

Analysis of multiomics datasets to elucidate the role of Tps1 in *Saccharomyces cerevisiae*

Master Thesis: Matthias Mattanovich

External supervisors: Jean Marie François
Jean-Luc Parrou

Internal supervisor: Jürgen Zanghellini



Universität für Bodenkultur Wien
University of Natural Resources
and Applied Life Sciences, Vienna

INSA

INSTITUT NATIONAL
DES SCIENCES
APPLIQUÉES
TOULOUSE



Content

Abstract	5
Introduction.....	6
Trehalose-6P synthase (Tps1).....	6
Hexokinase analysis.....	6
Interactomics.....	7
Interaction Models	8
Gene Ontology (GO) Analysis	8
Objectives of the project.....	9
Materials & Methods	10
Hexokinase analysis.....	10
Strains and culture conditions.....	10
Molecular biology techniques	12
qPCR Preparation	13
NMR.....	13
Phenotypical test for cell wall integrity.....	14
Interactomics.....	15
TAP-tag data analysis.....	15
Interaction Models	15
GO neighbourhood analysis	16
mixOmics	17
PLS	17
PLS-DA	17
Block-PLS-DA	17
Sparse analysis.....	18
Data	18
Microarray data.....	19
Scripts.....	20
Interaction network construction	20
Integration of microarray data	22
Gene ontology analysis.....	23
Preparation for mixOmics	24
Results	25
Determination of hexokinase activity using ³¹ P-NMR.....	25
TAP-tag data analysis.....	25

Interaction network analyses	28
Alternative names in BioGRID	28
Integration of microarray data into interaction networks	29
Gene Ontology analysis	30
Cell Wall.....	35
mixOmics	37
sPLA	37
sPLA-DA	38
Block-sPLA-DA	39
Discussion	41
Gene Ontology analysis	41
Interaction Modelling.....	41
Interactomics data.....	41
Problems with the BioGRID database inbuilt functions	42
Limits of the presented modelling approach	44
mixOmics	45
Cell wall	46
Outlook.....	47
The connection between sugar phosphorylation and Tps1	47
Interactomics analysis	48
References.....	49

Abstract

The trehalose-6P synthase (Tps1) is the first step of the trehalose metabolism in yeast. Deletion of *TPS1* leads to several phenotypic features, e.g. they cannot grow on fermentable sugars, addition of those sugars causes rapid ATP depletion, or sensitivity to heat and osmotic stress. These apparently unrelated features suggest a more complex mechanism behind this protein. To gain further insight into the causes of these effects the strain has been studied on multiple biological levels.

The growth defect on fermentable sugars might be due to a linkage of sugar uptake and its phosphorylation. *Saccharomyces cerevisiae* expresses three hexokinases involved in hexokinase processing: glucokinase 1 (Glk1), hexokinase 1 and hexokinase 2 (Hxk1 and 2). It has been shown that *TPS1* mutants lacking Hxk2 regain the ability to grow on glucose but not on fructose, another fermentable carbon source. This could not be explained so far and is assumed to be due to the function of Glk1 in glucose-, but not in fructose processing.

Since modern biology is data driven, methods to integrate multiple layers of information rapidly gain more importance. An effective way to approach such problems in systems biology is to use multi-omics information. So called “omics” studies try to capture the entirety of a functional layer of an organism. By combining them, effects that might appear uncorrelated from one perspective can be correlated or common causes identified.

Analysing physical protein-protein interaction allows conclusions about proximity and possible influences proteins have over another. Using interaction data, these interconnections can be modelled into a network. They can further be investigated by using gained information for subsequent analysis, e.g. using Gene ontologies. Results obtained from this modelling approach can be used as a guideline for a general closer investigation of a certain protein, exemplified by the unassigned influence of *TPS1* deletion on cell wall formation.

Introduction

Trehalose-6P synthase (Tps1)

In the yeast *Saccharomyces cerevisiae* the trehalose-6-phosphate synthase protein Tps1 catalyses the first step of the trehalose pathway by producing trehalose-6-phosphate (T6P) in a two-step process from glucose-6-phosphate and UDP-alpha-D-glucose. It is part of the trehalose synthase complex, together with its other subunits, Tps2, Tps3 and Tsl1, but has been described to exhibit more than just a catalytic function since many effects previously ascribed to trehalose-6-phosphate, especially in stress response, seem to depend on the presence of the enzyme itself rather than its product (Gibney et al., 2015).

Yeast mutants lacking the *TPS1* gene are unable to metabolize fermentative sugars (van de Poll and Schamhart, 1977a). The growth defect of $\Delta tps1$ strains is characterized by a rapid depletion of ATP and free organic phosphates in the cytosol and a hyper-accumulation of sugar phosphates (Thevelein and Hohmann, 1995). This might be due to the missing phosphate recycling function of Tps1 (van Heerden et al., 2014) although it cannot just be explained by the amount of phosphate that could be recovered from the trehalose pathway (Walther et al., 2013). It seems that the enzyme itself rather than the product is the crucial element for stress response (Vicente et al., 2018). Without Tps1, ATP depletes rapidly upon addition of glucose or fructose, probably due to the phosphorylation of the sugars, pushing them into glycolysis (Thevelein and Hohmann, 1995).

Hexokinase analysis

S. cerevisiae cells phosphorylate fermentable carbon sources, like glucose and fructose, using hexokinases while carbon sources that take more steps to be utilized, like galactose, do not depend on those genes. Fructose and glucose both can be processed by hexokinase 1 (Hxk1) and hexokinase 2 (Hxk2) while Glk1 is acting only on glucose. At least one of those enzymes is necessary for the cell to grow on the respective substrate (Lobo and Maitra, 1977). On glucose, Hxk2 is the predominantly present one while the expression of Hxk1 is repressed by increasing glucose concentrations (Clifton et al., 1993) but they are capable of sustaining growth to the same extent at the same rate. Glucokinase 1 shows the lowest level of specific activity in crude extracts.

Growth rate and -extent are reduced if only glucokinase 1 is present. According to Walsh et al. (1991), extracts of wild type yeast phosphorylated glucose at a maximum reaction rate (V_{\max}) of about 6 $\mu\text{mol A580 unit}^{-1} \text{ h}^{-1}$ while $\Delta h x k 1 \Delta h x k 2$ strains, expressing only Glk1, were reduced to a rate of about 1 $\mu\text{mol A580 unit}^{-1} \text{ h}^{-1}$. Comparing this to *in vivo* rates, wild type cells could phosphorylate glucose at about 2 $\mu\text{mol A580 unit}^{-1} \text{ h}^{-1}$ while the rate stayed the same as in the *in vitro* rates of the extracts for the Glk1-only cells. This indicates a rate limiting control mechanism in wild type cells while Glk1 appears to be sufficient to keep the rate at a fixed level. This mechanism is reported to be strongly regulated since raising the enzyme level per cell by increasing the gene copy number did not bring the metabolism up to levels possible in the wild type expressing Hxk1 and Hxk2. The exact mechanism behind the reaction is not known but it is indicated that the phosphorylation is linked to uptake since external glucose is preferentially used and internal glucose might not be directly available, allowing for the possibility that

the limitation, even under increased expression of Glk1, is related to a transport problem (Clifton et al., 1993).

This effect can be scavenged on glucose by limiting sugar uptake by deleting the *HXK2* gene. This deletion does not restore growth on fructose; double knock outs can only grow by consuming components of the complete medium (Hohmann et al., 1993). It does restore all secondary glucose related deficiencies that the mutant displays. The conclusion that this might indicate a regulating function of T6P on Hxk2 has been shown to be unlikely (Bonini et al., 2003). In Iynedjian (1998) it has been described that *HXK2* overexpressing pancreatic β cells show a similar phenotype as $\Delta tps1$ yeast cells, further indicating that unregulated activity of this enzyme is probably mainly responsible for the phenotype. While the mutant does not express the hexokinases in a significantly different amount compared to the wild type (van de Poll and Schamhart, 1977b), the rate of sugar phosphorylation exceeds the rest of glycolysis and is similar to *HXK2* overexpressing strains. This behaviour does not occur in strains that only express *GLK1* (Bonini et al., 2003).

Fructose can only be utilized by the hexokinases 1 and 2. The affinity to fructose compared to glucose is about 10 fold lower but comparing activity of the enzymes when comparing the different substrates (F/G), values of 4.2 for Hxk1 and 2.3 for Hxk2 have been published (Lobo and Maitra, 1977). Assuming, as postulated, that Hxk2 is dominant it can be assumed that the V_{max} for fructose will be 13.8 times higher in cell extracts and 4.6 times higher in vivo than the V_{max} for cells grown on glucose using only the glucokinase 1, suggesting a rate difference for these two substrates in *HXK2* deletion strains. This assumption is only numerical not considering arising biological responses.

Interactomics

In order to gain a closer insight into the influence of Tps1 onto the organism, laboratory experiments were combined with bioinformatics methods. Biologic activity is complex and most often it cannot be explained in simple, linear terms. Often the combined action of multiple partners causes an effect. Interactomics is a field of study focused on recording and displaying all, or a collective of, the interactions, e.g. genetic or protein-protein, that are going on in an organism (Gavin et al., 2002).

Protein-protein interactions can be studied by using one protein as a bait and analysing all its hits, i.e. the proteins that physically interact with it. One of the available methods is Tandem Affinity Purification (Rigaut et al., 1999). A tag, genetically linked to the target protein that is used as a bait, is used for two subsequent purification steps. After the first step it is enzymatically cleaved, enabling the second round. Two steps are performed to be as selective as possible to remove a background of proteins that would be purified as well in one of the selection steps.

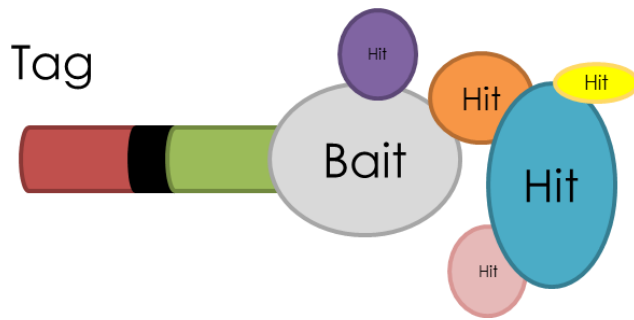


Figure 1: Schematic representation of a TAP-tagged bait protein and its hits. Also physical interaction partners of direct hits are recovered

The proteins are identified by mass spectrometry. Recovered proteins can be seen as hits, as physical interaction partners, of the bait, the protein under investigation. For published studies, this information is available in databases, e.g. the Saccharomyces Genome Database (SGD; www.yeastgenome.org), and summarized in a community effort in the Biological General Repository for Interaction Datasets (BioGRID; www.thebiogrid.org) (Oughtred et al., 2016a).

In the BioGRID, the genetic and physical interactions and chemical modifications can be checked specifically for an organism. For each annotated element, the respective interactions or modifications can be listed together with a variety of information, e.g. corresponding publications or type of experiment (Oughtred et al., 2016b).

Interaction Models

A network presentation function is available in the BioGRID (Chatr-aryamontri et al., 2017). The entire database can also be downloaded in a text file and processed and used using e.g. R (R Core Team, 2017). With this file, the information about the interactions can be presented numerically and visually, i.e. noted in adjacency matrices that can be translated into node/edge models allowing for automated analysis.

Gene Ontology (GO) Analysis

The Gene Ontology project (www.geneontology.org) facilitates gene analysis by putting them into context using ontologies, i.e. hierarchical categories describing properties. They are separated into three categories; biological processes, molecular functions and cellular compartments (Gene Ontology Consortium, 2001).

Providing a list of genes, e.g. interaction partners of proteins or subsets thereof, it can be tested if certain properties are overrepresented in this subset compared to the general distribution of properties in the entirety of annotated genes in a count based method (Draghici et al., 2006) where overrepresented terms are identified using Fisher's exact test (Lehmann and Romano, 2005).

Objectives of the project

This project combines micro- and molecular biological methods with bioinformatics data analysis and modelling to get a better understanding of the functions of Tps1, apart from initiating the trehalose metabolism, in the yeast *Saccharomyces cerevisiae*.

After extensive literature research, a new theory about the varying recovery of growth on different fermentable carbon sources and the growth defect in $\Delta tps1$ strains in general is set up. The investigation of this new possible explanation is started. A series of knock outs of the hexokinases are to be created and to be tested on their growth behaviour and reaction kinetics.

Apart from that, a multitude of existing data is analysed and combined with database entries to obtain undiscovered information by approaching the objective in a systemic way. A method able to combine 'omics data is applied and new methods are created to get more out of the data then by just using conventional analysis. This resulted in the creation of multiple scripts allowing the user to apply these new ways of treating data.

Following predictions that arise with these approaches, the effect of the deletion of *TPS1* on the synthesis of the cell wall is studied.

Literature research is presented along with a description of the developed and used methods. In the discussion, an overview of the limitations of the methods is given and follow up work is described in the outlook.

Unless explicitly stated otherwise, all work was performed by the author. All the R scripts were newly developed and only use existing packages for indicated purposes.

Materials & Methods

Hexokinase analysis

Strains and culture conditions

After deletion of the *HXK2* gene, *TPS1* mutant strains can grow on rich media with glucose and fructose, for the latter likely by consuming non-sugar components of the medium (Hohmann et al., 1993) but on minimal media only for glucose (unpublished data).

For the attempts to delete the hexokinases (*HXK1*, *HXK2* and *GLK1*), the vectors described in Laughery et al. (2015), carrying the *URA3* gene as a selection marker, and dedicated guide RNAs were used. The guide sequences and repair fragments were calculated using the online tool described in Mans et al. (2015) which also rendered diagnostic primers used to check the results. The *S. cerevisiae* strains CEN.PK 113-7D (van Dijken JP et al., 2000) and its derivative $\Delta tps1$ strain (Guillou et al., 2004) were used as backgrounds for the strain constructions.

All yeasts were either grown on YP-Galactose rich media or YNB-Galactose minimal media without selection pressure. If a selection pressure was necessary to maintain vectors for gene deletions, YNB-Galactose URA⁻, missing supplemented uracil, was used. The sterile filtered carbon source solutions were added after autoclaving. Liquid media corresponding to the ones indicated in the tables are missing the added agar.

Table 1: Composition and amounts of YP media, recipe for 100 mL.

YP Agar 100 mL	
Component	Amount
Yeast extract	2 g
Bacteropeptone	1 g
Agar-agar	2 g
dH ₂ O	95 mL
50 % carbon source solution	5 mL

Table 2: Composition and amounts of YNB media, recipe for 100 mL.

YNB Agar 100 mL	
Component	Amount
Yeast nitrogen base	0.171 g
Complete supplement mixture (CSM)	0.078 g
NH ₄ SO ₄	0.5 g
Agar-agar	2 g
dH ₂ O	95 mL
50 % carbon source solution	5 mL

Table 3: Composition and amounts of YNB URA- media, recipe for 100 mL.

