

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE QUÍMICA E BIOQUÍMICA



Metabolomic effects of single gene deletions in
Saccharomyces cerevisiae

João Mendonça da Luz

Mestrado em Bioquímica
Especialização em Bioquímica

Dissertação orientada por:
Doutor Carlos Cordeiro
Doutora Marta Sousa Silva

Resumo

Saccharomyces cerevisiae é um organismo modelo com cerca de 6000 genes. A maior parte destes genes podem ser eliminados sem comprometer a viabilidade da levedura, sendo uma vasta fração destas mutações silenciosas, não produzindo um fenótipo aparente observável. Mudanças fenotípicas associadas a muitas mutações podem apenas ser observadas com o crescimento em certos meios de cultura ou sob determinadas condições de *stress*. No entanto, variações significativas ocorrem no metabolismo intracelular das células mutadas, particularmente se estas mutações estiverem associadas a vias metabólicas chave. Os estudos destas variações normalmente envolvem a caracterização de atividades enzimáticas ou a quantificação de um pequeno número de metabolitos para obter uma pequena fração do metabolismo específico de uma estirpe mutante.

O desenvolvimento de plataformas analíticas que permitem a análise de perfis de metabolitos em grande escala, particularmente baseadas em espectrometria de massa, tem contribuído para uma caracterização mais completa do metaboloma de um organismo. A utilização de instrumentos de extrema resolução, tais como o espectrómetro de massa de ressonância ciclotrónica de ião com transformada de Fourier (FT-ICR, *Fourier-transform ion cyclotron resonance*), permite a deteção de milhares a dezenas de milhares de compostos e os mais recentes permitem até resolver a estrutura isotópica fina molecular, sendo particularmente relevantes na análise discriminatória de compostos com base em perfis químicos complexos.

Neste trabalho foi seguida uma abordagem de metabolómica global (também denominada *untargeted*) baseada em FT-ICR-MS para o estudo do impacto de deleções de um só gene em leveduras da espécie *Saccharomyces cerevisiae*. Para este efeito, cinco estirpes isogénicas desta espécie foram analisadas. Além da estirpe de referência, foram analisadas três estirpes mutantes com deleção de um gene relacionadas com o catabolismo do metilglioxal, um composto dicarbonilo, muito reativo e citotóxico implicado em diversas condições patológicas. Duas destas estirpes mutantes apresentavam deleções nos genes *GLO1* e *GLO2*, que codificam para os dois enzimas do sistema dos glioxalases, respetivamente o glioxalase I e o glioxalase II. Este sistema catalisa a degradação do metilglioxal de uma forma dependente do glutatióno. A terceira estirpe relacionada com o catabolismo do metilglioxal, apresentava uma deleção no gene *GRE3*, que codifica para o enzima aldose redutase da levedura, o principal atuante num processo alternativo de eliminação do metilglioxal que não depende do glutatióno. Finalmente, uma outra estirpe, deficiente no gene *ENO1*, que codifica para o enzima enolase 1, relacionado com a glicólise, foi também analisada como controlo.

As estirpes foram crescidas em iguais condições, tendo sido analisado o seu crescimento a 600nm e posteriormente o seu metaboloma por FT-ICR-MS. Não foi observada alteração de fenótipo de crescimento, tendo as cinco estirpes apresentado curvas de crescimento extremamente semelhantes, atingindo todas a fase estacionária de crescimento ao fim de 10 a 12 horas.

A extração dos metabolitos de todas as estirpes, em fase estacionária de crescimento, foi efetuada utilizando uma mistura de metanol/água (1:1) e os diferentes extratos foram de seguida analisados por FT-ICR-MS em modo positivo de ionização por *electrospray*. As listas de picos dos espectros obtidos foram depois alinhadas e utilizadas para a identificação dos metabolitos, utilizando bases de dados metabolómicas humana e de levedura, e para a obtenção das fórmulas de composição elementar, previstas com base numa série de regras heurísticas.

Realizaram-se depois contagens do número de metabolitos em cada amostra e em cada estirpe, e contruiu-se um diagrama de Venn com a distribuição dos números de metabolitos comuns e exclusivos para cada estirpe. Analisaram-se ainda as naturezas químicas das moléculas em cada estirpe, para as quais tinha sido possível prever uma fórmula química, construindo-se diagramas de Van Krevelen e um gráfico de séries de composição química.

Três métodos de análise estatística multivariada foram aplicados aos dados de metabolómica. Estes foram a análise de componentes principais (PCA – *principal component analysis*), a análise de agrupamento hierárquico aglomerativo (HCA- *hieararchical clustering analysis*) e a análise discriminatória por regressão de mínimos quadrados parciais (PLS-DA – *partial least squares discriminant analysis*). Os primeiros dois são métodos não-supervisionados, o que significa que não é considerada a existência de grupos previamente definidos pelos quais as amostras se distribuem (neste caso as estirpes). Isto permite-lhes fazer uma separação das amostras com base numa medida global de semelhança entre elas, assegurando que os resultados refletem o perfil químico das amostras. Já o terceiro, PLS-DA, é um método supervisionado que pretende maximizar a covariância entre grupos previamente definidos. Isto leva a uma separação que pode não refletir necessariamente as maiores diferenças entre as amostras, visto que é dada uma maior importância a algumas variáveis (metabolitos) de modo a permitir uma melhor separação entre os grupos pré-definidos (estirpe), independentemente de esses grupos corresponderem ou não à melhor forma de separar as amostras. No entanto, o PLS-DA é útil pois permite a identificação das variáveis que mais contribuem para a separação.

Neste trabalho, os dois métodos não supervisionados (PCA e HCA) demonstraram que era possível distinguir as estirpes com base nos seus perfis metabólicos, visto que amostras pertencentes à mesma estirpe apresentaram consistentemente um maior grau de semelhança metabólica entre elas do que com amostras pertencentes a estirpes diferentes. Além disso, revelaram-se também a existência de semelhanças entre as duas estirpes mutantes relacionadas com o sistema dos glioxalases (Δ GLO1 e Δ GLO2).

A aplicação do método PLS-DA, supervisionado, permitiu maximizar a separação entre as estirpes, o que se revelou extremamente semelhante às separações realizadas pelos dois métodos não-supervisionados. Esta concordância indica que a principal causa para as diferenças metabólicas entre as diferentes amostras se relaciona com a diferença de um só gene entre as estirpes, uma vez que a maximização da variância entre todas as amostras produz resultados semelhantes à maximização da covariância entre as estirpes. Através da análise das pontuações de importância da variável na projeção (*VIP scores – variable importances on projection scores*) calculadas para a separação por PLS-DA, identificaram-se os metabolitos que mais contribuíram para a separação, tendo o glutathione (GSH) emergido como o composto de maior importância, seguido de vários outros que apresentam, na generalidade, uma distribuição de abundâncias relativas semelhante.

A asserção da importância do glutathione está em concordância com os níveis de semelhança metabólica verificados pelos métodos de análise estatística não-supervisionados. O glutathione apresenta uma menor abundância relativa nas estirpes com deleções em genes que codificam para os enzimas do sistema dos glioxalases, visto estes enzimas serem essenciais para a regeneração dos níveis desta molécula. Assim sendo, e tendo em consideração a identificação do glutathione como o composto mais importante para a separação por PLS-DA, a qual é extremamente semelhante às separações pelos métodos não-supervisionados, é possível teorizar que as semelhanças verificadas entre as estirpes relacionadas com o sistema dos glioxilases são em larga parte devidas ao impacto da diminuição dos níveis de glutathione nas células.

Com esta abordagem de metabolómica *untargeted* baseada em FT-ICR-MS, foi possível distinguir entre cinco estirpes de levedura que diferiam umas das outras em apenas um gene e que não apresentavam quaisquer diferenças fenotípicas observáveis quando crescidas em condições normais.

Palavras chave: metabolómica; *Saccharomyces cerevisiae*, FT-ICR-MS; metilglioxal; glutationo.

Abstract

Saccharomyces cerevisiae is a model eukaryote with around 6000 genes. Most of these genes can be deleted without compromising yeast viability, with a vast fraction of these mutations being silent and not producing an apparent observable phenotype. Phenotypic changes associated with many mutations may only be observed in specific growth media or under certain stress conditions. Nevertheless, significant variations occur in the intracellular metabolism of mutated cells, particularly if these mutations are associated with key metabolic pathways. The studies that reveal these variations usually involve the characterization of an enzyme activity or the quantification of a small number of metabolites to obtain “metabolic snapshots” for a specific yeast mutated strain.

The development of analytical platforms allowing for the analysis of large metabolite profiles, particularly based on mass spectrometry, has contributed to a more thorough characterization of the organism’s metabolome. The use of extreme resolution instruments, like the Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometer, allows the detection of thousands to tens of thousands of compounds and the most recent ones are even able to resolve the isotopic fine molecular structure, being particularly relevant in sample discriminatory analysis based on complex chemical profiles.

An untargeted metabolomics approach based on FT-ICR-MS was applied to study of the impact of single-gene deletions in the yeast *Saccharomyces cerevisiae*. For this purpose, five isogenic strains belonging to this species were analysed. Besides the wild-type strain, we chose three null mutants involved in the methylglyoxal catabolism, a well characterized biochemical system in yeast. These mutants lack the genes coding for the main enzymes related with methylglyoxal catabolism, glyoxalase I, glyoxalase II and aldose reductase. Another strain lacking enolase 1 gene, related to glycolysis, was also analysed as control. All strains were grown under the same conditions, without any alteration in growth phenotype being reported. Afterwards, metabolite extraction was performed and the extracts were analysed through FT-ICR-MS. The identified metabolites were putatively annotated with names (using human and yeast metabolomic databases as reference) and with chemical formulas (predicted based on a set of heuristic rules).

Three multivariate statistical analysis methods were applied to the MS results. These were principal component analysis (PCA), agglomerative hierarchical clustering analysis (HCA) and partial least squares discriminant analysis (PLS-DA).

The two unsupervised methods (PCA and HCA) showed that it was possible to distinguish between the strains based on their metabolic profiles, despite the common genetic background. A higher degree of similarity between samples of the same strain was observed, as expected. Similarities between mutant strains related to the glutathione-dependent pathway of methylglyoxal catabolism were also observed.

The PLS-DA method, supervised, performed a separation between the samples that proved very similar to the ones performed by the two unsupervised methods. Through this method, the metabolites that contributed the most to the separation were identified, with glutathione (GSH) emerging as the compound with the greatest importance.

Through this approach, it was possible to accurately distinguish between five yeasts strains which differed from other solely in one gene and which did not present any observable phenotypic differences when grown under normal conditions.

Keywords: metabolomics; *Saccharomyces cerevisiae*; FT-ICR-MS; methylglyoxal; glutathione

Index

Resumo.....	I
Abstract	IV
Index.....	VI
Index of Figures	VIII
Index of Tables.....	IX
Abbreviations	X
Chapter 1 - Introduction	1
I. Science in the Age of Omics.....	1
Metabolomics: the apex of omics.....	2
II. Mass Spectrometry for Metabolomics.....	4
FT-ICR-MS: A powerful tool for untargeted metabolomics.....	4
III. The model organism: <i>Saccharomyces cerevisiae</i>	5
Single-gene deletions in yeast.....	6
IV. Methylglyoxal Metabolism	7
V. Objectives.....	9
Chapter 2 - Materials and methods.....	10
I. Yeast strains and conditions.....	10
II. Yeast growth measurements.....	10
III. Metabolite Extraction	11
IV. Mass spectrometry analysis.....	11
V. Data processing, annotation and analysis.....	11
VI. Multivariate statistical analysis	12
VII. Pathway mapping.....	13
Chapter 3 – Results	14
I. Yeast growth curves.....	14
II. Yeast chemical profile	16
III. Metabolite counts and annotation.....	19
IV. Van Krevelen plots and chemical composition series.....	21
V. Multivariate statistical analysis	23
VI. Pathway mapping.....	26
Chapter 4 – Discussion.....	28
Can an untargeted metabolomics approach using FT-ICR-MS distinguish between almost genetically identical yeasts?	28

Interpreting the differences	29
Chapter 5 - Conclusion.....	34
Bibliography.....	35
Annexes.....	41

Index of Figures

Figure 3.1: Absorbance vs time plot of the five yeast strains (a representative growth curve is shown, measured by OD _{600nm} readings).	14
Figure 3.2: Representative mass spectrum of an yeast extract (BY 4741), (top). Internal standard (Leucine-enkephalin) with fine isotopic structure (bottom).	17
Figure 3.3: Venn diagram showing the common and exclusive metabolites to each strain or combination of strains.	20
Figure 3.4: Van Krevelen diagram of the BY4741, ΔGRE3, ΔGLO1, ΔGLO2 and ΔENO1 strains, showing all the metabolites with assigned formulas positioned according to their Hydrogen/Carbon (H/C) and Oxygen/Carbon (O/C) ratios.	22
Figure 3.5: Bar plot showing the distribution of the metabolites with assigned formulas found in each strain by the various chemical composition series.	22
Figure 3.6: Principal components analysis of the fifteen yeast metabolite samples using the first two principal components.	24
Figure 3.7: Hierarchical clustering of the fifteen yeast metabolite samples.....	24
Figure 3.8: Partial least squares discriminant analysis of the fifteen yeast metabolite samples using principal components one and two.	25
Figure 3.9: Glutathione metabolism and methylglyoxal catabolism map. Compounds marked red were putatively identified by untargeted metabolomics; relative concentration bars for the putatively identified compounds are presented. The base map was obtained from KEGG (KEGG - Kyoto Encyclopaedia of Genes and Genomes, accessed on January 2020; (Kanehisa & Goto, 2000)). Methylglyoxal concentration was previously reported for BY4741, ΔGLO1, ΔGLO2 and ΔGRE3 (Gomes et al., 2005). The enzymes coloured in green were already characterized in <i>S. cerevisiae</i> . The enzymes GLO1 (glyoxalase I), GLO2 (glyoxalase II) and GRE3 (aldose reductase) are indicated. Peak intensities can be consulted in Annex 2.	27

Index of Tables

Table 3.1: Average values of the slopes of the log ₂ ABS vs time curves and first derivative peaks of the ABS vs time curves smoothed with the Savitzky-Golay filter.....	15
Table 3.2: Number of metabolites detected through FT-ICR-MS in each of the 15 samples (note: only those present in at least two samples were considered). Number of metabolites found at least once in each strain. Number of metabolites putatively identified by name using the YMDB and HMDB in MetaboScape. Number of metabolites for which a molecular formula was predicted using MetaboScape's SmartFormula algorithm.....	19
<i>Table 3.3: The ten metabolites that contribute the most to the separation between the strains. In the "Relative concentration" column, the five strains appear in the following order: BY4741-ΔGLO1-ΔGLO2- ΔGRE3- ΔENO1. 100% corresponds to the highest intensity signal for that metabolite; ■ 75-100%; ■ 50-75%; ■ 25-50%; ■ >25%; □ absent from the strain. Peak intensities can be consulted in</i>	

Abbreviations

FT-ICR-MS: Fourier-transform ion cyclotron resonance mass spectrometry

MS: Mass spectrometry

GLO1: Glyoxalase I

GLO2: Glyoxalase II

GRE3: NADPH-dependent Aldose reductase (from yeast)

ENO1: Enolase 1

ΔGLO1: Glyoxalase I null mutant

ΔGLO2: Glyoxalase II null mutant

ΔGRE3: Aldose reductase null mutant

ΔENO1: Enolase null mutant

ABS: Absorbance

YMDB: Yeast Metabolome Database

HMDB: Human Metabolome Database

KEGG: Kyoto Encyclopedia of Genes and Genomes

HCA: Hierarchical clustering analysis

PCA: Principal component analysis

PC1: Principal component 1

PC2: Principal component 2

PLS-DA: Partial least squares discriminant analysis

VIP: Variable importance on projection

Chapter 1- Introduction

I. Science in the Age of Omics

Ever since the introduction of the word “genomics” in 1986 (Kuska, 1998), and accelerating rapidly since the mid-1990s, the life sciences have experienced a proliferation of various “-omics” disciplines which take advantage of modern analytic and data-processing methods to examine large pools of biological molecules at the same time. These scientific areas include now, in addition to its originator, genomics, transcriptomics, proteomics, metabolomics, glycomics, lipidomics, and much more.

