de Ruijter, Jorg C.; Koskela, Essi V.; Nandania, Jatin; Frey, Alexander D.; Velagapudi, Vidya

Understanding the metabolic burden of recombinant antibody production in Saccharomyces cerevisiae using a quantitative metabolomics approach

Published in:
Yeast

DOI:
10.1002/yea.3298

Published: 01/04/2018

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Understanding the metabolic burden of recombinant antibody production in *Saccharomyces cerevisiae* using a quantitative metabolomics approach

Jorg C. de Ruijter[1], Essi V. Koskela[1], Jatin Nandania[2], Alexander D. Frey[1#] and Vidya Velagapudi[2#]

1 Department of Bioproducts and Biosystems, School of Chemical Engineering, Aalto University, Finland, Kemistintie 1, 02150 Espoo, Finland

2 Metabolomics Unit, Institute for Molecular Medicine Finland FIMM, University of Helsinki, Tukholmankatu 8U, Helsinki, Finland

Corresponding authors: alexander.frey@aalto.fi vidya.velagapudi@helsinki.fi

Phone: +358 50 411 65 06 Phone: +358 50 317 50 87

Fax: +358 9 462 373

Running title: Antibody expression affects cellular amino acid and redox metabolism

Keywords: Recombinant protein production, targeted metabolomics, antibody, metabolic burden

Abstract

The cellular changes induced by heterologous protein expression in the yeast *Saccharomyces cerevisiae* have been analyzed on many levels and found to be significant. However, even though high-level protein production poses a metabolic burden, evaluation of the expression host at the level of the metabolome has often been neglected. We present a comparison of metabolite profiles of a wild-type strain with those of three strains producing recombinant antibody variants of increasing size and complexity: a scFv fragment, a scFv-Fc fusion protein, and a full-length IgG molecule. Under producing conditions, all three recombinant strains showed a clear decrease in growth rate compared to the wild-type strain and the severity of the growth phenotype increased with size of the protein. The levels of 76 intracellular metabolites were determined using a targeted (semi) quantitative mass spectrometry based approach. Based on unsupervised and supervised multivariate analysis of metabolite profiles, together with pathway activity profiling, the recombinant strains were found to be significantly different from each other and from the wild-type strain. We observed the most prominent changes in metabolite levels for metabolites involved in amino acid and redox metabolism. Induction of the unfolded protein response was detected in all producing strains and is considered to be a contributing factor to the overall metabolic burden on the cells.

Introduction

The yeast *Saccharomyces cerevisiae* is a common expression host for the production of high value compounds, such as biopharmaceutical proteins and biofuels. The use of this yeast as an expression platform is favorable on account of the vast amount of scientific data available from single components to cellular level, combined with an intrinsic ease of manipulation and cultivation (Mattanovich et al., 2014). Often, it has been observed that recombinant protein production in microbial hosts induces significant changes to the host cell. The nature of these changes has been studied extensively, as they could be used as targets to improve the expression host. So far, changes related to heterologous protein production have been analyzed on genome, transcriptome, or proteome level. One example are the vast changes in the regulation of gene expression orchestrated by the unfolded protein response (UPR) induced by recombinant protein production. The UPR is the response of the cell to a high load of unfolded proteins in the endoplasmic reticulum (ER), which leads to an induction of expression of proteins related to many cellular processes, like protein folding, vesicular transport and the ER associated degradation pathway (Friedlander et al., 2000; Thibault et al., 2011).

However, a more general change that recombinant protein productions induces, is a significant metabolic burden on the host cell. This is the result of a redirection of resources from regular cellular activities towards the needs created by recombinant protein production (Glick, 1995). For *S. cerevisiae* this was shown to lead to a reduction in the maximum specific growth rate, a decreased biomass yield, and a lower respiratory capacity (Glick, 1995; Görgens et al., 2001; Karim et al., 2013; Kauffman et al., 2002; Kazemi Seresht et al., 2013). Noting that many cellular processes change when recombinant protein production is induced, it is to be expected that a shift in the direction of cellular resources to accommodate elevated protein production also alters the levels of intracellular metabolites. However, analysis of the effects of the induced metabolic burden has been mostly related to parameters like

growth rate, biomass yield, and carbon source consumption or by-product excretion (Görgens et al., 2001; Van Rensburg et al., 2012). The changes in intracellular metabolite levels caused by recombinant protein production in *S. cerevisiae* have often been neglected, even though such insights could lead to the identification of novel leads for metabolic engineering of recombinant strains. In one study, it was shown that long-term cultivation of a recombinant insulin producing *S. cerevisiae* strain eventually reduced the insulin production due to the cellular adaptation to the metabolic burden (Kazemi Seresht et al., 2013). This indicates that an evaluation of the intracellular metabolic alteration due to high level expression of recombinant proteins, provide us new insights on cellular adaptations to the stress, as currently not much is known about how the expression of heterologous proteins affects the allocation of cellular resources in *S. cerevisiae*.

Here, we studied the metabolic response of *S. cerevisiae* to production of recombinant human antibody variants. We overexpressed a scFv fragment, a scFv-Fc fusion protein, and a full IgG molecule in the yeast laboratory wild-type strain SS328 and measured 76 intracellular metabolites by using a high-throughput targeted (semi) quantitative mass spectrometry based metabolomics approach. The changes in metabolite levels from the three recombinant strains were compared to the wild-type and analyzed based on their specific roles in metabolic pathways together with changes in metabolic pathway activities found using a pathway activity profiling method. Our results demonstrate that significant differences in metabolite profiles were induced by recombinant protein production in the recombinant strains when compared with the wild-type strain, which to some degree can be ascribed to the induction of unfolded protein response.

Methods and Materials

All media components and reagents were obtained from Sigma-Aldrich (Helsinki, Finland), unless stated otherwise. Yeast nitrogen base without amino acids (YNB) was obtained from BD (Vantaa, Finland).

Strain generation

All *S. cerevisiae* strains used in this study were derived from the parental strain SS328 (ATCC® MYA193™) and are listed in Table 1. The lithium acetate method was used for yeast transformations. The sequences encoding the variable heavy (V_H) and light (V_L) domains were derived from the HyHEL-10 Fab molecule and were fused with the sequences encoding constant light (C_L) or constant heavy (C_H1, C_H2, and C_H3) domains derived from human IgG_1 to generate a scFv molecule, a scFv-Fc fusion protein, and a full-length IgG molecule, respectively. All sequences were codon-optimized for expression in yeast. The linker region for scFv comprised three repeats of the sequence GGGS. Generation of the scFv-Fc fusion is based on the cloning strategy described earlier (Powers et al., 2001). All proteins fused to the Matα prepropeptide were expressed under control of the *GAL1* promoter and *CYC1* terminator from low copy number plasmids.

An unfolded protein response reporter was generated based on a construct published earlier (Pincus et al., 2010). The *SacI ScaI* fragment from pDEP17 containing four repeats of the UPR responsive element fused to GFP was inserted into plasmid pRS413 creating pJR77. The reporter construct was co-transformed with constructs for expression of antibody variants.

Small scale cultivations

For recording growth curves, cultures of exponentially growing cells were diluted to OD_{600}= 0.02 in minimal medium (0.67 % YNB with supplementation of necessary amino acids and adenine) with 2 % raffinose, or in 2 % raffinose supplemented with 4 % galactose to induce protein expression. One

hundred microliters of these cultures were grown in a round-bottom microtiter plate, with continuous orbital shaking (425 rpm, 3 mm) at 30 °C in an Eon Microplate Spectrophotometer (BioTek, Winooski, USA). OD_{600} measurements were taken every 15 min for 48 h. The growth curves were analyzed with R-package *grofit* (Kahm et al., 2010), and the characteristic growth parameters were extracted from the best model fit. The experiment was conducted with four culture replicates.