YNB URA- Agar 100 mL	
Component	Amount
Yeast nitrogen base	0.171 g
CSM URA-	0.077 g
NH ₄ SO ₄	0.5 g
Agar-agar	2 g
dH ₂ O	95 mL
50 % carbon source solution	5 mL

Bacteria for plasmid multiplication were propagated Luria-Bertani (LB) broth.

Table 4: Composition and amounts of LB broth, recipe for 100 mL.

LB broth 100 mL	
Component	Amount
Peptone	1 g
Yeast extract	0.5 g
NaCl	0.5 g
dH ₂ O	Fill to 100 mL

Here, the modified version of the transformation protocol is provided. For the last trials, the amounts were varied (e.g. providing 10 times the amount of hybridized repair fragment described in literature). It was composed as a combination of the protocols described in the aforementioned publications outlining the experiment and the commonly used in house protocol, a modified version of (Gietz and Woods, 2002). The major change in regard to the original protocol is a reduced incubation time at 42 °C since the *TPS1* mutant strain that is used as one of the background strains is heat sensitive.

Molecular biology techniques

Transformation protocol

For **10** reactions

Day 1 :

1. Inoculate 5 mL of liquid media. 30 °C, 200 rpm, o/n

Day 2:

1. Measure OD₆₀₀
2. Inoculate 50 mL of liquid media. The OD₆₀₀ depends on your growth rate and your time. 2 or 3 doublings are preferred, correspondingly: inoculate to an OD₆₀₀ of 0.1 or 0.05. 30 °C, 200 rpm
3. The culture should reach an OD₆₀₀ in the range of 0.3/0.6 in about 4 h. Once there, harvest by centrifugation at 3000 g for 5 min and wash with 25 mL water
4. Resuspend in 1 mL water and transfer into a 1.5 mL tube
5. Centrifuge for 30 sec and discard the supernatant
6. Resuspend in 1mL water
7. Split into 10 x 100 µL, centrifuge for 30 sec and discard the supernatant

	1 reaction	
PEG 3350 (50 %, w/v)	240 µL	
LiAc (1.0 M)	36 µL	
Carrier DNA [2 mg/mL]	50 µL	
Plasmid [100 ng/µL]	10 µL	
hyb. Rep. Frag. [100 pmol/µL]	1 µL	
dH2O	14 µL	
8. total	360 µL	

9. Add mix to cells and vortex gently
10. Incubate at 30 °C, 200 rpm for 15 min
11. Incubate at 42 °C for 15 min
12. Centrifuge for 30 sec and remove the transformation mix
13. Add 1 mL of liquid media and incubate at 30 °C, 200 rpm for 4 h and
 - a. If it works badly: Centrifuge for 30 sec remove most of the supernatant, resuspend all of the cells and plate them directly on appropriate selective media
 - b. If it works well: plate 100 µL, centrifuge for 30 sec and plate the rest as a back up
14. Incubate at 30 °C for 3 to 6 days

qPCR Preparation

Differing expression levels of the hexokinases have been postulated with Hxk2 being dominant, Hxk1 glucose repressible and Glk1 having a lower V_{\max} but being responsible for the main part of glucose influx into glycolysis. To measure that, everything was prepared for a qPCR analysis. Comparing the relative amounts of transcripts for the hexokinases to those of reference genes allows also for comparison between wild type and mutant (Teste et al., 2009).

Samples for a preliminary qPCR analysis have been taken. The first one after growth over night, dilution to OD₆₀₀ of 1 and subsequent growth in exponential phase, if possible until OD₆₀₀ of 2. A part of the cultures have been left to grow overnight and harvested in the morning to gather the cells in a carbon source depleted state.

Table 5: Optical density at a wavelength of 600 nm at the time of sampling.

	Fresh [OD ₆₀₀]	o/n [OD ₆₀₀]
YNBGlu	2	10
YNBTre	1.94	10
YNBFru	1.55	6
YNBGal	0.97	12

Primers for *GLK1*, *HXK1* and *HXK2* (two sets of primers each) have been ordered together with primers for reference genes (one set each), *ACT1*, *ALG9*, *KRE11*.

This analysis could show differential regulation depending on carbon source availability. It would also display the described repression of *HXK1* on glucose (Clifton et al., 1993) that would ascribe higher importance to Glk1.

NMR

Nuclear Magnetic Resonance Spectroscopy (NMR) can be used to identify molecules in solution. For phosphorylation reactions, ³¹P-NMR proved to be convenient due to it being naturally dominantly abundant and applicable (Eicher et al., 2012).

NMR experiments were performed on a Bruker Avance III HD 800 MHz spectrometer equipped with a 5mm QCP cryogenic probe head. Analyses were performed at 293 K. Spectra were acquired and processed using the Bruker Topspin 3.5 software. All enzymatic reactions analyses were performed in 5 mm NMR tubes with 0.4 mM TSPD4 and 4 % D2O.

The device can also be used to acquire time series data by measuring entire spectra in fixed intervals. Using these spectra, the de- or increase of certain molecular species can be followed.

The enzyme is added after an initial measurement of the rest of the reaction mix. For technical reasons, the first measurement occurs after approximately three minutes. Subsequently every 30 seconds one measurement is acquired.

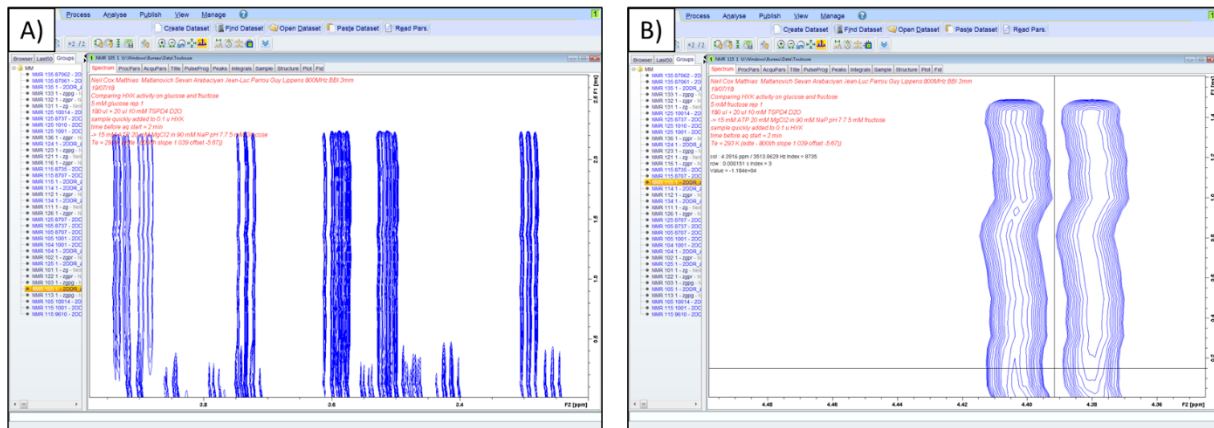


Figure 2: A) exemplary 2D time series spectrum showing the use and production of certain metabolites. X-Axis: chemical shift; Y-Axis: time (not in absolute values). B) 2D spectrum for ATP (left) and ADP (right).

Defining a range of chemical shift, the quantitative data obtained from all the spectra can be projected onto the time to follow the change in concentration of the respective metabolites.

The glucose utilisation was observed by selecting the common peaks described in the Human Metabolome database (<http://www.hmdb.ca/>) while for fructose samples the peaks changed after prolonged storage in a non-frozen state which can be explained by an equilibrium of tautomers (Barclay et al., 2012). These results can be compared with the observed ATP decrease and ADP increase. Since controlled conditions with a commercial enzyme were used in this assay, the ATP and ADP progressions were sufficient. The amount of substrate was compared between the sugar- and nucleotide peaks at the beginning and the end of the reaction.

Phenotypical test for cell wall integrity

The cell wall tests were performed by adding agents in defined concentrations, according to an in house protocol, to the media plates (see table 6). The strains to be compared, the wild type and the mutant, were grown over night at 30 °C and 200 rpm in 10 mL of YP-Galactose medium. The optical density of the cultures was measured at 600 nm. The cultures were centrifuged and pelleted at 4000 rpm (5 min, 4 °C) and re-suspended in sterile water to an optical density of 8 according to calculations from the previously mentioned measurements.

On the plates, four drops of each of the two cultures were deposited: undiluted, 1/10 diluted in sterile water, 1/100 diluted, 1/1000 diluted. The plates were incubated at 30 °C and pictures taken after 72 h and after 98 h to account for possible differences arising from a slower growth of the $\Delta tps1$ strain.

Table 6: List of the concentrations added to the plates for the substance sensitivity tests

Substance	Concentration 1	Concentration 2	Concentration 3
Calcofluor White	0.1 mg/mL	0.05 mg/mL	0.01 mg/mL
Congo Red	0.1 mg/mL	0.05 mg/mL	0.01 mg/mL
Caffeine	5 mM	1 mM	
SDS	0.01 %		

Interactomics

TAP-tag data analysis

The measurements were performed with three technical replicates each but no biological replicates, thus functional conclusions cannot be drawn confidently, but scripts and procedures can later on be applied to further experiments. The data obtained in a preliminary analysis, performed by Sevan Arabaciyian (unpublished data), was analysed to set up an analytical pipeline.

The TAP-tag technique was used to purify proteins that interact with the “bait” protein, in this case Tps1. Five samples representing different conditions were compared: the laboratory strain without a tag on the protein of interest was used to record the background noise and proteins that are artefacts of the technique. For two “acid” samples, HCl was added to one of them, bringing the pH down to 3, while a mix of HCl and acetic acid was added to a second sample to investigate the apoptosis inducing effect of acetic acid. Two “temperature” samples consisted of one completely untreated sample grown at 30 °C and one that was exposed to a heat-shock at 42 °C for one hour.

To allow for numerical comparison and data treatment, the median values of the mascot scores of the technical replicates were used for analysis since it is less prone to deviation by outliers. The negative control was subtracted from all results and negative values set to “NA” (and not used for further analysis).

Interaction Models

BioGRID processing

From the BioGRID database, all the organism specific interactions for *S. cerevisiae* were downloaded as a tab delimited text file. After loading it into R, the file can be reduced by separating the types of interaction, i.e. physical or genetic, and keeping only the information about the “Bait” and the “Hit” elements. The standard names of the genes/proteins are in the column “Official Symbol Interactor A” (Bait) and the names of the interactors are in the column “Official Symbol Interactor B” (Hit). Interactions described multiple times can be counted and filtered for, if higher confidence in the data through reproducibility is desired.

These columns were put into a large list where all interactors were assigned to the list element of the corresponding bait proteins. This list (or subsets of it) can be used for the interaction network creation. Since they are constructed out of the “Bait” Data, all they show is which element interacts with other elements of the subset when they are used as a bait. The same was done to create a list of the “Hit” data.

These lists can be used to select the elements for modelling. Later on they are used to construct the adjacency matrices as described in the general modelling part. Finally, by taking one adjacency matrix and comparing/correcting it with the other one we can obtain a combined model that contains the entire interaction information, hit and bait. These matrices can directly be used for the neighbourhood GO term enrichment script. In R, the resulting networks can be displayed using the *igraph* package (Csardi and Nepusz, 2006).

Construction

The construction of an interaction network using database information for a generic element (protein or gene) follows the described scheme:

A start protein (or gene) is picked and all its physical (or genetic) interaction partners are retrieved from a database (e.g. SGD or BioGRID). All partners listed here form the elements of the model. All the interaction information of all the elements is taken as well and reduced to the interactions regarding other interaction partners of the core protein (or gene).

Here, “A” is the core element interacting with “B”, “C”, “D” and “E”. The interactions are described by “0”s and “1”s in a square matrix. The columns of this matrix can be read as a description of the “Bait” nature of the respective partner, i.e. going down a column one can see which other elements of the model interact with them. On the other hand the “Hit” information is embedded into the rows; similarly one can see all the partners for which the respective element is a “Hit” by following a row.

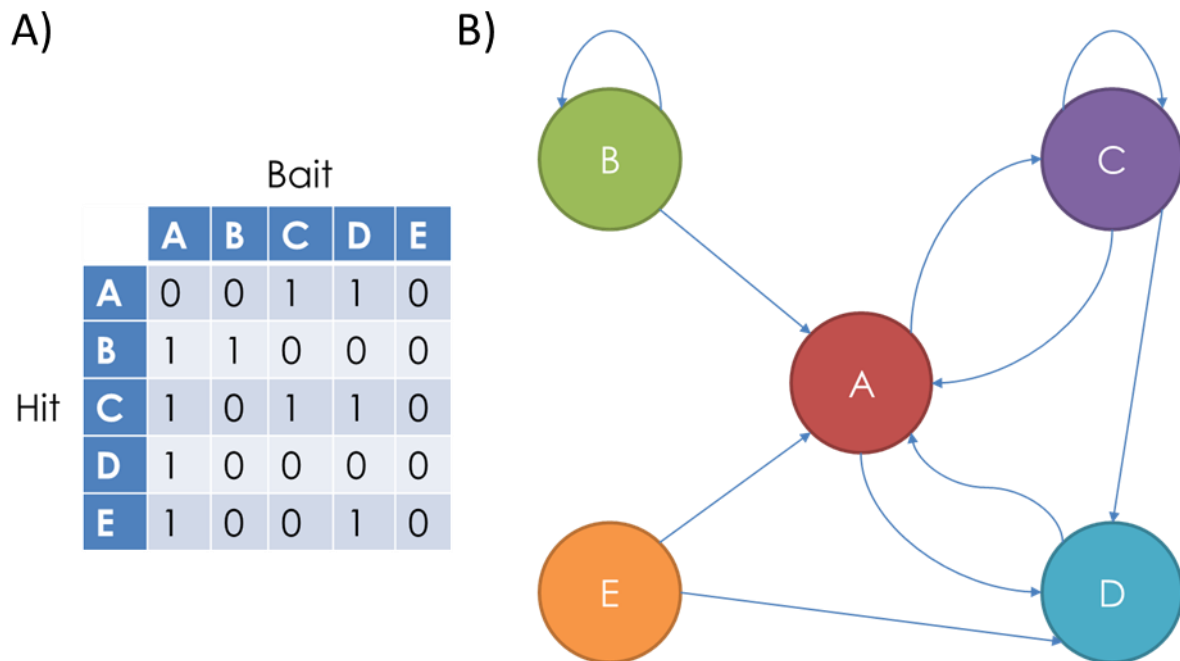


Figure 3: Exemplary adjacency matrix and corresponding interaction model. A) Following a column, a 1 indicates an interaction, i.e. a bait - hit relationship, between the element in the column- and the row-element. E.g. “B” is a bait just for itself but a hit for “A” and itself. B) Visualization of the described concept.

Differently, if an interaction model of a defined group is to be constructed, the model is limited to the members of that group while the interaction information is obtained from the database.