The continuous growth and diversification of these approaches over the course of the last three decades must be understood within the context of a paradigm shift in the biological sciences. As the evolution of technology made it possible, scientists began to distance themselves from the reductionist paradigm, which sought to isolate a single molecule or process and study it independently from the larger system, and move towards a more systems-based approach which tries, as much as possible, to consider the entire cell or organism as a whole made out of intrinsically interdependent parts (Kitano, 2002)

Understandably, most of the early excitement over this new paradigm was centred around genomics, and specifically the possibility of mapping the entire human genome (a project which reached completion as early as 2003). However, it soon became clear that, while the genome might be the origin of all biological information, it was not going to be giving us all the answers at the present time.

The complex regulatory processes involved in DNA translation, mRNA splicing and even post-translational modification of proteins make it difficult to assign a particular end product, let alone a biological function, to a gene. This creates the need to look for answers in other levels of the biological information transmission chain, such as RNA, proteins and metabolites, so that we may gain an understanding of how a certain phenotype comes to be.

Starting in the mid-1990s, transcriptomics emerged as a field which liberated the researcher from the ambiguities of regulation at the levels of transcription and splicing. By taking a snapshot of the sum of all mature RNA transcripts, the transcriptome, a closer relation with molecular end products and biological functions can be established. However, the transcriptome is still a long way from the phenotype, and a single RNA sequence detected in a transcriptomics experiment may still lead to the formation of different proteins. The first major reason for this is the fact that a single domain (polypeptide) may be found in more than one (quaternary structure) protein. The second major reason is the commonality of post-translational modifications (Khoury et al., 2011), which are not accounted for in the transcriptome. Finally, while transcriptomics may be performed quantitatively, thereby determining a gene’s transcription rate, the quantity of transcript doesn’t necessarily relate to protein activity, thus making it difficult to relate RNA with biological function (Gygi et al., 1999).

Proteomics presents itself as a solution capable of addressing some of these limitations. By focusing on the effectors of biological function (the proteins), it is possible to establish a much closer relationship with the phenotype, as both qualitative and quantitative studies can be performed to obtain a clear picture of the metabolic pathways inside the cell and their (highly regulated) activity. However, the technical challenges are significant, especially those that pertain to the proteome’s dynamic range. Furthermore, knowledge of the existence of a certain protein, and even possession of one or several structural models constructed with data from x-ray crystallography, Cryo-EM or NMR experiments, does not necessary

equate to a full understanding of its biological activity. Proteins oftentimes behave differently *in vivo* than they do *in vitro* as the presence of other, often unknown, biological factors may significantly regulate their activity, possibly even make them take on new substrates. While much outstanding work has been done for the purpose of studying proteins as close to their native conditions as possible (e.g. Native MS), there is no denying the fact that a prediction of protein function based on its physical and chemical characteristics and/or in its behaviour in *in vitro* assays will always remain just that, a prediction. In the end, even the characterization of large protein and ligand complexes falls short to this objective, given that it's impossible to preclude the possibility that different complexes and different interactions may occur *in vivo* with the same proteins.

Thus, to complement the data obtained from proteomics (which in turn came to complement the data obtained from genomics and transcriptomics), it's necessary to take it one step further and look not only at the effectors of biological activity, but at its products and intermediaries.

Metabolomics: the apex of omics

Unlike genes, mRNA and proteins, metabolites are direct chemical signatures of biological activity, and thus are the class of biomolecules with the closest relationship with the phenotype (see Figure 1.1), with no further regulated “downstream” processes to introduce additional variables.

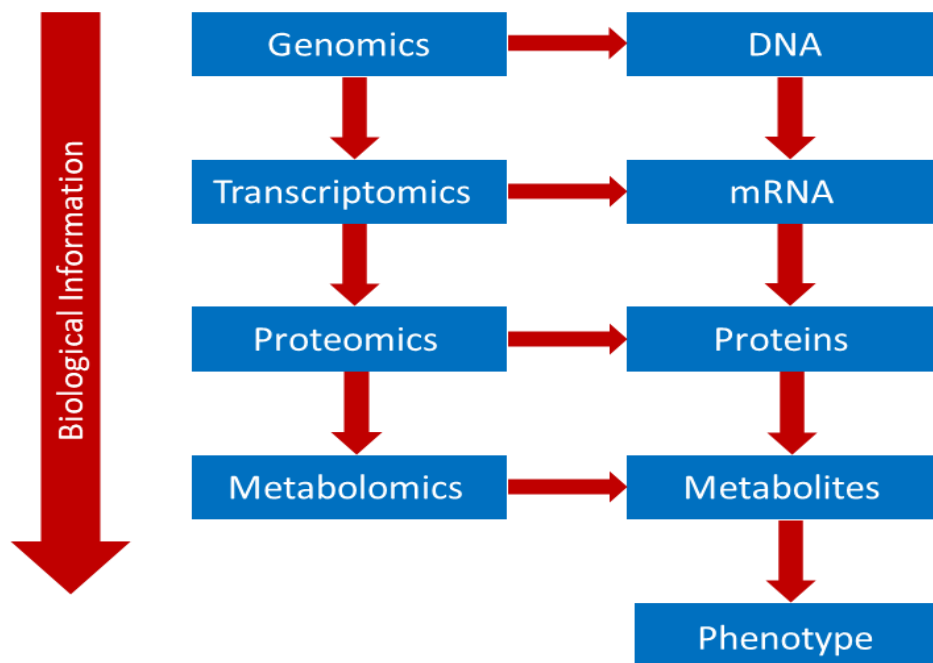


Figure 1.1: Metabolomics is closer to the phenotype than other 'omics'.

This makes them immensely valuable for scientific, clinical and biotechnological purposes, as a metabolomic “snapshot” of the cell can be used to make immediate predictions related to the phenotype without resorting to reductionist thinking. Additionally, in comparison with the genome, transcriptome and proteome, changes are amplified in the metabolome, which translates to greatly increased

sensitivity, if at the cost of making the researcher have to account for the metabolome's extremely dynamic state (Hollywood et al., 2006).

The idea that cells and organisms possessed a “metabolic profile” of sorts predates the omics era by far and can be said to go back to the 1940s, although quantitative studies had to wait until the 1970s for technological reasons. The development and maturation of both gas chromatography and coupling gas chromatography with mass spectrometry proved decisive in the early stages of metabolomics. However, despite the pioneering nature of these studies for their time, they were focused on the analysis of bodily fluids (such as blood and urine) and involved measuring only a relatively small number of metabolites (Gates & Sweeley, 1978).

Studies measuring large numbers of intracellular metabolites within the context of a systems-based approach to biochemistry would come only at around the turn of the century, as an extension of the already established genomics and proteomics (Fiehn, 2001; Roessner et al., 2001; Trethewey, 2001).

It is important to distinguish between the targeted and untargeted approaches to metabolomics, as these play complementary roles within the field and are adequate for different purposes. The term “targeted metabolomics” refers to experiments where only a specific set of metabolites, pre-identified as being of interest to the study at hand, are measured. This is adequate for studies where the researcher is only interested in a small set of molecules or in a particular pathway and has a well-defined starting hypothesis to work with, situations where the simplicity of only having to detect a short range of metabolites and of not producing very large amounts of data are appreciated. Examples include pharmacokinetic studies of drug metabolism (where only the disappearance of a molecule and its metabolic products matter) and studies intent on determining the effects of certain therapeutics or genetic modifications on the activity of a specific enzyme. While this approach certainly has its place in metabolomics, it risks incurring in reductionist thinking. “Untargeted metabolomics” on the other hand is a term which refers to experiments where the aim is to take as many simultaneous measurements of metabolites as possible. This has the effect of producing a much more global and complete metabolomic “snapshot” of the cell, making it ideal for studies without a clear hypothesis or dealing with less-known sections of metabolism or sections with many possible ramifications (Patti et al., 2012).

However, while it is undoubtedly the approach which better lines up with the new “systems” paradigm of the life sciences, it poses two significant technical challenges. Firstly, there is the difficulty of simultaneously measuring a wide range of metabolites with varied masses and chemical natures, requiring a technique that combines versatility and sensitivity with a high resolution and accuracy. And secondly, there is the difficulty in treating very large amounts of data and accessing information related to a wide variety of molecules.

Thus, in order to truly take the most out of the untargeted metabolomics approach, it's imperative to develop experimental procedures which take advantage of extremely high-resolution and high sensitivity techniques.

II. Mass Spectrometry for Metabolomics

The varied techniques that fall under the term of “Mass Spectrometry” (MS) have a long history dating back to the cathode rays experiments of the early 20th century (Aston, 1919) but had a relatively slow development. Initially the purview of physics (being used to calculate atomic masses), it is now found mostly in the domains of chemistry and biology (Griffiths, 2008). Together with Nuclear Magnetic Resonance (NMR) techniques, they form the core of modern metabolomics research (Zhang et al., 2012).

Mass Spectrometry is the measurement of the mass-to-charge ratio (m/z) of molecular ions, generally with the purpose of determining the chemical composition of an ionized sample. In many variants of the method, this is coupled with a preceding separation step, such as gas chromatography (GC), or high-performance liquid chromatography (HPLC).

The development of GC-MS was crucial in bringing about the onset of quantitative studies involving metabolites, and this importance has carried on to the age of omics. However, it has the obvious problem of excluding all non-volatile compounds. HPLC-MS and other techniques employing a liquid chromatography separation step are able to overcome this limitation, which has contributed to their growth in popularity (Zhang et al., 2012). However, even in this technique, the relatively strenuous purification process still inevitably leads to the exclusion of a significant number of metabolites (Gika et al., 2014).

FT-ICR-MS: A powerful tool for untargeted metabolomics

The ideal solution to the limitation posed by the exclusion of naturally occurring metabolites from MS spectra due to their retention in the previous purification step is to eliminate said purification step entirely. This also simplifies and dramatically speeds up the process, allowing for a much larger number of samples to be sequentially analysed in a short period of time.

However, it is far beyond the reach of the average mass spectrometer to produce results with enough resolution that the peaks corresponding to the various metabolites can be easily distinguished in a spectrum resulting from an untreated sample.

Fourier Transform Ion Cyclotron Resonance Mass Spectroscopy (FT-ICR-MS) is the only mass spectrometer capable of offering that amount of resolving power.

The origins of FT-ICR-MS date back to 1974, when Alan Marshall and Melvin Comisarow became the first researchers to apply Fourier Transform (FT) to Ion Cyclotron Resonance (ICR) data (Comisarow & Marshall, 1974). Fourier Transform is a mathematical transformation which decomposes a continuous function (such as an electromagnetic signal) into a series of frequencies. By the mid-1970s, it was already used to deconvolute the data in techniques such as Infrared spectroscopy (FTIR) and NMR spectroscopy (FT-NMR). Marshall and Comisarow’s great insight was realising that it could also be applied to ICR and harnessed into a functioning mass spectrometry technique (Griffiths, 2008). In its turn, Ion Cyclotron Resonance refers to the circular movement of ions in a uniform and static magnetic field. This movement was the basis of the “cyclotron” particle accelerator developed by Ernest Lawrence decades prior (Lawrence & Livingston, 1931), and from which the effect takes its name. The frequency, radius, velocity and energy of FT-ICR can be described as a function of mass, charge and magnetic field

strength. Thus, if one is able to excite the moving ions into a larger (and thus detectable) cyclotron radius, it is possible to measure cyclotron frequencies through detection of an oscillating signal in two parallel electrodes, which is then decomposed using the Fourier transform (Comisarow & Marshall, 1974).

The relationship between the angular cyclotron frequency (ω_c) and the m/z ratio when an ion is subject to a uniform magnetic field (B) can be described by the following equation (1.1):

$$\omega_c = \frac{zB}{m}$$

Equation 1.1: Angular cyclotron frequency of an ion with mass 'm' and charge 'z' when subject to uniform magnetic field B.

The high level of accuracy that is possible to achieve in measuring the cyclotron frequency allows FT-ICR-MS to offer much higher mass resolution and mass accuracy than other types of mass spectrometry. Furthermore, the technique produces spectra where a wide variety of m/z values are observable, thus allowing for the simultaneous identification of compounds of very different masses (Marshall & Chen, 2015).

Due to its extreme resolution and ultra-high mass accuracy, FT-ICR-MS is the most powerful analytical technique for an untargeted metabolome characterization. Additionally, FT-ICR-MS requires no chromatographic step or special preparation and thus offers a significant time advantage when used for untargeted metabolomics studies involving a large number of samples.

III. The model organism: *Saccharomyces cerevisiae*

One of the most ancient tools of biotechnology, *S. cerevisiae*, known as “the baker’s yeast”, is used for far more than just baking. Producing an extremely high alcoholic yield through its fermentation process, the yeast is used in brewing, winemaking and the production of just about any alcoholic beverage.

Since the species has always appeared associated to human activity across the whole of recorded history (it is believed to have been originally obtained from grapes), it’s extremely likely that its fermentative properties are at least partially the result of extensive artificial selection. Genetic diversity studies serve to corroborate this notion, as they show a strong correlation between the evolution of various yeast strains and human activity (Legras et al., 2007).

Possessing a dynamic lifecycle, *S. cerevisiae* is capable of growing in the form of both haploid and diploid cells. Thus, it often undergoes processes of meiosis and gamete fusion, allowing it to reproduce sexually and rip out the evolutionary benefits of a high degree of genetic recombination. Mating types are designated MAT α and MAT a , referring to their respective alleles of the MAT locus. Haploid cells of complementary mating types are fused into a diploid cell, which can later undergo meiosis and release new haploid cells through sporulation. There exist both homothallic and heterothallic strains of *S. cerevisiae*, with the first being able to reproduce sexually without the initial presence of both mating types, through the means of having certain cells change their mating type through DNA recombination. Regardless, both diploid and haploid cells regardless of type are capable of undergoing mitotic division, thus reproducing asexually (in this case through budding), (Herskowitz, 2017).

In addition to the biotechnological utilization of its fermentative capacity, *S. cerevisiae* has also become an immensely important organism for scientific research, as it has been the model organism for the study of the eukaryotic cell for many decades. Its short duplication time of about 90 minutes makes cell cultures of this species extremely easy to grow and maintain, thus allowing researchers to obtain large populations of cells in only a few hours of growth and in a small culture volume (Hanson, 2018). Furthermore, *S. cerevisiae* has the particularity of lending itself to the genetic manipulation with surprising ease for an eukaryotic organism, with the first reports of successful DNA transformation dating from the 1960s (Oppenoorth, 1962), and with reliable approaches having been developed by the 1970s (Beggs, 1978). This stimulated interest in yeast genetics, eventually leading to the sequencing of the entire *S. cerevisiae* genome by the mid-1990s, making it the first eukaryotic organism for which this was achieved (Goffeau et al., 1996). The high level of time and cost efficiency, genetical tractability, ease of transformation and high level of homology with human and other eukaryotic genomes combine to make *S. cerevisiae* the premier organism for studies in eukaryotic cells (Hanson, 2018).

Single-gene deletions in yeast

The high-level of genomic similarity between *S. cerevisiae* and humans, coupled with the fact that yeasts, as eukaryotes, possess much of the same molecular machinery involved in genetic expression as human beings, makes the baker's yeast the ideal model organism to investigate the effects of point mutations (such as single gene deletions) on the cell.

Studies of the effects of single gene deletions on the whole cell metabolism have been impaired by the inherent difficulty of measuring the wide array of disparate and oftentimes extremely subtle changes in biochemistry that this type of mutation may cause. Most studies so far have focused on growth, largely considered to be the "default" observable phenotype in yeast studies, as the sole indicator of change. In 2001, Raamsdonk and co-workers proposed an approach entitled Functional Analysis by Co-Response in Yeast (FANCY), which centred on the utilisation of metabolomic data to ascertain the functional impact of a point mutation (Raamsdonk et al., 2001). This approach was limited by the insufficient analytic capacity offered at the time by Nuclear Magnetic Resonance technology and by the fact that only a very a small portion of the metabolome could be analysed at one time.

The development of analytical platforms allowing the analysis of large metabolite profiles, contributed to a thorough characterization of an organism's metabolome. Nuclear magnetic resonance still is a widely used technique in metabolomics for its ability to quantify up to one hundred of the most abundant intracellular metabolites, particularly sugars, amino acids and intermediates of the tricarboxylic acid cycle (Kostidis et al., 2017). On the other hand, an increase in the use of mass spectrometry (MS) in metabolomics has been reported, particularly the use of extreme resolution instruments like the Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometer. These instruments allow for the detection of thousands to tens of thousands of compounds and the most recent ones are even able to resolve the isotopic fine molecular structure (Aharoni et al., 2002; Brown et. al, 2005; Maia et al., 2016; Nikolaev et. al, 2012), and may be used to study the effects of single gene deletions on the entirety of the cell metabolome.