The strains comprising the antibody variant expression construct and the UPR reporter were grown in identical conditions as described above. GFP expression was monitored using the appropriate filter set in Eon Microplate Spectrophotometer (BioTek, Winooski, USA).

Shake flask cultivation

Yeast cells were inoculated into precultures of 5 mL minimal medium. These overnight cultures were used to seed 30 mL of the main cultures at a starting OD_{600}=0.2. After 6 hours of cultivation at 30 °C and 180 rpm, a 40% galactose solution was added to a final concentration of 4 % to induce protein expression.

Cell collection for Western blotting

The cultures for analysis of protein expression were harvested after 18 hours of expression. A culture volume equivalent to 10 OD_{600} of cells was collected by centrifugation. The cell pellet was resuspended in 200 µL lysis buffer (SDS-Page sample buffer (62.5 mM Tris-HCl (pH 6.8), 2 % SDS, 10 % glycerol, 50 mM DTT, and 0.005 % bromophenolblue), supplemented with 1X cOmplete®, EDTA-free protease inhibitor cocktail (Roche, Basel, Switzerland) and 1 mM PMSF. Half of the volume of acid washed glass beads was added, and cells were lysed by vortexing for 10 minutes at 4°C using a Disruptor genie (Scientific industries). The samples were centrifuged for 5 min, 10,000 rcf at 4 °C, after which the supernatants were collected and heated for 5 minutes at 65°C.

10 µL of sample was loaded onto a 12.5 % SDS-polyacrylamide gel and the electrophoresis was run in SDS-Tris-glycine buffer. After the SDS-PAGE, proteins were transferred to a nitrocellulose membrane using western blotting. A 1:15,000 dilution of a goat anti-human IgG (Fc specific)-peroxidase labelled antibody (Sigma-Aldrich, Helsinki, Finland) was used for staining the scFv-Fc fusion protein and the full length-IgG. A 1:2,000 dilution of anti-tetra His antibody produced in mouse (Qiagen, USA) in combination with a 1:50,000 dilution of rabbit anti-mouse IgG (Fc specific)-peroxidase labelled (Sigma-Aldrich, Helsinki, Finland)) was used for staining the His-tag in the scFv fragment. Signal was detected with the Supersignal West Pico Chemiluminescent substrate kit (ThermoFisher scientific, Helsinki, Finland) following the manufacturer's instructions.

Collection and preparation of samples for metabolite analysis

The cultures for metabolome analysis were harvested in the late exponential phase ($OD_{600} \approx 2.2$).Seven replicate cultures were cultivated and processed for metabolite analysis. The fast filtration protocol for harvesting of the yeast cultures was modified from Kim and colleagues (Kim et al., 2013). In short, 2 mL of the *S. cerevisiae* cultures were vacuum filtrated through a nylon membrane filter (0.45 µm pore size, 30 mm diameter, Whatman, Piscataway, USA). The cell residue was washed with 10 mL of water, after which the filter with the cells was transferred to a 5 mL Eppendorf tube, which was flash frozen in liquid nitrogen. Samples were stored at -80 °C until metabolite extraction. Repeated sampling from various cultures showed that the sampling method was highly reproducible, with a variation of wet biomass weight well below 1%. The full procedure, from culture to flash freezing, was completed in less than 1 minute, so that metabolic changes during the harvesting procedure were minimized as much as possible.

An aliquot of 20 µl of internal standard mix(IST) was added to yeast cell samples, which were thawed step wise at -20 ^{o}C and +4 ^{o}C. A total of 960 µl of extraction solvent (90 % acetonitrile, 1 % formic

acid in H_20) was added and three cycles of extraction were carried out by vortexing for 2 min and sonicating for 1 min (settings: sweep mode, frequency 37, power 60, no heating). Tubes were incubated on ice for 10 min between vortexing and sonicating steps. After this, the tubes were centrifuged at 14,000 rpm for 15 min at +4 ˚C. An aliquot of 800 µl of the supernatant was transferred to Ostro 96-well plate (Waters Corporation, Milford, USA) and filtered by applying vacuum at a delta pressure of 400-500 mbar on Hamilton StarLine robot's vacuum station (Hamilton, Bonaduz, Switzerland). The clean extract was collected to a 96-well collection plate, which was placed under the Ostro plate. The collection plate was sealed and centrifuged for 10 min, 4,000 rpm, +4 ºC and placed in auto-sampler of the liquid chromatography system for injection. All samples were analyzed in a double random order i.e., first step at the metabolite extraction phase and the second step at the liquid chromatography (LC) injection order.

Instrumentation and analytical conditions

Sample analysis was performed on an ACQUITY ultra pressure liquid chromatography tandem mass spectrometry (UPLC-MS/MS) system (Waters Corporation, Milford, MA, USA) and the detection system, a Xevo® TQ-S tandem triple quadrupole mass spectrometer (Waters, Milford, MA, USA), was operated in both positive and negative polarities. A detailed description about the analytical conditions and instrument parameters is given elsewhere (Roman-Garcia et al., 2014).

Metabolomics data analysis

Metabolomics data analysis was carried out using a web-based comprehensive metabolomics data processing tool, MetaboAnalyst 2.0 (http://www.metaboanalyst.ca) (Xia et al., 2012, 2009). The non-transformed data was autoscaled i.e., mean centered and divided with standard deviation (SD). Dendrograms were plotted using Ward's linkage clustering algorithm and Pearson's correlation similarity measure. Dendrograms were visualized through heatmaps, where each colored cell on the

map corresponds to a concentration value. In order to explain the maximum separation among groups, unsupervised and supervised multivariate regression techniques, principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA), respectively, were performed. Variable importance in projection (VIP) is one of the important measures of PLS-DA, where it is a weighted sum of squares of the PLS loadings taking into account the amount of explained class variation in each dimension. Metabolites were ranked according to their VIP scores and usually metabolites with VIP scores greater than one are considered as the most significant contributors.

Statistical analysis

Graphpad prism (Version 7.01, Graphpad Software, La Jolla, USA) was used for statistical analysis of the measured metabolite concentrations. Mean and SD of the measured metabolite concentrations of all four strains were calculated, followed by baseline correction of the three recombinant strains with the wild-type. The baseline corrected data was log2 transformed to get the fold changes for the recombinant strains relative to wild-type.

PAPi analysis

Pathway activity profiling (PAPi) algorithm is described by(Aggio et al., 2010), and the analysis was conducted with R-software with the PAPi-package provided by the authors. KEGG IDs of the detected metabolites were mapped with MetaboAnalyst 3.0. The measured metabolite concentrations were used in the analysis and the wild-type sample was assigned as the reference condition. ANOVA-test was used to detect differentially regulated pathways, for which the activity scores were calculated.

Results

Recombinant protein expression reduced maximum growth rate and prolonged the lag phase

The *S. cerevisiae* laboratory wild-type (WT) strain SS328 was used as parental strain for all antibody variant producing strains. The recombinant strains harbor plasmids that encode genes for a scFv fragment (scFv strain), a scFv-Fc fusion protein (Fusion strain), and a corresponding full-length IgG molecule (IgG strain), respectively. To minimize changes in the host strains' physiology that are not related to the type of the produced antibody, all strains have the same genetic background and the expression constructs were identical except for the genes expressed. Expression was driven by the galactose inducible *GAL1* promoter and the polypeptides were directed to the secretory pathway using MATα signal peptide.