GO neighbourhood analysis

An automatized gene ontology term enrichment neighbourhood analysis was performed starting from the constructed interaction models. A neighbourhood “distance 1” consists of an element of the network and its direct interaction partners, forming a so-called clique. The interaction model in the

described case was constructed using the experimentally obtained list of physical interactors of Tps1. By iterating through a network, each clique around every element of the model was used for a GO enrichment analysis. The R package `topGO` was used (Adrian Alexa and Jörg Rahnenführer, 2018). Unless noted differently, a p-value of < 0.01 was used as a cut-off to identify overrepresented terms. The results are checked for frequency of occurrence, and separately, an html file with the cliques associated to the overrepresented terms is produced. This interactive table is linked to the SGD and allows for an in depth manual investigation (see Figure 15 in the Results section).

mixOmics

The R package `mixOmics` (Rohart et al., 2017) can be used to gain a deeper insight into omics data by combining multiple types of information. For example, gene expression data can be combined with protein abundance levels to find differences between treated- and control samples. Different types of data (e.g. factors) can be used as anchor points for the separation.

PLS

In a PLS (“partial least squares” or “projection to latent structures”) regression, multiple data matrices can be related. Similar to a principal component analysis, it will try to find a direction that explains the most variance in the distribution but due to the nature of this analysis this is done in a higher dimensional scale. By combining two data matrices, the multidimensional direction in one matrix is sought for that fits into the next matrix, explaining the maximum of the multidimensional variance in it (Wold et al., 2001).

PLS-DA

If the second information given is factorial, this method can be used to find the multidimensional directions that explain the biggest differences between the factors, giving a way to differentiate between them. This is then referred to as a PLS-discriminant analysis (Barker and Rayens, 2003).

Block-PLS-DA

As a combination of the previous methods, matrices carrying information about the same set of samples can be combined, effectively performing a PLS regression on two data matrices and subsequently discriminating between factors, searching for the biggest differences between the set types in the combined model.

Sparse analysis

In order to allow for a facilitated interpretation of the results, the amount of variables taken into account can be limited (Lê Cao et al., 2008), selecting and using a reduced amount of biologically meaningful variables.

Data

Gene expression levels obtained from microarrays were used for the preliminary analysis, as well as factorial values indicating the type of strain used and rates taken from the exponential part of the fermentations describing the use/production of sugars, cell mass or metabolites.

The results of a gene ontology term enrichment neighbourhood analysis can be included as additional information for the genes tested on the microarray. For each gene and its neighbourhood, a gene ontology enrichment analysis is run and the over-represented terms with a certain cut-off (here $p < 0.05$ using Fisher's exact test) are assigned to the gene. These categorical values can be made one dimensional in order to use them by applying the kmodes algorithm (Huang, Z, 1998) (similar to kmeans) with the `klAR` R package (Weihs et al., 2005).

	Gene expression								Type	OD600	Metabolite 1	Metabolite 2
TPS1_1	$\Delta tps1$
TPS1_2	$\Delta tps1$
TPS1_3	$\Delta tps1$
TPS1_4	$\Delta tps1$
WT_1	wt
WT_2	wt
WT_3	wt
WT_4	wt _{μ}
	GO terms	GO terms	GO terms	GO terms	GO terms	GO terms	GO terms	GO terms				
	GO terms (d = 1)											

Figure 4: Schematic representation of the data to be combined. Gene expression levels, factors describing the type and metabolic- or growth rates will be compared depending on the sample for which they were measured. GO terms should be included as additional information for the genes under investigation.

The data obtained from fermentations by Laurent Benbadis and Jean-Luc Parrou (unpublished) was screened thoroughly to find information that can be used for further analysis and gain of knowledge. Since most of it was just acquired for control purposes (e.g. the fermenter settings) only the OD600 and the metabolite levels at certain time points were considered. Through the measured values, a combination of models were fitted to approximate the course of these natural distributions.

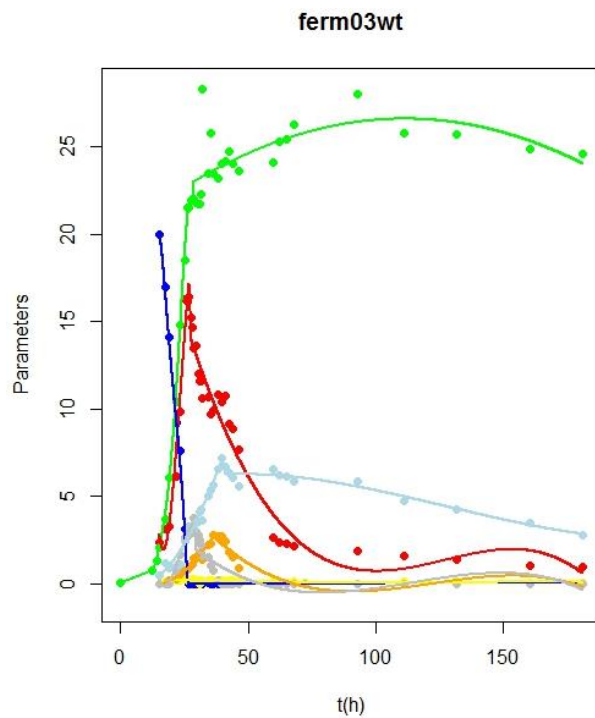


Figure 5: Exemplary graph for the curves fitted into the fermentation data

Here the result for one of the fermentations are shown as an example. At the latter development of the red curve one can see the problem of using high polynomials for that purpose as it starts to deviate from the estimated course of the data. These models can be used to calculate time series points for the experiments (not considered here).

The rates, e.g. substrate use or growth in the exponential phase can be used for `mixOmics`. At a later stage it might be possible to use time series information directly, i.e. the entire dataset instead of “just” the rates. The rates used were: the optical density (OD, as a representative of cell growth), galactose (as the substrate) and the produced metabolites trehalose, ethanol, acetate, glycogen and glycerol. Since trehalose was not measured in the $\Delta tps1$ samples it has been assumed to be absent in the mutants in order to use it for analysis.

Microarray data

Transcriptome analysis performed for samples taken in the exponential phase during the Fermentations mentioned before.

Table 7: General information about the samples.

Fermentation number	Date	Sample	
02	22. 01. 2008		TPS1_1
03	05.02. 2008	WT_3	
04	18. 03. 2008	WT_1	TPS1_2
05	31. 03. 2008	WT_2 WT μ -aéré_4	
07	?? 12. 2008		TPS1_3
08	12. 01. 2009		TPS1_4

The data has been background corrected by subtraction and normalized using the “loess” method, subsequently normalized between the lanes using the “Aquantile” method and then sorted to be used for `mixOmics` analysis. An interaction model has been calculated as well and was used as a base for finding overrepresented GO terms which possibly can be used as additional information for the gene names.

Scripts

The scripts that were developed for future users are described and their application is explained. All of them are written in R.

Interaction network construction

These functions make use of the downloaded BioGRID database, in this case the *Saccharomyces cerevisiae* specific information. Models can be constructed for:

1. a defined group of genes/proteins that the user enters themselves as a character vector
2. Distance = 1, i.e. for all interaction partners of a selected centre element
3. Distance = 2, i.e. for all interaction partners of a selected centre element and their respective interaction partners

Should the need arise, all functions could be quickly adjusted for any other organism.

“CROISSANT”

Database file preparation, **“CROISSANT”** (Concentration and Reduction Of Interaction Sample Size for Association Network Tables)

The file size of the downloaded BioGRID database file is reduced by saving limited versions of the database, i.e. only the names of the interaction partners (“Official Symbol Interactor A” and “Official Symbol Interactor B”) and only genetic or physical interactions respectively are being kept.

As input for the processing function, the name of the downloaded database file and the interaction type of interest are requested. The reduced version of the annotated physical and genetic data are saved separately. The output data frame has three columns, “Bait” and “Hit” where the protein names of the interactors are listed and “Count” where it is indicated how often the respective interaction occurs.

The second function takes the reduced database as an input and writes lists that associate each protein with its respective baits and hits. It saves those in two separate files which are the basis for the interaction modelling.

“PAIN”

Physical protein-protein interaction modelling, **“PAIN”** (Physical Association Interaction Networks)

Protein-protein interactions denoted as “physical” in the “Experimental System Type” in the downloaded file from the database are extracted. An interaction network is created based on these elements.

First the lists containing the associated baits and hits to every element have to be loaded. They are saved in a variable each which are subsequently used as input for the following function.

To generate the network, the selection has to be specified. The default value is set to “d1”, indicating that the programme will build an interaction network of distance one. A centre element has to be specified and all its interaction partners and their respective interactions with each other will be modelled. Other possible selections are “d2”, distance two, where all the interaction partners of a specified centre element and all the interaction partners of those interaction partners are modelled together, or “group” where the user specifies the element to be modelled themselves. The first two options are interesting for database exploration while the last option allows to investigate e.g. experimental results.

“FOIEGRAS”

Genetic interaction modelling, **“FOIEGRAS”** (Functional Organisation of Interaction Exhibiting Genes Representing Associated Segments)

Genetic interaction (can be positive, negative, ...) denoted as “genetic” in the “Experimental System Type” in the downloaded file from the database are extracted. An interaction network is created based on these elements.

This script works the same way as the previously described one, only specified for genetic interaction instead of physical protein-protein interaction. They were separated since they include a long loading step to make it easier for the user to keep a distinction.

“CREPE”

Database file preparation including an evidence limit, **“CREPE”** (Concentration and Reduction Enabled by Publication Enumeration)

The number of annotations are counted and files with an evidence limit are created if the user wants to have more confidence into the described interaction.

This filtering step is done to exclude interactions that are described by less than a certain number of publications to have more confidence through reproducibility. This can be applied before or after model construction. Doing it before limits the model to frequently described elements that are directly connected with the centre element even in the reduced list. Filtering the data after model construction keeps genes that are not directly connected but show frequently described interactions with other

elements of the unfiltered model. To construct models with filtering beforehand, the previously described construction scripts ("PAIN" and "FOIEGRAS") can be used by loading the limited interaction lists instead of the full ones.

The same approach for creating files containing the distinguished information of all respective hit- or bait elements associated to each element as in the first described script (CROISSANT) was used with the distinction that only interactions that were published more often than a specified number of times are taken into account. The reduced database file, the type of interaction and the numeric limit for the number of required publications are taken as inputs.

"QUICHE"

Interaction modelling using an evidence limit, "**QUICHE**" (**QU**antitative **I**ncident **C**urb for **H**igher **E**vidence)

This script constructs network models using a certain number of publications as a cut-off. It uses the output of CREPE as an input.

With this script, the adjacency matrix is constructed with all the partners, for which the respective type of interaction were described for the selected element, but only the interactions, that were published more often than the specified limit indicates, are included in the model. The same approach in a slightly varied version is used by the BioGRID database. For this function, the unfiltered interaction lists, the filtered lists and also the neighbourhood type ("group", "d1", "d2) and the respective selected element (centre element or entire group) are required input.

Integration of microarray data

"CANARD"

Integration of numerical data, in this case expression levels measured by microarrays, into interaction networks, "**CANARD**" (**C**ombined **A**nalysis of **N**umerical **A**nd **R**elationship **D**ata)

To have one dimensional expression values to work with, the fold changes between the average levels of the previously described *Δtps1* strains and the average of the wild type strains (excluding the highly deviating micro-aeration strain) were used to set up this analysis tool.

The combination can be done visually (see Figures 11 and 12 in the Results section) and numerically, for representation or further evaluation. Visually, a (multiple of the) absolute value of the respective differential expression of each gene in the network was assigned to the size of the nodes. Nodes with positive fold change values when taking the decimal logarithm (overexpression) are blue while negative ones (repression) are held in red. Numerically, the expression values are held separate from the network but can be referenced. For example, if one wants to check for possible network effects of the expression value it is possible to iterate through the network and check all the expression values of the cliques around each element. This can be combined by filtering steps, e.g. setting a limit for the standard deviation of the expression values of the genes in the respective cluster to reduce

heterogeneity to find clusters that behave in a uniform fashion or a limit for the mean to filter for strongly impacted clusters.

The first provided function helps to check to which the elements in the numerical experiment, in this case genes in the microarray experiments, overlap with the genes in the previously constructed interaction matrix using these genes as an input. Deviations might be caused by diverging annotation conventions (e.g. using minuses “-” instead of dots “.”) using them as inputs. The next one reduces the interaction matrix to its elements that are intersecting with the gene list. Further on the differential values for the groups in the array can be calculated and saved in a new variable. The loaded information and the range of the columns corresponding to the samples have to be specified. Everything is set now for a numerical data exploration. For the visual analysis or presentation, a function is given that takes the reduced interaction matrix and the differential values as input. Optionally the size of the nodes can be adjusted by a size factor.

Gene ontology analysis

Checks with the gene ontology information downloaded from the SGD database and the `topGO` R package if in the provided group of genes are any gene ontology terms overrepresented in comparison to the distribution in the entirety of genes. This can be done for:

- a defined group of genes/proteins that the user enters themselves
- Distance = 1, i.e. for all interaction partners of a selected centre element
- Distance = 2, i.e. for all interaction partners of a selected centre element and their respective interaction partners

“PATE”

Gene ontology file preparation “PATE” (Processing of the Associated Terms Enumeration)

For a gene ontology (GO) analysis, we need a file where all the ontologies are collected. To obtain that, the annotation file from the SGD can be downloaded and processed. For this purpose, a data frame consisting of two columns, one containing the elements name, one the references to the gene ontologies separated by commas if there are multiple, is constructed. The downloaded gene association file from the database, as a .csv file, is read, processed and the output file saved. This file is the base reference for the subsequent analysis.

“GOURMET”

Gene Ontology term overrepresentation analysis “GOURMET” (Gene Ontology - Unequal Representation of Multiple Element assigned Terms)

A different approach to get more information out of constructed interaction would be a gene ontology analysis. To make sense out of the structure given in the model, it is separated into groups. One way is to select a subset and analyse it using the online portal of the database. A more elaborate method

is to screen the model for any peculiar abnormalities. Here, a neighbourhood analysis is performed by selecting all the elements of the model one at a time and running a gene ontology enrichment analysis for it and all its interaction partners, also referred to as a clique. A p-value of 0.01 was set as a threshold. Functions that are overrepresented multiple times could give an indication of what would be impacted in a *TPS1* mutant strain. This is especially useful if there is not a lot known about the protein or gene and its importance.

The results are automatically saved in a file that can be presented using the subsequent function. The neighbourhoods for all the centre elements and the respective overexpressed GO terms are displayed and the element names are linked to the SGD. A second file is generated saving the occurrence frequency of the respective term, i.e. the amount of cliques in the neighbourhood of the centre element that over represent certain terms. Additionally graphs produced by the respective analysis for each element can be saved.

“ESCARGOT”

Html file generator **“ESCARGOT”** (Easy Sgd Coupled Access through R for Groups with Overrepresented Terms)

This file generator was created to present the results from the automatized gene ontology analysis. The automatically saved results file from the previous analysis can directly be used as an input and will be processed into an html file that links to the SGD database. Upon opening the file in a browser, all the centre proteins and their interaction partners are hyperlinks to their respective entries to present more information about them. For each clique that has overrepresented GO terms, the systematic name of the centre element, its standard name, the standard names of its interaction partners and the overrepresented terms are listed in an alphabetical (referring to the standard name) table. This script facilitates “classical” biological gain of knowledge by referring to information about the involved elements after the modern approach of using automatized comparisons.