IV. Methylglyoxal Metabolism

Being an inevitable side product of cell metabolism, methylglyoxal ($\text{CH}_3\text{C}(\text{O})\text{CHO}$) is a ubiquitous chemical compound found in all cells. While it can be formed through a variety of enzymatic and non-enzymatic pathways, the largest contributor to methylglyoxal concentration in eukaryotic cells is undoubtedly the glycolytic bypass resulting from the non-enzymatic decomposition of glyceraldehyde-3-phosphate (GAP) and dihydroxyacetone phosphate (DHAP), (Sousa Silva et al. , 2013). The two triose phosphates have, at a physiological pH, a high propensity to lose their α -carbonyl protons, thus forming an enediolate phosphate, which then may easily lose its phosphate group to form methylglyoxal (Richard, 1993; Richard, 1984).

Another common pathway involves the same enediolate phosphate dephosphorylation reaction, except the intermediate slips from the active site of TIM (triose phosphate isomerase/GAPketol-isomerase, EC.5.3.1.1), an enzyme which catalyses the interconversion between glycolysis' two triose phosphates using enediolate phosphate as an intermediate, rather than forming non-enzymatically (Richard, 1991).

Also noteworthy are the formation routes stemming from the catabolism of L-threonine (mediated by the enzyme SSAO, or semicarbazide-sensitive amine oxidase, EC 1.4.3.6.), (Lyles & Chalmers, 1992), the myeloperoxidase-dependent (donor:hydrogen-peroxide oxidoreductase, EC 1.11.1.7) oxidation of ketone bodies (Aleksandrovskii, 1992) and acetone oxidation by cytochrome P450 [substrate, reduced flavoprotein:oxygen oxidoreductase (RH-hydroxylating or -epoxidizing), EC 1.14.14.1].

Methylglyoxal is highly cytotoxic, as it possesses a reactive aldehyde group which modifies amino groups in a process known as glycation. Proteins are a common target of glycation, and this modification can often lead to loss of function. Other molecules possessing amino groups, such as nucleic acids and phospholipids, may also be liable to this type of modification (Sousa Silva et al., 2013).

Advanced Glycation End-Products (AGEs) are molecules which have become heavily glycated and have been shown to play a role in aging (Chaudhuri et al., 2018) as well as in the development of a variety of pathological conditions resulting from diabetes mellitus (Singh et al., 2014). Additionally, they have also been implicated in Alzheimer's disease (Srikanth et al., 2011), atherosclerosis (Yamagishi & Matsui, 2018), multiple sclerosis (Wetzels et al. , 2017) and more.

Comparatively with other glycation agents (e.g. glucose) methylglyoxal is much more reactive *in vivo*. Its targets are mostly guanidine groups of arginine residues, with which it reacts to form a hydroimidazolone derivate called MG-H, as well as argpyrimidine and THP (tetrahydropyrimidine). Alternatively, it can also react with lysine residues to form CEL [N ϵ -(carboxyethyl)lysine] and MOLDS (methylglyoxal-lysine dimers) or lead to the formation of cross-links between arginine and lysine residues [MODIC, methylglyoxal-derived imidazolium cross-linking]. The term MAGE (Methylglyoxal Advanced Glycation End-Products) specifically refers to proteins and other molecules glycated by methylglyoxal (Sousa Silva et al., 2013).

Several detoxification pathways involved in the elimination of methylglyoxal have been described. One of these pathways is the glyoxalase system. Typically thought of as a metabolic pathway composed of two enzymes, glyoxalase I (GLO1, lactoylglutathione methylglyoxallyase; EC 4.4.1.5) and glyoxalase II (GLO2, hydroxyacylglutathione hydrolase, EC 3.1.2.6), this system poses a significant challenge to modern biochemistry.

While its catalytic function is well defined – the glyoxalase system catalyses the formation of D-lactate from hemithioacetal (formed non-enzymatically from methylglyoxal and glutathione), with GLO1 first converting the substrate into an intermediate, S-(D)-Lactoylglutathione, and GLO2 then catalysing its hydrolysis into D-Lactate and Glutathione molecules – certain experimental findings are difficult to explain in light of this model (Sousa Silva et al., 2013). Factors casting doubt on the nature of the glyoxalase system as a linear metabolic pathway formed by two highly specific enzymes acting on a sequential manner include: the existence in *Escherichia coli* of a third glyoxalase enzyme (GLO3, EC 4.2.1.130) which catalyses the production of D-lactate from methylglyoxal without needing glutathione or any other co-factor (thus dispensing the formation and degradation of both hemithioacetal and S-D-Lactoylglutathione) (Misra et al., 1995); the apparent inconsequentiality of glyoxalase II deficiency (Rae et al., 1991; Valentine et. al, 1970); the phenotypically silent nature of null mutations of the glyoxalase enzyme genes (Gomes et al., 2005); and the absence or incompleteness of the system in some protozoan parasites (Sousa Silva et al., 2012).

Other than the glyoxalase system, the most prominent pathway for methylglyoxal elimination is its NADPH-dependent reduction to 1,2-propanediol in a two-step reaction catalysed by aldose reductase (GRE3; aldehyde reductase, EC 1.1.1.21), (Vander Jagt & Hunsaker, 2003).

A simplified scheme of both glyoxalase and aldose reductase pathways is shown in Figure 1.2.

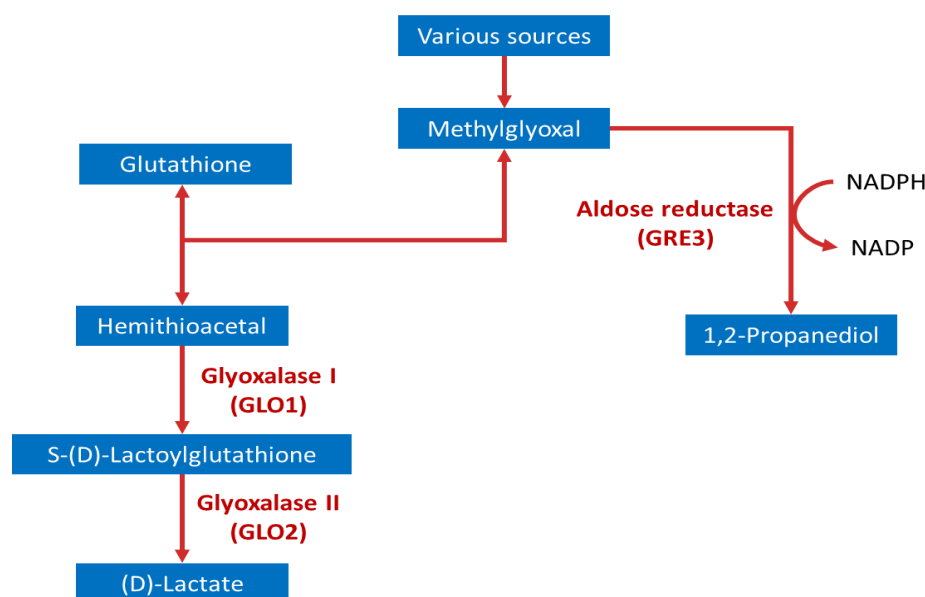


Figure 1.2: The two main pathways of methylglyoxal catabolism in eukaryotes (Sousa Silva et al., 2013).

Given the likelihood of a more complex metabolic network around methylglyoxal than what is currently known, and especially given the paradoxically silent nature of mutations deleting some of enzymes involved in its catabolism, this is one of the problems in biochemistry which may benefit the most from FT-ICR-MS based untargeted metabolomics analysis.

V. Objectives

The goal of this dissertation is to use Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR-MS) to discriminate single-gene yeast mutants, revealing a wide array of metabolic differences between them. The four selected null mutants share the same genetic background of the wild type strain (BY4741), except for the deletion of a single specific gene coding for an enzyme involved in methylglyoxal metabolism. The work followed an untargeted metabolomics approach using FT-ICR-MS, thus dispensing the need for any prior knowledge of the metabolic pathways which differ between the yeasts.

The goal was to demonstrate that this procedure could be useful to uncover new consequences of point mutations and reveal new facets of cell metabolism. I chose to work on this topic because, as explained in previous sections of this introduction, I believe that metabolomics holds the key for a more complete understanding of gene function in biology, as it provides the “missing link” between the preceding “omics” sciences (genomics, transcriptomics and proteomics) and the observable phenotype. More specifically, I consider untargeted metabolomics approaches based on extremely high-resolution analytic techniques, computational treatment of biological data and comprehensive online databases to hold the most potential in the mid to near term future of the biological sciences.

Chapter 2- Materials and methods

I. Yeast strains and conditions

Saccharomyces cerevisiae strains were obtained from the EUROSCARF collection (Frankfurt, Germany) and were kept at -80 °C. The yeast strains selected for this study were the BY4741 wild type strain (genotype: MATa; his3 Δ 1; leu2 Δ 0; met15 Δ 0; ura3 Δ 0) and the single-gene deletion mutants Δ GLO1 (isogenic to BY4741 with YML004c::KanMX4; lacks the glyoxalase I gene), Δ GLO2 (isogenic to BY4741 with YDR272w::KanMX4; lacks one of the two genes encoding glyoxalase II, the only one expressed in glucose), Δ GRE3 (isogenic to BY4741 with YHR104w::KanMX4; lacks the aldose reductase gene), and Δ ENO1 (isogenic to BY4741 with YGR254w::kanMX4; lacks one of the two genes encoding Enolase). Out of the four mutant strains, Δ GLO1, Δ GLO2 and Δ GRE3 are, when grown in a glucose medium, deficient in enzymes which participate in methylglyoxal catabolism, and consequently in the prevention of glycation (Bito et al., 1997; Cordero, 2001; Kimura, 1996), while the Δ ENO1 is not expected to have a significant effect on the levels of the glycolytic enzyme Enolase (Phosphopyruvate hydratase, EC 4.2.1.11), since cells grown on glucose contain 20-fold more ENO2 than ENO1, and thus serves as a control (Mcalister & Holland, 1982).

After being removed from their cold storage, the five strains were grown overnight in YPD medium (0.5% (w/v) yeast extract, 1% (w/v) peptone and 2% (w/v) D-glucose) agar slopes (containing 2% (w/v) agar) at 30 °C.

Following their initial growth in solid medium, the yeast cultures in agar slopes, were kept at 4 °C until further use. Then, a few colonies of each strain were transferred to YPD liquid medium (20 mL inside 100 mL Erlenmeyers), where they were grown for 18 h at 30°C and with a constant agitation of 160 rpm, in an IKA KS 3000 incubator.

The yeast cultures obtained at the end of the growth in liquid medium (at the stationary phase of growth) were then used to measure the growth phenotype of each strain and to extract the metabolites for FT-ICR-MS analysis.

II. Yeast growth measurements

Yeast growth assays were performed by diluting each of the previously mentioned liquid cultures to an absorbance of approximately 0.1 (at a wavelength of 640 nm), in a final volume of 2 mL. Cells were grown at 30 °C and the growth was followed at 640 nm for 14 h, with absorbance measurements every 150 s, in an Agilent 8453 diode array spectrophotometer (with temperature control and magnetic stirring).

The slopes of the log₂ *ABS* vs time plots were used as stand-ins for the actual duplication rates of the yeasts. This simplification was made possible by the fact that, in this study, the only purpose of measuring growth is to compare the different strains. Thus, with all measurements made in the exact same conditions for the five different yeasts, the use of an indirect measure of growth is adequate to compare the results.

The highest first derivative values of the $\log_2 ABS$ vs time plots were also taken as measures of the yeasts' growth phenotypes, being determined after smoothing the experimental values (without it, the amplification of data noise caused by derivatization would make the plots impossible to interpret) using the Savitzky-Golay filtering algorithm incorporated in the SciPy Python library (Virtanen et al., 2020) as the `savgol_filter` function. The filter was applied with a moving window size of 5 and a first-degree polynomial.

Four separate growth assays including all the five yeasts were performed for the purposes of this study, each in a different day and using a different original liquid medium culture for the same strain.

III. Metabolite Extraction

Metabolites were extracted from yeast cells collected at stationary phase of growth. Two mL were used to obtain cell pellets through a 5 min centrifugation at room temperature in a VWR Kinetic Energy 26 Joules Galaxy Mini Centrifuge. The supernatant was discarded and the cell pellet was washed twice with 1 mL of distilled water, with the pellets being resuspended through a 1 min vortexation and then again centrifuged during 5 min. The final cell pellets were stored at -80 °C for further metabolite extraction for MS analysis.

Metabolite extraction was performed by resuspending the cell pellets in 1 mL of a methanol (LC-MS grade, Merck, Germany)/water mixture (1:1), followed by three cycles of vortex / incubation on ice (1 min at each step). After a 5 min centrifugation at room temperature, the supernatants, which contained the metabolites, were recovered while the pellets were discarded. The metabolite extracts were immediately processed for FT-ICR-MS analysis.

IV. Mass spectrometry analysis

Metabolite extracts were diluted 1:100 in methanol/water (1:1). The standard leucine-enkephalin (peptide YGGFL, Sigma Aldrich) was added to all samples at a concentration of 0.1 $\mu\text{g/mL}$ and was used for online lock mass calibration ($[M + H]^+ = 556.276575$ Da). Formic acid (Sigma Aldrich, MS grade) was also added at a final concentration 0.1% (v/v). Samples were analysed by direct infusion in a SolariX XR 7-Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer (FT-ICR-MS, Brüker Daltonics, Germany), operating in positive electrospray ionization mode (ESI⁺). Three analysis, with three different extracts originating from three different cultures, were performed for each yeast strain. Samples were infused at 240 mL/min. Mass spectra were acquired in the mass range between 100 and 1200 m/z , with an acquisition size of 4M and 100 transients were accumulated for each sample.

V. Data processing, annotation and analysis

Raw data was analysed with the software MetaboScape 4.0 (Brüker Daltonics, Germany) using the T-ReX (Time aligned Region complete eXtraction) algorithm. All fifteen samples' peak lists were aligned in a single bucket table, with the intensities being normalized to that of the internal standard (leucine-

enkephalin). Metabolite m/z values appearing only in one of the fifteen samples were removed from the dataset as they were most likely to be experimental artifact of some sort, and thus non-informative.

The number of exclusive and common metabolites to each of the five yeast strains was calculated through simple python computing with the pandas data analysis library (<https://pandas.pydata.org/>, Reback et. al, 2020), with the results being used to draw a Venn diagram.

Possible molecular formulas for each mass were predicted using the SmartFormula function included in Bruker's MetaboScape 4.0 software. Parameters were set in accordance with a series of heuristic rules proposed in previous literature (Kind & Fiehn, 2007), with accommodations made to better account for the most likely compositions of the samples given their origin as cell extracts. As appropriate for our varied set of small masses, the minimum number of atoms of each element was set to 1 for hydrogen and carbon, and 0 to all others. The maximum number of atoms for each element was left at 39 for carbon, 72 for hydrogen, 20 for oxygen and nitrogen, 5 for chloride and sulphur, 8 for fluor, 9 for phosphate and 0 for bromide and silicon. Element ratios were allowed to variate between 0.2 and 3.1 for H/C; 0-1.3 for N/C; 0-1.2 for O/C; 0-0.3 for P/C; 0-0.8 for S/C and Cl/C; and 0-1.5 for F/C. A tolerance of 1.0 ppm was given for m/z values, electron configurations were required to be in accordance with the Lewis and Senior Chemical rules and an heuristic element count probability check was performed to filter out formulas with highly unlikely element counts.

The obtained compound formulas were used to plot H/C ratios against O/C ratios in a Van Krevelen diagram to highlight different compound families: lipids (H/C 1.5-2.0, O/C 0-0.3), carbohydrates (H/C 1.5-2.4, O/C 0.7-1), amino acids (H/C 1-2, O/C 0.2-0.8), nucleic acids (H/C 1.1-1.4, O/C 0.3-1), anthocyanins (H/C 0.5-1, O/C 0.4-0.8), and flavonols (H/C 0.6-1.1, O/C 0.3-0.5), (Lu et al., 2015; Petitgonnet et al., 2019). A chemical histogram of the CHO, CHON, CHOS and CHONS elemental composition series for each strain was also built.

Finally, a putative annotation of the metabolites with their respective names was performed in MetaboScape 4.0 using the Yeast Metabolome Database (YMDB), (Ramirez-Gaona et al., 2017), and the Human Metabolome Database (HMDB), (Wishart et al., 2007). The m/z tolerance for the annotation was set as follows: narrow below 0.1 ppm and wide below 1 ppm.