Previously, we had observed that the different strains showed variations in their growth patterns during the expression of the antibodies. In order to analyze differences in growth rate among the recombinant strains, growth of strains in microtiter plates was recorded for 48 hours. Without induction of protein production, all four strains displayed similar growth patterns (Figure 1A). Even though the differences between the μ_{max} values were statistically significant they were closely related, with WT growing only slightly faster than the recombinant strains (Table 2). However, when protein production was induced by the addition of galactose at the beginning of the cultivation, all three recombinant strains displayed a longer lag phase compared to the WT (Figure 1B). Additionally, the maximal growth rates of the strains were decreased compared with WT and this decrease became more accentuated with an increase of complexity and size of the produced protein (Table 2). For the largest protein expressed, the full-length IgG, the growth rate of the strain was only half of the control strain.

To confirm expression of the three constructs, cell extracts were prepared from cultivations after 18 hours of galactose induced protein expression. Western blot analysis of the cell lysates showed that all three strains displayed protein signals at the expected apparent molecular sizes, corresponding to the polypeptides with the processed and unprocessed MATα propeptide (Figure 2). In case of the full-length IgG molecule, most of the protein was found to be with the unprocessed MATα propeptide. Although, cell extract prepared from the scFv and Fusion strains also showed a large proportion of protein with the unprocessed MATα propeptide, in comparison to the IgG strain a higher proportion of the mature forms of these proteins could be detected (Figure 2).

Recombinant strains showed distinct metabolomics profiles compared to WT.

In order to analyze the alterations at metabolomics level, cells from the WT and from the antibody expressing strains were collected in the late logarithmic phase and 76 metabolites were measured using a targeted semi-quantitative mass spectrometry approach. Unsupervised multivariate analysis, Principal Component Analysis (PCA), was performed to reveal the clustering patterns among the four sample groups. The explained variance of principal component 1 (PC1) and PC2 were 47.7 % and 20.9 %, respectively, and it can be observed from Figure 3A that two principal components were sufficient to separate all the four strains from each other, with the Fusion strain being more similar to WT than the scFv and IgG strains. Supervised multivariate analysis, Partial Least Squares Discriminant Analysis (PLS-DA) was performed to identify the top variables that contributed for the group separation. The 20 most significantly changed metabolites in each of the recombinant strains, when compared to WT, were sorted according to their variable of importance in the projection (VIP) scores and the VIP plots are shown in Figure 3B for the scFv strain, Figure 3C for the Fusion strain, and Figure 3D for the IgG strain. The non-proteinogenic amino acid, ornithine, had the highest VIP-scores in all three

recombinant strains. Particularly in the scFv and IgG strains, amino acids had high VIP scores and were very numerous amongst the top20 metabolites.

We used two-way hierarchical clustering to identify changes and similarities among the different strains and metabolites. The dendrogram and heatmap visualized well the clear differences in the levels of the metabolites between the WT and the three recombinant strains and clusters of differentially regulated metabolites could be clearly distinguished (Figure 4A). Based on the metabolite profiles, the Fusion strain was slightly more similar to the WT than the other two recombinant strains. The mean metabolite concentrations were baseline corrected to WT, followed by log2 transformation to get the respective fold changes (Figure 4B). Out of the 76 measured metabolites, 30 metabolites showed a decrease in concentration for at least two out of three recombinant strains. An over 2-fold increase in concentration was observed for 13 metabolites in the scFv strain, and for seven and two metabolites in Fusion and IgG strains, respectively. 11 metabolites showed an over 2-fold decrease in concentration in the scFv strain, while for Fusion and IgG strains, the corresponding number was eight and ten metabolites, respectively.

Ornithine that had the highest VIP-scores in all three recombinant strains was decreased 3.48-, 1.85-, and 4.40-fold in the scFv, Fusion, and IgG strains, respectively, when compared to WT. From the amino acids with a high VIP-score, proline (2.17-, 1.68-, and 1.82-fold), tyrosine (3.06-, 0.61-, and 3.70-fold), and arginine (3.34-, 0.89-, and 3.01-fold) levels were decreased, while alanine showed an increase in concentration of 1.92-, 2.20-, and 1.19-fold for the scFv, Fusion, and IgG strains, respectively, when compared to WT. 4-aminobutanoate (GABA) is part of alanine, aspartate, and glutamate metabolism and its concentrations were significantly decreased in the recombinant strains: 3.55-, 4.42-, and 4.84-fold in the scFv, Fusion, and IgG strains, respectively. The TCA cycle intermediate succinate showed a 4.00-, 4.07-, and 3.62-fold increase in the scFv, Fusion, and IgG

strains, respectively. From the metabolites involved in tryptophan metabolism and *de novo* NAD-biosynthesis, L-kynurenine was in the top ten of VIP scores of all three strains, and was increased 2.51-, 4.06-, and 2.55-fold in the scFv, Fusion, and IgG strains, respectively. From the NAD synthesis pathway using direct incorporation of nicotinic acid, the metabolite nicotinic acid was increased 3.19-, 1.15-, and 1.91-fold in the scFv, Fusion, and IgG strains, respectively, although NAD was increased only 1.41-, 1.01-, and 0.88-fold in the scFv, Fusion, and IgG strains, respectively.

Metabolic pathway activity profiling identified amino acid and redox pathways to be activated in the production strains

In order to gain insights in pathway activity from the measured metabolite levels, we ran the Pathway Activity Profiling (PAPi) algorithm with the measured metabolite profiles (Aggio et al., 2010). The PAPi algorithm is designed to facilitate biological interpretation of metabolomics data by assigning an Activity Score (AS) to each recognized metabolic pathway. AS is inversely proportional to pathway activity. Based on the measured metabolites, the PAPi-algorithm recognized 159 possibly active pathways, of which 117 were found to be differentially regulated in the recombinant strains compared to WT using ANOVA (Supplementary table 1). As the algorithm takes into account the abundance of the metabolites, the AS gives a score for the predicted flux through the pathway and thus can be used to compare samples to each other. From the PAPi ANOVA-analysis, we selected a subset of the differentially regulated pathways based on their relevance to *S. cerevisiae*. The AS of these pathways are shown in Figure 5. Due to the inverse relationship of AS and predicted pathway activity, strains displaying a high AS for a certain pathway, the pathway is projected to be less active. The analysis showed that the recombinant strains have for most of the significantly changed pathways less metabolic activity than the WT. The differences between the predicted pathway activities of the different recombinant strains were relatively small, even though their growth rates were measurably different

under producing conditions (Table 2). This could indicate that only part of the decreased growth rates can be explained as an effect of metabolic changes, where others might be coming from other cellular responses, most probably the induction of the unfolded protein response (UPR) and its associated metabolic costs. The few pathways with higher predicted fluxes in the recombinant strains than in WT were mostly pathways connected to amino acid and redox metabolism, such as the pathway "glutathione metabolism" (Figure 5).

Comparing the amino acids profiles of yeast and antibody variants

As levels of most of the measured amino acids were significantly changed in the recombinant strains compared to the WT (Figure 4) corresponding to changes in predicted metabolic pathway activities in Figure 5, we wondered if the differences between the recombinant strains might originate from a varying amino acid composition of the expressed antibody variants compared to yeast protein.

Based on the amino acid sequence of the three antibody variants, the amino acid composition of the three constructs were calculated (Figure 6A), as reference the average amino acid composition in *S. cerevisiae* as reported by Karlin et al (Karlin et al., 2002) was used. The biggest differences appear to be a lower content of arginine, asparagine and aspartate, respectively, and a higher content of serine and threonine, in the recombinant proteins, when compared with the average composition of yeast protein. Based on the measured concentrations of the amino acids we calculated the fold changes in the scFv, Fusion, and IgG strains, relative to WT (Figure 6B). Most of the measured amino acids, 13 out of 18, showed a decrease in concentrations in the recombinant strains with the exception of methionine and alanine ,which were present at higher concentrations.