Preparation for mixOmics

“CHEVRE”

Categorical Clustering **“CHEVRE”** (Categorical Heuristic Evaluation to Reduce Elements)

After a gene ontology analysis of the genes tested in the microarray experiment and their interaction partners (distance 1) using a genetic interaction network as a reference, the result has to be made one dimensional in order to use it for a mixOmics analysis. To do this, the elements are clustered by similarity of the terms they display using the kmodes algorithm. Meaningful groups of genes, whose cliques over represent certain gene ontology terms, are associated by similarity of their categorical elements depending on how many of them overlap. These groups can be used as factors for the evaluation.

Results

Determination of hexokinase activity using ^{31}P -NMR

By using glucose or fructose as a substrate and a commercially available hexokinase mix (contains undefined amounts of Hxk1 and Hxk2 but predominantly the latter) as the enzyme we could see that the reaction is roughly 2 times faster on fructose compared to glucose, estimated by the ADP ratio when comparing fructose and glucose as substrates (taking the initial rates of the reactions, for fructose the first 3 sample points were used, for glucose until around 10 min).

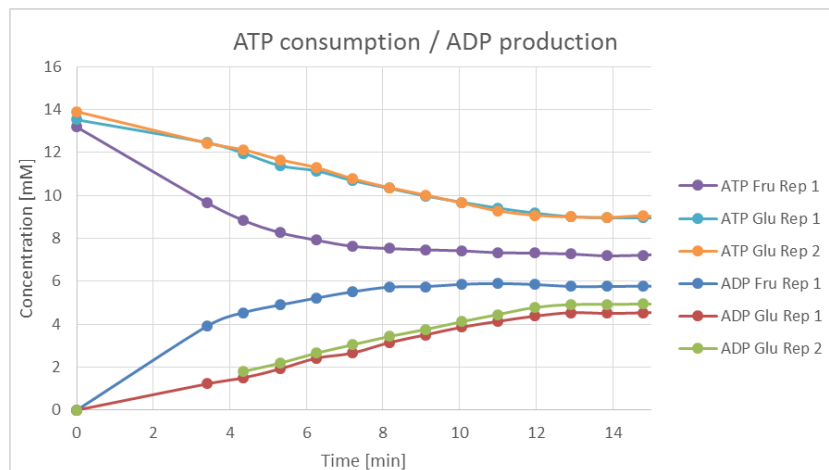


Figure 6: plotting the ATP and ADP concentrations measured at the specified time points

This fits for the behaviour of Hxk2 as described in (Lobo and Maitra, 1977), where the F/G ratio is said to be 2.3.

These results strengthen the idea that fructose causes ATP depletion in the cells by using all the ATP up too quickly.

TAP-tag data analysis

All the data analysed was obtained in previous experiments performed by Sevan Arabaciyan. The Mascot score was used as a semi-quantitative measure for protein abundance. The more (or less) abundant proteins can be separately listed for further manual analysis. Here, always a “stressed” sample, the heat shock sample exposed to 42 °C and the sample treated with acetic acid, was compared to its respective control sample, the one grown under normal conditions at 30 °C and the acid control sample grown at a pH value of 3.

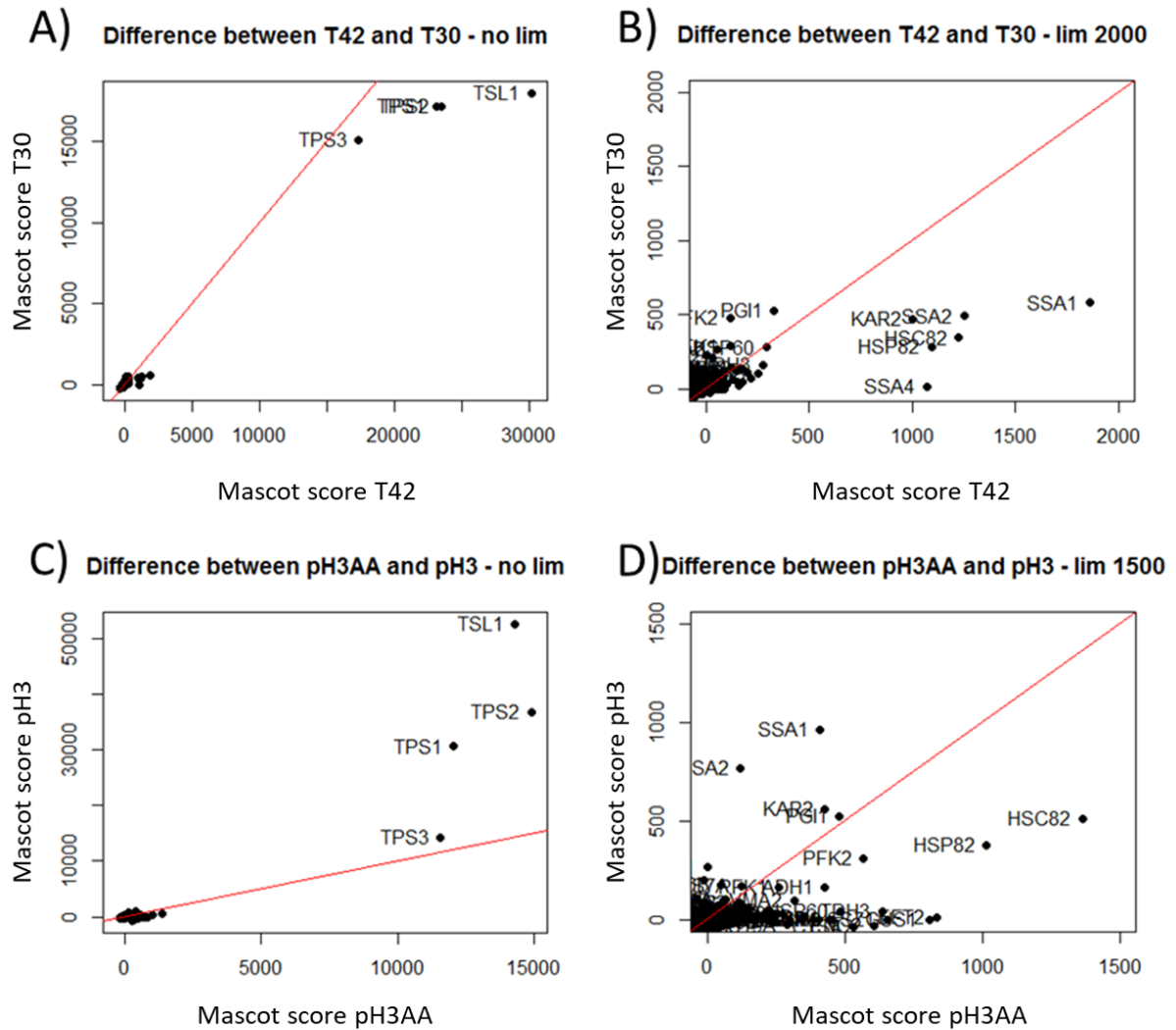


Figure 7: For a simple comparison of the samples, their abundance levels for proteins present in both conditions were plotted against each other. Above line: less abundant in testing condition. Below: more abundant in testing condition. A) + B) Temperature control and heat shock sample. Without and with axis limit, respectively. C) + D) Acid control and apoptosis inducing sample. Without and with axis limit, respectively. On all the axes, the mascot scores of the respective conditions are displayed.

Heat maps of the Mascot scores, as a semi-quantitative measure for abundance levels of the proteins interacting with the TAG-tagged Tps1 in the samples, were drawn. Since the TPS complex is far more interacting with Tps1 and thus far more abundant than all other proteins, comparing all protein levels does not make sense. Here a relative comparison displaying proteins that were present in at least three of the four tested conditions is displayed.

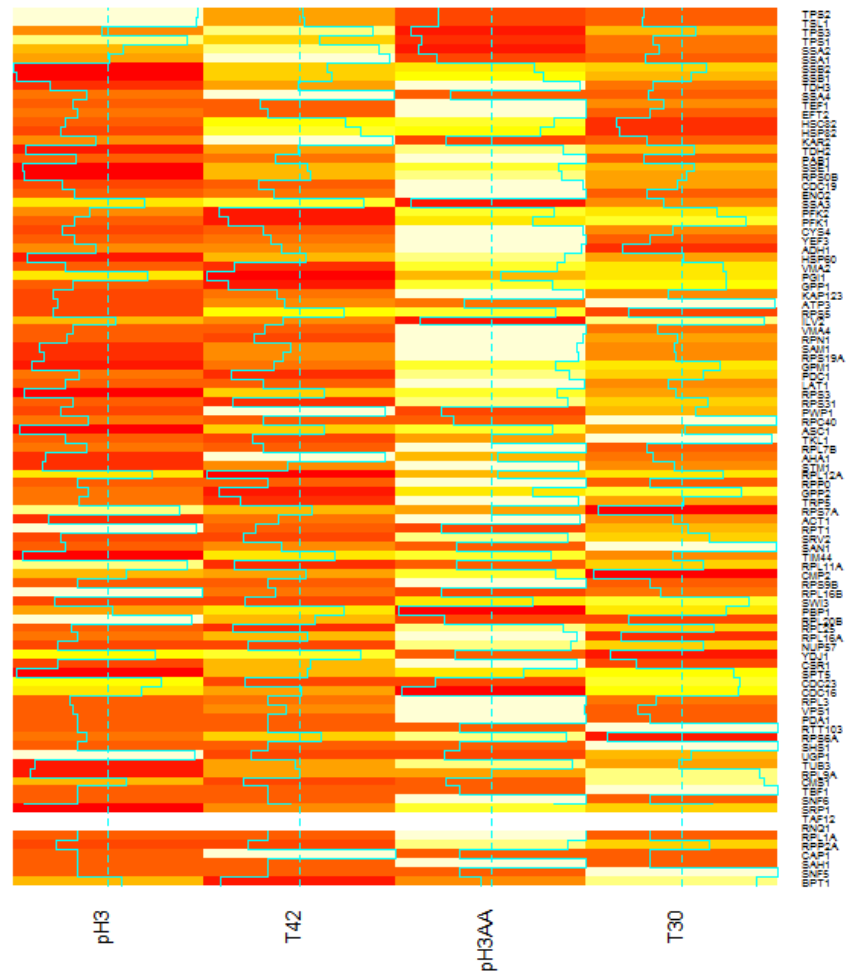
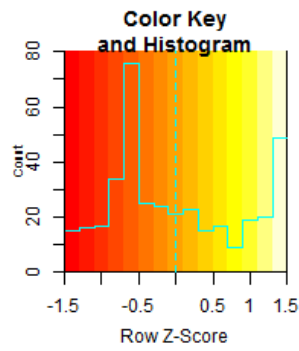


Figure 8: Heat map with row scaling for the corrected Mascot scores. The z-score represents the normalized deviation from the row mean value, also visualized by the blue line in the main graph. The box in the upper left corner shows the distribution of relative values.

In order to include data for proteins that were not present in most of the tested conditions, the clustering was performed separately. Conditions that display interaction partners that are not present in other conditions will cluster closer together. The data is generally scaled.

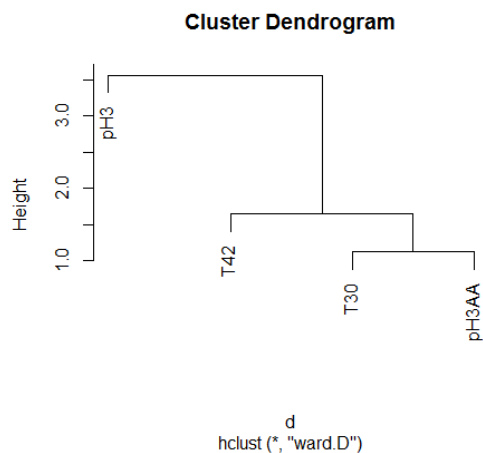


Figure 9: Cluster analysis using the "ward.D" method on generally scaled data

The sample grown under normal conditions clusters closest to the apoptosis inducing sample while the acid control sample shows the biggest distance to the other ones.

A Venn-diagram analysis illustrates the overlap in present proteins and also the uniquely expressed ones. All subsets can be listed separately for further manual analysis.

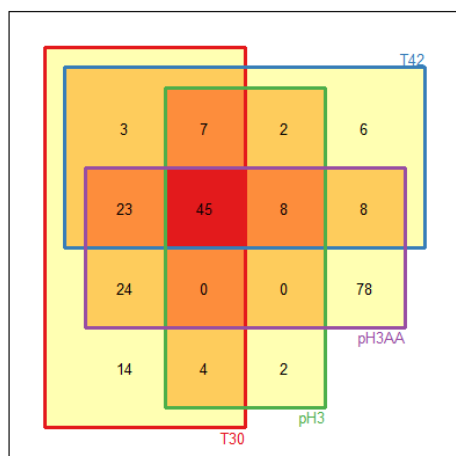


Figure 10: Venn diagram comparing the presence of proteins in the respective samples. The numbers in the boxes indicate how many proteins are shared by the conditions enclosing them (or are exclusively present in that condition)

Also looking at the distribution of the subsets, the normal conditions control sample shows a big similarity with the apoptosis inducing sample.

Interaction network analyses

Alternative names in BioGRID

First, all the names in the database were checked to make sure that no genes/proteins are annotated more than once using the standard or an alternative name. Everything was fine suggesting that the database is well curated automatically.

Integration of microarray data into interaction networks

Every presented exemplary model uses *Tsp1* as a centre. The aforementioned microarray data, consisting of four $\Delta tps1$ strains, three wild type strains grown under normal conditions and one wild type strain grown under micro-aeration conditions, was combined with the constructed interaction models using the “CANARD” script. In order to correlate them, the microarray results had to be reduced to one value for each element in the model. After normalization, the fold changes of the mean values of the $\Delta tps1$ strains in comparison to the wild type strains were calculated and used for further analysis. While the link with the genetic network was used to look for possible expressional influences the deletion might have, the protein-protein interaction network was used to represent possible functional correlations this would cause.

Integrating microarray data into interaction networks did not yield any remarkable results in this study. A more thorough and in depth microarray data analysis might allow for more sophisticated input. A visual analysis of the produced figures is only possible up to a certain extend. A graph with too many elements is not interpretable anymore due to the overlap of nodes and edges (see Figure 12). The physical interaction model is suited for an analysis (see Figure 11). Two out of four partners of the TPS complex (*TPS2* and *TPS3*) are clearly repressed while the fourth one, *TPS1*, seems not to change expression drastically. The unchanged value of *TPS1* in the *TPS1* inactivated strain might be explained by the way the gene was silenced. The part of the gene recognized by the microarray might still be expressed.