VI. Multivariate statistical analysis

Multivariate statistical analysis was performed with the bucket table from MetaboScape 4.0 (Bruker Daltonics, Germany), using the Python 3 programming language, including the aforementioned SciPy and pandas libraries, as well as NumPy (<https://numpy.org>, Harris et al., 2020) and matplotlib (<https://seaborn.pydata.org>, Hunter, 2007). The matplotlib-based data visualization library seaborn (Waskom et al., 2020) was also used, as was the NumPy, SciPy and matplotlib-dependent scikit-learn library for machine learning in Python (<https://scikit-learn.org/stable/>, Li & Phung, 2011).

Finally, the Metabolinks package (<https://github.com/aeferreira/metabolinks>), based on NumPy, pandas and matplotlib, and designed specifically for analysing data from high resolution mass spectrometry-based metabolomics, was used in the metabolomics data pre-treatment, to represent the data matrix, before the statistical analysis.

Before any statistical analysis was performed, the data was treated to remove the intensity values that were equal to zero and replace them with half of the minimum intensity. This was done to prevent the absence of a metabolite from being overemphasized in multivariate statistical analysis (considering that, when a metabolite is present, intensity values are generally quite high, and thus quite distant from zero) at the expense of smaller but possibly more biological significant differences in intensity between existing metabolites. Furthermore, Pareto scaling was applied to the data in order to ensure that large but proportionally small differences in the intensities of the most intense peaks don't overpower smaller but more proportionally relevant differences in the intensities of less intense peaks. Both of these preliminary data treatments relied on pandas and Metabolinks.

Two unsupervised methods were applied to investigate the metabolic profile differences between yeast samples. Sample Hierarchical Clustering (agglomerative) was performed using the SciPy Python library (`scipy.cluster.hierarchy.linkage`), with a an Euclidean distance (measured with `scipy.spatial.distance.pdist`) and using the average distances method (UPGMA algorithm). Principal Component Analysis (PCA) models were also built, using a function included in the Scikit-learn library (`sklearn.decomposition.PCA`) and only the first two components were used.

Additionally, a supervised technique, Partial Least Squares Discriminant Analysis (PLS-DA) was also performed with a function from the Scikit-learn library (`sklearn.cross_decomposition.PLSRegression`), with 5 components being used. The goal was to identify the compounds that discriminate between yeast samples. Variable Influence on Projection (VIP) scores were calculated with a NumPy-based Python script using the mathematical formulas described in literature (Galindo-prieto et al., 2012), and, based on them, the specific metabolites that contributed the most to the differentiation between the strains were determined.

VII. Pathway mapping

Pathway mapping of some of the putatively identified compounds was performed manually over a pathway map obtained from KEGG using the Mapper-Search & Color Pathway, using the KEGG identifiers and selecting the *S. cerevisiae* organism-specific pathways (KEGG - Kyoto Encyclopaedia of Genes and Genomes, accessed on January 2020; Kanehisa & Goto, 2000). The map was then manually changed and improved, to better reflect the set of metabolites identified in the samples. Colour gradient bars were added to the compounds to represent their relative concentrations in each yeast strain. Relative quantification of the compounds was performed by comparing the average of the peak intensity values in the 3 replicates between all strains; the highest value corresponded to 100% and the relative percentage of the others was calculated proportionally. In the case of methylglyoxal (one of the compounds added to the map) relative concentrations were estimated the same way based on data from previous research on the exact same yeast strains (Gomes et al., 2005).

Chapter 3– Results

I. Yeast growth curves

Five yeast strains, with identical genetic background, were analysed. BY4741 is a commonly used reference strain, developed in 1998 (Brachmann et al., 1998). The other 4 strains are single-gene mutants, isogenic for BY4741, developed by the EUROSCARF project. Δ GLO1, Δ GLO2 and Δ GRE3 lack the main enzymes involved in methylglyoxal catabolism (glyoxalase I, glyoxalase II and aldose reductase, respectively), while the Δ ENO1, lacking enolase I, served as control.

All yeast strains were grown in YPD medium, a complete yeast growing medium. Cultures grown either in large volumes (in Erlenmeyers) or small volumes (in cuvettes, inside a spectrophotometer) showed identical growth behaviour between the strains, with a brief “lag phase” followed by exponential growth (“log phase”) before finally plateauing into a stationary phase (Figure 2.1). All yeast strains reached the stationary phase after 12 hours and none of the single-gene deletion mutants revealed a growth deficiency (Figure 2.1).

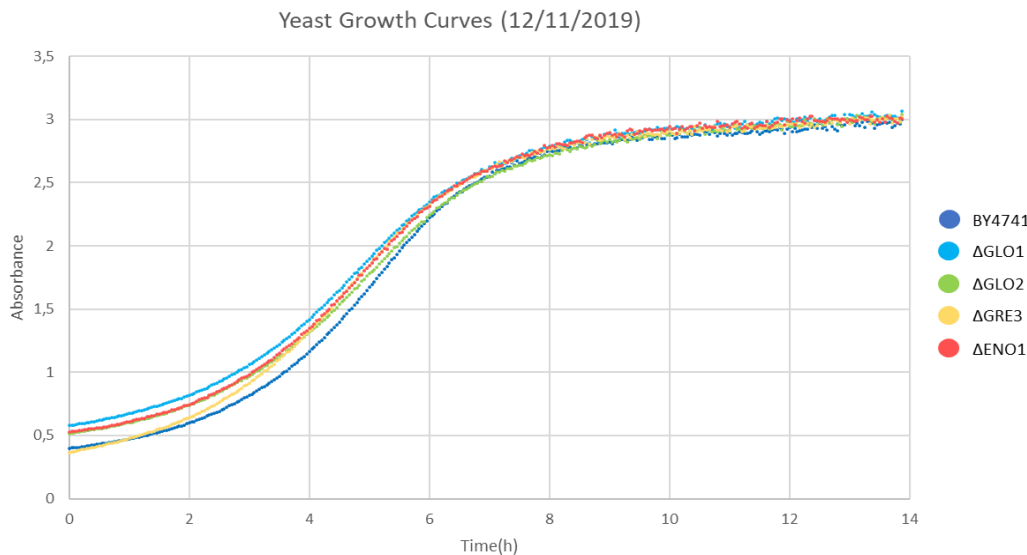


Figure 3.1: Absorbance vs time plot of the five yeast strains (a representative growth curve is shown, measured by OD_{600nm} readings).

The determination of relative growth indicators for each strain proved similarly unproblematic, with the application of a logarithmic scale to the absorbance axis resulting in nearly linear growth curves in the “log” phase, thus allowing for the use of a simple linear regression to determine their slopes (Table 2.1).

Savitzky-golay filtering of the original ABS vs time data was successful in reducing post-derivatization noise to such an extent that, in most cases, it was possible to discern a clear peak in the first derivative of ABS vs time plot. The averages between the first derivative peak values that were successfully determined can all be found in Table 2.1.

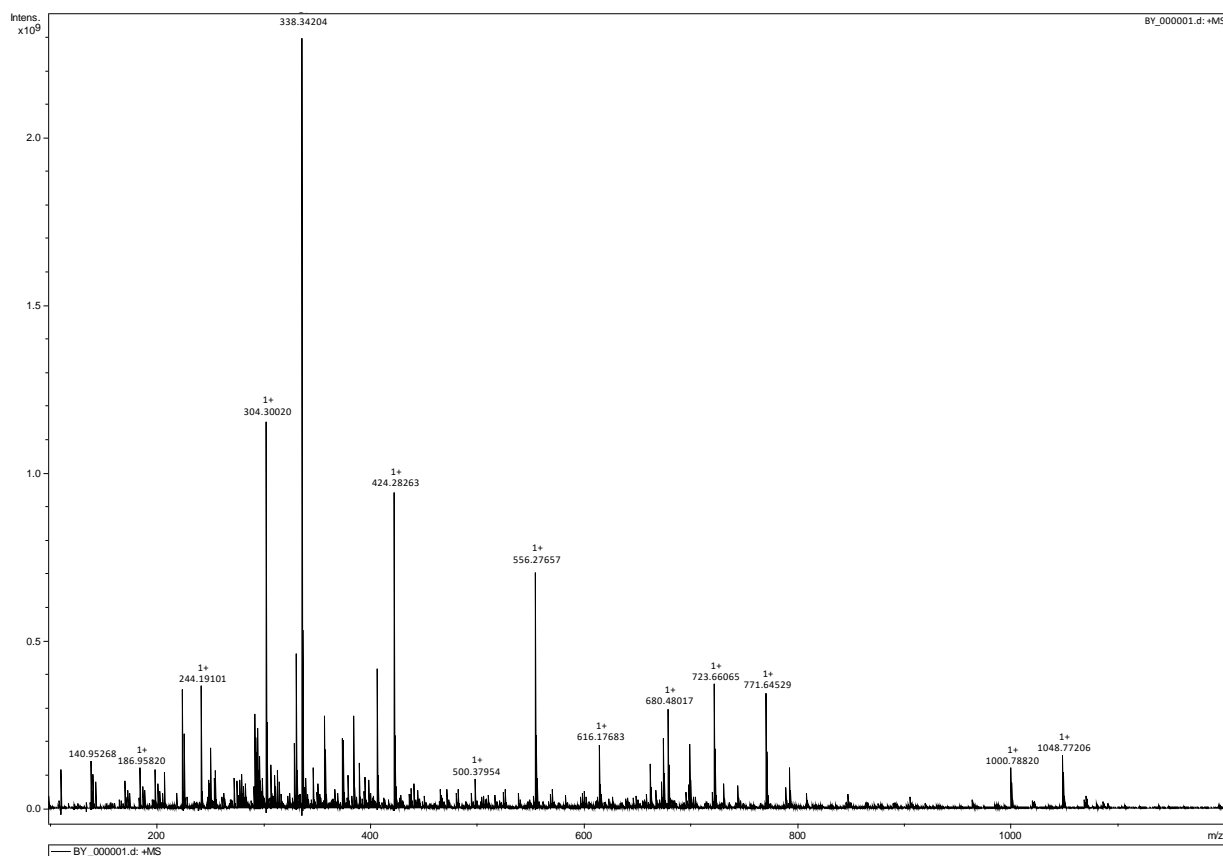
Strain	Slope Log ₂ Abs vs time (average)	First Derivative Peak (average)
BY4741	0.3842	0.025
ΔGLO1	0.3427	0.021
ΔGLO2	0.3699	0.022
ΔGRE3	0.3948	0.022
ΔENO1	0.3857	0.023

Table 3.1: Average values of the slopes of the log₂ ABS vs time curves and first derivative peaks of the ABS vs time curves smoothed with the Savitzky-Golay filter.

The average values of the slopes of the \log_2 ABS vs time curves varied only between 0.3427 and 0.3948, while the first derivative peak values of the ABS vs time curves showed even less variation (between 0.021 and 0.025).

II. Yeast chemical profile

An untargeted metabolomics analysis using FT-ICR-MS, by direct infusion and using electrospray ionization in positive (ESI+) ionization mode, was followed to analyse the chemical profiles in all yeast strains.



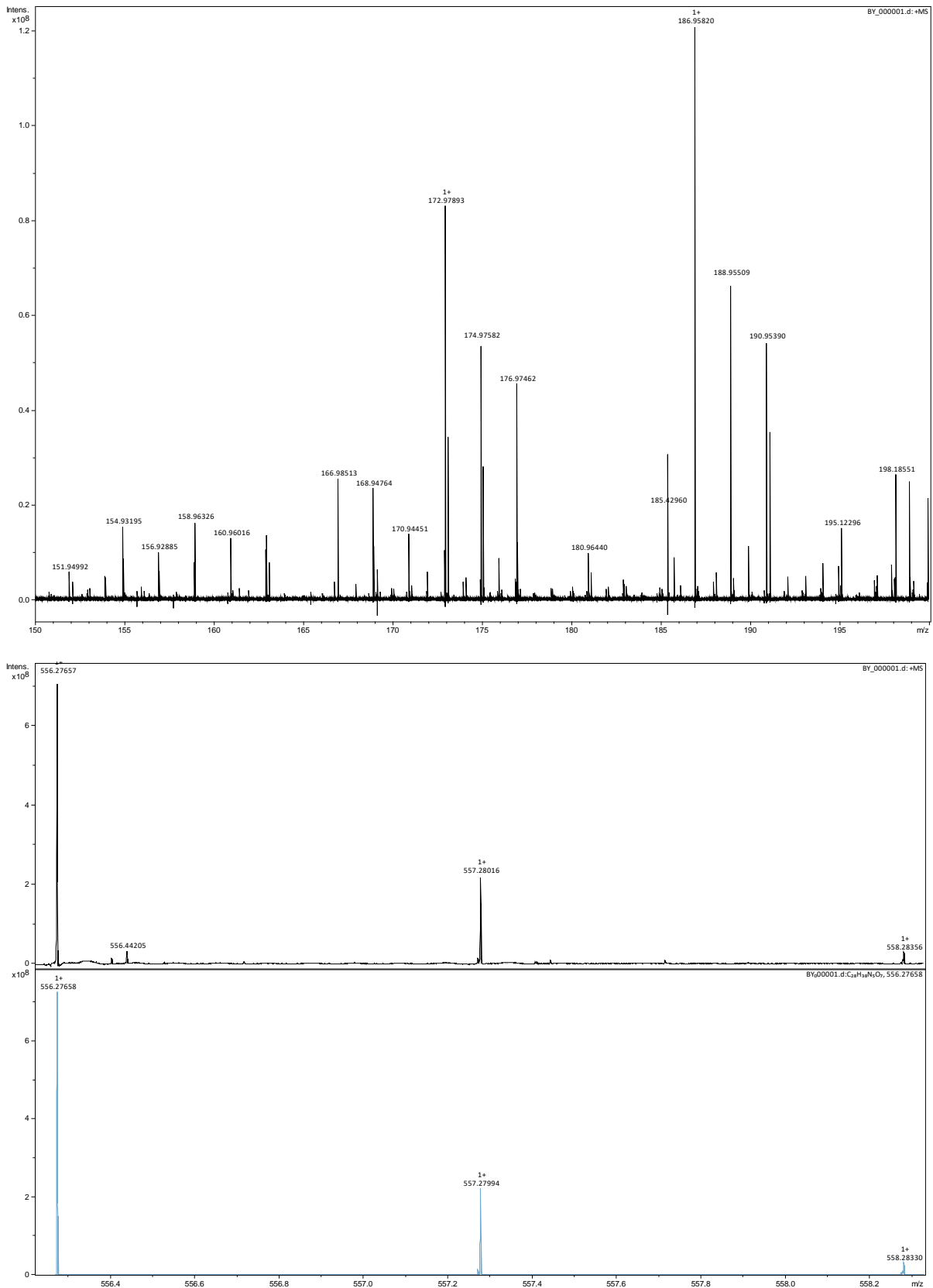


Figure 3.2: Representative mass spectrum of an yeast extract (BY4741), (top). Amplification in the 150-200 m/z zone (middle) Internal standard (Leucine-enkephalin) with fine isotopic structure (bottom).

The mass spectra showed extremely good resolution, as expected from FT-ICR-MS spectra, with a large number of clearly defined peaks spread over a vast range of m/z values (Figure 3.2).

III. Metabolite counts and annotation

The fifteen FT-ICR mass spectra were analysed using the MetaboScape software, being identified a total of 21175 distinct masses (m/z values). After filtering down the list only to those metabolites which registered a significant intensity signal (higher than 10^5) in at least two replicates (either of the same yeast strain or across different ones) the number was reduced to only 1047 metabolites, varying between 379 and 537 for the different samples. The distribution of unique masses (metabolites) by the various samples and strains can be seen in Table 3.2.

Strain	Replicate	Nr. of metabolites (samples)	Nr. of metabolites (strains)	Nr. of Metabolites putatively identified (with names)	Nr. of metabolites with assigned formulas
BY4741	BY (1)	494	657	155	656
	BY (2)	521			
	BY (3)	537			
Δ GLO1	Δ GLO1 (1)	475	602	151	601
	Δ GLO1 (2)	479			
	Δ GLO1 (3)	487			
Δ GLO2	Δ GLO2 (1)	468	622	145	621
	Δ GLO2 (2)	489			
	Δ GLO2 (3)	497			
Δ GRE3	Δ GRE3 (1)	470	652	164	651
	Δ GRE3 (2)	506			
	Δ GRE3 (3)	522			
Δ ENO1	Δ ENO1 (1)	379	556	136	555
	Δ ENO1 (2)	449			
	Δ ENO1 (3)	452			
Total			1047	201	1046

Table 3.2: Number of metabolites detected through FT-ICR-MS in each of the 15 samples (note: only those present in at least two samples were considered). Number of metabolites found at least once in each strain. Number of metabolites putatively identified by name using the YMDB and HMDB in MetaboScape. Number of metabolites for which a molecular formula was predicted using MetaboScape's SmartFormula algorithm.