UPR is activated in antibody expressing strains

The relatively low amounts of antibody variants produced by these strains probably can account directly neither for all of the metabolic changes nor for the effects on growth. Therefore, we have explored other sources for this additional metabolic burden. One of the major cellular reactions that can be triggered by the overexpression of proteins is the UPR.

Therefore, we have generated strains that in addition to the antibody variant expression plasmids harbored a GFP based UPR reporter. The strains were grown under non-inducing and inducing conditions and the GFP signal was continuously monitored. The GFP signals of the induced cultures were normalized to the non-induced cultures. Whereas in the case of the control cultures no increase in GFP signal was observed, expression of antibody variants induced GFP expression indicating that the UPR is activated, with the signal peaking around 16 hours of cultivation (Figure 7).

Discussion

Size of the expressed antibody fragment is an important factor in the decrease of maximum growth rate

The expression of the different antibody variants in *S. cerevisiae* imposed a serious burden on the host cells, resulting in a lag period after protein production was induced and an overall decrease in maximum growth rate. For expression of the scFv fragment, the lag period, followed by a general decrease in growth rate has been observed before (Kauffman et al., 2002). Moreover, several *S. cerevisiae* strains have been reported to show a decreased growth rate upon expression of recombinant proteins, for example for a recombinant insulin, xylanase, and cellulases (Görgens et al., 2001; Kazemi Seresht et al., 2013; Van Rensburg et al., 2012). Interestingly, in this study we also observed that the decrease in growth rate showed a correlation with the size and structural complexity of the expressed protein. The structure of all three proteins is solely based on varying numbers of the Ig-fold and the number of Ig-folds increases from two for the scFv fragment, eight for scFv-Fc fusion protein, to 12 for the full-length IgG molecule. We have found in a previous study that overexpression of certain folding catalysts increased IgG secretion efficiency, indicating that protein folding is indeed a bottleneck (de Ruijter et al., 2016). Moreover, the cell extracts of production cultures showed that the antibody variants were present with both processed and unprocessed Matα signal peptides (Figure 2). The presence of a relatively large fraction of the unprocessed polypeptides indicates that a large portion of the intracellular recombinant proteins is in the folding and maturation process. Protein folding has been proposed to be the most energy consuming process of the yeast secretory machinery (Feizi et al., 2013), which could explain the linear decrease of growth rate with protein size.

Differences in amino acid concentrations are not predictably correlated with amino acid composition of the produced antibody variant

When we compared changes in the pool of free amino acids in the recombinant strains with the observed differences in amino acid composition of the expressed proteins, no correlations between amino acid composition and free amino acid levels were detected (Figure 6). However, the data in Figure 6B showed that the expression of recombinant protein had a severe impact on the availability of the free amino acids, with changes ranging between a one- and two-fold increase for methionine and alanine to an over three-fold decrease for tyrosine and arginine. The decrease in concentration of most of the amino acids was reflected in an increased predicted relative flux of related metabolic pathways (Figure 5, for example "Phenylalanine, tyrosine and tryptophan biosynthesis"). Interestingly, alanine was less abundant in the fusion construct than in the scFv fragment, the full-length IgG molecule, or the average yeast protein (Figure 6A), but showed the highest increase in concentration of all three recombinant strains (a 2.0-fold increase compared with a 1.10- and 1.19-fold increase for the scFv and IgG strains, respectively). This partially explains the major differences in the PAPi analysis for the recombinant strains, as the AS of alanine metabolic pathways for the Fusion strain were closer to the WT (Figure 5), possibly reflecting the lower need of this amino acid. Although the primary and preferential N-source NH_4SO_4 was provided in sufficient amounts to sustain growth to much higher cell densities, the intracellular amino acid pool seemed to be limited. Therefore, a production strain could possibly be optimized by tuning amino acid levels through engineering of expression levels of enzymes in related metabolic pathways or through supplementation in the culture media. This latter approach has proven to be successful for some combinations of amino acids to improve the production of a xylanase (Görgens et al., 2005).

Recombinant protein production induces a carbon redistribution towards amino acid synthesis

Most of the amino acids with an increased concentration in the recombinant strains (alanine, valine, serine, glycine) were derived from glycolytic precursors, while amino acids derived from TCA intermediates (proline, arginine, aspartic acid, threonine) and the pentose phosphate pathway (phenylalanine, tyrosine, histidine), were predominantly decreased in concentration (Figure 6B). Most of the carbon influx to the cell starts at glycolysis, so it seems that most of the carbon is extracted early to the synthesis of simple amino acids, leaving little precursors to the other pathways. In the PAPi analysis, mainly the metabolism of more complicated amino acids had an increased predicted flux while pathways to produce energy, like the TCA cycle and oxidative phosphorylation, showed a decreased activity compared to the control (Figure 5). Energy formation might be limited by the use of available carbons to amino acid synthesis, thus resulting in a decreased growth rate. The limitation for growth and protein production seem to be found in the edges of complex biosynthetic pathways and in the decreased flux into energy formation.

Cellular redox balance is destabilized by recombinant protein production

The PAPi analysis predicted the pathway “glutathione metabolism” to have the one of highest increases in flux in the recombinant strains when compared with WT. Recombinant protein production leads to an increased demand for the formation of disulfide bonds, and this consumes reduced glutathione when the bonds are broken (Tyo et al., 2012). Additionally, glutathione can act as a buffer for oxidative stress in the ER (Cuozzo and Kaiser, 1999). Glutathione levels showed a 1.39-fold decrease in the IgG strain when compared with WT, and glutathione concentration was decreased 0.94- and 0.62-fold in the scFv and Fusion compared to WT, respectively. This indicates that the production of the antibody fragments imposes a burden on the ER redox balance.

A second metabolite that indicated oxidative stress and disturbances in redox-pathways was GABA, as the measured concentration of this metabolite was significantly lower in the recombinant strains and decreased more with an increased size of the expressed protein, with a 3.55-fold decrease for the scFv, 4.42-fold for the Fusion, and 4.84-fold for the IgG strain. Moreover, GABA was one of the determining metabolites contributing to difference among the strains, as can be seen from the VIP plots (Figure 3). GABA was shown to play an important role in oxidative stress tolerances as revealed in a study characterizing the enzyme glutamate decarboxylase, which converts glutamate to GABA (Coleman et al., 2001). In the same study, it was shown that an increased presence of glutamate decarboxylase can act as a buffer for redox changes in the cell through a downstream conversion from GABA to succinate. In our data, succinate was shown to be present at up to 4-fold higher levels in the recombinant strains. As redox metabolism seems to be a key difference between recombinant strains and WT in our metabolite analysis, it is a possible target for strain improvement. It could be helpful to increase the concentration of glutathione to elevate the redox buffer capacity of the cell, or, as demonstrated by Coleman et al (Coleman et al., 2001), upregulate the expression of glutamate decarboxylase, as this was shown to increase the oxidative stress tolerance of the cells.

Induction of UPR contributes to the overall metabolic burden on the yeast cells

The data from the growth and metabolomics analysis showed that the induction of recombinant antibody variants clearly had a large effect on the host cell. However, it has to be kept in mind that for the production of rather complex recombinant proteins the amount of protein produced is rather low compared to the total cellular protein produced, totaling often less than 1% of the total protein produced. For the yeast strains used in this study, for example concentration of full-length IgG molecule would reach a maximum of 1.3 mg IgG per gram of dry yeast cells (de Ruijter, 2016). This low fraction of produced recombinant proteins would indicate that other cellular processes also

contribute to the overall metabolic burden on the cells. One of the best described processes induced by recombinant protein production is the UPR, which regulates 381 genes and can thus cause a significant burden on the cells (Travers et al., 2000).