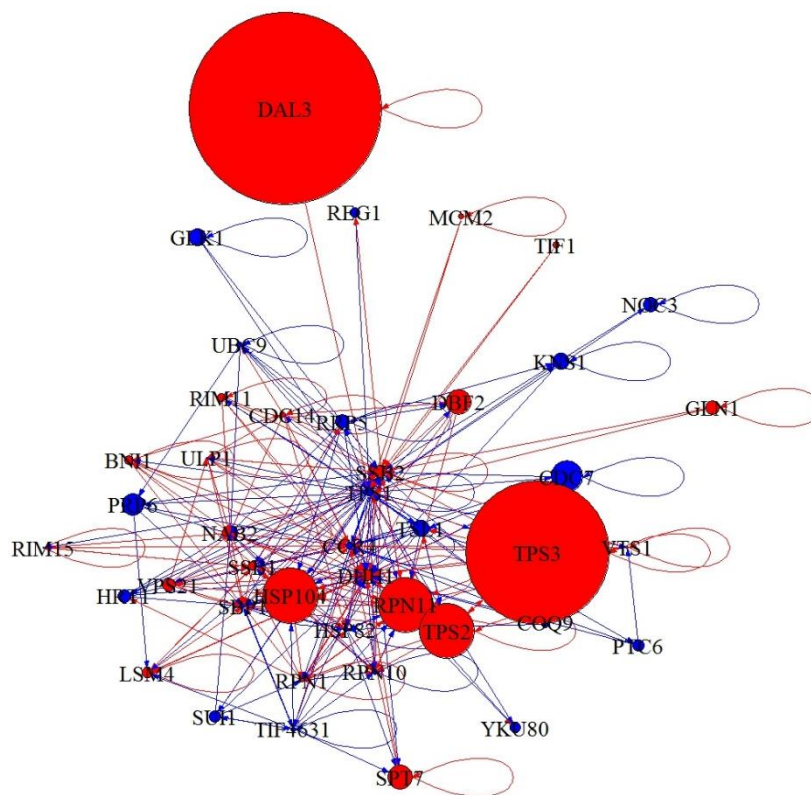


Figure 11: Microarray data integrated into a protein-protein interaction network for *Tps1* distance 1, size according to the absolute values of the \log_{10} of the fold changes of the mean of the mutant expression levels compared to the wild type. Blue indicates overexpression in the mutant, red repression.

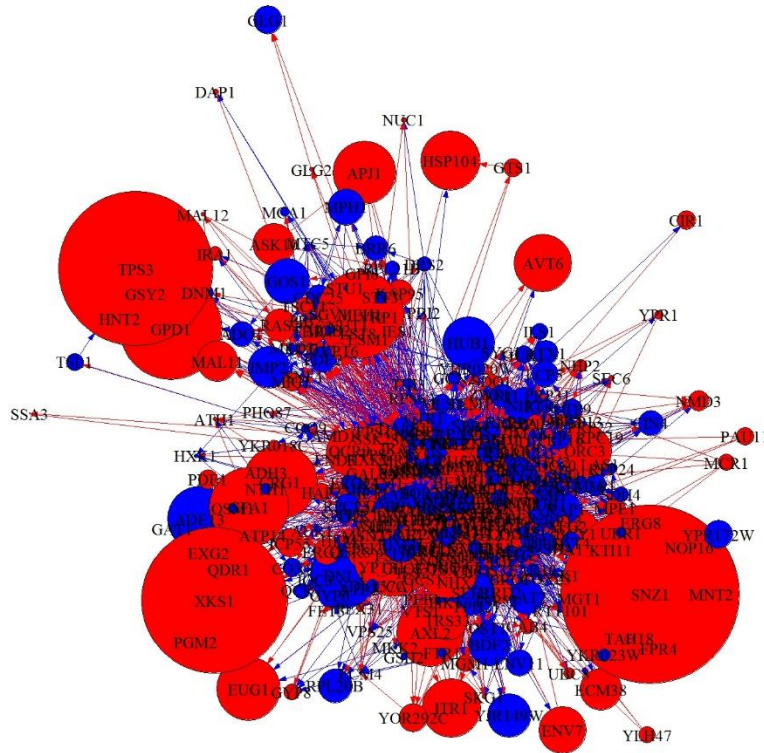


Figure 12: Microarray data integrated into a genetic interaction network for *Tps1* distance 1, size according to the absolute values of the \log_{10} of the fold changes of the mean of the mutant expression levels compared to the wild type. Blue indicates overexpression in the mutant, red repression. At this size of network, visual analysis becomes difficult. A tabular analysis of the results is preferred.

For future analysis and especially when regulatory elements are of concern, this method could provide further insight. Different approaches of combining networks with experimental data have been described in literature (Franceschini et al., 2012; Padi and Quackenbush, 2015).

Gene Ontology analysis

A gene ontology neighbourhood analysis was performed based on a physical interaction network constructed using the proteins obtained experimentally by TAP. The algorithm iterates through all the elements and takes all the interaction partners of each element and itself as a neighbourhood, also referred to as clique. Counting how often certain functions are overrepresented in cliques in this network provides functional information about the proximity of the element under investigation in the cell.

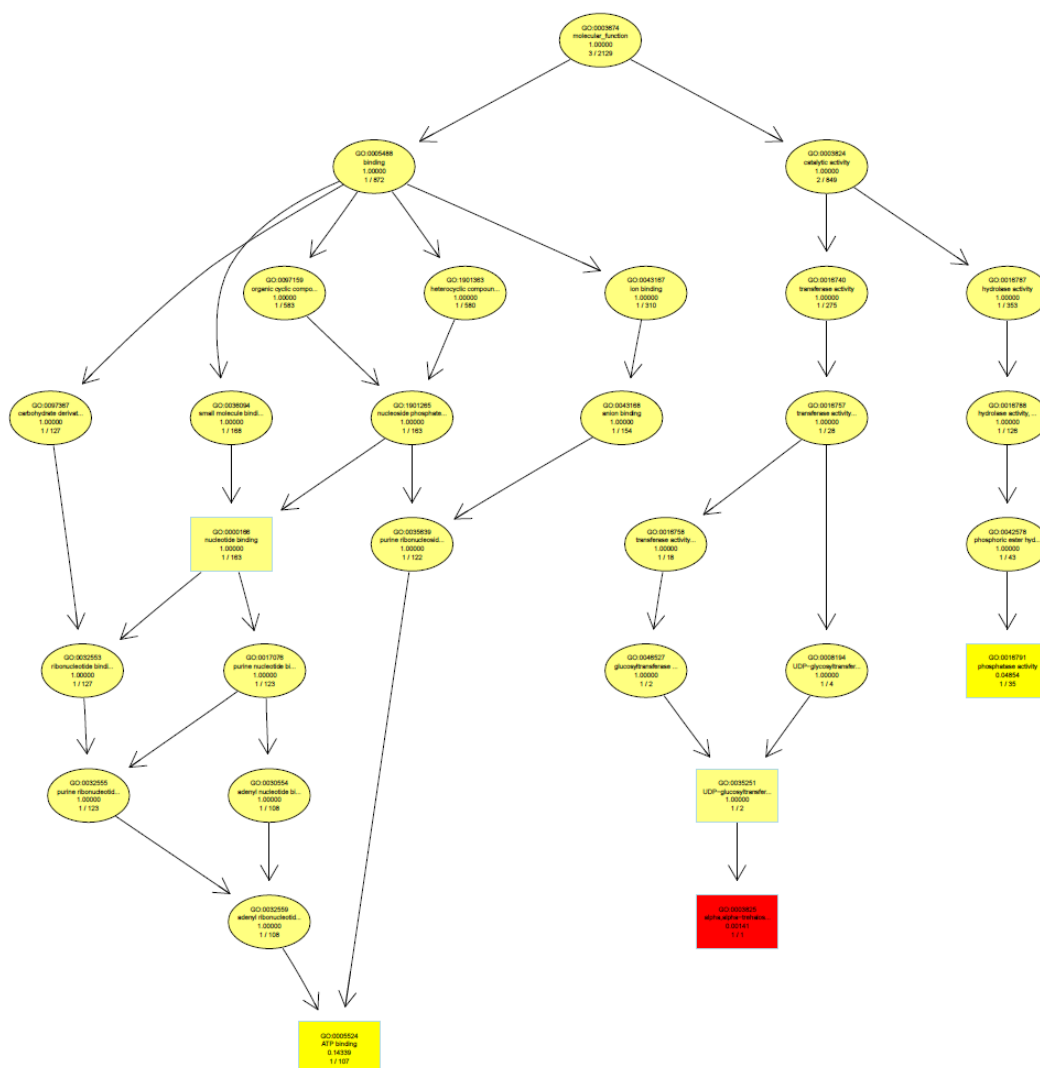


Figure 13: Exemplary GO tree for the TPS1 neighbourhood $d = 1$ showing the hierarchical structure of the GO terms (nodes). These graphs can be produced for all the enrichment analyses. The redness of the colour, which can change from pale yellow over multiple shade to orange and finally red, is related to the p-value and thus indicates the significance. Squares indicate the five GO-terms with the lowest p-values.

The GO terms are hierarchically structured following their function (see Figure 13). If a term is more abundant amongst the members of a tested group compared to the entirety of genes on the database it is considered to be enriched. The terms with p-values under a manually set threshold are collected and counted. Here the results for an analysis using an interaction network constructed using the partners identified in the TAP-tag analysis (Figure 14 and Table 8).

chorismate synthase activity	2
decanoate--CoA ligase activity	2
hydrogen-exporting ATPase activity, phos...	2
long-chain fatty acid-CoA ligase activit...	2
peptide-transporting ATPase activity	2
phenylalanine-tRNA ligase activity	2
porphobilinogen synthase activity	2
proteasome binding	2
structural constituent of ribosome	2
actin binding	1
imidazoleglycerol-phosphate dehydratase ...	1
phosphatidylinositol kinase activity	1
transcription factor activity, RNA polym...	1
unfolded protein binding	1
UTP:glucose-1-phosphate uridylyltransfer...	1

The most prevalent function in the cliques in the Tps1 based interaction model is the mannose-1-phosphate guanylyltransferase activity, an important step involved in cell wall synthesis. This is linked to the GDP-mannose pyrophosphorylase (mannose-1-phosphate guanyltransferase), which synthesizes GDP-mannose from GTP and mannose-1-phosphate and is encoded by the *PSA1* gene which is part of the experimentally obtained list of Tps1 interacting proteins. No interaction between these proteins is listed in the BioGRID database.

The results are afterwards summarized and presented in an html page linked to the SGD database.

Gene ontology neighbourhood (d = 1) analysis			
Molecular Function:			
Systematic_Name	Standard_Name	Interaction_Partners	Term
YDR214W	AHA1	AHA1 , HSC82 , HSP82 , KAP123 , PFK1 , RPC40 , RPN1 , RPT5 , SSA1 , SSB1 , SSB2	6-phosphofructokinase activity
YGL148W	ARO2	SSA1 , SSB2 , ARO2	chorismate synthase activity
YMR116C	ASC1	ASC1 , CLU1 , MIS1 , RPL25 , RPN1 , RPT5 , SAN1 , SSB2 , SSZ1	methylenetetrahydrofolate dehydrogenase ..., ribosome binding
YBR039W	ATP3	HSC82 , SSB2 , ATP3	ATP binding
YLL015W	BPT1	SSA1 , SSA2 , BPT1	peptide-transporting ATPase activity

Figure 15: Example for a result html page.

The same has been done for the database information alone (Figure 16 and Table 9).

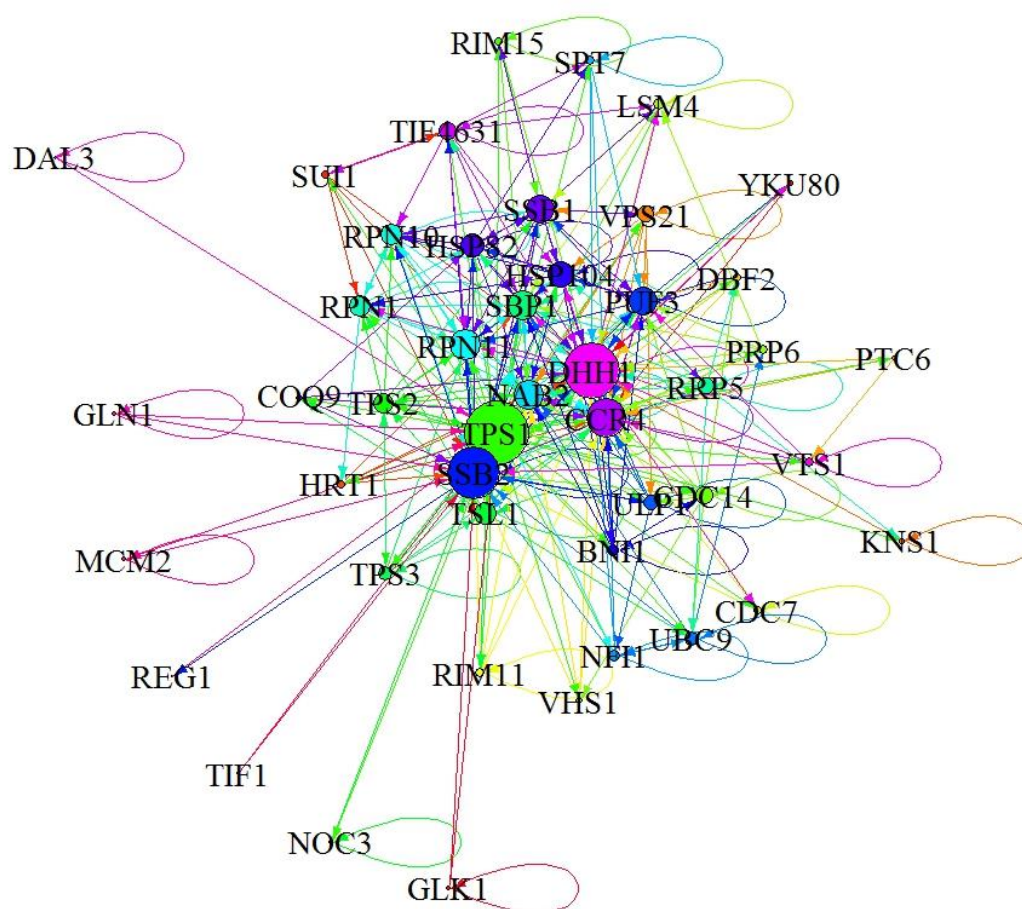


Figure 16: Interaction model of database $d = 1$ information using *TPS1* as a center.

Table 9: Overrepresented GO terms in the $d = 1$ model for *TPS1* and the percentage of cliques in this model they are overrepresented in

Overrepresented GO terms	Frequency [%]
thiol-dependent ubiquitin-specific protease activity	33
alpha,alpha-trehalose-phosphate synthase activity	15
U6 snRNA binding	15
SUMO transferase activity	11
protein serine/threonine kinase activity	9
ATP binding	7
[pyruvate dehydrogenase (lipoamide)] phosphorylation	4
protein serine/threonine phosphatase activity	2
ureidoglycolate lyase activity	2

In the model constructed purely from database information, thiol-dependent ubiquitin-specific protease activity was the most commonly found function while alpha alpha-trehalose-phosphate synthase activity, a function solely found in the members of the TPS complex, comes in second.

