on N-glycosylation level (number of sites/number of amino acids), as the glycosylation competes for carbons with amino acid biosynthesis.

In *ihGlycopastoris*, we introduced N-glycosylation for recombinant protein production. It should be noted that also native *P. pastoris* proteins undergo N-glycosylation, however, this was not taken into account. Effort has been made to model N-glycosylation of the proteome of *S. cerevisiae* (Feizi, Marras, Nielsen, manuscript in preparation), however, instrumental for this is the availability of genome-wide measurements of N-glycosylation. This extension is currently out of reach for *P. pastoris* as this data is unavailable. The absence of N-glycosylation of native *P. pastoris* in *ihGlycopastoris* also influences the FSEOF analysis, as the absence of N-glycans in *P. pastoris* biomass puts more weight on the presence of N-glycans in the recombinant protein. While this model is an improvement over previous genome-scale models of *P. pastoris* that do not take N-glycosylation into account, there is potential to further improve the model by describing processes involved in (recombinant) protein production that are currently not covered by our model. Processes such as protein folding, but also the remainder of the secretory pathway have been identified as having a severe effect on protein yields (Hohenblum et al., 2004). While inefficiencies of protein expression, such as aggregation, mis-targeting, and degradation are affecting protein yield, also the nature of the protein, e.g., its hydrophobicity, arrangement and composition of amino acids, affect protein yield. The absence of these processes in the current model are likely responsible for the discrepancy between the experimental and model predicted protein yields, indicating that future iterations of this model could further improve its application. Additionally, the current model does not take the micro- and macroheterogeneity of N-glycosylation into account, especially for the N-glycosylation of *P. pastoris*' native proteome. A more detailed incorporation would benefit from the challenging experimental measurements of the N-glycosylation of *P. pastoris* native protein. For *S. cerevisiae*, for which there is more information available, the protein secretory machinery has been modelled in more detail (Feizi et al., 2013). Nonetheless, *ihGlycopastoris* is an important step forward to gain a more accurate description of protein production in microbes.

References

Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J. 2013. The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLoS Comput Biol* 9:e1002980.

Barrigon JM, Valero F, Montesinos JL. 2015. A macrokinetic model-based comparative meta-analysis of recombinant protein production by *Pichia pastoris* under AOX1 promoter. *Biotechnol Bioeng* 112:1132–1145.

Brennan SO, Myles T, Peach RJ, Donaldson D, George PM. 1990. Albumin Redhill (-1 Arg, 320 Ala--Thr): A glycoprotein variant of human serum albumin whose precursor has an aberrant signal peptidase cleavage site. *Proc Natl Acad Sci* 87:26–30.

Bretthauer RK. 2003. Genetic engineering of *Pichia pastoris* to humanize N-glycosylation of proteins. *Trends Biotechnol* 21:459–462.

Çalik P, Inankur B, Soyaslan EŞ, Şahin M, Taşpınar H, Açıık E, Bayraktar E. 2010. Fermentation and oxygen transfer characteristics in recombinant human growth hormone production by *Pichia pastoris* in sorbitol batch and methanol fed-batch operation. *J Chem Technol Biotechnol* 85:226–233.

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

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

Cereghino J. 2000. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev* 24:45–66.

Chen W, Smeeckens JM, Wu R. 2014. Comprehensive analysis of protein N-glycosylation sites by combining chemical deglycosylation with LC-MS. *J Proteome Res* 13:1466–1473.

Chiba Y, Akeboshi H. 2009. Glycan engineering and production of “humanized” glycoprotein in yeast cells. *Biol Pharm Bull* 32:786–795.

Choi HS, Lee SY, Kim TY, Woo HM. 2010. In silico identification of gene amplification targets for improvement of lycopene production. *Appl Environ Microbiol* 76:3097–3105.

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

Cregg JM, Cereghino JL, Shi J, Higgins DR. 2000. Recombinant protein expression in *Pichia pastoris*. *Mol Biotechnol* 16:23–52.

Delic M, Valli M, Graf AB, Pfeffer M, Mattanovich D, Gasser B. 2013. The secretory pathway: Exploring yeast diversity. *FEMS Microbiol Rev* 37:872–914.

Dragosits M, Stadlmann J, Albiol J, Baumann K, Maurer M, Gasser B, Sauer M, Altmann F, Ferrer P, Mattanovich D. 2009. The effect of temperature on the proteome of recombinant *Pichia pastoris*. *J Proteome Res* 8:1380–1392.

Eskitoros MŞ, Ata Ö, Çalik P. 2014. Metabolic reaction network of *Pichia pastoris* with glycosylation reactions: Flux analysis for erythropoietin production. *J Chem Technol Biotechnol* 89:1675–1685.

Eskitoros MŞ, Çalik P. 2014. Co-substrate mannitol feeding strategy design in semi-batch production of recombinant human erythropoietin production by *Pichia pastoris*. *J Chem Technol Biotechnol* 89:644–651.

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

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

Gemmill TR, Trimble RB. 1999. Overview of N- and O-linked oligosaccharide structures found in various yeast species. *Biochim Biophys Acta* 1426:227–237.

Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, Bobrowicz P, Stadheim TA, Li H, Choi BK, Hopkins D, Wischniewski H, Roser J, Mitchell T, Strawbridge RR, Hoopes J, Wildt S, Gerngross TU. 2006. Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science* 313:1441–1443.

Hamilton SR, Gerngross TU. 2007. Glycosylation engineering in yeast: The advent of fully humanized yeast. *Curr Opin Biotechnol* 18:387–392.

Helenius A, Aebi M. 2001. Intracellular functions of N-linked glycans. *Science* 291:2364–2369.