The results from our UPR induction experiments (Figure 7) indicated that in all three recombinant strains the burden of the protein expression is high enough to activate the UPR. Unintuitively, the UPR-induction was highest in the scFv expressing strain, which is the protein with lowest complexity, while the UPR-induction was lower in the strains expressing the fusion and IgG. This lower induction level is possibly related to the more severe effects of the protein expression on cellular growth and viability, as this influences the generation of the GFP signal.

Induction of the UPR is dependent on activation of the sensor Ire1p and its downstream transcriptional activator Hac1p. However, some of the UPR target genes are under combined control of Ire1p/Hac1p and Gcn4p, respectively, which is the downstream transcriptional activator of the general amino acid control (GAAC) pathway (Herzog et al., 2013; Patil et al., 2004). It has been reported that around half of the UPR induced genes also rely on the presence of basal levels of Gcn4p. Gcn4p and its upstream kinase Gcn2 are activated among other stimuli by amino acid starvation. Thus, activation of the reporter in this study could be triggered by the accumulation of unfolded proteins in the ER, but could be accentuated by the low concentration of intracellular free amino acids.

Conclusions

The burden of protein expression as evidenced by the growth phenotypes of the strains increased with the size and complexity of the three antibody variants. This observation is very remarkable as the proteins differ only in the number of Ig-folds, which can fold independently from each other. Based on

the measurement of 76 intracellular metabolites using targeted semi-quantitative mass spectrometry, all three recombinant strains could be clearly distinguished from each other, and also from wild-type. Furthermore, the pathway activity profiling showed that the three recombinant strains had quite differently regulated metabolic pathways compared to wild-type. The most differentiating metabolic pathways were amino acid metabolism, and involvement in regulation of energy metabolism and redox homeostasis. Finally, there are indications that the induction of the UPR is a large contributing factor to the overall metabolic burden imposed on the host cells. Our approach of metabolic mapping was efficient in identifying major cellular consequences of recombinant protein production, pinpointing important areas for improving protein production by cellular and process engineering.

Acknowledgements

A Biocenter Helsinki (BCH) connecting scientists grant supported the work. Essi V. Koskela is recipient of a doctoral study grant from the School of Chemical Engineering, Aalto University, Finland.

Conflict of interest

The authors declare that they have no conflicts of interest.

References

Aggio, R.B.M., Ruggiero, K., Villas-Bôas, S.G., 2010. Pathway activity profiling (papi): From the metabolite profile to the metabolic pathway activity. Bioinformatics 26, 2969–2976. doi:10.1093/bioinformatics/btq567

Chumnanpuen, P., Nookaew, I., Nielsen, J., 2013. Integrated analysis, transcriptome-lipidome, reveals the effects of INO-level (INO2 and INO4) on lipid metabolism in yeast. BMC Syst. Biol. 7, S7. doi:10.1186/1752-0509-7-S3-S7

Coleman, S.T., Fang, T.K., Rovinsky, S.A., Turano, F.J., Moye-Rowley, W.S., 2001. Expression of a glutamate decarboxylase homologue is required for normal oxidative stress tolerance in *Saccharomyces cerevisiae*. J. Biol. Chem. 276, 244–50. doi:10.1074/jbc.M007103200

Cuozzo, J.W., Kaiser, C.A., 1999. Competition between glutathione and protein thiols for disulphide-bond formation. Nat. Cell Biol. 1, 130–135. doi:10.1038/11047

de Ruijter, J.C., Koskela, E. V., Frey, A.D., 2016. Enhancing antibody folding and secretion by tailoring the *Saccharomyces cerevisiae* endoplasmic reticulum. Microb. Cell Fact. 87. doi:10.1186/s12934-016-0488-5

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. doi:10.1371/journal.pone.0063284

Friedlander, R., Jarosch, E., Urban, J., Volkwein, C., Sommer, T., 2000. A regulatory link between ER-associated protein degradation and the unfolded-protein response. Nat. Cell Biol. 2, 379–384. doi:10.1038/35017001

Glick, B.R., 1995. Metabolic load and heterologous gene expression. Biotechnol. Adv.

doi:10.1016/0734-9750(95)00004-A

Görgens, J.F., Van Zyl, W.H., Knoetze, J.H., Hahn-Hägerdal, B., 2005. Amino acid supplementation improves heterologous protein production by *Saccharomyces cerevisiae* in defined medium. Appl. Microbiol. Biotechnol. 67, 684–691. doi:10.1007/s00253-004-1803-3

Görgens, J.F., van Zyl, W.H., Knoetze, J.H., Hahn-Hägerdal, B., Gorgens, J.F., van Zyl, W.H., Knoetze, J.H., Hahn-Hagerdal, B., 2001. The metabolic burden of the PGK1 and ADH2 promoter systems for heterologous xylanase production by *Saccharomyces cerevisiae* in defined medium. Biotechnol Bioeng 73, 238–45. doi:10.1002/bit.1056

Herzog, B., Popova, B., Jakobshagen, A., Shahpasandzadeh, H., Braus, G.H., 2013. Mutual cross talk between the regulators Hac1 of the unfolded protein response and Gcn4 of the general amino acid control of *Saccharomyces cerevisiae*. Eukaryot. Cell 12, 1142–1154. doi:10.1128/EC.00123-13

Kahm, M., Hasenbrink, G., Lichtenberg-Fraté, H., Ludwig, J., Kschischo, M., 2010. grofit : Fitting Biological Growth Curves with R. J. Stat. Softw. 33, 1–21. doi:10.18637/jss.v033.i07

Karim, A.S., Curran, K.A., Alper, H.S., 2013. Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. FEMS Yeast Res. 13, 107–116. doi:10.1111/1567-1364.12016

Karlin, S., Brocchieri, L., Trent, J., Blaisdell, B.E.E., Mrázek, J., 2002. Heterogeneity of Genome and Proteome Content in Bacteria, Archaea, and Eukaryotes. Theor. Popul. Biol. 61, 367–390. doi:10.1006/tpbi.2002.1606

Kauffman, K.J., Pridgen, E.M., Doyle, F.J., Dhurjati, P.S., Robinson, A.S., 2002. Decreased protein expression and intermittent recoveries in BiP levels result from cellular stress during heterologous protein expression in *Saccharomyces cerevisiae*. Biotechnol. Prog. 18, 942–950.

doi:10.1021/bp025518g

Kazemi Seresht, A., Cruz, A.L., de Hulster, E., Hebly, M., Palmqvist, E.A., Van Gulik, W., Daran, J.-M.M., Pronk, J., Olsson, L., 2013. Long-term adaptation of *Saccharomyces cerevisiae* to the burden of recombinant insulin production. Biotechnol. Bioeng. 110, 2749–2763. doi:10.1002/bit.24927

Kim, S., Lee, D.Y., Wohlgemuth, G., Park, H.S., Fiehn, O., Kim, K.H., 2013. Evaluation and optimization of metabolome sample preparation methods for *Saccharomyces cerevisiae*. Anal. Chem. 85, 2169–2176. doi:10.1021/ac302881e

Mattanovich, D., Sauer, M., Gasser, B., 2014. Yeast biotechnology: teaching the old dog new tricks. Microb. Cell Fact. 13, 34. doi:10.1186/1475-2859-13-34

Patil, C.K., Li, H., Walter, P., 2004. Gcn4p and novel upstream activating sequences regulate targets of the unfolded protein response. PLoS Biol. 2. doi:10.1371/journal.pbio.0020246

Pincus, D., Chevalier, M.W., Aragón, T., van Anken, E., Vidal, S.E., El-Samad, H., Walter, P., 2010. BiP binding to the ER-stress sensor Ire1 tunes the homeostatic behavior of the unfolded protein response. PLoS Biol. 8. doi:10.1371/journal.pbio.1000415