Cell Wall

Looking at the results of the GO enrichment analysis of the experimentally obtained interaction partners for Tps1, the molecular function overrepresented in the highest number of cliques was “mannose-1-phosphate guanylyltransferase activity” which suggests that Tps1 is in some way associated to genes involved in cell wall synthesis.

Since this data exploration approach is supposed to be used as an indicator of possible functions to explore, this result was used as a proof of principle.

The cell wall forms a protective layer made out of macromolecular components whose composition depends on genetic variations and growth conditions. Many genes, even if they are not directly involved in cell wall synthesis, can have a negative impact on the cells when they are altered (Lussier et al., 1997). A way to show differences between mutants of certain genes and the wild type is performing a spotting assay on agar plates containing media and drugs at varying concentrations to test for different phenotypes. Altered susceptibility to Calcofluor White and Congo Red, two anionic dyes, indicate alterations in chitin synthesis (Ram and Klis, 2006) by preventing nascent chains from co-crystallizing (Roncero et al., 1988). Changes do not have inherent positive or negative effects. Caffeine can enter cells with damaged cell walls and, as an analogue to purines, leads ultimately to cell death (Kuranda et al., 2006). Osmotic stress can be applied on the growing yeast by the addition of NaCl or sodium dodecyl sulphate to the media (Popolo et al., 2001).

So far, no influence of Tps1 on the synthesis of the wall has been described. Tests have been performed as a proof of principle of functions predicted using the methods described below.

Spotting assays on YP-Galactose plates containing drugs or substances affecting strains with defects in cell wall synthesis were performed with the wild type and the *TPS1* deletion strain. Pictures were taken after 3 and after 5 days to account for the slower growth of *TPS1* mutants. After five days, the extent of growth for both strains can be assumed to be about the same, as can be seen in the lowest concentrated plates for e.g. Calcofluor White, Caffeine or SDS.

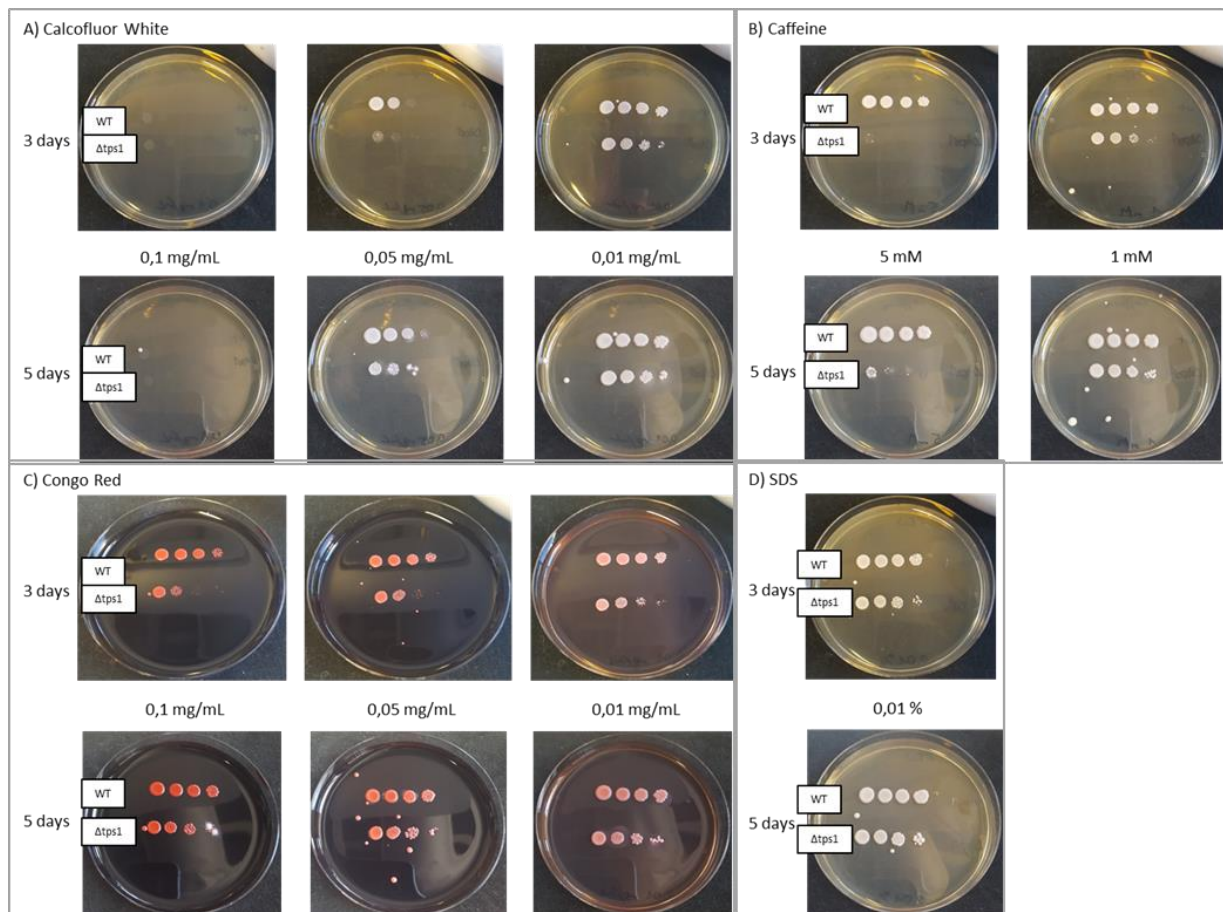


Figure 17: Drug sensitivity phenotypic tests on plates. Spotting assays on YP-Galactose plates containing different concentrations of different substances. Always the wild is type compared to a $\Delta tps1$ strain. Plates were substituted with A) different concentrations of Calcofluor White, B) different concentrations of Caffeine, C) different concentrations of Congo Red, D) SDS.

Calcofluor White and Congo Red bind to chitin. While differential reaction to the agents is not directly related to effects, cells that are more resistant to it than the wild type can be assumed to have problems in the chitin synthesis while a lower resistance mainly indicates a lower stress resistance in general. Lower resistance can be due to higher chitin levels which are a direct response to cell wall damage (Lagorce et al., 2003). Caffeine can enter into cells with a damaged cell wall and leads to cell death while cells with an intact wall are largely protected. SDS is used to test for susceptibility to osmotic stress.

Although there are differences visible on Calcofluor White and SDS, they are not strongly pronounced. Congo Red on the other hand seems to have a strong effect on the $\Delta tps1$ strain; a higher cell concentration seems to be necessary in order to allow for growth. The biggest effect is visible on Caffeine where a concentration of 5 mM completely inhibits growth of the $\Delta tps1$ strain after 3 days and only allows it at high cell concentrations after 5 days while the wild type is growing normally.

Effects on the cell wall have not been described so far for $\Delta tps1$ strains, it is advised to have a closer look into that in future projects. This phenotypic experiment supports the suggestion that *TPS1* deletion influences cell wall integrity.

The microarray data of wild type and *TPS1* mutant samples was analysed together with metabolite rates taken from the same fermentations the samples were taken from. Here, the results of the sparse analyses regarding the genes are displayed to allow for a more meaningful presentation. Because only a few genes are displayed, they cannot be related to each other. The PCA (principle component analysis) sample plots present the projection of the samples on the two major variants according to their multivariate models. The axes represent principle components in the datasets, maximising the variance in those directions. Block X refers to the genes, the microarray results, while block Y represents the rates obtained from the fermentation data processing. The numbers refer to the samples in the order presented in table 10.

Table 10: Samples used for mixOmics analysis.

Sample	Type
1	TPS1
2	TPS1
3	TPS1
4	TPS1
5	WT
6	WT
7	WT
8	WT_μ-aeration

In order to interpret the results, the sample plots have to be seen as a “map” for the circle plot. Arrows have been added for clarity. The position of the samples on the sample plot corresponds to positions on the circle plot. Genes or rates in certain locations in the circle plot can be understood to relate to clusters in a similar place in the sample plots.

sPLA

The source of the biggest variation, and thus principal component number one, is on the x-axis. Looking at the samples, this describes the difference between the wild type samples grown under normal conditions (5 – 7) and the micro-aeration sample (8) as they can be found on the outer edges of the plot. The second biggest source of variation arises from the difference between wild type and deletion strains. The distributions relate to the positioning of the samples relative to one another according to the descriptor (genes or rates) while the principal components are determined by the analysis for both blocks, thus they can be overlaid in the circle plot. Also the distributions in the sample plots follow a similar order, indicating that the distribution is sensible.

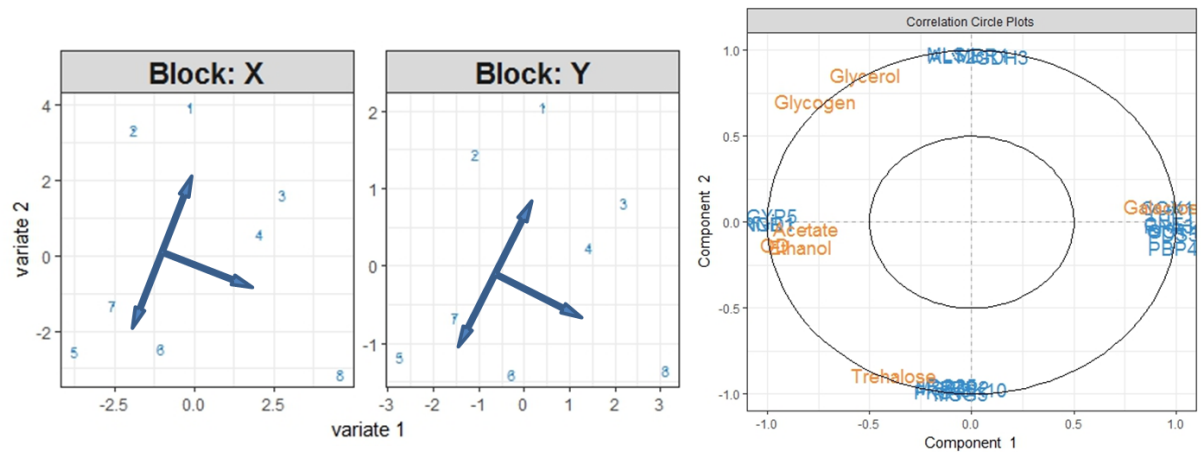


Figure 18: Results of sPLA analysis, sample plots for X (genes) and Y (rates) on the left, circle plot on the right.

Trehalose, since it is only present in the wild type samples, is located in the direction away from the deletion strains. Cell growth, acetate and ethanol production locate away from the micro-aeration condition closer to the wild type, glycogen and glycerol in the same direction but closer to the mutants. Galactose, since it is consumed it is a negative rate, indicated slow growth. It is located opposite of the optical density, associated to the micro-aeration sample and the $\Delta tps1$ samples.

sPLA-DA

For the discriminant analysis, the program tries to find the biggest differences between the named types of samples, namely wild type and deletion strain. Since it is looking for that, the principal component is the distinction between the types. On the genetic level, the micro-aeration had little influence while its effect is visible on the rate level.

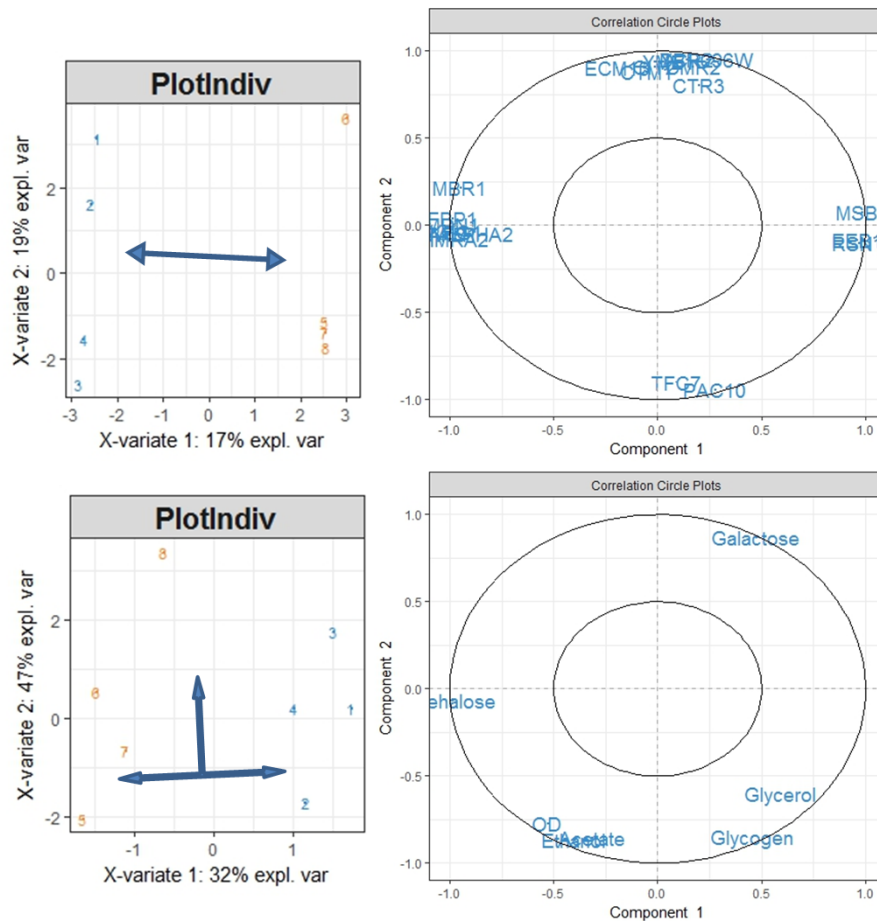


Figure 19: Results of sPLA-DA analysis, sample plots for the gene samples (up) and the rates (down) on the left, circle plot for the genes (up) and for the rates (down) on the right.

The rates orient themselves in a similar fashion compared to the sPLA.

Block-sPLA-DA

The block-sPLA-DA combines the integration of multiple 'omics' sets while looking for the biggest differences between the assigned types. The most important variant, due to the type of analysis, differentiates between wild type and deletion strain. Variant number two describes the difference of the micro-aeration sample to the strains grown under normal conditions.