Hohenblum H, Gasser B, Maurer M, Borth N, Mattanovich D. 2004. Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*. *Biotechnol Bioeng* 85:367–375.

Jordà J, de Jesus SS, Peltier S, Ferrer P, Albiol J. 2014. Metabolic flux analysis of recombinant *Pichia pastoris* growing on different glycerol/methanol mixtures by iterative fitting of NMR-derived ¹³C-labelling data from proteinogenic amino acids. *N Biotechnol* 31:120–132.

Jungo C, Rérat C, Marison IW, von Stockar U. 2006. Quantitative characterization of the regulation of the synthesis of alcohol oxidase and of the expression of

- recombinant avidin in a *Pichia pastoris* Mut⁺ strain. *Enzyme Microb Technol* 39:936–944.
- Jungo C, Marison I, von Stockar U. 2007a. 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* 128:824–837.
- Jungo C, Schenk J, Pasquier M, Marison IW, von Stockar U. 2007b. A quantitative analysis of the benefits of mixed feeds of sorbitol and methanol for the production of recombinant avidin with *Pichia pastoris*. *J Biotechnol* 131:57–66.
- Jungo C, Urfer J, Zocchi A, Marison I, von Stockar U. 2007c. Optimisation of culture conditions with respect to biotin requirement for the production of recombinant avidin in *Pichia pastoris*. *J Biotechnol* 127:703–715.
- Karagözü D, Kelleher DJ, Gilmore R. 1997. The highly conserved Stt3 protein is a subunit of the yeast oligosaccharyltransferase and forms a subcomplex with Ost3p and Ost4p. *J Biol Chem* 272:32513–32520.
- Keating SM, Bornstein BJ, Finney A, Hucka M. 2006. SBMLToolbox: An SBML toolbox for MATLAB users. *Bioinformatics* 22:1275–1277.
- Kerkhoven EJ, Lahtvee PJ, Nielsen J. 2015. Applications of computational modeling in metabolic engineering of yeast. *FEMS Yeast Res* 15:1–13.
- Kim B, Kim WJ, Kim DI, Lee SY. 2014. Applications of genome-scale metabolic network model in metabolic engineering. *J Ind Microbiol Biotechnol* 42:339–348.
- Kobayashi K, Nakamura N, Sumi A, Ohmura T, Yokoyama K. 1998. The development of recombinant human serum albumin. *Ther Apher* 2:257–262.
- Maccani A, Landes N, Stadlmayr G, Maresch D, Leitner C, Maurer M, Gasser B, Ernst W, Kunert R, Mattanovich D. 2014. *Pichia pastoris* secretes recombinant proteins less efficiently than Chinese hamster ovary cells but allows higher space-time yields for less complex proteins. *Biotechnol J* 9:526–537.
- Marx H, Mecklenbräuker A, Gasser B, Sauer M, Mattanovich D. 2009. Directed gene copy number amplification in *Pichia pastoris* by vector integration into the ribosomal DNA locus. *FEMS Yeast Res* 9:1260–1270.
- Nocon J, Steiger MG, Pfeffer M, Sohn SB, Kim TY, Maurer M, Rußmayer H, Pflügl S, Ask M, Haberhauer-Troyer C, Ortmayr K, Hann S, Koellensperger G, Gasser B, Lee SY, Mattanovich D. 2014. Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production. *Metab Eng* 24:129–138.
- Puxbaum V, Mattanovich D, Gasser B. 2015. Quo vadis? The challenges of recombinant protein folding and secretion in *Pichia pastoris*. *Appl Microbiol Biotechnol* 99:2925–2938.
- Samuelson J, Banerjee S, Magnelli P, Cui J, Kelleher DJ, Gilmore R, Robbins PW. 2005. The diversity of dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. *Proc Natl Acad Sci USA* 102:1548–1553.
- Sauer M, Branduardi P, Gasser B, Valli M, Maurer M, Porro D, Mattanovich D. 2004. Differential gene expression in recombinant *Pichia pastoris* analysed by heterogeneous DNA microarray hybridisation. *Microb Cell Fact* 3:17.
- Sohn SB, Graf AB, Kim TY, Gasser B, Maurer M, Ferrer P, Mattanovich D, Lee SY. 2010. Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production. *Biotechnol J* 5:705–715.
- Solà A, Maaheimo H, Ylönen K, Ferrer P, Szyperski T. 2004. Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*. *Eur J Biochem* 271:2462–2470.
- Solà A, Jouhten P, Maaheimo H, Sánchez-Ferrando F, Szyperski T, Ferrer P. 2007. Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates. *Microbiology* 153:281–290.
- Tortajada Serra M. 2012. Process development for the obtention and use of recombinant glycosidases: Expression, modelling, and immobilisation; Universitat Politècnica de València.
- Trimble RB, Atkinson PH, Tschopp JE, Townsend RR, Maley F. 1991. Structure of oligosaccharides on *Saccharomyces SUC2* invertase secreted by the methylotrophic yeast *Pichia pastoris*. *J Biol Chem* 266:22807–22817.
- Zhan H, Liu B, Reid SW, Aoki KH, Li C, Syed RS, Karkaria C, Koe G, Sitney K, Hayenga K, Mistry E, Savel L, Dreyer M, Katz BA, Schreurs J, Matthews DJ, Cheetham JC, Egrie J, Giebel LB, Stroud RM. 1999. Engineering a soluble extracellular erythropoietin receptor (EPObp) in *Pichia pastoris* to eliminate microheterogeneity, and its complex with erythropoietin. *Protein Eng* 12:505–513.
- Zhang W, Liu CP, Inan M, Meagher MM. 2004. Optimization of cell density and dilution rate in *Pichia pastoris* continuous fermentations for production of recombinant proteins. *J Ind Microbiol Biotechnol* 31:330–334.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.