Powers, D.B., Amersdorfer, P., Poul, M.-A., Nielsen, U.B., Shalaby, M.R., Adams, G.P., Weiner, L.M., Marks, J.D., 2001. Expression of single-chain Fv-Fc fusions in *Pichia pastoris*. J. Immunol. Methods 251, 123–135. doi:10.1016/S0022-1759(00)00290-8

Roman-Garcia, P., Quiros-Gonzalez, I., Mottram, L., Lieben, L., Sharan, K., Wangwiwatsin, A., Tubio, J., Lewis, K., Wilkinson, D., Santhanam, B., Sarper, N., Clare, S., Vassiliou, G.S., Velagapudi, V.R., Dougan, G., Yadav, V.K., 2014. Vitamin B_{12}-dependent taurine synthesis regulates growth and bone mass. J. Clin. Invest. 124, 2988–3002. doi:10.1172/JCI72606

Thibault, G., Ismail, N., Ng, D.T.W., 2011. The unfolded protein response supports cellular robustness as a broad-spectrum compensatory pathway. Proc. Natl. Acad. Sci. U. S. A. 108, 20597–602. doi:10.1073/pnas.1117184109

Travers, K.J., Patil, C.K., Wodicka, L., Lockhart, D.J., Weissman, J.S., Walter, P., 2000. Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. Cell 101, 249–258. doi:S0092-8674(00)80835-1 [pii]

Tyo, K.E.J., Liu, Z., Petranovic, D., Nielsen, J., 2012. Imbalance of heterologous protein folding and disulfide bond formation rates yields runaway oxidative stress. BMC Biol. 10, 16. doi:10.1186/1741-7007-10-16

Van Rensburg, E., Den Haan, R., Smith, J., Van Zyl, W.H., Görgens, J.F., 2012. The metabolic burden of cellulase expression by recombinant *Saccharomyces cerevisiae* Y294 in aerobic batch culture. Appl. Microbiol. Biotechnol. 96, 197–209. doi:10.1007/s00253-012-4037-9

Xia, J., Mandal, R., Sinelnikov, I. V, Broadhurst, D., Wishart, D.S., 2012. MetaboAnalyst 2.0--a comprehensive server for metabolomic data analysis. Nucleic Acids Res. 40, W127--33. doi:10.1093/nar/gks374

Xia, J., Psychogios, N., Young, N., Wishart, D.S., 2009. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucleic Acids Res. 37, W652--60. doi:10.1093/nar/gkp356

Tables

Table 1: An overview of *Saccharomyces cerevisiae* strains used in this study.

Strain name	**Genotype**	**Description**
WT	MATα*ade2*-101, *his3*Δ200, *lys2*-80, *ura3*-52	Laboratory wild type strain SS328
ScFv	SS328 with pAF10(*URA3*)	ScFv fragment expression plasmid
Fusion	SS328 with pAX512(*URA3*)	ScFv-Fc fusion expression plasmid
IgG	SS328 with pAX538(*URA3*)	Full-length IgG expression plasmid
WT-UPR	WT with pJR77 (*HIS3*)	Laboratory wild type strain SS328with UPR reporter
ScFv-UPR	ScFv with pJR77 (*HIS3*)	ScFv fragment expression plasmidwith UPR reporter
Fusion-UPR	Fusion with pJR77 (*HIS3*)	ScFv-Fc fusion expression plasmidwith UPR reporter
IgG-UPR	IgG with pJR77 (*HIS3*)	Full-length IgG expression plasmidwith UPR reporter

Table 2: Maximum growth rates of strains under normal (raffinose) and protein production inducing (raffinose + galactose) conditions

Strain	**Antibody (fragment)**	**Molecular weight**	**Ig folds**	**μ_{max}(h^{-1}) (raffinose)[(a)]**	**μ_{max}(h^{-1}) (raffinose + galactose)[(a)]**
WT	N/A: WT control	N/A	N/A	0.161±0.00143	0.095±0.0009.58
ScFv	ScFv fragment	26.1 kDa	2	0.151±0.000852	0.087±0.000715*
Fusion	ScFv-Fc fusion	106.0 kDa	8	0.152±0.000663	0.060±0.000409**
IgG	Full IgG	144.5 kDa	12	0.148±0.00066	0.048±0.000534**

(a) *P<0.01; **P<0.001 (versus WT in the same condition, by Student's two-tailed *t* test)

Supplemental table 1: A list of differentially regulated pathways as determined with ANOVA in the PAPi analysis. Means of the calculated activity scores are shown for each sample. P-values for the significance of the difference are as indicated, calculated by students *t*-test performed as two-sided, non-paired and with unequal variance between the experimental conditions.

Figure Legends

Figure 1 Growth of WT and recombinant strains in non-inducing (a) and protein expression inducing (b) minimal media. Exponentially growing cells were diluted to an OD_{600}=0.02 in fresh media and grown in microtiter plates for 48 hours at 30 °C with continuous shaking, while OD_{600} was measured every 15 min. For visualization, a logistic growth curve was fitted to the average OD_{600} of four replicate cultivations and represented in (a) and (b). For WT on galactose (in (b)) this model was not applicable.

Figure 2 Western blots of cell extracts from recombinant strains. Cell extracts were prepared after 18 hours of protein expression from the strains expressing the scFv fragment (a), the scFv-Fc fusion protein (b), and the full-length IgG (c). The scFv fragment was probed using a mouse anti-tetra his antibody together with a rabbit anti-mouse peroxidase labelled antibody. The Fc portion of the scFv-Fc fusion protein and the full-length IgG were probed with a goat anti-human (Fc-specific)-peroxidase labelled antibody.

Figure 3 Multivariate analysis of metabolomics profiles from recombinant strains and wild-type. (a)PCA scores plot showed clustering of replicate samples within each strain (n = 7, except for the Fusion strain n = 6), as well as separation between the strains. Two principal components (PC1 and PC2) were sufficient to separate all the four strains from each other. (b-d) PLSDA-VIP plot showing pairwise comparison between WT and antibody variant expressing strains (scFv (b), Fusion(c), and IgG (d)). The metabolites were ranked according to their increasing importance to group separation. Metabolites with VIP scores greater than one were considered as the most significant contributors. For each metabolite, the color code represents the relative amount of the measured compound (high to low) in the displayed pair.

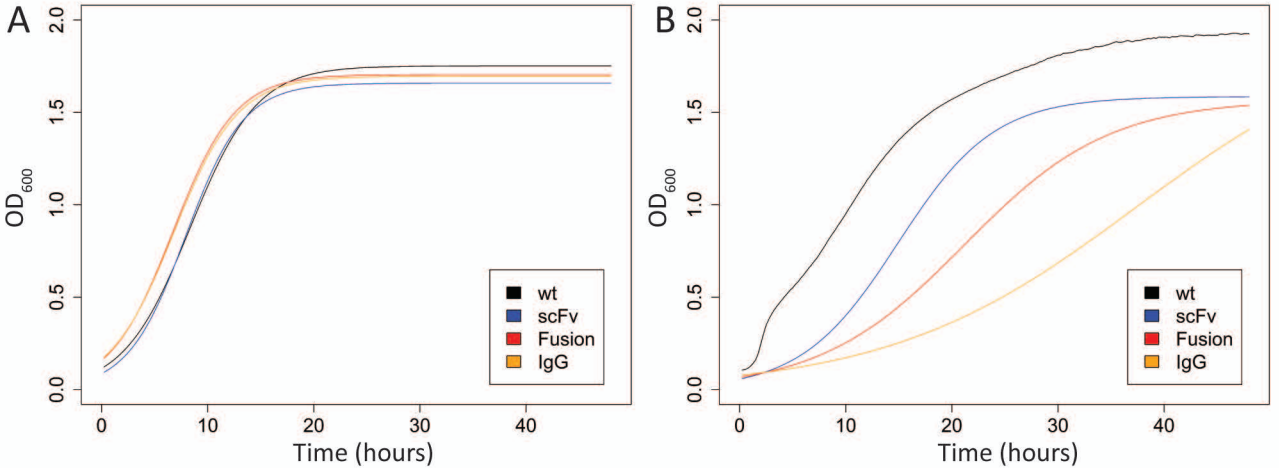
Figure 4 Metabolomics profiling of recombinant strains compared to wild-type. (a) Two-way hierarchical cluster analysis of metabolomics data. Dendrogram and heat-map overview of non-transformed and autoscaled metabolite levels. Columns represent individual samples and rows represent metabolites. Several blocks of co-regulated metabolites emerge that are shared to different extents among the different strains. Red and blue in cells reflect high and low metabolite levels, respectively, as indicated in the scale bar. The order of metabolites is as in (b). (b) Changes in measured metabolite concentrations expressed as log2-transformed fold changes in the recombinant strains compared to the WT. WT was set to zero. The color code for the strains is as in (a).