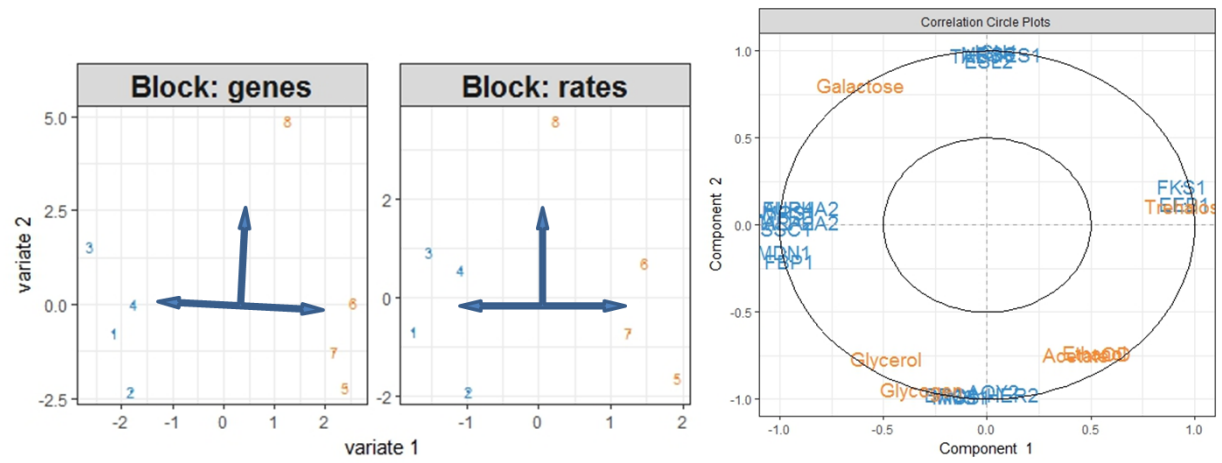


Figure 20: Results of Block-sPLA-DA analysis, sample plots for X (genes) and Y (rates) on the left, circle plot on the right.

The rates are again oriented similar to the distribution after the sPLA.

Discussion

Gene Ontology analysis

While distance = 1 analyses are of interest as exemplified by the cell wall experiments, the distance = 2 analysis is not very fruitful since there are some highly interacting proteins in the model and any protein connected to one of them will automatically be connected to almost all of the others and thus all of the $d = 2$ models look basically the same. It remains to be shown that they could be interesting for bigger models that are not based on data from a single element.

The statistics of the implemented R package `topGo` was checked by Serguei Sokol for a few cliques at random using the `phyper` function. Other sources of error were also tested, e.g. if all proteins are only presented in the database by only one name. Name convention differences between the BioGRID and Gene Ontology databases were taken care of.

An R script was written as well that produces an html file for a more convenient manual analysis of the results. The “Systematic Name”, the “Standard Name”, the standard names of the “Interaction Partners” and the overrepresented GO terms are presented. The standard names link directly to their respective database entry.

The biggest problem of this approach remains. Database entries, especially of high throughput studies, are usually based on experiments performed using glucose as a substrate while the experimental data here (and especially the follow-up experiments) used galactose. This inherently favours the presence of proteins or overexpression of genes involved in galactose processing or related processes, thus distorting the result. Overrepresentation of functional terms linked to this alternative carbon source are to be expected.

Interaction Modelling

Interactomics data

A prominent general problem in this field of study is reproducibility. Most interactions are only described in one study and cannot be shown again in subsequent experiments. This is usually attributed to the nature of these studies. Most are high throughput screenings testing multiple proteins in a standardized setting which opens up the possibility of biological errors arising in the set-up or just technical errors due to these studies having a higher risk of unspecific results than focused studies.

This problem causes the need for a filtering step. For proteins that are described in multiple publications, it makes sense to restrict the number of interactions to those described in (at least) more than one study.

Problems with the BioGRID database inbuilt functions

As for the results, every presented exemplary model uses Tsp1 as a centre.

Using the “evidence level” approach with the presented method, i.e. setting a limit for a minimum number of publications mentioning interactions, those that have not been observed repeatedly are excluded from the interaction model construction. Applying the filtering step prior to model construction limits it to direct interactions with the centre element that have been reported more often than the defined limit. This approach is very restrictive and only highlights the direct proximity of the element(s) under investigation.

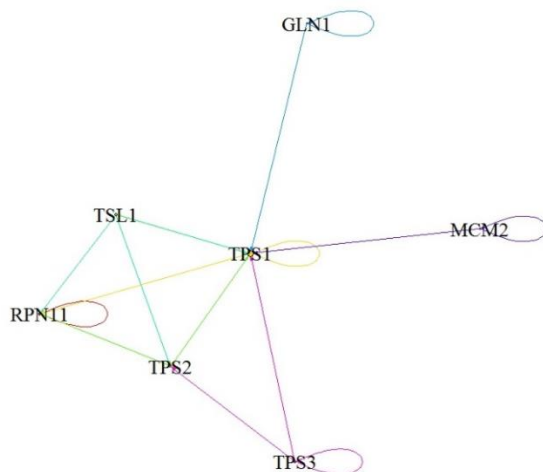


Figure 21: Tps1 distance 1 protein-protein interaction model with filtering for interactions that have been described in at least 2 publication prior to model construction.

Constructing the interaction model before filtering the data, i.e. taking all described interactors into account for model construction but only connecting repeatedly reported interactors, gives a reduced model in broader context. Not only the direct proximity but also interactors that are separated from the centre after model reduction are displayed.

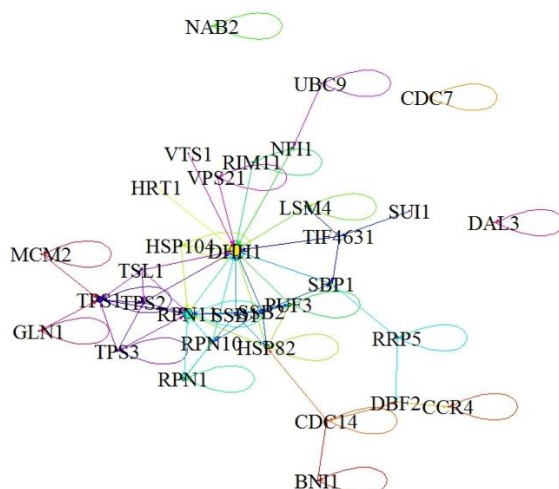


Figure 22: Tps1 distance 1 protein-protein interaction model with filtering for interactions that have been described in at least 2 publication prior to model construction.

Here the physical interactions network on the database is given to show concerns about it.

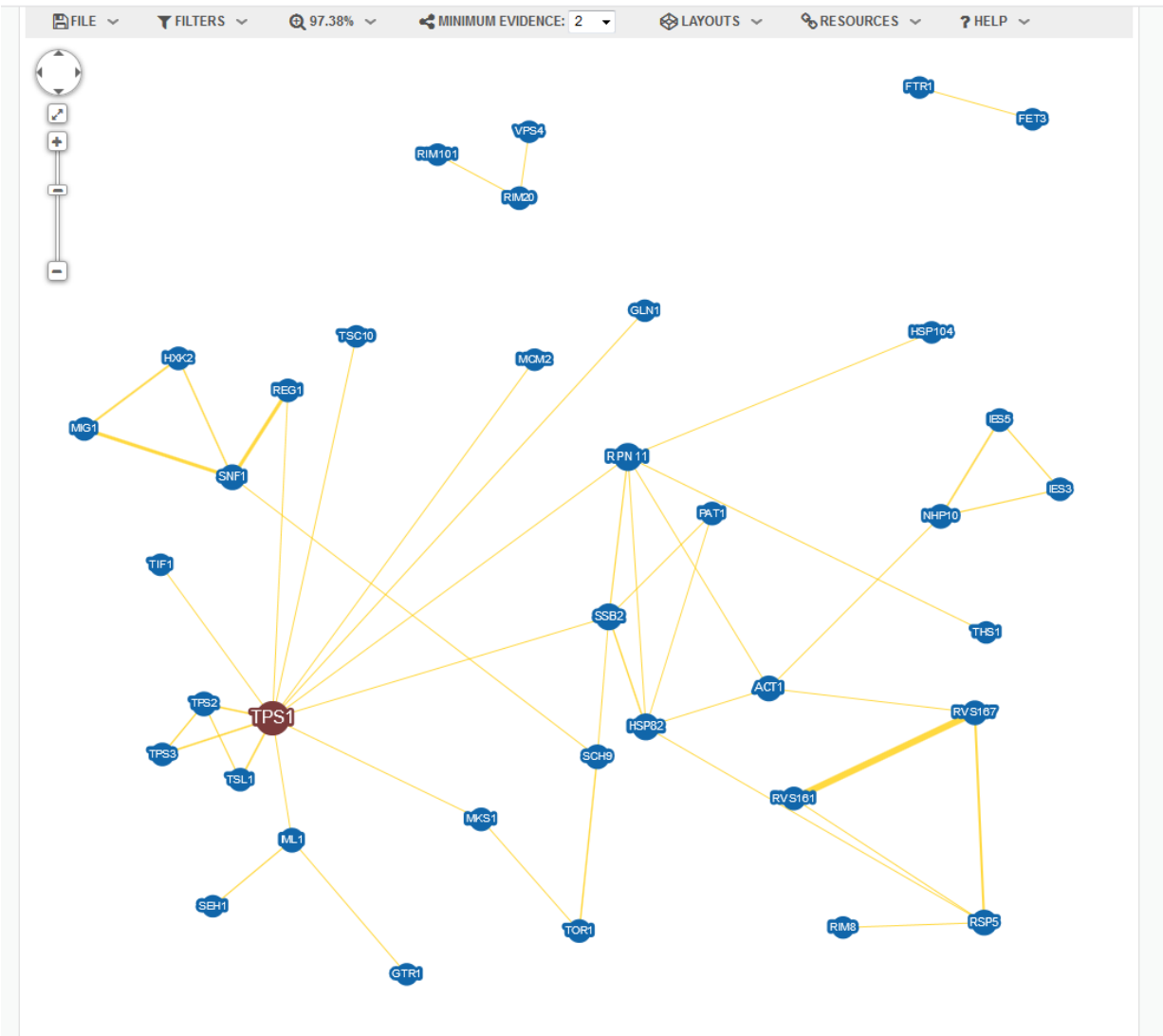


Figure 23: *Tps1* distance 1 protein-protein interaction model with filtering for interactions that have been described in at least 2 publication constructed by the BioGRID database.

BioGRID will also display genes (or in that case proteins) that show genetic interactions or chemical modifications and their respective interactions with each other even though they do not have any physical interactions connected to the physical network of the gene of interest. This means that those connections would not only be absent with the set evidence limit, they should also be absent in a model of the set type of interaction at all.

Some genes are also inexplicably missing. Like for example a physical interaction of “TPS1” with “DHH1” is described, listed in the “Interactions” section of the database, shows up in the model constructed from the downloaded data but is missing in the network constructed by the database in the browser.

DGK1	BAIT	S. cerevisiae	Negative Genetic	Costanzo M (2016)	High	-0.3234	BioGRID	
DHH1	BAIT	S. cerevisiae	Affinity Capture-RNA	Jungfleisch J (2017)	High	-	BioGRID	
DHR2	BAIT	S. cerevisiae	Negative Genetic	Costanzo M (2016)	High	-0.2964	BioGRID	

Figure 24: Protein-protein interaction entry of "DHH1" with "TPS1" in the database as an example.

Limits of the presented modelling approach

Here, TPS1 has been used as an example for a distance = 2 model. Graphical analysis becomes impossible but numerical analysis is still possible. Gene ontology enrichment analysis is still possible although one has to ask themselves if it makes any sense at this level. The following is just presented to show how far we can go, not to suggest any conclusions.

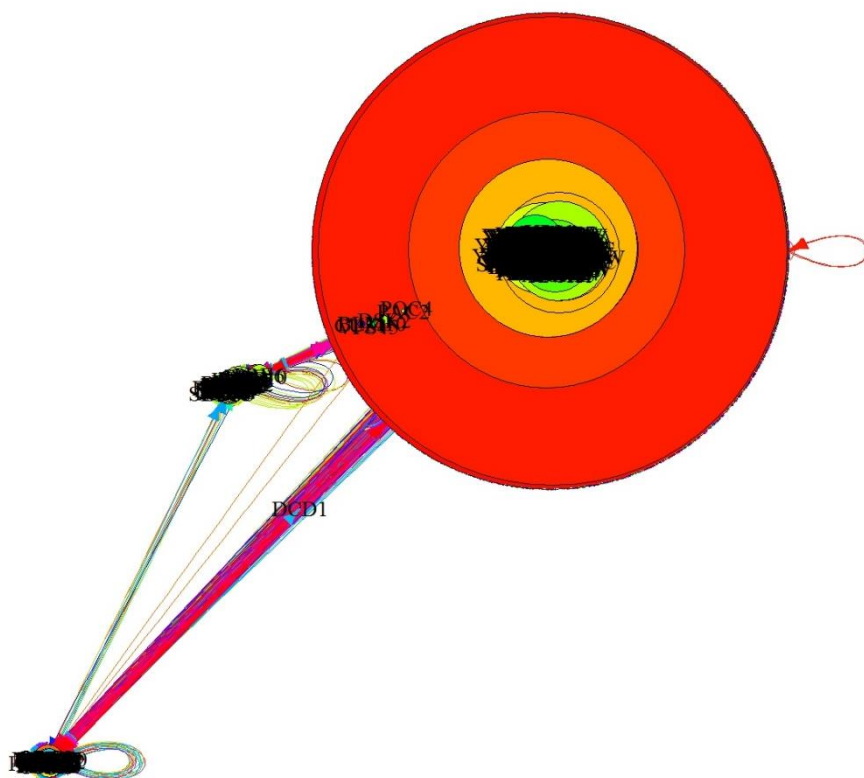


Table 11: Reduced list of the most interacting proteins in the model, how many other proteins they interact with as counts, and percentage of the number of proteins in the model.

	Dominant proteins					
Names	DHH1	SSB2	CCR4	ISW1	NAB2	RPN11
Int. Partners [counts]	1735	1611	1232	692	609	543
Int. Partners [%]	74	68	52	29	26	23

Table 12: Reduced list of overrepresented GO terms in the $d = 2$ model for TPS1 and the number of time they occur as counts and percentage of the number of proteins in the model.

Overrepresented GO terms	Frequency [counts]	Frequency [%]
thiol-dependent ubiquitin-specific prote...	341	14
ATP binding	230	10
RNA polymerase III activity	104	4
protein serine/threonine kinase activity	90	4
mRNA binding	81	3

mixOmics

The given set of data is not the most suitable one for evaluation and it might be difficult to draw conclusions from it. There are a lot of missing values, e.g. for the metabolites, since every time a different set was used. One of the wild type fermentations was conducted under micro-aeration conditions so it might have to be excluded altogether. It is not an easy set of data but it is interesting to tackle all those problems and to try to integrate the functional terms, the interaction models and the time series information.

It was however possible to see expected results, which speaks for the validity of the method. All of the distributions of the rates are how one would assume them to be; the trehalose production orients itself close to the only samples that produce trehalose, the formation rates of metabolic products are lowest on the micro-aeration samples and the growth indicator (OD and galactose) lie opposite of one another, saying that the remaining substrate is highest in mutant strains and micro-aeration samples while cell growth is higher in normal conditions. The results of this microarray data analysis were not examined in great extent. A limited, subjective, manual analysis deemed the result to be sensible.