The number of metabolites present in each strain varied between 675 and 556. The largest count, 657 metabolites, corresponds to the wildtype strain (BY4741), and while Δ GRE3, with 652 metabolites, does not deviate much from that number, Δ GLO1 and Δ GLO2 do, with 602 and 622 metabolites respectively. The control mutant strain Δ ENO1, however, was the one to deviate the most, having only 556 metabolites, by far the smallest count of all.

Molecular formula prediction was successful, with only one mass (587.133027 Da) not being assigned a formula. The number of metabolites with putative formulas assigned varied between 656 (for BY4741) and 555 (for Δ ENO1), (Table 3.2).

Putative metabolite identification (by name) was only possible for 201 metabolites, which corresponded to an entry in either YMDB or HMDB. Most of the metabolites with assigned names (108 out of the 201) were common to all strains.

A Venn diagram was built based on the number of common and exclusive metabolites to each of the five strains. A total of 313 metabolites were common to all yeasts, with the remaining being exclusive either to a single strain or to a combination of two or more. The number of metabolites exclusive to a single strain ranged between 36 (for Δ GLO1) and 80 (for Δ GRE3). (Figure 3.2).

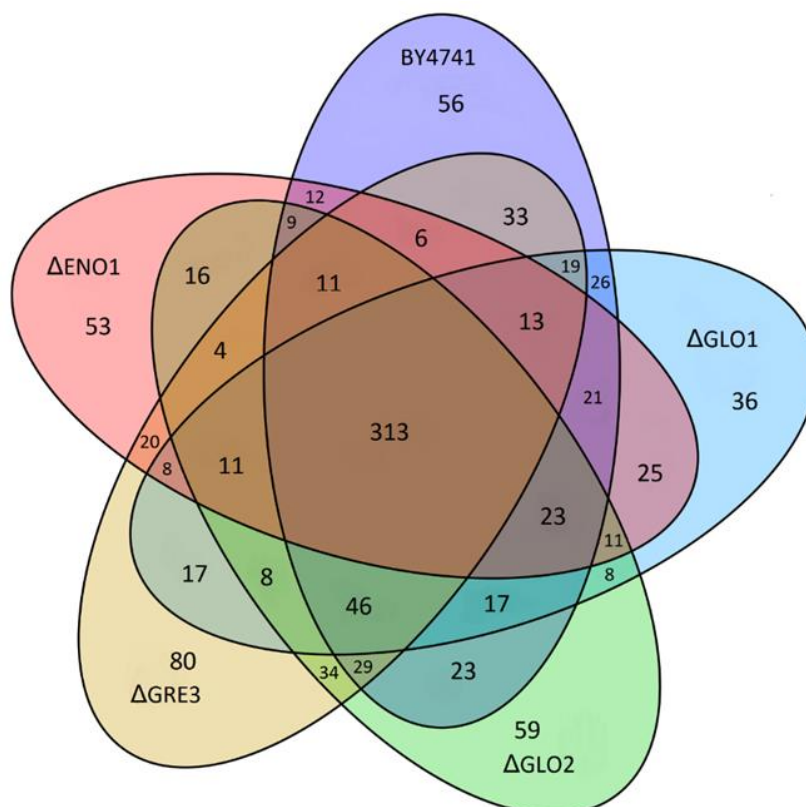
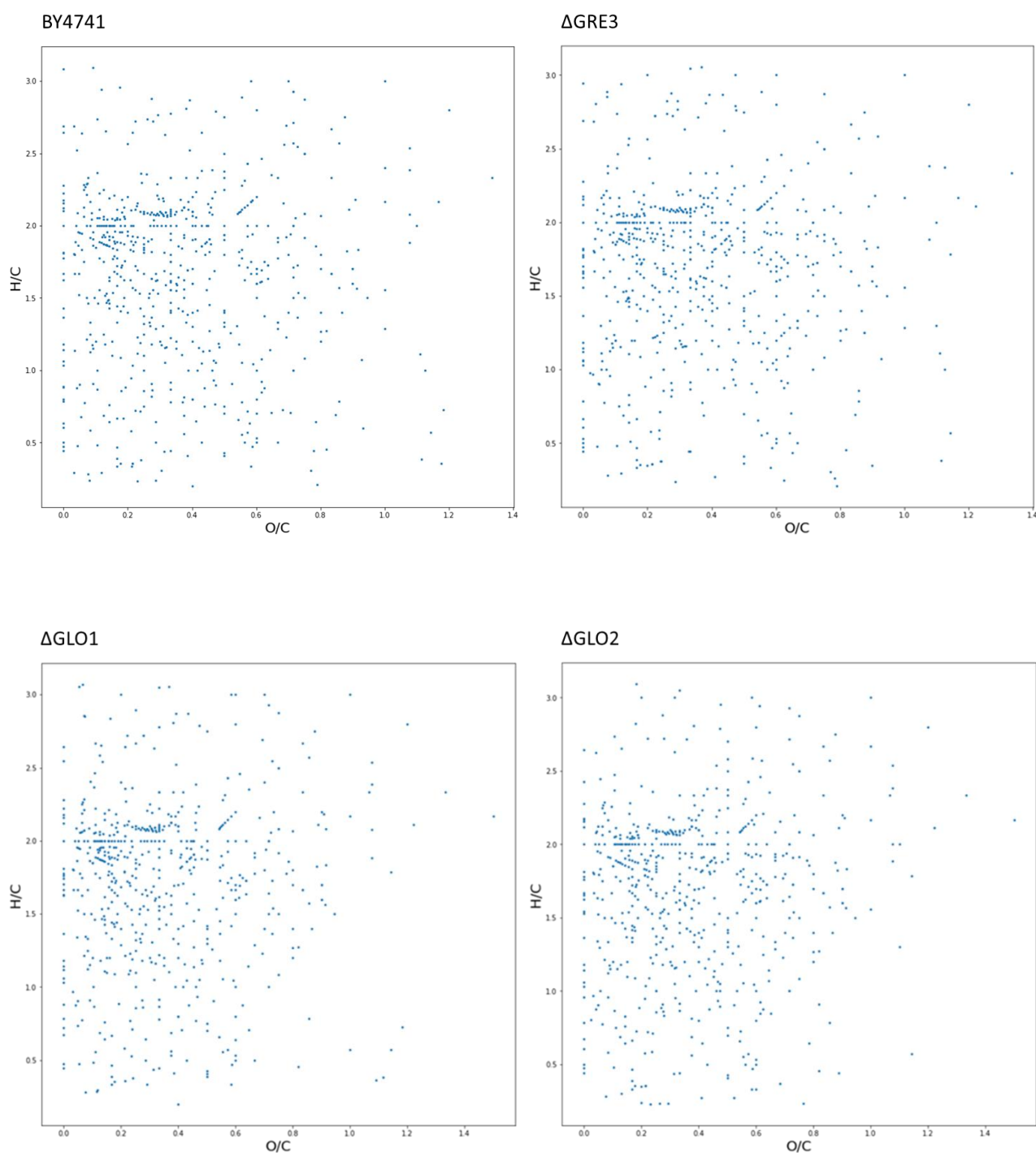


Figure 3.3: Venn diagram showing the common and exclusive metabolites to each strain or combination of strains.

IV. Van Krevelen plots and chemical composition series

The chemical nature of the metabolites with an assigned molecular formula was analysed by counting the number of atoms of each of the six most abundant elements in biological molecules (carbon, hydrogen, oxygen, nitrogen, sulphur and phosphorous). These counts were used to build a Van Krevelen (VK) plot for each of the five strains, by plotting H/C ratios against O/C ratios (Figures 3.4) and a chemical histogram showing the number of metabolites by strain in each elemental composition series (Figure 3.5).



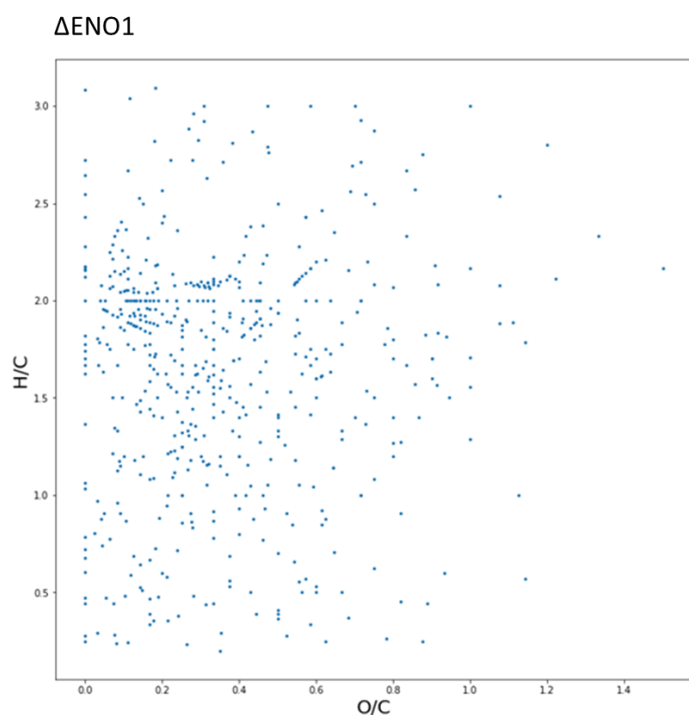


Figure 3.4: Van Krevelen diagram of the BY4741, Δ GRE3, Δ GLO1, Δ GLO2 and Δ ENO1 strains, showing all the metabolites with assigned formulas positioned according to their Hydrogen/Carbon (H/C) and Oxygen/Carbon (O/C) ratios.

Van Krevelen diagrams are a very useful tool in metabolomics to visualize the chemical composition of complex mixtures by plotting the H/C ratio against the O/C ratio for every molecule in the mixture (Krevelen, 1950). VK plots for all strains analysed seem very similar in appearance, with the lipids (H/C 1.5-2.0, O/C 0-0.3) and amino acids (H/C 1-2, O/C 0.2-0.8) regions being the densest in all cases (Figure 3.4).

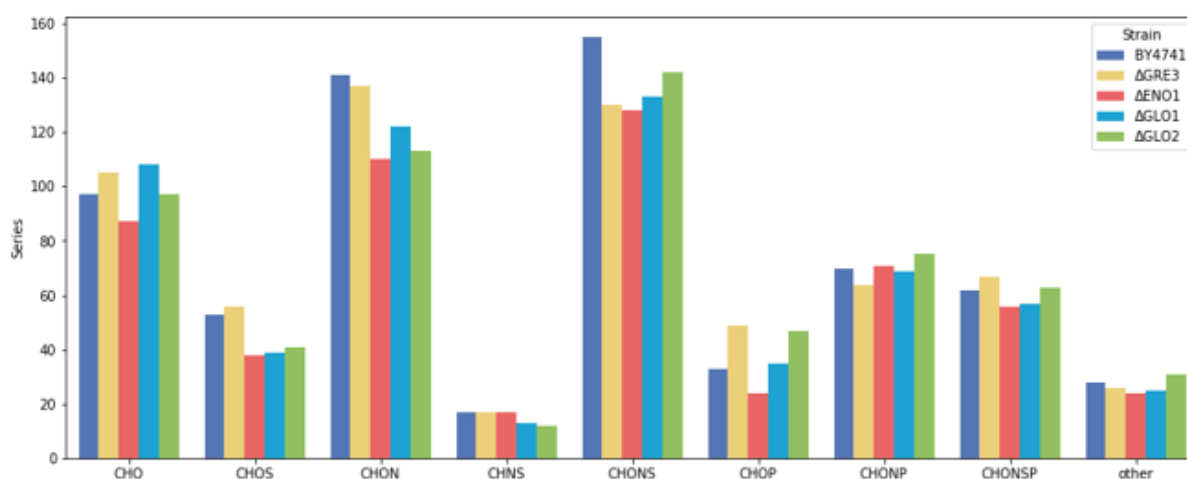


Figure 3.5: Bar plot showing the distribution of the metabolites with assigned formulas found in each strain by the various chemical composition series.

The elemental composition series revealed some differences between the strains (Figure 3.5). The largest differences occurred in the CHONS, CHON (possibly amino acids) and CHO (carbohydrates, lipids) series, which are also, in this order, those that have the largest numbers of metabolites in all strains. It may be of interest, however, to note that BY4741 is the yeast strain with the largest overall number of metabolites mostly due to the CHONS and CHON series, seeing as it is not the largest in any of the others. Another interesting observation is that Δ GRE3, close to the wildtype in the overall metabolite counts as well in most series counts, suffers a noticeable decrease in CHONS, where its numbers dwindle behind those of the GLO mutants (Figure 3.5). A decrease in CHOS compounds was observed in Δ GLO1, Δ GLO2 and Δ ENO1 (Figure 3.5).

V. Multivariate statistical analysis

Principal component analysis (PCA) and hierarchical clustering analysis (HCA), both unsupervised methods, were applied to the untargeted metabolomics data to validate data reproducibility and to evaluate the degree of metabolic similarity between the different samples and strains.

In the PCA score plots (Figure 3.6), a clear separation of the five strains was observed, with low variability between different samples of the same strain. Hierarchical clustering further confirmed this separation (Figure 3.7). In the resulting dendrogram, and consistent with the PCA results, samples of the same strain clustered firstly with each other, and then into two groups: one with the reference BY4741 and Δ GRE3 strains, and the other with both mutants for glyoxalase pathway enzymes, Δ GLO1 and Δ GLO2, and Δ ENO1 strains. The Δ GLO1 and Δ GLO2 strains appeared at a shorter Euclidean distance from each other than other strains, and the Δ ENO1 strain appeared closer to the GLO mutants than to the other two strains. Similarly, BY4741 and Δ GRE3 were also closer to each other than they were to the other strains, but the distance between them was still significantly larger than both the distances between Δ GLO1 and Δ GLO2 and between the GLO mutants and Δ ENO1. It was also evident that the reproducibility of the data is very high, as evidenced by the clustering of replicates together in both plots (Figures 3.6 and 3.7).

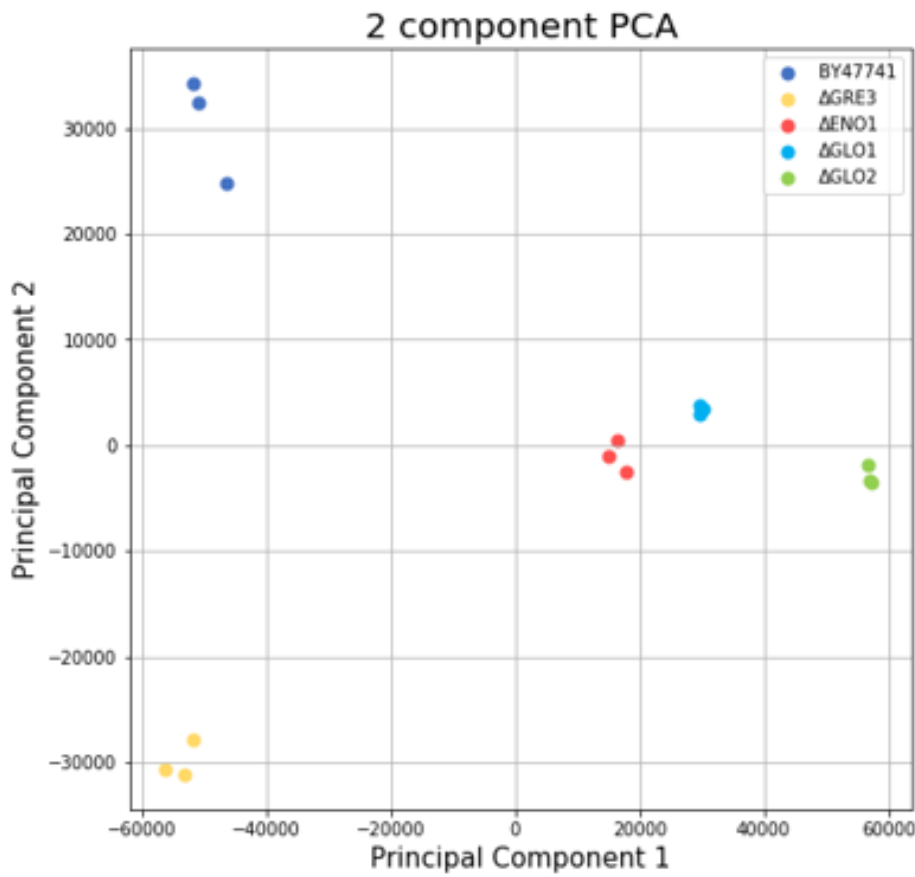


Figure 3.6: Principal components analysis of the fifteen yeast metabolite samples using the first two principal components.