Figure 5 Pathway Activity Profiling (PAPi) analysis of the strains. Calculated Activity Scores (AS) are shown for each selected pathway. The AS is inversely proportional to the predicted metabolic flux. In general, the AS of the included metabolic pathways were more similar among the three antibody variants expressing strains, and more distant to the AS for the WT.

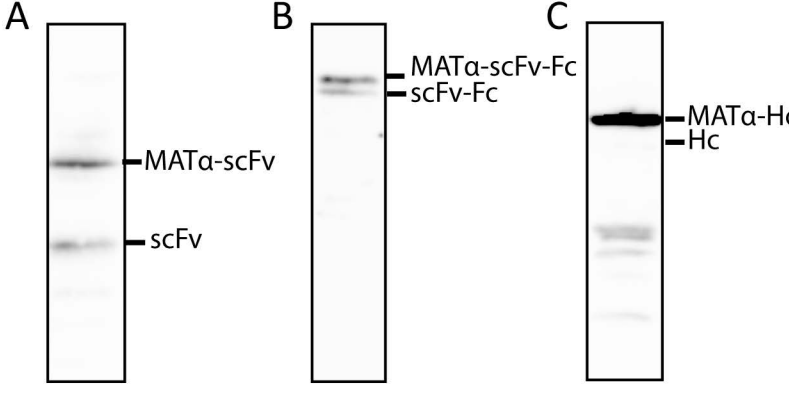
Figure 6 (a) The amino acid composition of the antibody variants was compared with the amino acid composition of the average yeast protein as reported by Karlin et al (Karlin et al., 2002). (b) The relative changes in intracellular amino acid content of the three antibody variants expressing strains compared to the WT.

Figure 7 Induction of unfolded protein response (UPR) contributes to metabolic burden. Fluorescence of GFP under an UPR-responsive promoter was recorded in the three antibody variants expressing strains and WT under non-inducing and inducing conditions. GFP signals were normalized to the respective signals of the non-induced culture and are expressed as relative fluorescence (arbitrary units). Induction of UPR peaks around 16 hours of cultivation.