For the functional information, a problem, like always when it comes to clustering, is the selection of a suitable number of clusters. One cluster overshadows the others while some have only one member. In the end, the number of clusters is a subjective decision. The validity of summarising functional terms into clusters and using them for further analysis remains to be shown. It is possible that this would create nonsensical groups without any benefit for the analysis rather than functional clusters.

Another idea to expand the analysis would be to include the network information of the interaction model. Gene expression can be assumed to be quite different in wild type and mutant strains. Database information about interactions is assumed to represent the wild type state. Looking at a network that would be constructed for one strain under one condition would be different than one constructed for another one. If it could be managed to include the models and “assign” them to the wild type samples, changes arising in the $\Delta tps1$ strain could be connected, e.g. seemingly uncorrelated perturbations in two genes might be linked and this piece of information could be indicated hereby.

At a later stage it might be possible to include the data from the fitted models for the fermentations directly as “time series” information instead of reducing the curves to rates. This could be crucial considering that effects on metabolite levels might arise delayed or in a specific fashion, e.g. sudden changes of rates or loss of correlations between metabolites, which is not being fully displayed in the exponential growth phase.

Cell wall

The cell wall synthesis perturbation experiments were promising but one should keep in mind that they are based on preliminary data. Should the same functional overrepresentation be present in an exhaustive analysis, more thorough experiments highlighting the involvement of the Tps1 protein or the effect of its absence are advised.

The model based solely on database information does not indicate an involvement of Tps1 in cell wall synthesis. This can be caused by the fact that available interactomics information for this protein is almost exclusively based on high throughput experiments leaving a larger margin of error and possibly overlooking key interactions compared to a more focused approach.

Outlook

The connection between sugar phosphorylation and Tps1

A collection of all variations of deletions for the genes encoding the three hexokinases will be created for the wild type and the *TPS1* deletion strain. All variations can grow on galactose.

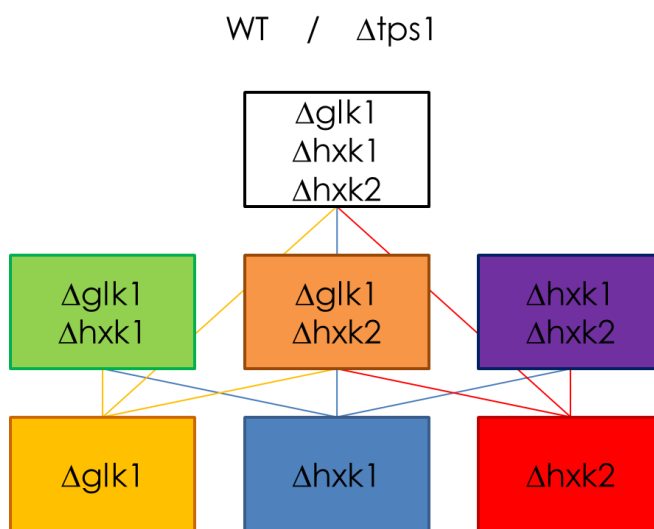


Figure 26: All combinations of the kinase knock outs should be prepared, each in a wild type and a *Δtps1* strain. The triple deletion should not be viable on fermentable carbon sources.

These knock outs will be used for subsequent analysis. Exposing *Δtps1Δglk1Δhxxk2* cells to glucose media will reveal if the strain can recover growth. This would indicate that recovery is linked to the deletion of *HXXK2*.

If it is suggested that growth of the *Δtps1Δhxxk2* depends on glucokinase 1, the hypothesis that the growth defect is due to the V_{max} of hexokinases on fructose being too high gains credibility.

To validate this idea, the following experiments could be performed:

qPCR experiments could unravel the expression of the hexokinases in various situations, e.g. comparing *HXXK1* and *GLK1* levels in *Δtps1Δhxxk2* cells that can grow on glucose to see if, as is assumed, *GLK1* is more abundant, especially in regard of *TPS1* mutant versus wild type. Expression levels are also interesting for the role of the enzymes. Their involvement in the utilisation of non-fermentable carbon sources remains unknown.

By measuring the ATP utilisation rate in a *TPS1* mutant strain expressing only *GLK1* and normalizing it by the cultures' total protein content, a relative measure for a growth-sustaining hexokinase rate can be obtained.

By studying the reaction behaviour of Hxxk1 and Hxxk2 on fructose and the change in rate upon addition of an inhibitor, e.g. 6-desoxyglucose, it can be calculated how much of this inhibitor is necessary to artificially lower the activity of the hexokinases on fructose to the level of the glucokinase 1 on glucose. A spotting assay on a minimal media plate using fructose as a carbon source and the respective relative amount of inhibitor to lower hexokinase activity can reveal if this approach restores growth. A positive result would be an indication that the reaction velocity of hexokinases using fructose as a substrate

brings $\Delta tps1$ cells irreversibly out of balance and, further on, that the $\Delta tps1$ phenotype is, at least in part, a hexokinase rate problem.

A sugar pulse experiment in which the mutant strain is confronted with an excess of glucose or fructose could also be performed. A dramatic loss of ATP in response to fructose but not to glucose would be expected.

Interactomics analysis

The interactomics experiment will be repeated with biological controls. The new data can be evaluated with the scripts and methods developed in this project. It would make sense to be more restrictive with the proteins assumed to be interaction partners. After subtracting the mascot scores of proteins, those that are close to 0 can either be real interactants or technical remnants. An exclusive strategy seems feasible. Possible conclusions from functional analyses, like e.g. the described cell wall synthesis experiments, should be tested to verify the applicability of these methods.

References

- Adrian Alexa, and Jörg Rahnenführer (2018). topGO: Enrichment Analysis for Gene Ontology. R Package Version 2301.
- Barclay, T., Ginic-Markovic, M., Johnston, M.R., Cooper, P., and Petrovsky, N. (2012). Observation of the keto tautomer of D-fructose in D(2)O using (1)H NMR spectroscopy. *Carbohydr. Res.* **347**, 136–141.
- Barker, M., and Rayens, W. (2003). Partial least squares for discrimination. *J. Chemom.* **17**, 166–173.
- Bonini, B.M., Van Dijck, P., and Thevelein, J.M. (2003). Uncoupling of the glucose growth defect and the deregulation of glycolysis in *Saccharomyces cerevisiae* Tps1 mutants expressing trehalose-6-phosphate-insensitive hexokinase from *Schizosaccharomyces pombe*. *Biochim. Biophys. Acta* **1606**, 83–93.
- Chatr-aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N.K., O'Donnell, L., Oster, S., Theesfeld, C., Sellam, A., et al. (2017). The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* **45**, D369–D379.
- Clifton, D., Walsh, R.B., and Fraenkel, D.G. (1993). Functional studies of yeast glucokinase. *J. Bacteriol.* **175**, 3289–3294.
- Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, 1695.
- van Dijken JP, null, Bauer, null, Brambilla, null, Duboc, null, Francois, null, Gancedo, null, Giuseppin, null, Heijnen, null, Hoare, null, Lange, null, et al. (2000). An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb. Technol.* **26**, 706–714.
- Draghici, S., Sellamuthu, S., and Khatri, P. (2006). Babel's tower revisited: a universal resource for cross-referencing across annotation databases. *Bioinformatics* **22**, 2934–2939.
- Eicher, J., Snoep, J., and Rohwer, J. (2012). Determining Enzyme Kinetics for Systems Biology with Nuclear Magnetic Resonance Spectroscopy. *Metabolites* **2**, 818–843.
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., et al. (2012). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41**, D808–D815.
- Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., Cruciat, C.-M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147.
- Gene Ontology Consortium (2001). Creating the gene ontology resource: design and implementation. *Genome Res.* **11**, 1425–1433.
- Gibney, P.A., Schieler, A., Chen, J.C., Rabinowitz, J.D., and Botstein, D. (2015). Characterizing the in vivo role of trehalose in *Saccharomyces cerevisiae* using the *AGT1* transporter. *Proc. Natl. Acad. Sci.* **112**, 6116–6121.

- Gietz, R.D., and Woods, R.A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* 350, 87–96.
- Guillou, V., Plourdeowobi, L., Parrou, J., Goma, G., and Francois, J. (2004). Role of reserve carbohydrates in the growth dynamics of. *FEMS Yeast Res.* 4, 773–787.
- van Heerden, J.H., Wortel, M.T., Bruggeman, F.J., Heijnen, J.J., Bollen, Y.J.M., Planqué, R., Hulshof, J., O’Toole, T.G., Wahl, S.A., and Teusink, B. (2014). Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. *Science* 343, 1245114.
- Hohmann, S., Neves, M.J., de Koning, W., Alijo, R., Ramos, J., and Thevelein, J.M. (1993). The growth and signalling defects of the *ggs1* (*fdp1/byp1*) deletion mutant on glucose are suppressed by a deletion of the gene encoding hexokinase PII. *Curr. Genet.* 23, 281–289.
- Huang, Z (1998). Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Min. Knowl. Discov.* 283–304.
- Ilyedjian, P.B. (1998). Glycolysis, turbo design and the endocrine pancreatic beta cell. *Trends Biochem. Sci.* 23, 467–468.
- Kuranda, K., Leberre, V., Sokol, S., Palamarczyk, G., and François, J. (2006). Investigating the caffeine effects in the yeast *Saccharomyces cerevisiae* brings new insights into the connection between TOR, PKC and Ras/cAMP signalling pathways. *Mol. Microbiol.* 61, 1147–1166.
- Lagorce, A., Hauser, N.C., Labourdette, D., Rodriguez, C., Martin-Yken, H., Arroyo, J., Hoheisel, J.D., and François, J. (2003). Genome-wide Analysis of the Response to Cell Wall Mutations in the Yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* 278, 20345–20357.
- Laughery, M.F., Hunter, T., Brown, A., Hoopes, J., Ostbye, T., Shumaker, T., and Wyrick, J.J. (2015). New vectors for simple and streamlined CRISPR-Cas9 genome editing in *Saccharomyces cerevisiae*. *Yeast Chichester Engl.* 32, 711–720.
- Lê Cao, K.-A., Rossouw, D., Robert-Granié, C., and Besse, P. (2008). A Sparse PLS for Variable Selection when Integrating Omics Data. *Stat. Appl. Genet. Mol. Biol.* 7.
- Lehmann, E.L., and Romano, J.P. (2005). *Testing Statistical Hypotheses* (New York, NY: Springer New York).
- Lobo, Z., and Maitra, P.K. (1977). Physiological role of glucose-phosphorylating enzymes in *Saccharomyces cerevisiae*. *Arch. Biochem. Biophys.* 182, 639–645.
- Lussier, M., White, A.M., Sheraton, J., di Paolo, T., Treadwell, J., Southard, S.B., Horenstein, C.I., Chen-Weiner, J., Ram, A.F., Kapteyn, J.C., et al. (1997). Large scale identification of genes involved in cell surface biosynthesis and architecture in *Saccharomyces cerevisiae*. *Genetics* 147, 435–450.
- Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G.A., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J.A., and Daran, J.-M.G. (2015). CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 15.
- Oughtred, R., Chatr-aryamontri, A., Breitkreutz, B.-J., Chang, C.S., Rust, J.M., Theesfeld, C.L., Heinicke, S., Breitkreutz, A., Chen, D., Hirschman, J., et al. (2016a). BioGRID: A Resource for Studying Biological Interactions in Yeast: Table 1. *Cold Spring Harb. Protoc.* 2016, pdb.top080754.

Oughtred, R., Chatr-aryamontri, A., Breitskreutz, B.-J., Chang, C.S., Rust, J.M., Theesfeld, C.L., Heinicke, S., Breitskreutz, A., Chen, D., Hirschman, J., et al. (2016b). Use of the BioGRID Database for Analysis of Yeast Protein and Genetic Interactions. *Cold Spring Harb. Protoc.* 2016, pdb.prot088880.

Padi, M., and Quackenbush, J. (2015). Integrating transcriptional and protein interaction networks to prioritize condition-specific master regulators. *BMC Syst. Biol.* 9.

van de Poll, K.W., and Schamhart, D.H. (1977a). Characterization of a regulatory mutant of fructose 1,6-bisphosphatase in *Saccharomyces carlsbergensis*. *Mol. Gen. Genet.* MGG 154, 61–66.

van de Poll, K.W., and Schamhart, D.H. (1977b). Characterization of a regulatory mutant of fructose 1,6-bisphosphatase in *Saccharomyces carlsbergensis*. *Mol. Gen. Genet.* MGG 154, 61–66.

Popolo, L., Gualtieri, T., and Ragni, E. (2001). The yeast cell-wall salvage pathway. *Med. Mycol.* 39 Suppl 1, 111–121.

R Core Team (2017). R: A language and environment for statistical computing. R Found. Stat. Comput. Vienna Austria URL <https://www.R-project.org/>.

Ram, A.F.J., and Klis, F.M. (2006). Identification of fungal cell wall mutants using susceptibility assays based on Calcofluor white and Congo red. *Nat. Protoc.* 1, 2253–2256.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Séraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 17, 1030–1032.

Rohart, F., Gautier, B., Singh, A., and Lê Cao, K.-A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* 13, e1005752.

Roncero, C., Valdivieso, M.H., Ribas, J.C., and Durán, A. (1988). Isolation and characterization of *Saccharomyces cerevisiae* mutants resistant to Calcofluor white. *J. Bacteriol.* 170, 1950–1954.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675.

Teste, M.-A., Duquenne, M., François, J.M., and Parrou, J.-L. (2009). Validation of reference genes for quantitative expression analysis by real-time RT-PCR in *Saccharomyces cerevisiae*. *BMC Mol. Biol.* 10, 99.

Thevelein, J.M., and Hohmann, S. (1995). Trehalose synthase: guard to the gate of glycolysis in yeast? *Trends Biochem. Sci.* 20, 3–10.

Vicente, R.L., Spina, L., Gómez, J.P.L., Dejean, S., Parrou, J.-L., and François, J.M. (2018). Trehalose-6-phosphate promotes fermentation and glucose repression in *Saccharomyces cerevisiae*. *Microb. Cell Graz Austria* 5, 444–459.

Walsh, R.B., Clifton, D., Horak, J., and Fraenkel, D.G. (1991). *Saccharomyces cerevisiae* null mutants in glucose phosphorylation: metabolism and invertase expression. *Genetics* 128, 521–527.

Walther, T., Mtimet, N., Alkim, C., Vax, A., Loret, M.-O., Ullah, A., Gancedo, C., Smits, G.J., and François, J.M. (2013). Metabolic phenotypes of *Saccharomyces cerevisiae* mutants with altered trehalose 6-phosphate dynamics. *Biochem. J.* 454, 227–237.

Weihls, C., Ligges, U., Luebke, K., and Raabe, N. (2005). klaR Analyzing German Business Cycles. In Data Analysis and Decision Support, D. Baier, R. Decker, and L. Schmidt-Thieme, eds. (Berlin/Heidelberg: Springer-Verlag), pp. 335–343.

Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58, 109–130.