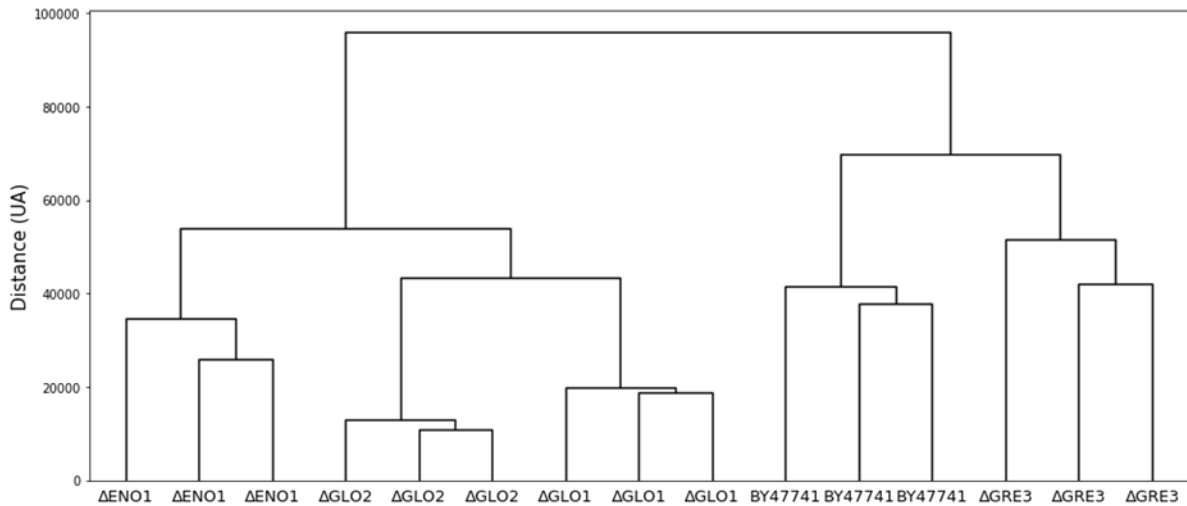


Figure 3.7: Hierarchical clustering of the fifteen yeast metabolite samples.

The general trend of separation between the strains suggests that multivariate statistical analysis of the FT-ICR-MS metabolic profiles of the samples is enough to discriminate between the strains.

To identify the metabolites that contributed the most for this separation, a Partial Least Squares Discriminant Analysis (PLS-DA) model was fitted to the MS intensity data, building a system of components that maximized covariance between the groups (in this case, the five different yeast strains). While PLS-DA is frequently used as a classifier, the main purpose here is to identify the most important compounds for discrimination between the strains. The model showed very good performance, with a score plot of the first two components indicating that the predictor component was able to discriminate between the five groups (Figure 3.8). It is also noteworthy that the first two components were very similar to those in the PCA (Figure 3.6), thus reinforcing the conclusion that the metabolic differences between the strains overpower any random variations between the samples.

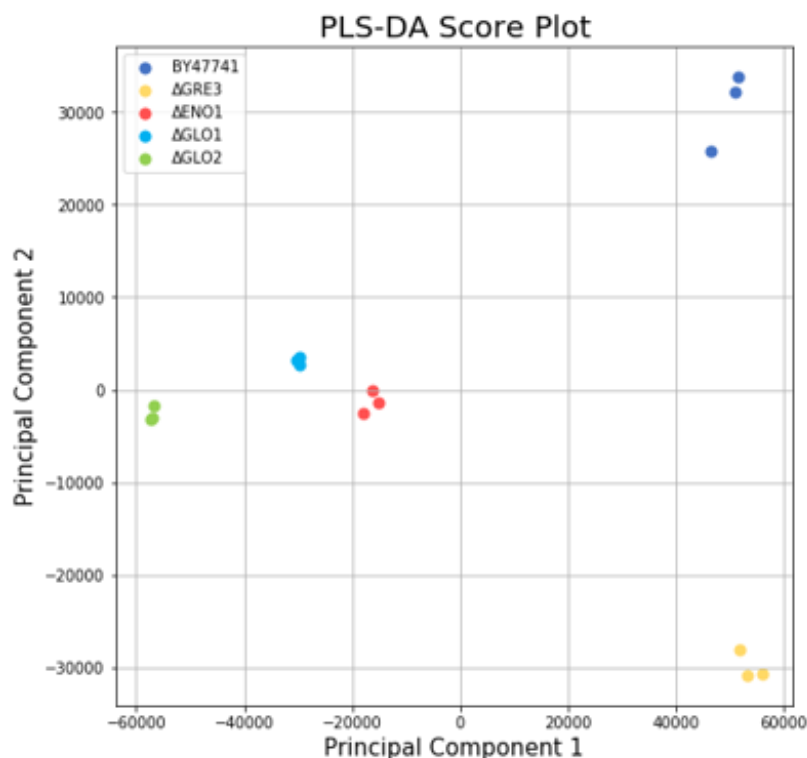


Figure 3.8: Partial least squares discriminant analysis of the fifteen yeast metabolite samples using principal components one and two.

The variable importance on projection (VIP) scores are a measure of the contribution of each individual metabolite for the separation between groups performed by the PLS-DA algorithm. The ten metabolites with the highest VIP scores can be seen in table 3.3, together with their relative concentrations in each of the five strains.
















Position	Mass (Da)	Metabolite Name	Molecular Formula	VIP Score	Relative concentration in the yeast strains
1	307.0838	Glutathione	C ₁₀ H ₁₇ N ₃ O ₆ S	8.417995046	
2	493.3168	PC(16:1(9Z)/0:0)	C ₂₄ H ₄₈ NO ₇ P	5.993469554	
3	624.0873	N/A	C ₁₄ H ₂₈ N ₁₀ O ₁₀ S ₄	5.587837143	
4	257.1029	Glycerophosphocholine	C ₈ H ₂₀ NO ₆ P	4.837523121	
5	324.1057	N/A	C ₁₂ H ₂₀ O ₁₀	4.177424151	
6	337.3345	N/A	C ₂₂ H ₄₃ NO	4.153906335	
7	254.2246	Hypogeic acid	C ₁₆ H ₃₀ O ₂	4.077281087	
8	385.3192	Pentadecanoylcarnitine	C ₂₂ H ₄₃ NO ₄	3.77589036	
9	398.1372	N/A	C ₁₅ H ₂₂ N ₆ O ₅ S	3.484675636	
10	451.2699	PE(16:1(9Z)/0:0)	C ₂₁ H ₄₂ NO ₇ P	3.172850261	

Table 3.3: The ten metabolites that contribute the most to the separation between the strains. In the "Relative concentration" column, the five strains appear in the following order: BY4741- Δ GLO1- Δ GLO2- Δ GRE3- Δ ENO1. 100% corresponds to the highest intensity signal for that metabolite;  75-100%;  50-75%;  25-50%;  >25%;  absent from the strain. Peak intensities can be consulted in Annex 1.

Glutathione leads the list, with a VIP score of 8.417995046, significantly larger than the second in the list, the lipid PC(16:1(9Z)/0:0). Six of these ten compounds have been putatively assigned a name and all of them have been assigned formulas. Only one, Pentadecanoylcarnitine, is not present in all five strains.

VI. Pathway mapping

Glutathione was the most important metabolite that contributed for the separation between the five strains. Hence, a pathway analysis was performed, using the pathways involved in the metabolism of glutathione and the methylglyoxal catabolism systems (Figure 3.9).

Chapter 4 – Discussion

Saccharomyces cerevisiae is a model eukaryote with around 6000 genes (Goffeau et al., 1996). The majority of these genes can be deleted without compromising yeast viability, being a vast fraction of these mutations silent and not producing an apparent observable phenotype (Kuepfer et al., 2005; Wagner, 2000). The four null mutants used in this work had previously been shown to be phenotypically silent in *S. cerevisiae* under normal growth conditions, as none of the enzymes appears to be essential for cell viability and mutants show no particular impairment to growth when compared to isogenic wildtype strains (Bito et al., 1997; Cordero, 2001; Kimura, 1996; Pietkiewicz et al., 2009). Growth assays performed at the beginning of this study served to confirm that assertion.

All yeast strains analysed had identical growth behaviours, reaching the stationary phase of growth after 10-12 hours of growth. None of the mutant yeast strains suffered a growth defect in comparison with the wildtype strain (Figure 3.1; Table 3.1). The values of the slopes and first derivative peaks showed great uniformity between the different strains (Table 3.1).

Can an untargeted metabolomics approach using FT-ICR-MS distinguish between almost genetically identical yeasts?

The extreme resolution and mass accuracy offered by Fourier-transform ion cyclotron resonance mass spectrometry allows for the simultaneous detection of a very large number of metabolites at once, thus providing a far more comprehensive profile of cell metabolism than could be achieved with less sensitive techniques.

The five *S. cerevisiae* strains examined in this study were genetically identical, except for a single gene and did not show any phenotypic differences whatsoever when grown under these conditions (Figure 2.1). However, FT-ICR-MS-based untargeted metabolomics uncovered a wide array of metabolic differences which can be seen in a variety of indicators such as the number of metabolites identified in each strain (Table 3.1) and the number of metabolites exclusive to each strain or to any possible combination of strains (Figure 3.1).

The most important indicators of the success of this experimental approach are the unsupervised multivariate statistical analysis results (Figures 3.6-7), which display a high degree of coherence with each other and strongly suggest that most of the differences between the fifteen samples are related with the point mutations. It is undeniable that metabolite samples belonging to the same strain have a higher degree of statistical similarity with each other than with samples belonging to different strains, as they always cluster together in the HCA dendrogram and appear at close positions in the PCA score plot. PCA and HCA are both unsupervised methods, meaning that the algorithms that perform them do not consider the existence of different groups in the data (in this case, the five strains), and thus do not operate in a way that seeks to maximize the separation between *a priori* defined groups. This can at times translate into a poorer performance in distinguishing between different groups when compared to supervised methods such as PLS-DA. However, unsupervised methods such as PCA and HCA ensure that any groupings that form in their resulting plots reflect a global measure of statistical similarity, with all variables (in this case, the intensity of each individual metabolite) being given equal consideration and no separation between the samples being forced. Thus, the PCA and HCA reflect the chemical

composition of the samples as far as it can be determined through FT-ICR-MS, and in this case prove that samples of the same strain are more similar to each other than to samples of different strains and that there is a clear separation between the strains.

It is then clear that an untargeted metabolomics approach based on FT-ICR-MS can be used to distinguish between nearly identical yeasts in a coherent manner that definitively relates to the single-gene differences between them.

An interesting question to be answered in future research can be determining where lies the upper limit of this discriminatory capacity of FT-ICR-MS-based untargeted metabolomics. That is, having proven its capacity to distinguish between single-gene mutants of the same yeast, can it also, on the other end of the spectrum, be used - for example - to distinguish between different species of yeast in a way that has some correlation with the genome or phenotype? If one were to take metabolic samples from a series of yeasts and analyse them in the same way as I did for this dissertation, would they appear in the multivariate statistical analysis plots grouped according to their *genus* or some other *taxum*?

Interpreting the differences

While simply determining the existence or not of metabolic differences between the strains is a relatively straightforward task when the right statistical tools are available, making sense of these differences in a way that proves biologically informative can be enough work for a lifetime.

Metabolite counts and the numbers of common and exclusive metabolites between the strains can provide valuable clues, as can the PCA, HCA and especially the PLS-DA results. While more research will be necessary to explore every possible implication, the results of this study reveal some interesting, and sometimes unexpected, insights about the impacts of the point mutations analysed.

It seems unlikely that the differences between the strains are mostly due to the effects of methylglyoxal-dependent glycation. The glyoxalase and aldose reductase pathways are known to be equally important in the prevention of methylglyoxal-dependent glycation, with Δ GRE3 and Δ GLO1 having very similar intracellular methylglyoxal concentrations, while Δ GLO2 lies closer to the wildtype in terms of both methylglyoxal concentration and the presence of glycation end products (Gomes et al., 2005). As such, any metabolic differences related to the process of glycation should translate to a high degree of similarity between Δ GRE3 and Δ GLO1, since both of these mutations have the reported effect of increasing methylglyoxal concentrations and glycation levels to a similar degree. Similarly, given that deletion of the glyoxalase II enzyme does not immediately lead to a large increase in methylglyoxal concentration or the number of glycated proteins, Δ GLO2 should be relatively closer to the reference strain than Δ GRE3 and Δ GLO1. This does not appear to be the pattern of inter-strain similarity between the yeasts in this study. In fact, all potential indicators consistently point out to a higher degree of similarity between Δ GLO1 and Δ GLO2, while Δ GRE3 leans closer to BY4741.

Δ GRE3 has a nearly identical number of putatively identified metabolites to that of the wildtype strain (652 for Δ GRE3 and 657 for BY4741; Table 3.1) and the two strains share 33 metabolites in exclusive with each other (Figure 3.1). They also cluster together in the hierarchical clustering dendrogram, standing at a significantly smaller Euclidean distance from each other than the other strains (Figure 3.7). While at a first glance, the PCA score plot (Figure 3.6) may appear to put in question this assumption of a relatively higher degree of similarity between Δ GRE3 and BY474, the two strains' PC1 scores are

extremely similar, with the apparently large distance between the two being solely due to their PC2 scores, which are nearly the inverse of each other. Thus, keeping in mind that the first principal component generally conserves more of the original information, the PCA plot may not necessarily contradict the HCA one. This is not the first time that Δ GRE3 has been reported to lean closer to the reference strain in comparison with other methylglyoxal catabolism-related mutants, with growth studies showing that the Δ GRE3 has a similar growth phenotype to BY474 even in the presence of an external source of methylglyoxal, while other strains showed marked differences (Ponces Freire et al., 2003). However, it should be noted that these results refer specifically to a situation where the yeasts were grown in the presence of an external source of methylglyoxal, and that, as previously stated, Δ GRE3 showed a higher methylglyoxal concentration and a larger number of glycated proteins compared to the reference strain under normal growth conditions (Gomes et al., 2005).

Multivariate statistical analysis also consistently shows that the GLO mutant strains, Δ GLO1 and Δ GLO2, are significantly closer to each other than they are to the wildtype. The two strains cluster together in the hierarchical clustering dendrogram, with the Euclidean distance between them being so small that it's comparable to the distances between BY4741 and Δ GRE3's different samples (Figure 3.7). They also appear closer to each other in the PCA score plot, although here the proximity is far more apparent in first principal component (Figure 3.6). Δ GLO1 and Δ GLO2's metabolite counts (602 and 622 respectively) are lower than the wildtype (657), (Table 3.1). In previous research, Δ GLO2 showed neither a particularly heightened sensitivity to an external methylglyoxal source (Ponces Freire et al., 2003) nor, as previously stated, an immediate and large increase in glycation levels and methylglyoxal concentration when compared with the reference strain (Gomes et al., 2005).

While methylglyoxal itself, and consequently glycation, cannot explain the consistent grouping of BY4741 with Δ GRE3 and of Δ GLO1 with Δ GLO2, it seems likely that glutathione can. In addition to its role in methylglyoxal elimination through the glyoxalase pathway, glutathione plays number of other important roles in cell metabolism, including as an antioxidant to eliminate reactive oxygen species and modulating the activity of a wide variety of proteins by bonding with their cysteine residues in a process known as S-glutathionylation (Pompella et al., 2003). In fact, it is known that changes in glutathione concentration could lead to a wide array of metabolic effects in yeast (Gergondey et al., 2017).

Relative abundances of glutathione followed a pattern consistent with its participation in methylglyoxal catabolism. It seems to be most abundant in the wildtype (BY4741) strain and in the aldose reductase-deficient mutant strain (Δ GRE3), while being less abundant in the Δ GLO1 strain and even less in Δ GLO2 (Figure 3.9). This is most likely due to the non-enzymatic formation of hemithioacetal from glutathione and methylglyoxal. Normally, this is first step of the glyoxalase pathway (Figure 1.2), but in a situation where one of the glyoxalase enzymes is absent, it can be expected a decrease in glutathione concentration, as this thiol becomes sequestered in the form of either hemithioacetal (in Δ GLO1) or S-(D)-Lactoylglutathione (in Δ GLO2) instead of being regenerated at the end of the glyoxalase pathway. While the non-enzymatic formation of hemithioacetal can occur in the reverse direction, the GLO1-catalyzed formation of S-(D)-Lactoylglutathione is not reversible, and this makes the lower glutathione abundance more pronounced in Δ GLO2 (Sousa Silva et al., 2013).