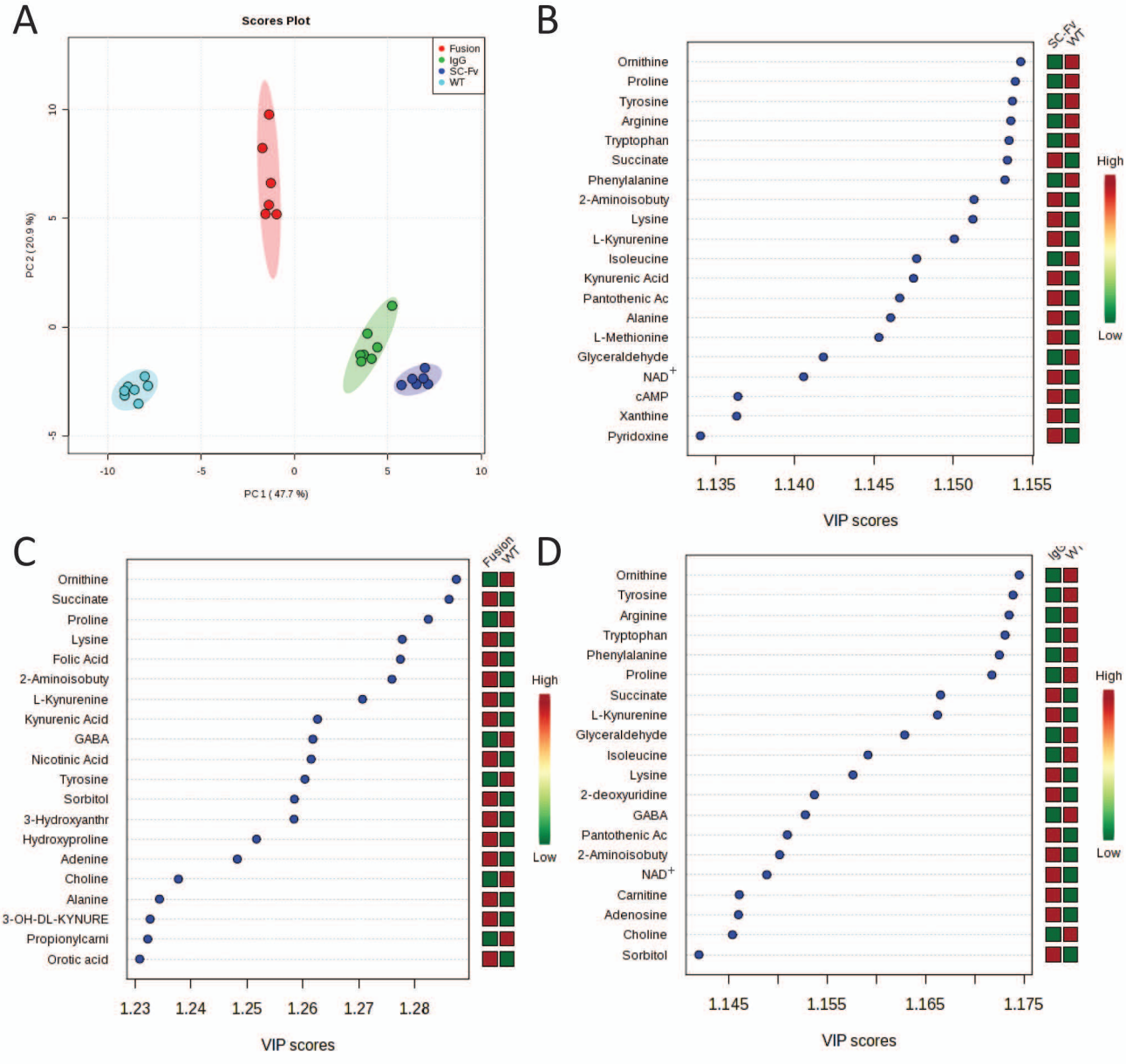
De Ruijter et al.,
Fig. 1



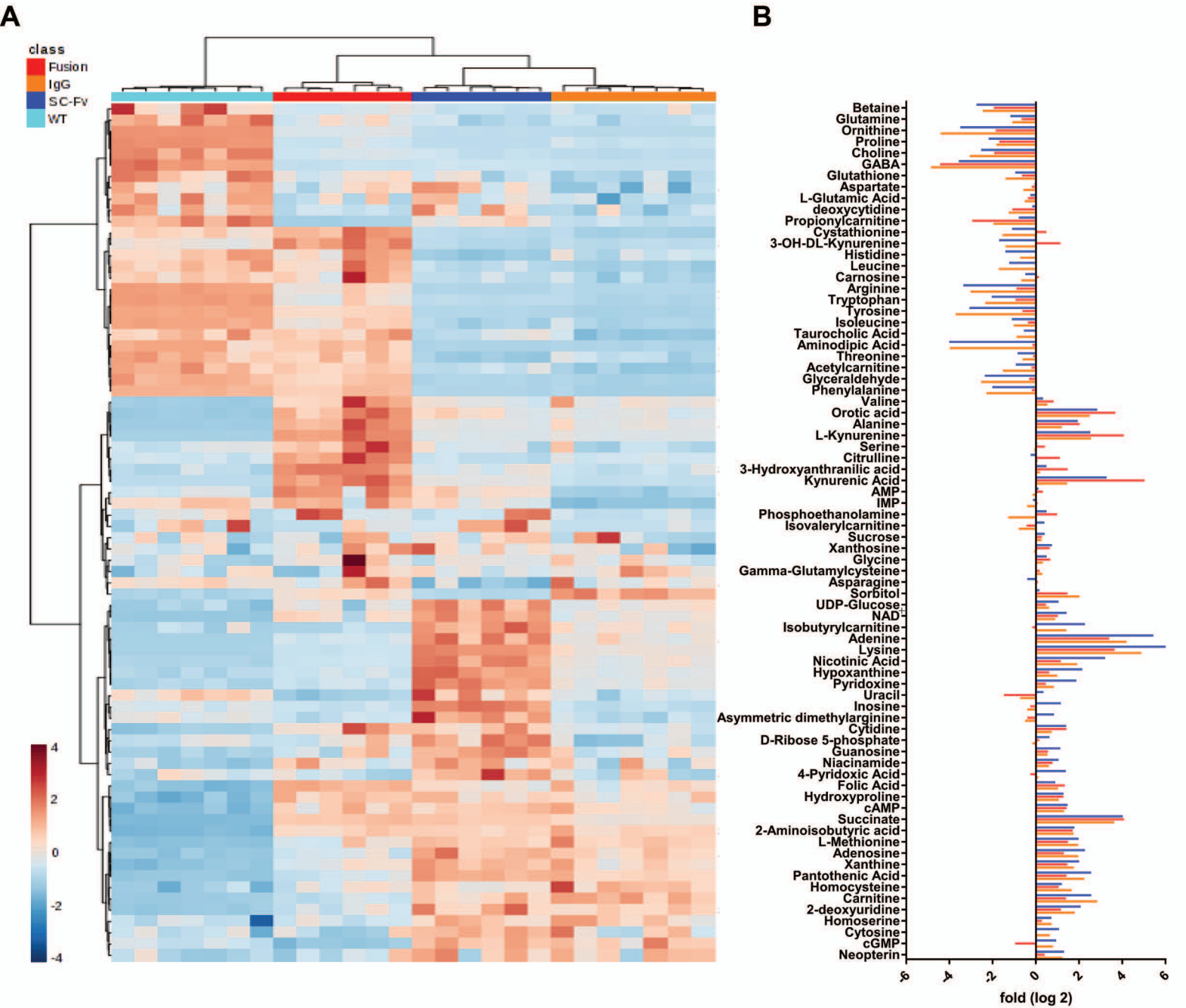
De Ruijter et al.,
Fig. 2



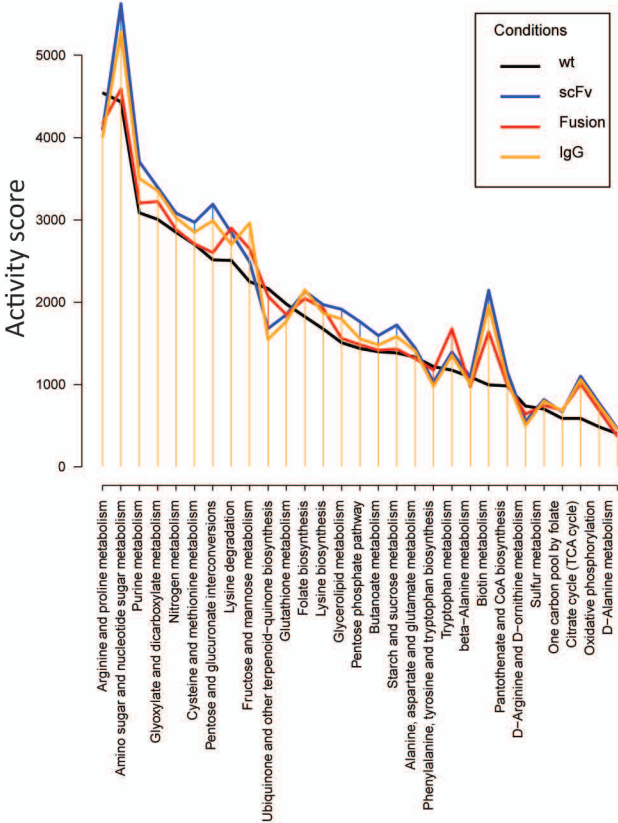
De Ruijter et al.,
Fig. 3



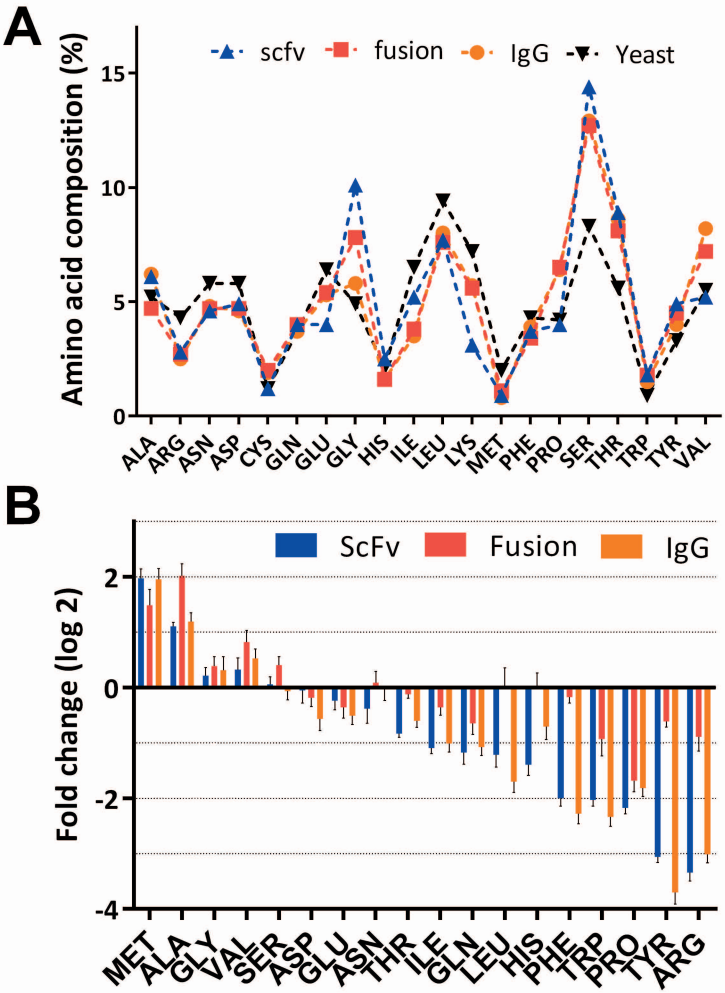
De Ruijter et al.,
Fig. 4



De Ruijter et al.,
Fig. 5



De Ruijter et al.,
Fig. 6



De Ruijter et al.,
Fig. 7