Other metabolites closely related to glutathione followed its variation pattern, being most abundant in the reference and in the aldose reductase-deficient strains and less in the glyoxalase mutants, with Δ GLO2 frequently registering even lower abundances than Δ GLO1. These include oxidized glutathione (GSSG), L- γ -glutamylcysteine and cysteine glutathione disulphide. S-(D)-Lactoylglutathione showed a higher abundance in the reference strain than in Δ GLO2, where it was supposed to accumulate the most,

being undetected in the Δ GRE3 strain, where it should have been present at a similar relative abundance to that found in the wildtype since this strain has the glyoxalase enzyme which catalyses the conversion of hemithioacetal into S-(D)-Lactoylglutathione.

It is possible to conclude that the separation of the reference and methylglyoxal-related mutant strains into two groups, one comprising the Δ GLO1 and Δ GLO2 strains, and the other the Δ GRE3 and BY4741 strains, closely relates with the variation of the relative abundances of glutathione and its closely related metabolites.

This assignment of a central role of glutathione for yeast strain separation is also supported by the PLS-DA results. Unlike PCA and HCA, PLS-DA is a supervised technique, meaning that it takes into account the existence of *a priori* groups (in this case, the strains) to which the samples belong, and deliberately works to maximize the covariance between them. This approach makes it useful as a classifier to assign new unknown samples to a series of previously defined groups, but a rather poor choice for evaluating whether the similarities and differences between samples belonging to different groups are truly due to them belonging to these groups. Using a supervised technique such as PLS-DA for such a purpose would be extremely misleading, as the algorithm inherently works towards evidencing inter-group differences (Ruiz-Perez et al., 2018). Nevertheless, PLS-DA is extremely useful in the identification of the compounds which contribute the most to the separation between different groups. Thus, having the supervised PLS-DA separation be extremely similar to the unsupervised PCA and HCA separations, as clearly happens in this study (Figures 3.6-8), is a very important and useful information, as it tells us that the most important metabolites for the separation of the strains also play an important part in a “blind” separation of the samples. This has the effect of opening new possibilities for interpreting the data, seeing as it allows for potential biological meanings of the metabolic differences between the strains to be deduced by taking a closer at specific metabolites.

Analysing the PLS-DA VIP scores, the ten most important metabolites for the separation between the strains were identified (Table 3.3). Consistent with the grouping pattern observed in the other two multivariate statistical analysis methods, glutathione appears on top of that list, having by far the highest VIP score. It is also the metabolite with the greatest peak intensity after normalization to leucine-enkephalin, but this is unlikely to be the reason for its prominence given that Pareto scaling prevents differences in very large variables from overpowering subtler but potentially more important differences in variables that register overall smaller values.

The compound with the second highest VIP score is the lipid PC(16:1(9Z)/0:0), (YMDB ID: YMDB01184, HMDB ID: HMDB0010383), a lysophospholipid found on mitochondria and the endoplasmic reticulum where it acts as membrane stabilizer, energy source and energy storer. PC(16:1(9Z)/0:0) is present in all strains, but it is far more abundant in BY4741 and Δ GRE3, much like glutathione. Its derivative, glycerophosphocholine (YMDB ID: YMDB00309; HMDB ID: HMDB0000086), is associated with the same functions, and is the fourth most important compound for the strain separation, even though it presents a more balanced distribution than PC(16:1(9Z)/0:0) while still presenting the same glutathione-like pattern of being more abundant in BY4741 and Δ GRE3 and less so in the GLO mutants. Another chemically and functionally similar compound, PE(16:1(9Z)/0:0), (tenth on the VIP scores list) presents an identical distribution of relative abundances to PC(16:1(9Z)/0:0). The two remaining compounds with high VIP score putatively identified were hypogeic acid (HMDB ID: HMDB0002186) and pentadecanoylcarnitine (HMDB ID: HMDB0062517), also membrane lipids (Ramirez-Gaona et al., 2017; Wishart et al., 2007). These two present different distributions from the previous three, with hypogeic acid being approximately equally abundant in all

strains except in Δ GLO2, and pentadecanoylcarnitine not being detected in Δ GLO2 or Δ GRE3 and being more abundant in Δ GLO1 than in the reference strain.

While none of these five lipids has a readily apparent reason to differ between the strains, a potential explanation for the pattern in PC(16:1(9Z)/0:0), glycerophosphocholine and PE(16:1(9Z)/0:0) may be related to the role of glutathione in combatting lipid peroxidation, which appears listed as a significant biological process in all of these lipids' HMDB profiles (Wishart et al., 2007b). *S. cerevisiae* has glutathione peroxidase (GPx) enzymes which use glutathione as a co-factor and electron donor to reduce membrane lipids (Avery & Avery, 2001). It is possible that the decrease in glutathione abundance in Δ GLO1 and Δ GLO2 might have caused an increase in lipid peroxidation which resulted in decreased abundances of these lipids. Methylglyoxal and glycation have also been reported to increase lipid peroxidation in *S. cerevisiae* (Tupe et al., 2019), so it is possible that the higher methylglyoxal concentration found in Δ GLO1 contributed to the decrease in relative abundance of these membrane compounds. However, given that Δ GRE3, which under the conditions of this work has a methylglyoxal concentration comparable to that of Δ GLO1 and higher than that of the reference strain, showed no decrease in the relative abundances of these membrane lipids, and that Δ GLO2 did despite being similar to the wildtype in terms of methylglyoxal concentration and glycation levels (Gomes et al., 2005), glutathione still appears more likely to be the direct cause of these differences between the strains than methylglyoxal.

All indicators seem to consistently point out to the centrality of glutathione in the differentiation between the strains, suggesting that fluctuations in its concentrations have a larger impact on the metabolome than fluctuations in the concentrations of any other molecules, including methylglyoxal. In future research, it may be interesting to perform glutathione quantification in each of the strains and apply the same FT-ICR-MS untargeted metabolomics approach to other strains with mutations in genes related with glutathione metabolism without being restricted to its role in the glyoxalase system.

However, as of this moment, it would still be extremely premature to assert without reservations that most of the metabolic differences found between the strains are directly related to glutathione metabolism and its far reaching consequences on cell metabolism and redox balance as a whole, or even any pathway in particular.

The control mutant yeast strain used in this work (Δ ENO1) produced very surprising results which cannot be easily explained based on our current understanding of Enolase 1 or any of the methylglyoxal catabolism-related enzymes, and thus may complicate the task of attributing similarities and differences observed between other strains to particular biological causes.

ENO1 and *ENO2* are the two genes encoding the two isoforms of the glycolytic enzyme enolase. In yeasts grown in glucose, almost all enolase found in the cell corresponds to the *ENO2* peptide, meaning that the deletion of the *ENO1* gene has no consequences at the level of glycolysis (Mcalister & Holland, 1982). Enolase 2 is a known target of methylglyoxal-dependent glycation (Gomes et al., 2006), but Enolase 1 is not. Thus, there is no readily apparent reason why Δ ENO1 should differ substantially from the reference strain or present any notable similarities with any of the other mutants. However, the deletion of ENO1, not only has a very substantial impact on metabolism (with only 556 metabolites it has, by far, the smallest metabolite count of any strain, and the largest difference from the reference strain; Figure 3.1), but also it appears to have a higher degree of similarity with Δ GLO1 and Δ GLO2 than with the other two strains. All three multivariate statistical analysis methods support this. In the hierarchical clustering dendrogram (Figure 3.7), Δ ENO1 appears at only a slightly larger Euclidean

distance from Δ GLO1 and Δ GLO2 than the two GLO mutants are from each other. In the PCA score plot (Figure 3.6), Δ ENO1 appears even closer to Δ GLO1 than Δ GLO2 is, although this is mostly due to the second principal component. Additionally, several of the compounds with high PLS-DA VIP scores showed variations in abundance between Δ ENO1 and the wildtype. Glutathione, PC(16:1(9Z)/0:0) and PE(16:1(9Z)/0:0) were all less abundant in Δ ENO1 than in BY4741, with the Δ ENO1 mutant approaching the relative abundances of the glyoxalase mutants. Interestingly, this wasn't the case for glycerophosphocholine, which was approximately equally abundant between Δ ENO1, BY4741 and Δ GRE3. The same thing happened for Hypogaeic acid (which, as previously stated, was equally abundant in all strains other than Δ GLO2), while pentadecanoylcarnitine seemed less abundant in Δ ENO1 than in the wildtype.

Not all differences between the strains are properly understood, and they are unlikely to be all due to glutathione and methylglyoxal metabolism alone. It is perfectly possible that many of them are due to genetic or physical interactions of the deleted genes and their proteins that are not yet known or understood. Therefore, there may be a convenient biological explanation for the consistent finding of a high degree of similarity between Δ ENO1 and the GLO mutants.

However, the fact remains that, while it has been proven that an accurate separation between the strains through an FT-ICR-MS-based untargeted metabolomics approach is possible, there is still no conclusive proof that, based on the similarity indicators used in this work, the degree of biological similarity between the different strains can be predicted. Future research will benefit from the investigation of a wider number of mutant strains to increase the robustness of this approach.

Chapter 5 - Conclusion

In this dissertation I have demonstrated that an untargeted metabolomics approach based on extremely high-resolution Fourier-transform ion cyclotron resonance mass spectrometry can accurately distinguish between phenotypically identical single-gene deletion mutants of the same yeast, revealing a vast array of metabolic differences between them.

FT-ICR-MS's unparalleled sensitivity and mass determination accuracy were fundamental for the successful application of this approach, as they allowed for the simultaneous identification and relative quantification of a large number of metabolites. Through multivariate statistical analysis, it was possible to prove that yeast metabolite samples belonging to different cultures of the same strain consistently showed a higher degree of similarity between them than with yeast metabolite samples extracted from cultures of different strains, with supervised and unsupervised methods performing very similar separations. Putative identification of metabolites using the YMDB and HMDB databases was also extremely useful, as it provided clues as to what may be the potential biological meaning of the differences observed between the strains.

Glutathione appears to play a central role in driving the metabolic differences between the strains. Multivariate statistical analysis divided the strains in two groups in a way that correlates most closely with the relative abundances of glutathione and its closely related metabolites, with the wildtype and Δ GRE3 strains forming one group, while the two glyoxalase system mutants (where regeneration of glutathione is impaired) formed another. Glutathione was also identified by PLS-DA as the most important metabolite for the separation between the strains, with several other important compounds presenting relative abundances in each strain similar to those of glutathione.

More research will be necessary to confirm this assignment of a central role to glutathione for the emergence of metabolic differences between these strains, including performing glutathione quantification in each one of them and investigating the effects of other mutations related to glutathione metabolism but not necessarily to the glyoxalase system.

However, the unexpected finding of similarities between the control mutant Δ ENO1 and the glyoxalase system mutants cannot be explained solely based on our present understanding of glutathione or methylglyoxal metabolism. While it is possible that a convenient biological explanation may exist, this raises the question of whether there may be limits to this approach's capacity to evaluate the degree of metabolic similarity between different strains and derive meaningful biological conclusions from that information. Future research may benefit from the use of a larger number of mutant strains.

Bibliography

- Aharoni, A., De Vos, C. H. R., Verhoeven, H. A., Maliepaard, C. A., Kruppa, G., Bino, R., & Goodenowe, D. B. (2002). Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry. *OMICS A Journal of Integrative Biology*, 6(3), 217–234. <https://doi.org/10.1089/15362310260256882>
- Aston, F. W. (1919). LXXIV. A positive ray spectrograph. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 38(228), 707–714. <https://doi.org/10.1080/14786441208636004>
- Avery, A. M., & Avery, S. V. (2001). Saccharomyces cerevisiae Expresses Three Phospholipid Hydroperoxide Glutathione Peroxidases. *Journal of Biological Chemistry*, 276(36), 33730–33735. <https://doi.org/10.1074/jbc.M105672200>
- Beggs, J. D. (1978). Transformation of yeast by a replicating hybrid plasmid. *Nature*, 275(5676), 104–109. <https://doi.org/10.1038/275104a0>
- Bito, A., Haider, M., Hadler, I., & Breitenbach, M. (1997). Identification and phenotypic analysis of two glyoxalase II encoding genes from Saccharomyces cerevisiae, GLO2 and GLO4, and intracellular localization of the corresponding proteins. *Journal of Biological Chemistry*, 272(34), 21509–21519. <https://doi.org/10.1074/jbc.272.34.21509>
- Brachmann, C. B., Davies, A., Cost, G. J., Caputo, E., Li, J., Hieter, P., & Boeke, J. D. (1998). Designer deletion strains derived from Saccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast*, 14(2), 115–132. [https://doi.org/10.1002/\(SICI\)1097-0061\(19980130\)14:2<115::AID-YEA204>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1097-0061(19980130)14:2<115::AID-YEA204>3.0.CO;2-2)
- Brown, S. C., Kruppa, G., & Dasseux, J. L. (2005). Metabolomics applications of FT-ICR mass spectrometry. *Mass Spectrometry Reviews*, 24(2), 223–231. <https://doi.org/10.1002/mas.20011>
- Comisarow, M. B., & Marshall, A. G. (1974). Fourier transform ion cyclotron resonance spectroscopy. *Chemical Physics Letters*, 25(2), 282–283. [https://doi.org/10.1016/0009-2614\(74\)89137-2](https://doi.org/10.1016/0009-2614(74)89137-2)
- Cordero, R. R. O. (2001). Deletion of the GRE3 Aldose Reductase Gene and Its Influence on Xylose Metabolism in Recombinant Strains of Saccharomyces cerevisiae Expressing the xylA and XKS1 Genes, 67(12), 5668–5674. <https://doi.org/10.1128/AEM.67.12.5668>
- Fiehn, O. (2001). Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comparative and Functional Genomics*, 155–168.
- Galindo-prieto, B., Eriksson, L., & Trygg, J. (2014). Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS). *Journal of Chemometrics*, 28(8), 623–632. <https://doi.org/10.1002/cem.2627>
- Gates, S. C., & Sweeley, C. C. (1978). Quantitative metabolic profiling based on gas chromatography. *Clinical Chemistry*, 24(10), 1663–1673. <https://doi.org/10.1093/clinchem/24.10.1663>
- Gergondey, R., Garcia, C., Marchand, C. H., Lemaire, S. D., Camadro, J. M., & Auchère, F. (2017). Modulation of the specific glutathionylation of mitochondrial proteins in the yeast Saccharomyces cerevisiae under basal and stress conditions. *Biochemical Journal*, 474(7), 1175–1193. <https://doi.org/10.1042/BCJ20160927>
- Gika, H. G., Wilson, I. D., & Theodoridis, G. A. (2014). LC-MS-based holistic metabolic profiling.

- Problems, limitations, advantages, and future perspectives. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 966, 1–6. <https://doi.org/10.1016/j.jchromb.2014.01.054>
- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., ... Oliver, S. G. (1996). Life with 6000 Genes. *Science*, 274(October), 546–567. <https://doi.org/10.1126/science.274.5287.546>
- Gomes, R. A., Sousa Silva, M., Vicente Miranda, H., Ferreira, A. E. N., Cordeiro, C. A. A., & Freire, A. P. (2005). Protein glycation in *Saccharomyces cerevisiae*. Argpyrimidine formation and methylglyoxal catabolism. *FEBS Journal*, 272(17), 4521–4531. <https://doi.org/10.1111/j.1742-4658.2005.04872.x>
- Gomes, R. A., Vicente Miranda, H., Silva, M. S., Graça, G., Coelho, A. V., Ferreira, A. E., ... Freire, A. P. (2006). Yeast protein glycation in vivo by methylglyoxal: Molecular modification of glycolytic enzymes and heat shock proteins. *FEBS Journal*, 273(23), 5273–5287. <https://doi.org/10.1111/j.1742-4658.2006.05520.x>
- Griffiths, J. (2008). A brief history of mass spectrometry. *Analytical Chemistry*, 80(15), 5678–5683. <https://doi.org/10.1021/ac8013065>
- Gygi, S. P., Rochon, Y., Franz, B. R., & Aebersold, R. (1999). Correlation between Protein and mRNA Abundance in Yeast. *Molecular and Cellular Biology*, 19(3), 1720–1730. <https://doi.org/10.1128/mcb.19.3.1720>
- Hanson, P. K. (2018). *Saccharomyces cerevisiae*: A Unicellular Model Genetic Organism of Enduring Importance. *Current Protocols in Essential Laboratory Techniques*, 16(1), 1–15. <https://doi.org/10.1002/cpet.21>
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Herskowit, I. (2017). Life Cycle of the Budding Yeast *Saccharomyces cerevisiae*. *Molecular Ecology Resources*, 17(3), 353–355. <https://doi.org/10.1111/1755-0998.12611>
- Hollywood, K., Brison, D. R., & Goodacre, R. (2006). Metabolomics: Current technologies and future trends. *Proteomics*, 6(17), 4716–4723. <https://doi.org/10.1002/pmic.200600106>
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Jeff Reback, Wes McKinney, jbrockmendel, Joris Van den Bossche, Tom Augspurger, Phillip Cloud, ... Marco Gorelli. (2020). pandas-dev/pandas: Pandas. Zenodo. <https://doi.org/10.5281/zenodo.3509134>
- Kanehisa, M., & Goto, S. (2000, January 1). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*. Oxford University Press. <https://doi.org/10.1093/nar/28.1.27>
- Khoury, G. A., Baliban, R. C., & Floudas, C. A. (2011). Proteome-wide post-translational modification statistics: Frequency analysis and curation of the swiss-prot database. *Scientific Reports*, 1, 1–5. <https://doi.org/10.1038/srep00090>
- Kimura, A. (1996). Identification of the Structural Gene for Glyoxalase I from *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 271(42), 25958–25965. <https://doi.org/10.1074/jbc.271.42.25958>

- Kind, T., & Fiehn, O. (2007). Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics*, 8, 1–20. <https://doi.org/10.1186/1471-2105-8-105>
- Kitano, H. (2002). Systems biology: A brief overview. *Science*, 295(5560), 1662–1664. <https://doi.org/10.1126/science.1069492>
- Kostidis, S., Addie, R. D., Morreau, H., Mayboroda, O. A., & Giera, M. (2017). Quantitative NMR analysis of intra- and extracellular metabolism of mammalian cells: A tutorial. *Analytica Chimica Acta*, 980, 1–24. <https://doi.org/10.1016/j.aca.2017.05.011>
- Krevelen, D. W. Van. (1950). Graphical-Statistical Method for the Study of Structure and Reaction Processes of Coal. *Fuel*, 29, 269–228.
- Kuepfer, L., Sauer, U., & Blank, L. M. (2005). Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Research*, 15(10), 1421–1430. <https://doi.org/10.1101/gr.3992505>
- Kuska, B. (1998). Beer, Bethesda, and Biology: How “Genomics” Came Into Being. *Journal of the National Cancer Institute*, 90(2), 93. <https://doi.org/10.1093/jnci/90.2.93>
- Lawrence, E. O., & Livingston, M. S. (1931). The production of high speed protons without the use of high voltages [3]. *Physical Review*, 38(4), 834. <https://doi.org/10.1103/PhysRev.38.834>
- Legras, J. L., Merdinoglu, D., Cornuet, J. M., & Karst, F. (2007). Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Molecular Ecology*, 16(10), 2091–2102. <https://doi.org/10.1111/j.1365-294X.2007.03266.x>
- Li, H., & Phung, D. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Lu, Y., Li, X., Mesfioui, R., Bauer, J. E., Chambers, R. M., Canuel, E. A., & Hatcher, P. G. (2015). Use of ESI-FTICR-ms to characterize dissolved organic matter in headwater streams draining forest-dominated and pasture-dominated watersheds. *PLoS ONE*, 10(12), 1–21. <https://doi.org/10.1371/journal.pone.0145639>
- Lyles, G. A., & Chalmers, J. (1992). The metabolism of aminoacetone to methylglyoxal by semicarbazide-sensitive amine oxidase in human umbilical artery. *Biochemical Pharmacology*, 43(7), 1409–1414. [https://doi.org/10.1016/0006-2952\(92\)90196-p](https://doi.org/10.1016/0006-2952(92)90196-p)
- Maia, M., Monteiro, F., Sebastiana, M., Marques, A. P., Ferreira, A. E. N., Freire, A. P., ... Sousa Silva, M. (2016). Metabolite extraction for high-throughput FTICR-MS-based metabolomics of grapevine leaves. *EuPA Open Proteomics*, 12, 4–9. <https://doi.org/10.1016/j.euprot.2016.03.002>
- Marshall, A. G., & Chen, T. (2015). 40 years of Fourier transform ion cyclotron resonance mass spectrometry. *International Journal of Mass Spectrometry*, 377(1), 410–420. <https://doi.org/10.1016/j.ijms.2014.06.034>
- Mcalister, L., & Holland, M. J. (1982). Targeted Deletion of a Yeast Enolase Structural Gene. *Journal of Biological Chemistry*, 257(12), 7181–7188.
- Mehmood, T., Liland, K. H., Snipen, L., & Sæbø, S. (2012). Chemometrics and Intelligent Laboratory Systems A review of variable selection methods in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems*, 118, 62–69. <https://doi.org/10.1016/j.chemolab.2012.07.010>
- Misra, K., Banerjee, A. B., Ray, S., & Ray, M. (1995). Glyoxalase III from *Escherichia coli*: A single novel enzyme for the conversion of methylglyoxal into D-lactate without reduced glutathione.

Biochemical Journal, 305(3), 999–1003. <https://doi.org/10.1042/bj3050999>

- Nikolaev, E. N., Jertz, R., Grigoryev, A., & Baykut, G. (2012). Fine structure in isotopic peak distributions measured using a dynamically harmonized fourier transform ion cyclotron resonance cell at 7 T. *Analytical Chemistry*, 84(5), 2275–2283. <https://doi.org/10.1021/ac202804f>
- Oppenoorth, W. F. F. (1962). Transformation in yeast: Evidence of a real genetic change by the action of DNA. *Nature*. <https://doi.org/10.1038/193706a0>
- Patti, G. J., Yanes, O., & Siuzdak, G. (2012). Innovation: Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology*, 13(4), 263–269. <https://doi.org/10.1038/nrm3314>
- Petitgonnet, C., Klein, G. L., Roullier-Gall, C., Schmitt-Kopplin, P., Quintanilla-Casas, B., Vichi, S., ... Alexandre, H. (2019). Influence of cell-cell contact between *L. thermotolerans* and *S. cerevisiae* on yeast interactions and the exo-metabolome. *Food Microbiology*, 83(April), 122–133. <https://doi.org/10.1016/j.fm.2019.05.005>
- Pietkiewicz, J., Gamian, A., Staniszevska, M., Pietkiewicz, J., Gamian, A., & Staniszevska, M. (2009). Inhibition of human muscle-specific enolase by methylglyoxal and irreversible formation of advanced glycation end products. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 24(2), 356–364. <https://doi.org/10.1080/14756360802187679>
- Pompella, A., Visvikis, A., Paolicchi, A., De Tata, V., & Casini, A. F. (2003). The changing faces of glutathione, a cellular protagonist. *Biochemical Pharmacology*, 66(8), 1499–1503. [https://doi.org/10.1016/S0006-2952\(03\)00504-5](https://doi.org/10.1016/S0006-2952(03)00504-5)
- Ponces Freire, A., Ferreira, A., Gomes, R., & Cordeiro, C. (2003). Anti-glycation defences in yeast. *Biochemical Society Transactions*, 31(6), 1409–1412. <https://doi.org/10.1042/bst0311409>
- Raamsdonk, L. M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M. C., ... Oliver, S. G. (2001). A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnology*, 19(1), 45–50. <https://doi.org/10.1038/83496>
- Rae, C., Board, P. G., & Kuchel, P. W. (1991). Glyoxalase 2 deficiency in the erythrocytes of a horse: 1H NMR studies of enzyme kinetics and transport of S-lactoylglutathione. *Archives of Biochemistry and Biophysics*, 291(2), 291–299. [https://doi.org/10.1016/0003-9861\(91\)90137-8](https://doi.org/10.1016/0003-9861(91)90137-8)
- Ramirez-Gaona, M., Marcu, A., Pon, A., Guo, A. C., Sajed, T., Wishart, N. A., ... Wishart, D. S. (2017a). YMDB 2.0: A significantly expanded version of the yeast metabolome database. *Nucleic Acids Research*, 45(D1), D440–D445. <https://doi.org/10.1093/nar/gkw1058>
- Ramirez-Gaona, M., Marcu, A., Pon, A., Guo, A. C., Sajed, T., Wishart, N. A., ... Wishart, D. S. (2017b). YMDB 2.0: A significantly expanded version of the yeast metabolome database. *Nucleic Acids Research*, 45(D1), D440–D445. <https://doi.org/10.1093/nar/gkw1058>
- Richard, J. P. (1993). Mechanism for the formation of methylglyoxal from triosephosphates. *Biochemical Society Transactions*, 21(2), 549–553. <https://doi.org/10.1042/bst0210549>
- Richard, John P. (1991). Kinetic Parameters for the Elimination Reaction Catalyzed by Triosephosphate Isomerase and an Estimation of the Reaction's Physiological Significance. *Biochemistry*, 30(18), 4581–4585. <https://doi.org/10.1021/bi00232a031>
- Richard, John P. (1984). Acid-Base Catalysis of the Elimination and Isomerization Reactions of Triose Phosphates. *Journal of the American Chemical Society*, 6(3), 4926–4936.
- Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., & Fernie, A. R. (2001). Metabolic profiling allows comprehensive phenotyping of genetically or environmentally

- modified plant systems. *Plant Cell*, 13(1), 11–29. <https://doi.org/10.1105/tpc.13.1.11>
- Ruiz-Perez, D., Guan, H., Madhivanan, P., Mathee, K., & Narasimhan, G. (2018). So you think you can PLS-DA? (pp. 1–1). Institute of Electrical and Electronics Engineers (IEEE). <https://doi.org/10.1109/iccabs.2018.8542038>
- Singh, V. P., Bali, A., Singh, N., & Jaggi, A. S. (2014). Advanced glycation end products and diabetic complications. *Korean Journal of Physiology and Pharmacology*. Korean Physiological Soc. and Korean Soc. of Pharmacology. <https://doi.org/10.4196/kjpp.2014.18.1.1>
- Sousa Silva, M., Ferreira, A. E. N., Gomes, R., Tomás, A. M., Ponces Freire, A., & Cordeiro, C. (2012). The glyoxalase pathway in protozoan parasites. *International Journal of Medical Microbiology*, 302(4–5), 225–229. <https://doi.org/10.1016/j.ijmm.2012.07.005>
- Sousa Silva, M., Gomes, R. A., Ferreira, A. E. N., Ponces Freire, A., & Cordeiro, C. (2013). The glyoxalase pathway: The first hundred years... and beyond. *Biochemical Journal*, 453(1), 1–15. <https://doi.org/10.1042/BJ20121743>
- Srikanth, V., Maczurek, A., Phan, T., Steele, M., Westcott, B., Juskiw, D., & Münch, G. (2011). Advanced glycation endproducts and their receptor RAGE in Alzheimer's disease. *Neurobiology of Aging*, 32(5), 763–777. <https://doi.org/10.1016/j.neurobiolaging.2009.04.016>
- Trethewey, R. N. (2001). Gene discovery via metabolic profiling. *Current Opinion in Biotechnology*, 12(2), 135–138. [https://doi.org/10.1016/S0958-1669\(00\)00187-7](https://doi.org/10.1016/S0958-1669(00)00187-7)
- Tupe, R. S., Vishwakarma, A., Solaskar, A., & Prajapati, A. (2019). Methylglyoxal induces glycation and oxidative stress in *Saccharomyces cerevisiae*. *Annals of Microbiology*, 69(11), 1165–1175. <https://doi.org/10.1007/s13213-019-01498-z>
- Valentine, W. N., Paglia, D. E., Neerhout, R. C., & Konrad, P. N. (1970). Erythrocyte glyoxalase II deficiency with coincidental hereditary elliptocytosis. *Blood*, 36(6), 797–808. <https://doi.org/10.1182/blood.v36.6.797.797>
- Vander Jagt, D. L., & Hunsaker, L. A. (2003). Methylglyoxal metabolism and diabetic complications: Roles of aldose reductase, glyoxalase-I, betaine aldehyde dehydrogenase and 2-oxoaldehyde dehydrogenase. *Chemico-Biological Interactions*, 143–144, 341–351. [https://doi.org/10.1016/S0009-2797\(02\)00212-0](https://doi.org/10.1016/S0009-2797(02)00212-0)
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Wagner, A. (2000). Robustness against mutations in genetic networks of yeast. *Nature Genetics*, 24(4), 355–361. <https://doi.org/10.1038/74174>
- Waskom, M., Botvinnik, O., Gelbart, M., Ostblom, J., Hobson, P., Lukauskas, S., ... Brunner, T. (2020). mwaskom/seaborn. Zenodo. <https://doi.org/10.5281/zenodo.592845>
- Wetzels, S., Wouters, K., Schalkwijk, C. G., Vanmierlo, T., & Hendriks, J. J. A. (2017). Methylglyoxal-derived advanced glycation endproducts in multiple sclerosis. *International Journal of Molecular Sciences*, 18(2). <https://doi.org/10.3390/ijms18020421>
- Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., ... Querengesser, L. (2007a). HMDB: The human metabolome database. *Nucleic Acids Research*, 35(SUPPL. 1). <https://doi.org/10.1093/nar/gkl923>
- Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., ... Querengesser, L. (2007b).

HMDB: The human metabolome database. *Nucleic Acids Research*, 35(SUPPL. 1), 521–526. <https://doi.org/10.1093/nar/gkl923>

Yamagishi, S., & Matsui, T. (2018). Role of Hyperglycemia-Induced Advanced Glycation End Product (AGE) Accumulation in Atherosclerosis. *Annals of Vascular Diseases*, 11(3), 253–258. <https://doi.org/10.3400/avd.ra.18-00070>

Zhang, A., Sun, H., Wang, P., Han, Y., & Wang, X. (2012). Modern analytical techniques in metabolomics analysis. *Analyst*, 137(2), 293–300. <https://doi.org/10.1039/c1an15605e>

Annexes

Mass	Name	Formula	BY474 1	ΔGRE3	ΔENO1	ΔGLO1	ΔGLO2
307.08 38	Glutathione	C10H17N3O6 S	1,51E+ 09	1,22E+ 09	7,11E+ 08	5,91E+ 08	2,90E+ 08
493.31 68	PC(16:1(9Z)/0:0)	C24H48NO7P	4,14E+ 08	4,80E+ 08	8,84E+ 07	8,30E+ 07	5,80E+ 07
624.08 73	N/A	C14H28N10O 10S4	4,92E+ 08	5,87E+ 08	3,82E+ 08	2,68E+ 08	1,32E+ 08
257.10 29	Glycerophosphoch oline	C8H20NO6P	2,08E+ 08	2,06E+ 08	1,99E+ 08	1,18E+ 08	6,10E+ 07
324.10 57	N/A	C12H20O10	7,22E+ 06	1,11E+ 07	4,50E+ 07	8,61E+ 06	1,99E+ 06
337.33 45	N/A	C22H43NO	4,95E+ 06	2,84E+ 06	5,35E+ 06	2,09E+ 07	4,45E+ 05
254.22 46	Hypogeic acid	C16H30O2	5,98E+ 07	4,81E+ 07	6,23E+ 07	5,73E+ 07	1,44E+ 07
385.31 92	Pentadecanoylcarni tine	C22H43NO4	6,32E+ 06	N/A	4,40E+ 06	1,80E+ 07	N/A
398.13 72	N/A	C15H22N6O5 S	1,55E+ 08	1,13E+ 08	4,32E+ 07	5,14E+ 07	4,31E+ 07
451.26 99	PE(16:1(9Z)/0:0)	C21H42NO7P	1,01E+ 08	1,24E+ 08	2,20E+ 07	1,77E+ 07	1,63E+ 07

Annex 1: Peak intensities of the ten metabolites with the highest VIP scores in PLS-DA.

Mass	Name	Formula	BY4741	ΔGRE3	ΔENO1	ΔGLO1
307.0838	Glutathione	1,51E+09	1,22E+09	7,11E+08	5,91E+08	2,90E+08
612.1521	Oxidized glutathione	3,28E+07	3,18E+07	1,30E+07	1,46E+07	5,79E+06
250.0624	gamma-Glutamylcysteine	3,46E+07	2,81E+07	2,25E+07	1,73E+07	9,39E+06
379.1049	S-Lactoylglutathione	2,43E+06	0,00E+00	1,67E+06	0,00E+00	7,01E+05
426.0879	Cysteineglutathione disulfide	3,30E+06	2,92E+06	0,00E+00	1,34E+06	7,08E+05
222.0674	L-Cystathionine	5,69E+07	4,91E+07	1,82E+07	1,91E+07	1,80E+07
384.1215	S-Adenosylhomocysteine	7,82E+06	6,41E+06	2,29E+06	2,69E+06	2,17E+06

Annex 2: Peak intensities of the metabolites involved in glutathione metabolism